



**T. C.  
BURSA ULUDAĞ ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ  
İŞLETME ANABİLİM DALI  
SAYISAL YÖNTEMLER BİLİM DALI**

**ÖZ-ÖRGÜTLEMELİ HARİTALAR İLE KENT  
MARKALAŞMASINDA KENT SAKİNLERİNİN YEŞİL KİMLİK  
PROFİLLERİNİN BELİRLENMESİ: “YEŞİL BURSA” ÖRNEĞİ**

**(DOKTORA TEZİ)**

**Gülcan PETRİÇLİ**

**BURSA - 2020**





**T. C.  
BURSA ULUDAĞ ÜNİVERSİTESİ  
SOSYAL BİLİMLER ENSTİTÜSÜ  
İŞLETME ANABİLİM DALI  
SAYISAL YÖNTEMLER BİLİM DALI**

**ÖZ-ÖRGÜTLEMELİ HARİTALAR İLE KENT  
MARKALAŞMASINDA KENT SAKİNLERİNİN YEŞİL KİMLİK  
PROFİLLERİNİN BELİRLENMESİ: “YEŞİL BURSA” ÖRNEĞİ**

**(DOKTORA TEZİ)**

**Gülcan PETRİÇLİ  
ORCID: 0000-0001-6296-6183**

**Danışman:  
Doç.Dr. A.Gül EMEL**

**BURSA - 2020**

## YEMİN METNİ

Doktora Tezi Çalışması olarak sunduğum “Öz-Örgütlemeli Haritalar ile Kent Markalaşmasında Kent Sakinlerinin Yeşil Kimlik Profillerinin Belirlenmesi: ‘Yeşil Bursa’ Örneği” başlıklı çalışmanın bilimsel araştırma, yazma ve etik kurallarına uygun olarak tarafımdan yazıldığına ve tezde yapılan bütün alıntıların kaynaklarının usulüne uygun olarak gösterildiğine, tezimde intihal ürünü cümle veya paragraflar bulunmadığına şerefim üzerine yemin ederim.

Tarih ve İmza

30.09.2020



Adı Soyadı : Gülcan PETRİÇLİ

Öğrenci No : 711314006

Anabilim Dalı : İşletme-Sayısal Yöntemler

Programı : Doktora

Statüsü : Yüksek Lisans  Doktora

Sanatta Yeterlik



**SOSYAL BİLİMLER ENSTİTÜSÜ**  
**DOKTORA İNTİHAL YAZILIM RAPORU**

**BURSA ULUDAĞ ÜNİVERSİTESİ**  
**SOSYAL BİLİMLER ENSTİTÜSÜ**  
**İŞLETME ANABİLİM DALI BAŞKANLIĞI'NA**

Tez Başlığı / Konusu: Öz-Örgütlemeli Haritalar ile Kent Markalaşmasında Kent Sakinlerinin Yeşil Kimlik Profillerinin Belirlenmesi: 'Yeşil Bursa' Örneği


Yukarıda başlığı gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam 379 sayfalık kısmına ilişkin, 30/09/2020 tarihinde şahsım tarafından Turnitin\* adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan özgünlük raporuna göre, tezimin benzerlik oranı % 3'tür.

Uygulanan filtrelemeler:

- 1- Kaynakça hariç
- 2- Alıntılar hariç/dahil
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Bursa Uludağ Üniversitesi Sosyal Bilimler Enstitüsü Tez Çalışması Özgünlük Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

	30.09.2020		Tarih ve İmza
<b>Adı Soyadı:</b>	Gülcan Petriçli		
<b>Öğrenci No:</b>	711314006		
<b>Anabilim Dalı:</b>	İşletme-Sayısal Yöntemler		
<b>Programı:</b>	Doktora		
<b>Statüsü:</b>	Y.Lisans	<b>Doktora</b>	Sanatta Yeterlik

**Danışman**

**Doç.Dr. A.Gül EMEL/30.09.2020**



\* Turnitin programına Bursa Uludağ Üniversitesi Kütüphane web sayfasından ulaşılabilir.

## ÖZET

Yazar Adı ve Soyadı	: Gülcan PETRİÇLİ
Üniversite	: Bursa Uludağ Üniversitesi
Enstitü	: Sosyal Bilimler Enstitüsü
Anabilim Dalı	: İşletme
Bilim Dalı	: Sayısal Yöntemler
Tezin Niteliği	: Doktora Tezi
Sayfa Sayısı	: xxv + 347
Mezuniyet Tarihi	: 14/09/2020
Tez Danışman(lar)ı	: Doç.Dr. A.Gül EMEL

### ÖZ-ÖRGÜTLEMELİ HARİTALAR İLE KENT MARKALAŞMASINDA KENT SAKİNLERİNİN YEŞİL KİMLİK PROFİLLERİNİN BELİRLENMESİ: “YEŞİL BURSA” ÖRNEĞİ

İleri teknoloji ve hızlı kentleşme, insanı doğadan ayırmakta, doğal kaynakların israfına ve çevresel zararlara neden olmaktadır. Sürdürülebilirlik, bu tip faaliyetlerden kaynaklanan sorunlara çözüm olarak ele alınmakta, küresel sürdürülebilirlik için de yerel çözümler önerilmektedir. Kent markalaşması ise kent vatandaşlarının yanı sıra yatırımcı ve işletmeler için de kenti cazip hale getirmekte, kentin yaşanabilirlik ve refah düzeyini artırmaktadır. Sürdürülebilirlik ve yeşil kent markalaşması faaliyetleri, kente ve markasına kimlik veren ve onu canlı tutan "vatandaş" tan başlamalıdır. Bu tezin amacı, tutum, insan-doğa ilişkisi ve sürdürülebilir tüketim davranışlarını dikkate alarak vatandaşların ekolojik kimlik profillerini keşfetmektir. Çalışmanın genelinde, veri madenciliği bilgi keşfi süreci benimsenmiş ve uygulanmaya çalışılmıştır. Ülkemizin önemli kentlerinden Bursa, on yıldır “Avrupa Yeşil Başkenti (AYM)” sistemine uyum çalışmaları yapmaktadır. Fakat, AYM olmak için henüz ön seçim aşamasını geçememiştir. Sürece katkıda bulunabilmek adına Bursa kenti çalışmanın uygulama alanı seçilmiştir. Veriler, çevresel endişe (tutum), insan-doğa ilişkisi ve çevre dostu davranışı ölçmek için tasarlanmış ve Türkçeye uyarlanmış olan üç ölçme aracı ile toplanmıştır. İlk aşamada, tutum ve sürdürülebilir tüketim davranışı arasındaki ilişkide doğa ile ilişkinin aracılık rolü Kısmi En Küçük Kareler Yapısal Eşitlik Modellemesi ile incelenmiştir. İkinci aşamada, veriler Süper Öz-Örgütlemeli Haritalar, Ward bağlantılı hiyerarşik ve k-medoid kümeleme algoritmaları ile kümelenebilir. Kümeler içsel ve dışsal değerlendirme kriterleri ile değerlendirilmiş ve en uygun küme sayısı beş olarak tespit edilmiştir. Her bir küme, yeşil kimlik profillerini tanımlamak için kutu bıyık grafikleri, çubuk grafikler ve karar ağaçları ile incelenmiştir. Ne yazık ki, sürdürülebilirlik ile uyumlu tamamen “yeşil” bir tipoloji tespit edilememiştir. Böylece, yerel yönetimlerin kentin sürdürülebilirliği için vatandaşların bakış açısını ve davranışlarını değiştirecek politikalar ve faaliyetler de tasarlaması gereği ortaya çıkmış ve “Yeşil Bursa” yeşil vatandaşlar ile elde edilebilir sonucuna varılmıştır.

#### Anahtar Sözcükler:

Öz-Örgütlemeli Haritalar, Kısmi En Küçük Kareler Yapısal Eşitlik Modellemesi, PERMANOVA, Kent Markalaşması, Sürdürülebilir Kentler, Çevre Dostu Davranış

## ABSTRACT

Name and Surname : Gülcan PETRİÇLİ  
University : Bursa Uludag University  
Institution : Social Science Institution  
Field : Business Administration  
Branch : Quantitative Methods  
Degree Awarded : PhD  
Page Number : xxv + 347  
Degree Date : 14/09/2020  
Supervisor (s) : Assoc. Prof. A. Gül EMEL

### **DETERMINING GREEN IDENTITY PROFILES OF URBAN RESIDENTS IN CITY BRANDING WITH SELF-ORGANIZING MAPS: CASE OF “GREEN BURSA”**

Advanced technology and rapid urbanization have caused waste of natural resources and environmental damage by disconnecting human from nature. Sustainability is considered as a solution to problems arising from such activities, and local solutions are suggested for global sustainability. Urban branding, on the other hand, makes the city attractive for investors businesses and citizens, and increases the livability and welfare level of the city. Sustainability and green city branding activities should start from the "citizen", who gives the city and its brand an identity and keeps them alive. The purpose of this thesis is to discover ecological identity profiles by taking attitudes, human-nature relationship and sustainable consumption behaviors of citizens into consideration. Data mining and knowledge discovery process has been adopted and implemented throughout the study. Bursa, one of the important cities of our country, has been working on adaptation to the "European Green Capital (EGC)" system for the last ten years. However, it has not yet passed the pre-selection stage to become an EGC. In order to contribute to the process, the city of Bursa has been chosen as the application area of the study. The data were collected with three adapted measurement scales that are designed for measuring environmental concerns (attitude), human-nature relationship and environmentally friendly behavior. In the first stage, the mediating role of the nature-relatedness in the relationship between attitude and sustainable consumption behavior was investigated by Partial Least Squares Structural Equation Modeling. In the second stage, the data were clustered with Super Self-Organizing maps, Ward-linked hierarchical and k-medoid clustering algorithms. Clusters were evaluated with internal and external evaluation criteria and the most appropriate cluster number was determined to be five. Each cluster was investigated with box-whiskers plots, bar charts and decision trees in order to describe green identity profiles. Unfortunately, a completely “green” typology compatible with sustainability has not been identified. Therefore, the need for local governments to design policies and activities that will change the perspective and behavior of citizens for the sustainability of the city emerged and it was concluded that “Green Bursa” can be achieved with green citizens.

#### **Keywords:**

Self-Organizing Maps, Partial Least Square Structural Equation Modelling, PERMANOVA, City Branding, Sustainable Cities, Pro-Environmental Behavior

## ÖNSÖZ

Bu tez çalışması hiç şüphesiz üç isimli bir çalışmadır. Her bir kelimesinde eşimin desteği ve oğlumun kısmen ertelenmiş oyunları bulunmaktadır. Bu yüzden öncelikle, sevgili eşim İsmail'e ve günışığım oğlum Can'a teşekkür ederim.

Her zaman bana güvenen, hayatım boyunca maddi ve manevi olarak yanımda olan fakat ayaklarımın üzerinde tek başıma durmamı da öğreten annem ve babama, bizi tüm koşullar altında destekleyen tüm Petriçli ve Yalnızca ailelerine, çok ama çok uzaklardayken bile hep yanımda ve kalbimde olan yol arkadaşım Gamze'ye, özel sektörde çalışırken Tezsiz Yüksek Lisans programına başladığımda birlikte uzun bir yola çıktığımı bilmeden hayatı paylaşarak çoğaldığım(ız) Elvan, Merve ve Sinem'e sonsuz teşekkürü borç bilirim.

Lisansüstü eğitim hayatım boyunca her zaman yanımda olan, her şeyin zorlaştığı ya da ters gittiği zamanlarda önümü açan, akademik desteğini esirgemeyen, beni her zaman iyiyi bulmaya teşvik eden, Bursa Uludağ Üniversitesindeki varlık sebebim değerli danışmanım Doç.Dr. A.Gül EMEL'e çok teşekkür ederim.

Doktora tez çalışmamda bilgisi ve yol göstericiliği ile bana her daim destek olan, güler yüzlü ve dinamik hocam sayın Prof.Dr. Çağatan TAŞKIN'a ayrıca teşekkür ederim.

Endüstri Mühendisliği bölümündeki Veri Madenciliği dersini dışarıdan aldığım ilk günden beri akademik anlamda kendisinden çok şey öğrendiğim, yüzyüze görüşemediğimiz zamanlarda bile yoğun temposuna rağmen bazen telefonla bazen sanal ortamda benimle devamlı irtibatla kalan ve her görüşmemizde ufkumu açan güler yüzlü, naif ve alçak gönüllü hocam Doç.Dr. Tülin İNKAYA'ya çok teşekkür ederim.

Bu tez çalışmasını sonuna kadar okuyan ve kıymetli fikirlerini benimle paylaşan Doç.Dr. Aysun KAPUCUGİL İKİZ ve Doç.Dr. Emre İPEKÇİ ÇETİN'e ayrıca teşekkür ederim.

Verileri sahadan toplayan PROSİS Danışmanlık ekibine ve Dilek ŞEN'e ve Bursa Büyükşehir Belediyesi Çevre Koruma ve Kontrol Dairesi Başkanlığına tarafımla paylaşmış olduğu bilgiler için teşekkürü bir borç bilirim.



Gerek tez ve gerek ise araştırma görevliliğim süresince hayatı benimle paylaşan, gerektiğinde dinleyen, gerektiğinde ise yol gösteren ve beni motive eden tüm Bursa Uludağ Üniversite’li yol arkadaşlarıma teşekkürlerimi ve sevgilerimi sunarım.

Önsözü, “ben” ile bitirmek istiyorum. Lisede yatılı devlet okuluna gidişim ile birlikte yaklaşık 25 yıldır yoldayım, gerek Türkiye’de ve gerek ise yurt dışında. Bu yüzden “Yol”ların benim hayatımda ayrı bir yeri var. “Her bir yola çıkış, çıkılacak yeni yolların sorumluluğunu da getirir” diyor Oruç Auroba... evet, bu bir bitiş değil, yeni bir yola çıkış.

“Yol”da karşılaştığım ve karşılaştığım herkese selamlarımla...

Bolca sevgi ve doğa ile kalın!

*Temmuz 2020*

*Nilüfer, Bursa*

## İÇİNDEKİLER

YEMİN METNİ .....	ii
DOKTORA İNTİHAL YAZILIM RAPORU .....	iii
ÖZET .....	iv
ABSTRACT .....	v
ÖNSÖZ.....	vi
İÇİNDEKİLER.....	viii
KISALTMALAR.....	xiv
TABLolar .....	xvi
ŞEKİLLER .....	xix
EKLER .....	xxiv
GİRİŞ.....	1

### BİRİNCİ BÖLÜM

#### (KENT MARKALAŞMASI ve SÜRDÜRÜLEBİLİRLİK İLİŞKİSİ)

1. SÜRDÜRÜLEBİLİRLİK ve KENTLEŞME.....	7
2. SÜRDÜRÜLEBİLİRLİK, ENERJİ ve TÜKETİM .....	13
3. KENT MARKALAŞMASI.....	17
4. ÇEVRE DOSTU DAVRANIŞ.....	25

### İKİNCİ BÖLÜM

#### (VERİ MADENCİLİĞİ ve BİLGİ KEŞFİ)

1. VERİ MADENCİLİĞİ.....	34
2. BİLGİ KEŞFİ SÜRECİ.....	36
2.1. Veri Seçimi.....	36
2.2. Veri Önleme.....	40
2.2.1. Tekrarlı Veri.....	41

2.2.2.	Eksik Veri .....	41
2.2.3.	Hatalı Veri.....	43
2.2.4.	Aykırı Veri .....	43
2.3.	Veri Dönüştürme .....	54
2.3.1.	Veri Düzleştirme .....	54
2.3.2.	Veri Normalleştirme .....	54
2.3.3.	Veri Ayırıklaştırma .....	56
2.3.4.	Veri İndirgeme .....	57
2.4.	Veri Madenciliği.....	67
2.4.1.	Sınıflandırma.....	67
2.4.2.	Kümeleme .....	69
2.4.3.	Birliktelik Kuralları.....	69
2.4.4.	Tahminleme .....	70
2.5.	Sonuçların Yorumlanması .....	70
3.	KEŞİF SÜRECİNİ OLUMSUZ ETKİLEYEN FAKTÖRLER.....	70
4.	ÖLÇME DÜZEYLERİ .....	73
4.1.	Sınıflayıcı Ölçme Düzeyi .....	74
4.2.	Sıralayıcı Ölçme Düzeyi .....	76
4.3.	Eşit Aralıklı Ölçme Düzeyi .....	77
4.4.	Oranlı Ölçme Düzeyi.....	78
5.	BENZERLİK VE BENZEŞMEZLİK (YAKINLIK VE UZAKLIK) ÖLÇÜTLERİ .....	78
5.1.	Tek Özniteliğe Sahip Gözlemlerin Benzerlik-Benzeşmezlik Ölçütleri.	80
5.2.	Tüm Öznitelikleri Nominal Olan Gözlemler İçin Benzerlik- Benzeşmezlik Ölçütleri .....	82

5.3. Tüm Öznitelikleri İkili Türde Olan Gözlemler İçin Benzerlik-Benzeşmezlik Ölçütleri .....	84
5.4. Tüm Öznitelikleri Ordinal Türde Olan Gözlemler İçin Benzerlik-Benzeşmezlik Ölçütleri .....	86
5.5. Tüm Öznitelikleri Nümerik Türde Olan Gözlemler İçin Benzerlik-Benzeşmezlik Ölçütleri .....	86
5.6. Farklı Türde Özniteliklere Sahip Gözlemler İçin Benzerlik-Benzeşmezlik Ölçütleri .....	88
5.7. Yüksek Korelasyonlu Özniteliklere Sahip Gözlemler İçin Benzerlik-Benzeşmezlik Ölçütleri .....	90

## ÜÇÜNCÜ BÖLÜM

### (KÜME ANALİZİ)

1. KÜME ANALİZİ.....	98
1.1. Kümeleme Algoritmaları.....	100
1.2. Uygun Küme Sayısının Tespiti .....	107
1.3. Küme Kalitesinin İncelenmesi .....	110
1.3.1. İçsel Değerlendirme Kriterleri .....	111
1.3.2. Dışsal Değerlendirme Kriterleri.....	113
1.4. Küme Analizini Olumsuz Etkileyen Faktörler .....	119
2. KARMA VERİ KÜMELEME ALGORİTMALARI .....	120
2.1. Bölümleme Esaslı Karma Veri Kümeleme Algoritmaları.....	121
2.2. Hiyerarşik Karma Veri Kümeleme Algoritmaları.....	123
2.3. Model Esaslı Karma Veri Kümeleme Algoritmaları.....	124
2.4. Yapay Sinir Ağları.....	125
3. ÖZ-ÖRGÜTLEMELİ HARİTALAR .....	126
3.1. Temel YSA Kavramları.....	126

3.2.	Öz-Örgütlemeli Haritalar ve Öğrenme Algoritmaları .....	129
3.3.	Öz-Örgütlemeli Haritaların Başarımı .....	142
3.4.	Öz-Örgütlemeli Haritaların Doğrulanması.....	145
3.5.	Karma Veri Kümelemede Kullanılan Öz-Örgütlemeli Harita Çeşitleri	148

## DÖRDÜNCÜ BÖLÜM

### (UYGULAMA)

1.	ARAŞTIRMA TASARIMI VE SÜRECİ .....	160
2.	TEZ ÇALIŞMASININ UYGULAMA AŞAMASINDA KULLANILAN DİĞER YÖNTEMLER.....	167
2.1.	Güvenirlilik Analizi .....	167
2.2.	KEKK-YEM Yöntemi.....	168
2.3.	Hiyerarşik Küme Analizi.....	169
2.4.	k-Temsilci Küme Analizi .....	172
2.5.	Permütasyonel Çok Değişkenli Varyans Analizi .....	173
2.6.	Karar Ağaçları .....	177
3.	PİLOT ARAŞTIRMANIN ANALİZİ .....	182
3.1.	Birinci Pilot Araştırmanın Aykırı Veri Analizi (Veri Önışleme).....	183
3.2.	Birinci Pilot Araştırmanın Geçerlik ve Güvenirlilik Analizi.....	187
3.2.1.	ÇE Ölçeğine Ait Analiz Sonuçları .....	187
3.2.2.	Dİ Ölçeğine Ait Analiz Sonuçları .....	189
3.2.3.	STD Ölçeğine Ait Analiz Sonuçları.....	191
3.3.	İkinci Pilot Araştırmanın Aykırı Veri Analizi (Veri Önışleme).....	194
3.4.	İkinci Pilot Araştırmanın Geçerlik ve Güvenirlilik Analizi .....	195
3.4.1.	ÇE Ölçeğine Ait Analiz Sonuçları.....	195
3.4.2.	DİÖ'ye Ait Analiz Sonuçları .....	198

3.4.3. STD Ölçeğine Ait Analiz Sonuçları.....	201
3.5. Pilot Araştırma Süreci Hakkında Genel Değerlendirme .....	204
4. ARAŞTIRMA MODELİNİN DEĞERLENDİRİLMESİ .....	205
4.1. Veri Önışleme.....	205
4.2. Yapısal Analiz .....	207
4.2.1.Güvenirlık ve Geçerlık Analizleri .....	208
4.2.2. Model Analizi .....	213
5. PROFİL ÇIKARMA .....	220
5.1. Veri İndirgeme (Boyut Azaltma) .....	220
5.2. Veri Dönüştürme .....	221
5.3. Küme Analizi .....	222
5.3.1. Çevrimiçi Öğrenme Algoritması ile Eğitilen Haritaların Analizleri	226
5.3.2. Yığın Öğrenme Algoritması ile Eğitilen Haritaların Analizleri ....	242
5.4. Kimlik Tanımlama .....	259
5.4.1. Betimsel İstatistikler ile Tanımlama .....	259
5.4.2. Karar Ağaçları ile Tanımlama.....	275
6. GENEL DEĞERLENDİRME.....	282
SONUÇ.....	287
KAYNAKLAR.....	298
EKLER .....	319
Ek 1 .....	319
Ek 2.....	321
Ek 3.....	325
Ek 4.....	326
Ek 5.....	331

Ek 6.....	332
Ek 7.....	333
Ek 8.....	341
Ek 9.....	342
EK 10.....	343
Ek 11.....	344
Ek 12.....	345
Ek 13.....	346
ÖZGEÇMİŞ.....	347

## KISALTMALAR

<b>Kısaltma</b>	<b>Bibliyografik Bilgi</b>
AB	Avrupa Birliđi
AFA	Açımlayıcı Faktör Analizi
ART	Adaptif Rezonans Teorisi
AVE	Açıklanan Ortalama Varyans
AYB	Avrupa Yeşil Başkenti
BBB	Bursa Büyükşehir Belediyesi
BHÖÖH	Büyüyen Hiyerarşik ÖÖH
BKÖÖH	Büyüyen Karma ÖÖH
BM	Birleşmiş Milletler
BÖÖH	Büyüyen ÖÖH
CART	Sınıflandırma ve Regresyon Ağaçları
CR	Birleşik Güvenirlilik
ÇDD	Çevre Dostu Davranış
ÇEÖ	Çevresel Endişe Ölçeđi
DEK	Dünya Enerji Konseyi
DFA	Dođrulayıcı Faktör Analizi
DİÖ	Dođayla İlişki Ölçeđi
DÖÖASA	Düzeltilmiş Öz Örgütlemeli Artımlı Sinir Ađı
EID	Environmental Identity
FA	Faktör Analizi
FNKÖÖH	Frekans Nöronlu Karma ÖÖH
GGKÖÖH	Genelleştirilmiş Görselleştirme Kaynaklı ÖÖH
GIGO	Garbage in, Garbage out
GKÖÖH	Görselleştirme Kaynaklı ÖÖH
GÖÖT	Genişletilmiş Öznitelik Odaklı Tümevarım
GÖÖH	Genelleştirilmiş ÖÖH
HCPC	Hierarchical Clustering On Principal Components
HTMT	Heterotrait-Monotrait Ratio
IQR	Kartiller Arası Mesafe
KAM	Kümeler Arası Mesafeler
KDD	Knowledge Discovery Database
KEKK	Kısmi En Küçük Kareler
KEKK-YEM	Kısmi En Küçük Kareler Esaslı Yapısal Eşitlik Modelleme
KMO	Kaiser-Mayer-Olkin
KOÖÖH	Kategorik Olasılıksal ÖÖH
KÖÖASA	Karma Öz Örgütlemeli Artımlı Sinir Ađı
KT	Kareler Toplamı



MKM	Minimum Kovaryans Mesafesi
MSA	Measure Of Sampling Adequacy
NH	Nicemleme Hatası
NKÖÖH	Nümerik-Kategorik ÖÖH
OKSS	Ortalama Karesel Standart Sapma
OKT	Ortak Kaynakların Trajedisi
OKTH	Olasılıksal Karma Topolojik Haritalar
ÖM	Örüntü/Öznitelik Matrisleri
ÖÖASA	Öz Örgütlemeli Artımlı Sinir Ağı
ÖÖH	Öz Örgütlemeli Haritalar
PERMANOVA	Permütasyonel Çok Değişkenli Varyans Analizi
RK	R-Kare
SMM	Sağlam Mahalanobis Mesafesi
SÖÖH	Süper Öz-Örgütlemeli Haritalar
STD	Sürdürülebilir Tüketim Davranışları
TBA	Temel Bileşenler Analizi
TH	Topografik Hata
VAF	Açıklanan Varyans
VIF	Varyans Arttırıcı Faktör
ViSOM	Visualization-Induced SOM
VM	Veri Madenciliği
VTBK	Veri Tabanlarında Bilgi Keşfi
YEM	Yapısal Eşitlik Modeli
YKR	Yarı Kısmi R-Kare
YSA	Yapay Sinir Ağları

## TABLÖLAR

<b>Bilgi</b>	<b>Sayfa</b>
<b>Tablo 1.</b> Mevcut müşterilere ait veri seti	68
<b>Tablo 2.</b> Tek öznitelikli gözlemler için benzerlik/benzeşmezlik ölçütleri	81
<b>Tablo 3</b> Örnek 9'a ait veri seti	82
<b>Tablo 4</b> İkili türdeki öznitelikler için kontenjans tablosu	84
<b>Tablo 5</b> Örnek 10'a ait veri	85
<b>Tablo 6.</b> Örnek 10'a ait örnek bir çözüm	85
<b>Tablo 7.</b> Örnek 11'e ait veri	89
<b>Tablo 8.</b> Örnek 11'e ait örnek bir çözüm	89
<b>Tablo 9.</b> STD ölçeğinin Çevreye Duyarlı Gıda Tercihi boyutunda yer alan genişletilerek güncellenen ifadeler	183
<b>Tablo 10.</b> Birinci pilot veri setinin ÇE ölçeği için KMO ve Barlett testi sonuçları	187
<b>Tablo 11.</b> Birinci pilot veri setinin ÇE ölçeği için varyans açıklama yüzdeleri	187
<b>Tablo 12.</b> Birinci pilot veri setinin ÇE ölçeği için faktör yükleri	188
<b>Tablo 13.</b> Schultz tarafından farklı çalışmalarda kullanılan ÇE ölçeği boyutlarının ait alfa değerleri	188
<b>Tablo 14.</b> Birinci pilot veri setinin Dİ ölçeği için KMO ve Barlett testi sonuçları	189
<b>Tablo 15.</b> Birinci pilot veri setinin ÇE ölçeği için varyans açıklama yüzdeleri	189
<b>Tablo 16.</b> Birinci pilot veri setinin Dİ ölçeği için birinci faktör analizi faktör yükleri	190
<b>Tablo 17.</b> Birinci pilot veri setinin Dİ ölçeği için ikinci faktör analizi varyans açıklama yüzdeleri	191
<b>Tablo 18.</b> Birinci pilot veri setinin Dİ ölçeği için ikinci faktör analizi faktör yükleri	191
<b>Tablo 19.</b> Birinci pilot veri setinin STD ölçeği için KMO ve Barlett testi sonuçları	192
<b>Tablo 20.</b> Birinci pilot veri setinin STD ölçeği için ikinci faktör analizi varyans açıklama yüzdeleri	192
<b>Tablo 21.</b> Birinci pilot veri setinin STD ölçeği için ikinci faktör analizi faktör yükleri	193
<b>Tablo 22.</b> İkinci pilot veri setinin ÇE ölçeği için KMO ve Barlett testi sonuçları	196
<b>Tablo 23.</b> İkinci pilot veri setinin ÇE ölçeği için varyans açıklama yüzdeleri	196
<b>Tablo 24.</b> İkinci pilot veri setinin ÇE ölçeği için faktör yükleri	197
<b>Tablo 25.</b> İkinci pilot veri setinin ÇE ölçeği için ikinci faktör analizi varyans açıklama yüzdeleri	197
<b>Tablo 26.</b> İkinci pilot veri setinin ÇE ölçeği için faktör yükleri	198
<b>Tablo 27.</b> İkinci pilot veri setinin Dİ ölçeği için KMO ve Barlett testi sonuçları	199

<b>Tablo 28.</b> İkinci pilot veri setinin Dİ ölçeği için ikinci faktör analizi varyans açıklama yüzdeleri	199
<b>Tablo 29.</b> İkinci pilot veri setinin Dİ ölçeği için birinci faktör analizi faktör yükleri	200
<b>Tablo 30.</b> İkinci pilot veri setinin Dİ ölçeği için ikinci faktör analizi varyans açıklama yüzdeleri	200
<b>Tablo 31.</b> İkinci pilot veri setinin Dİ ölçeği için ikinci faktör analizi faktör yükleri	200-201
<b>Tablo 32.</b> İkinci pilot veri setinin STD ölçeği için KMO ve Barlett testi sonuçları	201
<b>Tablo 33.</b> İkinci pilot veri setinin STD ölçeği için varyans açıklama yüzdeleri	202
<b>Tablo 34.</b> İkinci pilot veri setinin STD ölçeği için faktör yükleri	203
<b>Tablo 35.</b> Orijinal ölçekler için ölçme modelinin sonuçları	209-210
<b>Tablo 36.</b> Revize edilmiş ölçekler için ölçme modelinin sonuçları	210-211
<b>Tablo 37.</b> Ayrışma geçerliği (Fornel-Larcker kriteri)	212
<b>Tablo 38.</b> Ayrışma geçerliği (HTMT kriteri)	212
<b>Tablo 39.</b> Çoklu doğrusallık (VIF kriteri)	213
<b>Tablo 40.</b> Araştırma modeli katsayıları I (Çevresel endişenin STD üzerindeki direkt etkisi)	215-216
<b>Tablo 41.</b> Araştırma modeli katsayıları II (Çevresel endişenin Doğayla ilişki üzerindeki etkisi)	216
<b>Tablo 42.</b> Araştırma modeli katsayıları III (Doğayla ilişkinin STD üzerindeki etkisi)	217
<b>Tablo 43.</b> Araştırma modeli katsayıları III (Çevresel endişenin STD üzerindeki dolaylı etkisi)	218-219
<b>Tablo 44.</b> Harita 11 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri	231
<b>Tablo 45.</b> Harita 149 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri	233
<b>Tablo 46.</b> Harita 227 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri	234
<b>Tablo 47.</b> Harita 11, harita 149 ve harita 227'nin küme sayısı tespit etme prosedürünün iki, üç ve dört numaralı aşamaları kapsamında karşılaştırılması	237
<b>Tablo 48.</b> Harita 2 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri	247
<b>Tablo 49.</b> Harita 132 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri	249
<b>Tablo 50.</b> Harita 95 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri	250
<b>Tablo 51.</b> Harita 2, harita 132 ve harita 95'in küme sayısı tespit etme prosedürünün iki, üç ve dört numaralı aşamaları kapsamında karşılaştırılması	254
<b>Tablo 52.</b> Harita 227 ve harita 132'nin tüm kriterler bağlamında karşılaştırılması	258
<b>Tablo 53.</b> Kural çıkarımında kullanılan karar ağacının hata matrisi	276

<b>Tablo 54.</b> Kural çıkarımda kullanılan 629 numaralı ağacın çapraz doğrulama sonuçları	277
<b>Tablo 55.</b> Karar ağacı bölümlene kriterlerinin önem dereceleri	279
<b>Tablo 56.</b> Veri normalleştirme yöntemleri	319-320
<b>Tablo 57.</b> 200 örneklemlik pilot veri setinin ÇE ölçeği için faktör yükleri	326
<b>Tablo 58.</b> 200 örneklemlik pilot veri setinin Dİ ölçeği için birinci faktör analizi faktör yükleri	327
<b>Tablo 59.</b> 200 örneklemlik pilot veri setinin Dİ ölçeği için ikinci faktör analizi faktör yükleri	328
<b>Tablo 60.</b> 200 örneklemlik pilot veri setinin STD ölçeği için ikinci faktör analizi faktör yükleri	329-330
<b>Tablo 61.</b> Mahalanobis ve SMM ölçütleri ile temizlenen ikinci pilot veri setine ait FA sonuçları ve Cronbach alfa değerleri	331
<b>Tablo 62.</b> Gözlem sayısı 352 olan ana araştırmaya ait veri setinin ÇE ölçeği için varyans açıklama yüzdeleri	333
<b>Tablo 63.</b> Gözlem sayısı 352 olan ana araştırmaya ait veri setinin ÇE ölçeği için faktör yükleri	334
<b>Tablo 64.</b> Gözlem sayısı 352 olan ana araştırmaya ait veri setinin ÇE ölçeği için varyans açıklama yüzdeleri	335
<b>Tablo 65.</b> Gözlem sayısı 352 olan ana araştırmaya ait veri setinin Dİ ölçeği için faktör yükleri	336
<b>Tablo 66.</b> Gözlem sayısı 352 olan ana araştırmaya ait veri setinin ÇE ölçeği için varyans açıklama yüzdeleri	337
<b>Tablo 67.</b> Gözlem sayısı 352 olan ana araştırmaya ait veri setinin STD ölçeği için faktör yükleri	339
<b>Tablo 68.</b> Karar ağacında kullanılan 227 numaralı haritanın TBA analizinin sonucunda özniteliklerin birinci ve ikinci temel bileşenler üzerindeki yükleri	345

## ŞEKİLLER

Bilgi	Sayfa
Şekil 1. İnsan davranışlarının sürdürülebilirlik ve kent markalaşması ile ilişkisi	2
Şekil 2. Araştırmanın motivasyonu	8
Şekil 3. Kenti oluşturan doğal, sosyal ve ekonomik çevreler arasındaki ilişkiden doğan olumlu ve olumsuz etkiler	11
Şekil 4. Kotler'in mekân pazarlama süreci	20
Şekil 5. Schwartz'ın çembersel değer modeli	30
Şekil 6. Bilgi keşfi süreci	36
Şekil 7. Örnek 2'nin histogram ve yoğunluk grafiği	46
Şekil 8. Örnek 2'nin kutu-bıyık diyagramında bıyık uzunluğunun kartiller arası mesafenin 1.5 katı kuralı (a) dikkate alınarak ve (b) alınmayarak tespit edilmesi	46
Şekil 9. Örnek 2'nin teorik kartiller dikkate alınarak tespit edilen aykırı veriler grafiği	47
Şekil 10. Örnek 2'nin veri setinden aykırı gözlemler çıkartıldıktan sonra elde edilen histogram ve yoğunluk grafiği	48
Şekil 11. Örnek 2'nin veri setinden aykırı gözlemler çıkartıldıktan sonra elde edilen kutu-bıyık diyagramı	48
Şekil 12. Örnek 2'nin veri setinden aykırı gözlemler çıkartıldıktan sonra teorik kartiller dikkate alınarak elde edilen dağılım grafiği	49
Şekil 13. Örnek 3'ün saçılım grafiği veri setinden aykırı gözlemler çıkartıldıktan sonra, çıkartılan gözlemler yerine regresyon ile yeni veri elde edilen	52
Şekil 14. Örnek 3'ün veri setinden aykırı gözlemler çıkartıldıktan sonra, çıkartılan gözlemler yerine regresyon ile yeni verinin elde edilmesi	53
Şekil 15. Yanlılık, doğruluk ve kesinliğin grafiksel gösterimi	72-73
Şekil 16. Farklı h değerleri (söz konusu görselde q ile sembolize edilen değerler) için hesaplanan uzaklıkların geometrik anlamı	87
Şekil 17. Mahalanobis uzaklığının geometrik gösterimi	93
Şekil 18. (2x2) boyutlu bir matrisin determinantının geometrik gösterimi	95
Şekil 19. Mahalanobis mesafesi ve SMM'nin geometrik gösterimi	97
Şekil 20. Hata matrisi	117
Şekil 21. Yapay bir nöronun/düğümün yapısı	127
Şekil 22. Örnek 14'e ait YSA	128
Şekil 23. Örnek bir ÖÖH'nin yapısı	131
Şekil 24. Örnek bir ÖÖH'nin eğitim süreci	131
Şekil 25. Örnek 14'ün ÖÖH ile eğitiminin R Studio çıktısı	138
Şekil 26. $m_i$ kazanan nöronun sinaptik ağırlığı olmak üzere yığın algoritmasının bir iterasyonu	141
Şekil 27. Gauss komşuluk fonksiyonu	145
Şekil 28. Araştırma süreci	161-162
Şekil 29. Grup içi varyansın merkez nokta dikkate alınarak ve dikkate alınmayarak hesaplanabilirliğinin geometrik gösterimi	175

<b>Şekil 30.</b> Grup içi, gruplar arası ve genel kareler toplamının şematik gösterimi	176
<b>Şekil 31.</b> Gaus gürültüsü eklenmiş sinüs fonksiyonunun CART algoritması ile tahmininde (a) ağaç derinliği (x-ekseni) ile minbucket (y-ekseni) miktarının ve (b) cp değerinin öğrenmeye etkisi	180
<b>Şekil 32.</b> x ekseninde bir sınıftan olma olasılığı ve y ekseninde endeks değeri olmak üzere entropi değeri $\frac{1}{2}$ ile ölçeklendirildiğinde iki sınıflı bir karar problemi için endeks değerlerinin benzerliği	181
<b>Şekil 33.</b> Gözlem sayısı 200 olan endişe veri setinin aykırı veri analizi öncesi ve sonrası $\alpha = 0.05$ , serbestlik derecesi = 12, $\chi^2 = 21.026$ için hesaplanan Mahalanobis mesafeleri	183
<b>Şekil 34.</b> Gözlem sayısı 200 olan endişe veri setinin aykırı veri analizi sonrası $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinantı $-7.87$ için SMM	184
<b>Şekil 35.</b> Gözlem sayısı 200 olan ilişki veri setinin aykırı veri analizi öncesi ve sonrası $\alpha = 0.05$ , serbestlik derecesi = 21, $\chi^2 = 32.670$ için hesaplanan Mahalanobis mesafeleri	184
<b>Şekil 36.</b> Gözlem sayısı 200 olan ilişki veri setinin aykırı veri analizi sonrası $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinantı $-29.1$ için SMM	184
<b>Şekil 37.</b> Gözlem sayısı 200 olan davranış veri setinin aykırı veri analizi öncesi ve sonrası $\alpha = 0.05$ , serbestlik derecesi = 27, $\chi^2 = 40.113$ için hesaplanan Mahalanobis mesafeleri	185
<b>Şekil 38.</b> Özniteliklere ait iki değişkenli saçılım grafikleri ve histogramları	186
<b>Şekil 39.</b> Gözlem sayısı 300 olan endişe veri setinin aykırı veri analizi sonrası $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinantı $-6.74$ için SMM	194
<b>Şekil 40.</b> Gözlem sayısı 300 olan ilişki veri setinin aykırı veri analizi sonrası $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinantı $-20.2$ için SMM	195
<b>Şekil 41.</b> Gözlem sayısı 300 olan davranış veri setinin aykırı veri analizi sonrası $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinantı $-26.6$ için SMM	195
<b>Şekil 42.</b> Gözlem sayısı 692 olan endişe veri setinin aykırı veri analizi sonrası $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinantı $-7.57$ için SMM	206
<b>Şekil 43.</b> Gözlem sayısı 692 olan ilişki veri setinin aykırı veri analizi sonrası $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinantı $-16.9$ için SMM	206
<b>Şekil 44.</b> Gözlem sayısı 692 olan davranış veri setinin aykırı veri analizi sonrası $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinantı $-23.5$ için SMM	206
<b>Şekil 45.</b> Gözlem sayısı 352 olan ana araştırmaya ait veri setinin endişe, ilişki ve davranış ölçekleri için Mahalanobis mesafelerinin normallik testi	207
<b>Şekil 46.</b> Multitrait-multimethod (MTMT) matrisi	212

Şekil 47. Araştırma modeli	214
Şekil 48 Aracılık ilişkisi	215
Şekil 49. ÇE boyutlarını, STD boyutları üzerindeki etkisinde, DI boyutlarının anlamlı bulunan aracılık etkileri	219
Şekil 50. Çevrimiçi öğrenme algoritmaları ile eğitilen haritalardan eleme sonrası kalan 15 haritanın nicemleme hatasının nöron sayısına göre değişimi	226
Şekil 51. Harita 11'in (a) hücrelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe	227
Şekil 52. Harita 149'un (a) hücrelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe	228
Şekil 53. Harita 227'nin (a) hücrelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe	229
Şekil 54. Harita 11 için (a-b) küme içi kareler toplamının küme sayısına göre değişimi ve (c-d) ortalama silüet endeksinin küme sayısına göre değişimi	230
Şekil 55. Harita 149 için (a-b) küme içi kareler toplamının küme sayısına göre değişimi ve (c-d) ortalama silüet endeksinin küme sayısına göre değişimi	232
Şekil 56. Harita 227 için (a-b) küme içi kareler toplamının küme sayısına göre değişimi ve (c-d) ortalama silüet endeksinin küme sayısına göre değişimi	234
Şekil 57. Harita 11'in (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına k-temsilli kümeleme algoritması ile elde edilen $k=4$ için küme bölümlenmelerinin yansıtılması	235
Şekil 58. Harita 149'un (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına k-temsilli kümeleme algoritması ile elde edilen $k=5$ için küme bölümlenmelerinin yansıtılması	235
Şekil 59. Harita 227'nin (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına k-temsilli kümeleme algoritması ile elde edilen $k=5$ için küme bölümlenmelerinin yansıtılması	234
Şekil 60. Harita 11'in TBA anilizi grafiği ve $k = 4$ için k-temsilli küme analizi ile elde edilen küme bölümlenmesinin TBA üzerine yansıtılması	236
Şekil 61. Harita 149'un TBA anilizi grafiği ve $k = 5$ için k-temsilli küme analizi ile elde edilen küme bölümlenmesinin TBA üzerine yansıtılması	236
Şekil 62. Harita 227'nin TBA anilizi grafiği ve $k = 5$ için k-temsilli küme analizi ile elde edilen küme bölümlenmesinin TBA üzerine yansıtılması	237
Şekil 63. Harita 11'de $k=4$ için elde edilen kümelerin varyanslarının homojenliği testi sonuçları	238
Şekil 64. Harita 11'de $k=4$ için elde edilen kümelerin çok değişkenli varyans analizi sonuçları	239
Şekil 65. Harita 11'de $k=4$ için elde edilen kümelerin çok değişkenli varyans analizine ait gruplar arası ikili karşılaştırma sonuçları	239

<b>Şekil 66.</b> Harita 227’de $k=5$ için elde edilen kümelerin varyanslarının homojenliği testi sonuçları	240
<b>Şekil 67.</b> Harita 227’de $k=5$ için elde edilen kümelerin çok değişkenli varyans analizi sonuçları	240
<b>Şekil 68.</b> Harita 227’de $k=5$ için elde edilen küme bölümlerinin çok değişkenli varyans analizine ait gruplar arası ikili karşılaştırma sonuçları	241
<b>Şekil 69.</b> Yığın öğrenme algoritmaları ile eğitilen haritalardan eleme sonrası kalan 58 haritanın nicemleme hatasının nöron sayısına göre değişimi	243
<b>Şekil 70.</b> Harita 2’nin (a) hücrelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe	244
<b>Şekil 71.</b> Harita 132’nin (a) hücrelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe	244
<b>Şekil 72.</b> Harita 95’in (a) hücrelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe	245
<b>Şekil 73.</b> Harita 2 için (a-b) küme içi kareler toplamının küme sayısına göre değişimi ve (c-d) ortalama silüet endeksinin küme sayısına göre değişimi	246
<b>Şekil 74.</b> Harita 132 için (a-b) küme içi kareler toplamının, (c-d) ortalama silüet endeksinin küme sayısına göre değişimi	248
<b>Şekil 75.</b> Harita 95 için (a-b) küme içi kareler toplamının, (c-d) ortalama silüet endeksinin küme sayısına göre değişimi	250
<b>Şekil 76.</b> Harita 2’nin (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına k-temsili kümeleme algoritması ile elde edilen $k=3$ için küme bölümlerinin yansıtılması	251
<b>Şekil 77.</b> Harita 132’nin (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına hiyerarşik kümeleme algoritması ile elde edilen $k=6$ için küme bölümlerinin yansıtılması	251
<b>Şekil 78.</b> Harita 95’in (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına k-temsili kümeleme algoritması ile elde edilen $k=4$ için küme bölümlerinin yansıtılması	252
<b>Şekil 79.</b> Harita 2’nin TBA anilizi grafiği ve $k = 3$ için k-temsili küme analizi ile elde edilen küme bölümlerinin TBA üzerine yansıtılması	253
<b>Şekil 80.</b> Harita 132’nin TBA anilizi grafiği ve $k = 6$ için k-temsili küme analizi ile elde edilen küme bölümlerinin TBA üzerine yansıtılması	253
<b>Şekil 81.</b> Harita 95’in TBA anilizi grafiği ve $k = 4$ için k-temsili küme analizi ile elde edilen küme bölümlerinin TBA üzerine yansıtılması	253
<b>Şekil 82.</b> Harita 132’de $k=6$ için elde edilen küme bölümlerinin varyanslarının homojenliği testi sonuçları	255
<b>Şekil 83.</b> Harita 132’de $k=6$ için elde edilen küme bölümlerinin çok değişkenli varyans analizi sonuçları	255
<b>Şekil 84.</b> Harita 132’de $k=6$ için elde edilen küme bölümlerinin çok değişkenli varyans analizine ait gruplar arası ikili karşılaştırma sonuçları	256
<b>Şekil 85.</b> Harita 95’te $k=4$ için elde edilen küme bölümlerinin varyanslarının homojenliği testi sonuçları	257



<b>Şekil 86.</b> Harita 95’te $k=4$ için elde edilen küme bölümlerinin çok değişkenli varyans analizi sonuçları	257
<b>Şekil 87.</b> Harita 95’te $k=4$ için elde edilen küme bölümlerinin çok değişkenli varyans analizine ait gruplar arası ikili karşılaştırma sonuçları	258
<b>Şekil 88.</b> İkili ve kategorik özniteliklerin küme esaslı sütun grafikleri	269-270
<b>Şekil 89.</b> Demografik-Nümerik özniteliklerin küme esaslı sütun grafikleri	270
<b>Şekil 90.</b> Ölçek-Nümerik özniteliklerin küme esaslı sütun grafikleri	271
<b>Şekil 91.</b> Tüm nümerik özniteliklerin küme esaslı kutu-bıyık grafikleri	272-274
<b>Şekil 92.</b> Karar ağacı karmaşıklık grafiği	277
<b>Şekil 93.</b> Kural çıkarmada kullanılan budanmış karar ağacı	279
<b>Şekil 94.</b> Özniteliklerin (a) TBA faktör yükleri ve (b) karar ağacı önem derecelerinin karşılaştırılması	284
<b>Şekil 95.</b> ÇEO’ye ait yamaç-birikinti grafiği	333
<b>Şekil 96.</b> ÇEO’ye ait faktörlerin üç boyutlu uzayda temsili	334
<b>Şekil 97.</b> DİÖ’ye ait yamaç-birikinti grafiği	335
<b>Şekil 98.</b> DİÖ’ye ait faktörlerin üç boyutlu uzayda temsili	337
<b>Şekil 99.</b> STD ölçeğine ait yamaç-birikinti grafiği	338
<b>Şekil 100.</b> STD ölçeğine ait faktörlerin üç boyutlu uzayda temsili	338
<b>Şekil 101.</b> Çevrimiçi öğrenme algoritması ile eğitilmiş olan 293 haritanın nicemeleme hatasının nöron sayısına göre değişimi	341
<b>Şekil 102.</b> Yığın öğrenme algoritması ile eğitilmiş olan 293 haritanın nicemeleme hatasının nöron sayısına göre değişimi	342
<b>Şekil 103</b> Çevrimiçi öğrenme algoritması ile eğitilmiş olan 15 haritanın iterasyon esaslı yakınsama grafiği	343
<b>Şekil 104.</b> Yığın öğrenme algoritması ile eğitilmiş olan 15 haritanın iterasyon esaslı yakınsama grafiği	344
<b>Şekil 105.</b> Kural çıkarmada kullanılan karar ağacının budanmamış versiyonu	346

## EKLER

Bilgi	Sayfa
<b>Ek 1</b> RStudio cluster. Sim kütüphanesinde yer alan 13 farklı veri normalleştirme yöntemine ait bilgiler	316
<b>Ek 2</b> Soru formu	318
<b>Ek 3</b> Katılımcılar, aileleri ve meskenlerine ait bazı betimsel istatistikler	322
<b>Ek 4</b> Birinci pilot veri seti (200 gözlemlik) için Oblimin eğik döndürme yöntemi ile elde edilen FA sonuçları	323
<b>Ek 5</b> Mahalanobis ve SMM ölçütleri ile temizlenen ikinci pilot veri setine ait FA sonuçları ve Cronbach alfa değerleri	328
<b>Ek 6</b> Ana araştırma veri seti için çok değişkenli normal dağılımın Mardia katsayıları	329
<b>Ek 7</b> Ölçme modellerinin SPSS yazılımı kullanılarak elde edilen güvenilirlik ve geçerlik analizleri sonuçları	330
<b>Ek 8</b> Çevrimiçi öğrenme algoritması ile eğitilmiş olan 293 haritanın nicemeleme hatasının nöron sayısına göre değişimi grafiği	338
<b>Ek 9</b> Yığın öğrenme algoritması ile eğitilmiş olan 293 haritanın nicemeleme hatasının nöron sayısına göre değişimi	339
<b>Ek 10</b> Çevrimiçi öğrenme algoritması ile eğitilmiş olan 15 haritanın iterasyon esaslı yakınsama grafiği	340
<b>Ek 11</b> Yığın öğrenme algoritması ile eğitilmiş olan 15 haritanın iterasyon esaslı yakınsama grafiği	341
<b>Ek 12</b> Karar ağacında kullanılan 227 numaralı haritanın TBA analizinin sonucunda özniteliklerin birinci ve ikinci temel bileşenler üzerindeki yükleri	342
<b>Ek 13</b> Kural çıkarmada kullanılan karar ağacının budanmamış versiyonu	343

*“Yeryüzünün güzelliğini seyre dalanlar,  
yaşam sürdükçe tükenmeyecek güç kaynakları bulurlar.”*

*–Rachel Carson  
Çevre Aktivisti ve Biyolog*

## GİRİŞ

Sürdürülebilirlik, aralarında metabolik bir ilişki olan sosyal, ekonomik ve doğal çevre boyutlarından oluşan bir paradigmadır. Hızlı kentleşme, doğal kaynakların kontrolsüz tüketimi ve çevre kirliliği ile doğan sürdürülebilir kalkınma paradigması, sürdürülebilir enerji yönetimi ve sürdürülebilir kentleşmeyi toplumlar ve gelecek kuşakların refahı için gerekli kılmıştır. Yakın ya da uzak coğrafyalardaki sosyal, ekonomik ve doğal çevrenin, insanlar da dâhil olmak üzere sahip olduğu kaynaklar&süreçler ve atıklar da dâhil olmak üzere ürettiği çıktılar ülkeden ülkeye ve hatta kentten kente değişmektedir. Dolayısı ile her bir kent sürdürülebilirlik çalışmalarında ayrı bir metabolizma olarak ele alınmalıdır.

Çeşitli organizasyonlar bu bağlamda yerel yönetimlere ışık tutacak sistemler tasarlamıştır. Bu sistemlerden bir tanesi olan Avrupa Yeşil Başkenti (AYB), Avrupa Birliği (AB) enerji komisyonu tarafından geliştirilmiştir. Bu sistem kapsamında AB, AYB'yi uygulamaya çalışan kentleri, kentlerin başvuruları doğrultusunda her yıl değerlendirmekte ve en başarılı kente ödül vermektedir. AYB ödülü için aday kentler, 12 farklı AYB parametresi hakkında komisyona sundukları gerçekleştirilmiş ve gerçekleştirilecek olan faaliyet planları dikkate alınarak değerlendirilmektedir. Bu parametreler, iklim değişikliği ile mücadele, iklim değişikliğine uyum, sürdürülebilir ulaşım, sürdürülebilir arazi kullanımı, doğa ve biyo-çeşitlilik, hava kalitesi, gürültü yönetimi, katı atık yönetimi, su ve atık su yönetimi, eko-yenilik, enerji performansı ve bütünleşik çevre yönetimi şeklindedir. Aday kentlerin parametreler için sundukları stratejiler ve faaliyetler üç boyut göz önüne alınarak değerlendirilmektedir. Bu boyutlardan ilki ise kent sakinleri ile iletişim, onların söz konusu parametreler ile ilişkilendirilmesi, bu süreçte gerçekleşen faaliyetlerde ne derecede rol aldıkları, kent sakinlerinin davranış değişikliğinden ötürü kaynaklanan etkililik, sakinlerin bu süreçte neler öğrendiği ve gelecekte yapılması planlanan değişiklikler şeklindedir. Dolayısı ile 12 değerlendirme parametresi nicel türde olsa da her bir parametrenin nitel türde insan boyutu yer almaktadır.<sup>1</sup>

---

<sup>1</sup> Diğer boyutlar finansal işbirlikleri ve kapasite hakkındadır.



söylemindeki gibi bazen geriye gitmek, ilerlemektir.<sup>2</sup>, Teknoloji ve kentleşme ile kendi kendini doğadan koparmış olan insan ve gelecek nesilleri için öze-doğaya dönüş, kurtuluşun tek çaresidir. Çünkü insanın kalbi doğadan uzaklaşırsa katılaşır ve canlı varlıklara olan saygının yok olması çok geçmeden insana olan saygının da yok olmasına neden olur (Dikilen Ayı Luther'dan akt. Louv, 2008)<sup>3</sup>

İnsanın kendisini doğanın kısmen, tamamen ya da hiçbir parçası olarak değerlendirip, çevre dostu davranışta (ÇDD) bulunma(ma) durumu sürdürülebilirlik psikolojisi altında çalışılmaktadır. İnsan-doğa ilişkisi ise ÇDD'yi açıklamaya çalışan en son yaklaşımlardandır. Bu yaklaşım biyofili hipotezine dayanmakta olup, insan-çevre arasındaki duyuşsal (duygusal), bilişsel ve fiziksel ilişki ağını incelemektedir. Hipoteze göre, doğa ile yoğun ilişki içinde olan insanlar, doğanın tüm bileşenlerine ve insanlara karşı daha olumlu bir eğilime sahiptir, empati yapar, saygı duyar, korur ve dolayısıyla davranışlarını olumlu yönde şekillendirir.

İnsan-doğa ilişkisine ek olarak değerler ve tutumlar da davranışları öngörmede (ve değişikliği tetiklemede) başarılıdır. Çünkü değerler, önceliklendirme sistemine sahiptir ve tutumlar değerler ile şekillenmektedir. ÇDD yazınında değerlere bağlı olarak egoistik, altruistik ve biyosferik olmak üzere üç tutum önem arz etmektedir. Bu özelliği ile tutumlar, farklı (ve hatta benzer) sosyo-demografik özelliklere sahip bireyleri ya da toplulukların davranışlarını açıklama ve tanımlamada sıklıkla kullanılmaktadır.

Bu bağlamda, söz konusu tez çalışmasının amacı, AYB markasına sahip olabilmek için Bursa halkının sürdürülebilir tüketim davranışlarını, değerlere bağlı olarak çevresel endişelerini (tutumları) ve doğayla ilişkilerini dikkate alarak ekolojik kimlik profillerini keşfetmektir. Çünkü hem kent markalaşması hem de sürdürülebilirlik çalışmalarının ortak paydası insan olup, her iki çalışmanın da ilk basamağı söz konusu kentte yaşayan vatandaşların analiz edilerek sürece dâhil edilmesidir. Bu aşama Kotler'in kent

---

<sup>2</sup> "When going back makes sense, you are going ahead." Cümlenin tam çevirisi, geri dönmenin mantıklı olduğu zamanlarda/durumlarda, aslında ilerliyorsunuzdur.

<sup>3</sup> Dikilen Ayı Luther "Luther Standing Bear" bir Kızılderili yazardır. Orijinal cümle "Man's heart, away from nature, becomes hard; the Lakota knew that lack of respect for growing-living things soon led to lack of respect for humans too." şeklindedir.

markalaşmasında, kent sakinlerinin katılımı ön şartının kısmen gerçekleştirilmiş olmasını sağlayacaktır.

Bu güdü ile alan yazınında tutumu, insan-doğa ilişkisini ve ÇDD'yi ölçtüğü belirtilen ve Türkçeye uyarlanmış olan üç ölçme aracı veri toplama amacı ile kullanılmıştır. Bu ölçekler Nisbet, Zelenski, Murphy (2009) çalışmasında geliştirilmiş olan ve Çakır, vd. (2015) çalışmasında Türkçeye adapte edilmiş olan doğa ile ilgili duyusallığı, bilişselliği ve deneyimi ölçen Doğayla İlişki Ölçeği (DİÖ), W. P. Schultz (2001) tarafından, Schwartz'ın değer teorisi esas alınarak, geliştirilmiş olan biyosferik, altruistik ve egoistik tutumu ölçen Çevresel Endişe Ölçeği (ÇEÖ) ve Başar (2016) tarafından OECD'nin (Organisation for Economic Co-operation and Development) hane halkı sürdürülebilir tüketim göstergeleri dikkate alınarak geliştirilmiş olan, ÇDD'yi atık oluşturma ve geri dönüşüm, kişisel ulaşım tercihi, evsel enerji talebi, evsel su tüketimi ve çevreye duyarlı gıda tercihi şeklinde çok boyutlu olarak ölçen Hane Halklarının Sürdürülebilir Tüketim Davranışları (STD) ölçeğidir.

Araştırmanın gözlem birimi insan olup, yaklaşımı aşağıdan-yukarıdır. Toplanan verilerin analizi iki aşamada gerçekleştirilmiştir. İlk aşamada doğa ile ilişkinin tutum ve davranış arasındaki aracı etkisine ait yapısal eşitlik modeli (YEM) incelenmiştir. Biyofili hipotezine göre bireyin doğayla ilişkisi arttıkça daha çok ÇDD'de bulunur ve aynı zamanda biyosferik tutuma sahip bireyler daha çok ÇDD sergilerler. Bu bağlamda ilgili yazınında, tutum (çevresel endişe), doğayla ilişki ve ÇDD arasındaki ilişkinin bütünsel olarak ele alınmadığı gözlemlenmiştir. Fakat diğer taraftan biyofilik ya da altruistik tutuma sahip olan bireyler farklı kaygılar dolayısı ile özellikle enerji tüketiminde olumsuz davranışlar sergileyebilmektedir. Örneğin sağlık ve konfor amacı ile gereğinden fazla enerji tüketmek gibi. Başka bir ifade ile alan yazınında teorik olarak olumlu çevresel tutumun, olumlu çevresel davranışlar doğuracağı ifade edilmekte ve fakat bu durumun her koşulda genellenemediği belirtilmektedir. Bu yüzden yapısal model bu aşamada doğrulayıcı yerine keşfedici yaklaşımla incelenmiş ve Kısmi En Küçük Kareler (KEKK) yöntemi ile analiz edilmiştir. Analizler için Smart-PLS yazılımı kullanılmıştır.

İkinci aşamada ise vatandaşların ekolojik kimlik profilleri keşfedilmiştir. Veri seti karma nitelikte olduğu için karma veri kümeleme yöntemlerinden Süper Öz-Örgütlemeli Haritalar (SÖÖH) (Wehrens, Kruisselbrink, 2018a) kullanılmıştır. En uygun küme

bölümlemesinin tespiti için SÖÖH ile elde edilen küme analizi çıktıları, Ward bağlantılı hiyerarşik, k-temsili küme analizi, içsel değerlendirme kriterleri ve parametrik olmayan permütasyonel çok değişkenli varyans analizi (PERMANOVA) yöntemleri ile ayrıca test edilmiştir. Bu sürecin sonunda kesinleştirilmiş olan küme bölümlemeleri, betimsel istatistikler ve sınıflandırma ve regresyon ağaçları (CART) ile tanımlanarak yeşil kimlik profilleri anlamlandırılmıştır.

Kimlik profillerinin tespiti, bu tez çalışmasının devamında gerçekleştirilmesi planlanan çalışmalarda tasarlanacak olan kent markasına ait kimliğin, kent sakinlerinin kimliği ile uyumlu, onu destekleyen ve güçlendiren nitelikte olmasını sağlayacaktır. Böylelikle kent ve kent sakini arasında var olması gerekli olan sembiyotik ilişkinin bağları da şekillendirilmiş olacaktır.

Bu tez çalışmasının gerçekleştirilmesinde veri madenciliği bilgi keşfi süreci benimsenerek uygulanmaya çalışılmıştır. Veri toplamada kullanılan ölçekler Türkiye örnekleminde ilk defa kent sakinleri gibi genel bir örneklem kitlesi üzerinde uygulanmış ve ölçeklerin Türkiye için genellenebilirliği sağlanmıştır. Analizler sonucunda, doğa ile ilişkinin tutumlar ile davranış arasındaki aracı etkisinin ilk defa bu tez çalışması ile incelendiği düşünülmektedir.<sup>4</sup> Bursa merkez ilçelerinde yaşayan vatandaşlar, doğa ile ilişkileri, tutumları, davranışları, enerji tüketimleri ve sosyo-demografik özellikleri dikkate alınarak kümelenebilir. Böylelikle her bir kümenin nitelikleri tespit edilmiş ve kent sakinlerinin ekolojik kimlik profilleri tanımlanmıştır. Ayrıca (sürdürülebilir) kent markalaşması bu çalışma ile alışlagelen turizm bağlamında araştırılmamış, yerine kent sakini bağlamında incelenmiştir. Son olarak kent sakinlerinin sürdürülebilir tüketim davranışları, tutumları ve doğa ile duyuşsal, bilişsel ve deneyimsel ilişkileri dikkate alınarak, vatandaşların (gelecekteki) markalaşma sürecine dâhil edilebilirliği sağlanmıştır.

Tez, giriş ve sonuç kısımları ile dört ana bölümden meydana gelmektedir. İlk bölüm, sürdürülebilirlik kapsamında kentleşme, enerji, kent markalaşması ve ÇDD hakkında yazın taramasını içermektedir. İkinci bölümde, tezin uygulama kısmında

---

<sup>4</sup> Bu kanıya varmak için ilgili yazın taraması yapılmış ve bütünsel olarak modelin testini yapan bir çalışmaya rastlanılmamıştır. Ek olarak ilgili ölçeği geliştirmiş olan ekipten Elisabeth Lisa Nisbet ile irtibata geçilmiştir (çünkü ölçek, Nisbet'in tezinin bir çıktısıdır).



yararlanılan veri madenciliđi (VM) süreci ele alınmaktadır. Üçüncü bölüm, veri analizinde kullanılan temel yöntem SÖÖH'ler ile küme analizi ve veri seti de karma veri olduđu için karma veri kümeleme algoritmaları ve SÖÖH'ler hakkındadır. Son olarak dördüncü bölüm ise uygulama olup, tüm araştırma ve analiz sürecini içermektedir.

## BİRİNCİ BÖLÜM

### (KENT MARKALAŞMASI ve SÜRDÜRÜLEBİLİRLİK İLİŞKİSİ)

#### 1. SÜRDÜRÜLEBİLİRLİK ve KENTLEŞME

Birleşmiş Milletlere göre (BM) 2050 yılı itibari ile dünya nüfusunun %68'inin kentlerde yaşaması beklenmektedir. Günümüz Avrupası için bu oran şimdiden %74'e ulaşmıştır (UN, 2018). Türkiye için ise 2012 yılında %77,3 iken 2018 yılı için %92,3'e ulaşmıştır.<sup>5</sup> C40 Büyük Kentler İklim Liderlik Grubunun 2017 yılı raporuna göre ise dünyadaki 96 büyük kent, dünya nüfusunun 700 milyondan fazlasını barındırmakta, dünya ekonomisinin %25'ini elinde tutmakta ve bu kentlerin kişi başı yıllık ortalama sera gazı salınımı 5,3 tonu bulmaktadır (C40, 2000). BM İnsan Yerleşimleri Programının araştırmalarına göre ise kentler, yeryüzündeki birincil enerji kaynaklarının %70'ini doğrudan tüketmekte ve yeryüzündeki sera gazı salınımının %50-60'luk kısmı bu tüketimden kaynaklanmaktadır. Kent sakinlerinin dolaylı enerji tüketimleri de dikkate alındığında salınım oranı %80'i bulmaktadır (UN HABITAT, 2018). Kentlerin enerji talebindeki artış ve çevre kirliliği, kent nüfusunun artışı, kentlerin ekonomik ve sosyal faaliyetlerinin artışı ve kentlerin yapay çevresinin (gri alt yapı) <sup>6</sup> çarpıklaşmasından kaynaklanmaktadır. (Bibri, Krogstie, 2017; Egger, 2006). Bu neden ile kentlerin hızlı büyümeye adapte olması ve onu yönetmesi gerekmektedir. Fakat oluşturulacak yönetim sistemi sadece iktisadi ve teknolojik nitelikli olmamalıdır. Kendini düzeltebilme yeteneği olan, ekolojik olarak duyarlı, vatandaşların etkin katılımını sağlayan, sosyal ve esnek bir yönetim sistemi şeklinde olmalıdır (Bruntland, 1987). Bu kapsamda sürdürülebilir kalkınma paradigması Şekil 2.'deki gibi görselleştirilerek bu tez çalışmasının motivasyonunu olmuştur. Bu paradigma, yerel yönetimlere insani, çevresel, ekonomik ve teknolojik boyutları içeren çok yönlü ve uzun vadeli bir perspektif sağlamaktadır (Babahan, Güler, 2015).

---

<sup>5</sup> 2012 yılına göre "il ve ilçe merkezleri" ile "belde ve köyler" nüfuslarındaki büyük farklılıkların ana nedeni, 6360 sayılı Yasa uyarınca yapılan idari bölünüş değişiklikleridir. Bu yasa ile büyükşehir niteliğinde olan kentlerin köyleri büyükşehir bağlı mahalleler haline getirilmiştir.

<sup>6</sup> Built environment (Grey infrastructure)



**Şekil 2.** Araştırmanın motivasyonu

Sürdürülebilirlik gelecek nesiller için iyi bir yaşam anlamına gelmektedir. Sürdürülebilirlik, yeryüzündeki ekosistemleri korumak ve kuşaklararası sosyal adaleti sağlamak sureti ile yaşam koşullarının ve refahın iyileştirilmesi kısaca kalkınma olarak tanımlanmaktadır. Her ne kadar “sürdürülebilirlik” günümüzde popüler bir kavram olsa da günümüz toplumları ve yönetimleri kalkınmayı hala iktisadi açıdan ele almakta ve faaliyetlerini bu perspektifte planlayıp gerçekleştirmektedir. Hâlbuki iktisadi kalkınma anlayışı bir önceki yüzyıla aittir. Günümüz için, iktisadi parametrelerin yanı sıra sosyo-kültürel ve çevre ile ilgili konuları da dikkate alan ve insan merkezli bir yapıya sahip olan sürdürülebilir kalkınma benimsenmelidir. Çorakçı (1989)<sup>7</sup>’dan aktaran (Babahan, Güler, 2015)<sup>7</sup>’e göre, BM Dünya Çevre ve Kalkınma Komisyonunun 1987 yılındaki Bruntland Raporu’nda<sup>7</sup> sürdürülebilir kalkınma tanımlanmış, enerji güvenliği, halkın katılımı, yerel yönetimlerin ve sivil toplum kuruluşlarının güçlendirilmesi, kentleşme, çevre sorunları ve çevre yönetimi gibi konular sürdürülebilir kalkınma kapsamına dahil edilmiştir. Söz konusu kapsam genişlemesine paralel olarak BM, sürdürülebilir kalkınmayı ekonomik, sosyal ve çevresel sürdürülebilirlik olarak birbiri ile ilişkili şekilde üç boyut altında incelemeye başlamıştır. Diğer bir deyiş ile sürdürülebilir kalkınma insan ve doğa faktörü dikkate alınarak sürdürülebilir bir hayat tasarlandığı zaman başarılabilir olmaktadır.

Sürdürülebilir kent veya kentleşme kavramı ise ilk defa sürdürülebilir kalkınma ile birlikte kullanılmıştır (Karakurt Tosun, 2009). Günümüzde de BM’nin Agenda 2030 faaliyet planında sürdürülebilir kentler ve toplumlar, sürdürülebilir global kalkınmanın amaçlarından bir tanesi olarak yer almaktadır. İlk tanımı sağlıklı, planlı ve yaşanabilir kentlerdir. Sürdürülebilir kent, European Union (2018)<sup>7</sup>’e göre, tükettiği malzeme ve enerji miktarı ve ürettiği atıkları, çevresel kapasitesini aşmayan kenttir. Kentsel

<sup>7</sup> Agenda 21 faaliyet planı olarak da adlandırılmaktadır. Bu faaliyet planının güncellenmiş versiyonu Agenda 2030’dur.

tüketimin, kentin doğal çevresinin sağlamış olduğu orman, toprak ve su kaynaklarına eşit veya bu kaynaklardan daha az olması ve ayrıca tüketim sonucu oluşan atıkların insanlar ve ekosistemin diğer bileşenleri için çevrenin kaynak sağlama yeteneğini olumsuz etkilememesi gerekmektedir. Kent yönetimleri, hem kaynakları ve atıkları yönetmeli, hem kontrollü bir şekilde yeşil, mavi ve gri altyapı yatırımları yapmalı ve hem de işverenlerin bölge kaynaklarını kullanarak yeşil ürünler üretmesini teşvik etmelidir. AB söz konusu çok yönlü yönetim sisteminde, kent sakinlerinin davranış ve yaşam tarzının önemli bir role sahip olduğunu belirtmektedir.

Sürdürülebilir kentleşmede amaç, toplum, kent ve doğa arasında yeni ve dinamik bir denge kurmaktır. Katılım, eğitim ve yenilik sürdürülebilir toplumun itici güçleridir. kentler ancak bu şekilde yapılandırılır ve yönetilirse insanlığın yaşam alanı doğal döngünün içindeki yerini tekrar kazanabilir (Rogers, 1997) ve gelecek nesillere daha sosyal, ekonomik ve ekolojik açıdan dengeli bir dünya bırakılabilir. Bu nedenle, çeşitli kuruluşlar sürdürülebilir kentler ile ilgili değerlendirme ve derecelendirme sistemleri geliştirmiştir. Bu sistemler, sosyal, ekonomik ve ekolojik çevresini iyileştiren kentlerin bilinirliğini arttırarak diğer kentlere yol gösteren bir kılavuz niteliğindedir. Bu sistemlerden bir veya birkaçının kentlere uyarlanarak dönüşümün sağlanması, söz konusu kentlere özellikle ekolojik ve sosyal açıdan değer katacak olan bir süreçtir.

Bu sistemlerin ortak özelliği, sistem parametrelerinin insan, çevre, enerji ve ekonomi kapsamında oluşturulmuş olmasıdır. Bunlardan bazıları, City Blueprints, Urban Metabolism Framework, European Green Capital Award, Reference Framework for Sustainable Cities, STAR Community Rating System, Urban Sustainability Indicators, Eurostat Sustainable Development Indicators, Indicators for Sustainability, European Green Leaf Award, Climate + Development Program, SDEWES Index, LEED-ND ve NABERS.

Sürdürülebilir kentler, kent planlamasında kentsel ekoloji, ekonomi ve sosyoloji dikkate alındığı ve vatandaşlar motive edilerek sürece dahil edildiği zaman yaratılabilir. Çünkü ekolojik ve sosyal çözümler birbirlerini güçlendirir ve daha sağlıklı, daha canlı, daha açık fikirli kentler kurmayı sağlar. İnsanın yarattığı çevre krizi ise küresel (makro) ölçekte bir problemdir. Fakat kriz, her bir kentin ve vatandaşların bakış açısı ile ele alındığı zaman yönetilebilir ve problemlere çözüm bulunabilir. Çünkü çevre ile ilgili

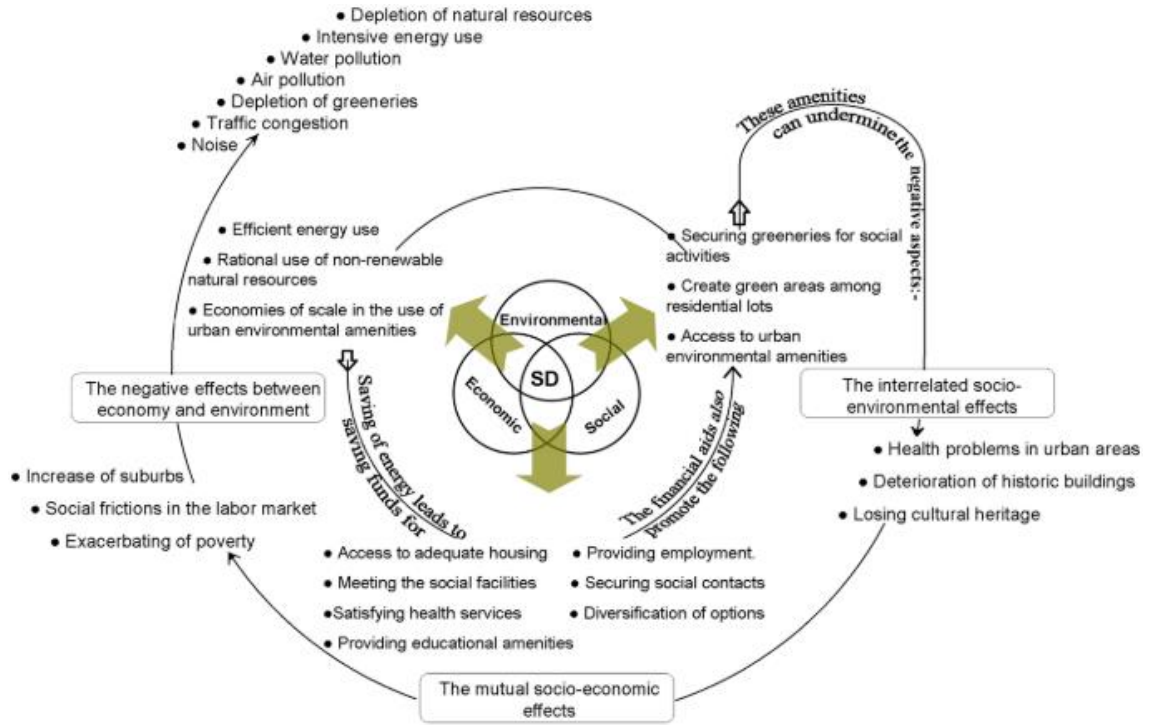
problemler sosyal sorunlar ile ilişkilidir. Çevrenin iyileştirilmesine yönelik politikalar vatandaşların sosyal yaşamını da iyileştirmektedir (Rogers, 1997). Dolayısı ile, sürdürülebilir kentleşme için kentin metabolizmasının anlaşılması gerekmektedir.

Canlılar hayatta kalmak için gıda, su ve oksijeni enerjiye dönüştürerek ihtiyaçlarını giderirler veya bir kısmını daha sonra kullanmak için depolarlar. Dönüşüm ve kullanım süreçlerinin ise karbondioksit, ter ve dışkı gibi çeşitli atıkları oluşur. Aynı şekilde, kentler de hayatta kalmak için vatandaşlarına bakım hizmeti verir, barınak sağlar, mal ve hizmet üretir, büyür ve meydana gelen atıkları ortadan kaldırır. Bu faaliyetlerin tümü için ise enerji, malzeme, su ve gıdaya ihtiyaç duyar. Metabolizmanın varlığını sürdürebilmesi beyin, organlar ve enzimler arasında işbirliği gerektirmektedir. Kentsel metabolizmanın varlığı ise kentin yönetim politikaları, altyapısı ve vatandaşlarının eşgüdümü/uyumlu olması ile sağlanır (Chrysoulakis, de Castro ve Moors, 2014'ten aktaran European Union, 2018).

Diğer taraftan, dünyanın birçok ülkesinde insanlar kırsaldan kente göç etmektedir. Bu yüzden daha büyük kentler daha fazla gıdaya, suya ve enerjiye ihtiyaç duymakta ve bu durum kentlerin daha fazla atık (kirli gaz, katı atık ve atık su) üretmesine neden olmaktadır (Kennedy, Cuddihy, Engel-yan, 2007). Kentlerin varlığı ekosisteme bağlıdır. Fakat diğer taraftan kentler, kaynak ve arazi kullanımı ve ürettiği atıklardan kaynaklanan kirlilik nedeni ile bağlı olduğu ekosistemin varlığını tehdit etmektedir. Bu döngü, yerel ve küresel ölçekte çevreyi değiştirerek, doğanın dengesini ve iklimi olumsuz etkilemektedir (Grimm, et al., 2008). Bu etkileşim ise, toplumların ekonomik faaliyetlerini ve halk sağlığını etkileyen yeni sorunlara yol açmaktadır. Nüfus yoğunluğu ve sosyo-ekonomik dengesizlik daha da artmakta ve altyapı sorunları ortaya çıkmaktadır. Bu nedenle, sürdürülebilir kentler bağlamında kentsel metabolizmanın anlaşılması önem arz etmektedir. Karar vericiler, kentsel metabolizmanın kent dokusuna sosyal ve fiziksel olarak pratikte nasıl uygulanabileceğini (vatandaşlar ve iş çevresi ile) işbirliği yaparak anlamak ve planlamak zorundadır (Karvounis, 2015).

Hassan, Lee (2015)'in sürdürülebilirlik ve kentleşme alanındaki yazın taraması çalışmasında ise kentsel metabolizma doğal, sosyal ve ekonomik çevrelerden oluşan bir sistem olarak nitelendirilmiş ve Şekil 3.'teki gibi görselleştirilmiştir. Yazarlara göre çevreler arası etkileşimin olumlu sonuçlarının toplamı (iç çember), olumsuzların

toplamından (dış çember) büyük olduğu zaman bir kentin sürdürülebilir olduğu ifade edilmektedir.<sup>8</sup> Dominski (1993) sürdürülebilir kentlerin ütopyik bir proje olmadığını fakat sürdürülebilir kentleşme için gerekli olan dönüşümün uzun vadede gerçekleşebileceğini belirtmiş ve dönüşümü üç aşama ile seviyelendirmiştir. “3 R” olarak adlandırılan bu yaklaşım: Azalt “Reduce”, Yeniden Kullan “Reuse” ve Geri Dönüştür “Recycle” şeklindedir. Yazar dönüşümün başarısını, bir parçası olduğumuz sistemi (metabolizmayı) ve ona olan ihtiyacımızı anlamamıza ve dönüşüm sürecine istekli olmamıza bağlamıştır.



Şekil 3. Kenti oluşturan doğal, sosyal ve ekonomik çevreler arasındaki ilişkiden doğan olumlu ve olumsuz etkiler (Hassan, Lee, 2015)

AB'nin konu hakkındaki genel kanısı da alan yazın ile paralel olarak, sürdürülebilirliğin kentin ancak sosyal, ekonomik, çevresel ve yönetsel bileşenleri eşgüdümlü çalıştığı zaman başarılabilir olduğu yönündedir. Ayrıca, kentlerin doğal kaynak tüketimi ve atık üretimi ile alakalı ayak izlerini azaltmaları ve arazi kullanım verimliliğini (özellikle greyfield ve brownfield arazilerinin yeniden kullanımı) arttırmaları çevresel etkinin en aza indirilmesi açısından önemlidir. İnsanlar ve çevreleri arasında sürdürülebilir dinamik bir ilişki kurmak (ve çevre dostu işletmelerin teşvik edilmesini

<sup>8</sup> Zira bu tespit, bu tez çalışmasının Giriş bölümünde sunulan AB'nin sürdürülebilir kent tanımı ile örtüşmektedir.

sağlamak) hem uzun vadede yaşam kalitesini hem de ekonomik refahı sağlamak için gerek şart olarak belirtilmektedir. Çünkü ekonomik verimlilik, eğitim, sağlık, güvenlik, gıda, su, ulaşım, temiz hava ve elektriğe kolay erişme ihtiyacı olan sağlıklı ve mutlu vatandaşlara bağlıdır (European Union, 2018).

Richard Rogers, yüksek teknoloji mimari, sürdürülebilir binalar ve kentleşmenin öncülerindendir. Rogers'a göre çevresel açıdan sürdürülebilir ve sosyal açıdan eşit kentler yaratabilmek için toplumun modern teknoloji ve iletişimden faydalanması ve aynı zamanda vatandaşları sürece dâhil etmesi gerekmektedir. Amaç, çevreye zarar veren ve insanları birbirine yabancılaştıran kentler yerine hem çevreyi hem de toplumu birleştiren ve besleyen kentler inşa etmek olmalıdır (Rogers, 1997).

Rogers kentsel sürdürülebilirliğin yoğun kent “dense city” modeli<sup>9</sup> ile sağlanabileceğini belirtmektedir. Yoğun kentler entegre planlama yolu ile enerji verimliliğini artırarak daha az kaynak tüketmeyi, daha az kirlilik üretmeyi ve kentleşmenin kırsal alana doğru büyümesini önleyerek, insanların fiziksel ve sosyal açıdan yakın ilişki içinde olmasını sağlar. Fizik olarak en geniş alanlı bir yapılanmaya sahip olmadığı için kent içi ulaşım erişilebilir olup şahsi araç kullanımını en aza indirger. Böylelikle ulaşım için kullanılan enerji azalır, daha az sayıda şahsi araç, daha az trafik ve daha iyi hava kalitesi elde edilmiş olur. Ayrıca kent sakinlerini bisiklet vb araçları kullanmaya ya da yürümeye teşvik eder. Daha iyi hava kalitesi ise klima kullanımını azaltabilir. Daha az yol ve daha çok yeşil alana sahip bu kentler yaz aylarında doğal gölgelendirmeye sahiptir ve bu yüzden kent merkezinin sıcaklığı ortalama sıcaklığa biraz daha yakındır.<sup>10</sup> Çünkü kentsel peyzaj düzenlemeleri kentlerin ısı akışını azaltmakta ve yağmuru emerek kentsel yağışların olumsuz etkilerini bertaraf edebilmektedir. Ek olarak

---

<sup>9</sup> Rogers'ın Dense City modeline benzer olarak alan yazınında Compact City modeli de yer almaktadır. Sürdürülebilir kentleşmede bu modellerden sonra Ubiquitous Eco-City ve Zero-Carbon Eco-City modelleri geliştirilmiş olup bu modeller, sürdürülebilirliğin sosyal ve ekonomik adalet kavramları ile örtüşmemektedir (Hassan, Lee, 2015).

<sup>10</sup> Bir kent merkezinin sıcaklığı o kentin ortalama sıcaklığının 1-2°C üstünde seyreder. Bu durum insan faaliyetleri dolayısıyla oluşan ısı akışından “anthropogenic heat flux” meydana gelir ve termodinamiğin ikinci yasası ile bilimsel olarak açıklanmaktadır. Söz konusu ısınma kent merkezlerinde ısı adalarının oluşmasına neden olmakta ve kent klimatolojisi alanı kapsamında çalışıla gelmektedir. Daha fazla bilgi için Flanner (2009) ve C.S.B., vd. (2015) çalışmalarına başvurulabilir.

kentin gürültü seviyesini düşürmekte, kent sakinlerine psikolojik açıdan olumlu katkılar sağlamakta ve kentsel yaban hayatı desteklemektedir (Rogers, 1997).

Sürdürülebilir bir kent inşa etmek, toplumun fiziksel, sosyal ve ekonomik ihtiyaçlarını oluşturan ve onları diğer/üst çevreler ile ilişkilendiren tüm faktörleri göz önünde bulunduran bütüncül bir planlama gerektirir. Bu tür bir planlama, nüfus, enerji, su, ulaşım, topografya, istihdam ve en önemlisi yerel teknoloji ve kültürün analizini gerektirir. Rogers'a göre sürdürülebilir bir kent (Rogers, 1997).

- Adalet, gıda, barınak, eğitim, sağlık ve umudun adil bir şekilde dağıldığı ve vatandaşların da yerel yönetime katıldığı adil bir kenttir.
- Sanat, mimarlık ve peyzajın hayal gücünü tetikleyen ve insan ruhunu harekete geçiren güzel bir kenttir.
- Açık fikirliliğin ve deneyimin insan kaynaklarının potansiyelini tam olarak harekete geçirdiği ve değişime hızlı bir şekilde yanıt verdiği yaratıcı bir kenttir.
- Yeşil ve yerleşik alanlar arasında bir dengeye ve güvenli ve verimli altyapı ve binalara sahip olan ve çevresel etkisi az olan ekolojik bir kenttir.
- Hem yüz yüze hem de elektronik veri alışverişine olanak sağlayan erişilebilir ve hareketli bir kenttir.
- Mahalleleri ve toplulukları bütünleştirerek insanları birbirine yakınlaştıran, kırsal bölgeyi koruyan, yoğun ve çok merkezli bir kenttir.
- Şehir hayatını besleyen, canlandıran ve ilham veren çeşitli etkinliklerin gerçekleştirildiği farklı bir kenttir.

## 2. SÜRDÜRÜLEBİLİRLİK, ENERJİ ve TÜKETİM

Enerji alanında sürdürülebilirlik ise Dünya Enerji Konseyinin (DEK) belirttiği üzere enerjinin yönetimi, verimli kullanımı, talebinin kontrol altına alınması, yeryüzünde adaletli olarak dağıtılması ve karbon salınımının azaltılması gibi birçok faaliyeti kapsamaktadır. Bu faaliyetlerin başarı ile yürütülmesi ve sonuçlandırılmasında insan davranışlarının değişimi önemli rol oynamaktadır (Davoudi, Dilley, Crawford, 2013; European Commission Directorate-General for Energy, 2019; Frederiks, Stenner, Hobman, 2015a; Steg, Vlek, 2009; Oskamp, 2000). Çünkü çevre problemlerinin çoğu ve enerji tüketimi insan davranışlarının bir sonucudur ve çözümü, yine insan davranışlarına



bağlıdır (Oskamp, 2000). Fakat davranış değişimini ve ÇDD'yi<sup>11</sup>, tetikleyecek tek tip bir çözüm önerisi veya (Frederiks, Stenner, Hobman, 2015b) sihirli bir formül bulunmamaktadır. Örneğin, benzer sosyo-demografik özelliklere sahip bireyler, farklı tüketim davranışları sergileyebilmekte (Druckman, Jackson, 2008) ve dolayısı ile enerji tüketim miktarı ülkeden ülkeye, kentten kente ve hatta aynı binada yaşayan haneler arasında bile farklılaşabilmektedir (Xie, Ouyang, Gao, 2016). Bu yüzden, sürdürülebilir enerji yönetimi küresel, bölgesel veya ulusal düzeyden, insan odaklı yerel düzeye (kent) indirgenir ise başarılıdır.

Sürdürülebilir kentler için sürdürülebilir enerji yönetimi hayati önem taşımaktadır çünkü doğal kaynaklar sınırlı olmasına rağmen bu kaynaklar kullanılarak elde edilen birincil ve ikincil enerji kaynaklarına olan talep gün geçtikçe artmaktadır (Becker, et al., 1981) ve talep artışı, nüfus artışından çok daha hızlı gerçekleşmektedir (Egger, 2006). Dünya nüfusu 1860-1991 yılları arasında dört kattan fazla artmış olmasına rağmen enerji talebi yıllık 1 milyar megavat saatten 93 milyar megavat saate çıkarak tam 93 kat artmıştır. DEK'in yayınladığı Dünya Enerji Senaryoları raporundaki senaryolara göre, 2050 yılında dünyanın toplam enerji ihtiyacının, 2013 yılı dünya enerji tüketiminden %27 veya %61 daha fazla olacağı öngörülmüştür (The European Commission Directorate General For Energy, 2013). BM İnsan Yerleşimleri Programına göre ise kentlerin faaliyetlerini yürütebilmesi için kesintisiz bir şekilde enerjiye ihtiyaç duyulmaktadır. Dünya çapında tüketilen birincil enerjinin %75'i kentler tarafından tüketilmekte ve bu tüketim ise dünyadaki toplam sera gazlarının %50-60'ına tekabül etmektedir. Hanelerin enerji tüketimleri sonucunda oluşan sera gazı salınımları da dikkate alındığında bu oran %80'i bulmaktadır (UN HABITAT, 2018). C40'ın raporuna göre ise kentlerdeki ısınma ve serinleme amaçlı enerji tüketiminden kaynaklanan sera gazı emisyonu, toplam kentsel emisyonun %40'ına tekabül etmektedir (Hidalgo, 2017). Örneğin yakın coğrafyamızdaki elektrik tüketimi incelendiğinde AB üye ülkelerinde 1990-2009 yılları arasında elektrik tüketiminin toplam enerji tüketimi içindeki payı %21,0 a yükselirken, AB'ye üye olmayan ülkeler için bu oran %68,3 olarak tespit edilmiştir. AB'ye üye olmayan ülkelerin

---

<sup>11</sup> Terimin alan yazınında çeşitli versiyonları bulunmaktadır: Çevreci davranış, sürdürülebilir davranış, yeşil davranış, çevreye duyarlı davranış, çevresel davranış, ekolojik davranış, koruma (conservation) davranışı

tüketimindeki hızlı artışın nedeni olarak ise Türkiye'nin söz konusu yıllar arasında daha modernize bir ekonomiye geçmesi olarak açıklanmıştır (European Environmental Agency, 2013). Söz konusu bulgu, Çin'in 1990-2014 yılları arasındaki ekonomik büyümesiyle birlikte elektrik tüketiminin artması durumu ile benzeşmektedir (Khanna, Guo, Zheng, 2016; Xie, Ouyang, Gao, 2016). Mesken tipi enerji talebindeki hızlı artış ilgili yazında, toplum refahına bağlı olarak mesken alanının büyümesi, tek başına yaşayan birey sayısındaki artıştan dolayı mesken başına düşen hane halkı sayısının azalması (Hiller, 2014) ve yine refahın artışına bağlı olarak tüketicinin satın alma gücünün ve dolayısı ile enerji talebinin artması şeklinde açıklanmıştır (Becker, vd., 1981; Khanna, Guo, Zheng, 2016). Dolayısı ile insanlık, tarih öncesi ilkel hayata geri dönmediği ve ülkemiz veya yakın coğrafya veya dünya ekonomisi kötüye gitmediği sürece, enerji talebinin nüfus ve refah ile birlikte günden güne artacağı olasıdır. Özellikle bu durum gelişmekte olan ülkeler için geçerlidir (Tiwari, 2000).

Dünya, hanelerin enerji tüketimleri ile meydana gelen sera gazı salınımını azaltma amaçlı politikalar oluşturmayı ve uygulamayı tartışmaktadır (İngiltere örneği için (Huebner, et al., 2015)). Ülkemiz ekonomisini, kaynak kıtlığını ve ekolojisini dikkate alınca, ülkemizde de mesken tipi enerji talebinin yönetilmesi son derece önem teşkil etmektedir. AB bu kapsamda 90'lı yıllar itibariyle çeşitli aksiyonlar geliştirmeye başlamıştır. Örneğin AB enerji politikaları, DEK'in yayınladığı üçlü açmaz – trilemma – (enerji güvenliği, enerji eşitliği ve çevresel sürdürülebilirlik) ile paralel olarak üç amaç arasında denge kuracak şekilde oluşturulmuştur (The European Commission Directorate General For Energy, 2013). Söz konusu amaçlar rekabet edebilirlik, enerji arzının güvenliği ve sürdürülebilir kalkınma temelinde çevrenin korunması şeklindedir. Gerek dünya gerek ise ülkemizin geleceği için bahsi geçen amaçlar doğrultusunda talebin olumlu şekilde yönlendirilmesi kaçınılmaz bir gerekliliktir. Talebi yönetebilmek için ise etkin politikalar oluşturulmalı ve bunun için ise talebi oluşturan birimlerin (hanelerin) kapsamlı bir şekilde mikro seviyede analiz edilmesi gerekmektedir.

Mesken tipi enerji talebinde hanelerin yaşam standartlarını düşürmeden enerji tüketimlerinin azaltılması sürdürülebilir kentler için önem arz etmektedir. Alan yazına göre enerji tüketimini en çok etkileyen faktörler sosyo-demografik faktörlerdir. Fakat günümüz çalışmaları incelendiğinde, benzer sosyo-demografik özelliklere sahip

hanelerin enerji taleplerinde farklılıklar olduğu ve bu farklılıkların enerji tüketim davranışlarından kaynaklandığı (Peschiera, Taylor, 2012), dolayısı ile talep yönetiminde talebi oluşturan tüketici alışkanlıklarının/davranışlarının analiz edilerek değiştirilmesi gerektiği belirtilmektedir (Chicco, et al., 2004; Räsänen, Ruuskanen, Kolehmainen, 2008). Diğer bir deyiş ile davranış değişimi için söz konusu davranışları meydana getiren alt kırımların tespiti ve ölçülmesi önem teşkil etmektedir (Frederiks, Stenner, Hobman, 2015b). Aynen, ünlü yönetim bilimci Peter F. Drucker'ın da belirttiği üzere “Ölçmezsen, yönetemezsin!”<sup>12</sup>

Tüketici davranışları; bir ürünü/hizmeti satın alma, kullanma ve imha etme gibi bütün faaliyetleri ve aynı zamanda bu faaliyetlere neden olan, belirleyen veya sonucu olan duygusal, bilişsel ve davranışsal tepkileri inceleyen bir alandır (Kardes, Cronley, Cline, 2011). Bir davranışın oluşumu ise bireylerin algısı, öğrenmesi, motivasyonu, tutumu, benlik ve kişiliği, değerleri ve kültürüne bağlı olarak uzun vadede şekillenmektedir (Koç, 2016). Fakat enerji tüketimi, direkt bir davranış olmayıp tüketicilerin hane içindeki diğer davranışlarının bir sonucu olarak tanımlanmaktadır (Van Raaij, Verhallen, 1983). Bireyin veya hane halkının enerji tüketim davranışları ise tutumlar, değerler, normlar veya alışkanlıklar gibi psikolojik faktörler ve mesken özellikleri veya hanedeki elektrikli alet sayısı gibi fiziksel faktörlerden (Abrahamse, Steg, 2009; Grantham, 2010) ve sosyo-kültürel, ekonomik, politik, yasal çevre vb. gibi durumsal/dışsal faktörlerden etkilenmektedir. Dolayısıyla söz konusu karmaşık yapıdan ötürü, enerji tüketim davranışını direkt tespit etmek veya ölçmek zorlaşmaktadır.

Enerji tüketim davranışını ve bu davranışların tüketime etkisini tespit etmeyi zorlaştıran bir diğer unsur ise bireylerin (tüketicilerin) tutumları ile davranışlarının birbiriyle tutarsız olmasıdır. Söz konusu tutarsızlık ilgili yazınında tutum-davranış boşluğu “attitude-behavior gap/value-action gap” olarak tanımlanır. Faiers, Cook, Neame (2007)'e göre tutum, bireylerin bir nesneye (olaya, kavrama, kişiye, kuruma vb. gibi) karşı sahip olduğu değerlendirici düşünceleri ve davranışlarıdır “evaluative views and behaviors”. Örneğin tüketicinin olaylar karşısında çevreye duyarlı olduğunu düşünmesi ve bunu ifade ederek olumlu bir tutum sergilemesi fakat bu tutumu davranışlarına (çeşitli

---

<sup>12</sup> “If you can't measure it, you can't manage it”

nedenlerle) yansıtması gibi (Caruana, Carrington, Chatzidakis, 2016; Frederiks, Stenner, Hobman, 2015a). İngiltere’de gerçekleştirilen bir tüketici araştırmasının sonuçları da bu boşluğun bir kanıtı olarak sunulmuştur (Young, vd., 2010). Department for Environment, Food and Rural Affairs (2011) raporuna göre ise tüketicilerin %30’u kendilerini çevreye duyarlı olarak tanımlamış ama bu duyarlılığı tüketim davranışlarına yansıtamadıklarını belirtmiştir.

Tüketici davranışlarının değişimi ise alışkanlıkların değişimi ile gerçekleştirilebilmektedir. Alışkanlıklardaki köklü değişimler, uzun vadeli bir dizi uygulamalar bütünü ile gerçekleştirilmektedir. Örneğin, eğitimler ve sosyal kampanyalar düzenlemek, denetimler yapmak, hanenin anlık enerji tüketim bilgilerini hane ile paylaşmak, aylık hedefler koymak gibi uygulamaların tüketici alışkanlıklarında değişim yaratabileceği belirtilmektedir. Bu ve benzeri uygulamalar, manevi baskı kullanılarak davranışların değiştirilmesi kapsamında değerlendirilmektedir (Tao Zhang, Siebers, Aickelin, 2012).

### **3. KENT MARKALAŞMASI**

İlgüner, Asplund (2011) markayı planlı bir pazarlama çabası süreci sonucunda erişilen bir kazanç veya ödül olarak tanımlamaktadır. Marka bir logo, slogan ya da broşürden öte aslında bir değerler bütünüdür. Dolayısı ile bir kent markası da bir kentin fiziki çevresinden ziyade kentin topyekün mücadelesi olarak tanımlanmaktadır. Gerçekçi, uygulanabilir, birleştirici, güç verici ve stratejik olmalıdır.

Kent markalaşması, ilgili yazında mekan pazarlama, kent pazarlama ya da destinasyon pazarlama olarak da ifade edilmekte olup, hepsinin amacı birbiri ile benzeşmektedir, söz konusu kentin çekiciliğini/cazibesini arttırarak (Busch, Anderberg, 2015), onu diğer kentlerden farklılaştırmaktır (Ergan, Akyol, Kucukaltan, 2014). Günümüz kent sakinleri ve işletmeler kent yönetimlerinden temel ihtiyaçların karşılanmasına ek olarak kentin daha fazla iş olanağı sunmasını ve refahın arttırılmasını talep etmektedir. Kent yönetimleri bu talepleri yerine getirebilmek için hem kendi ülkelerinde hem de başka ülkelerde bulunan kentlerle önemli bir yarış içine girmektedir. Bu bağlamda kentlerin daha fazla fırsat yaratabilmesi ve yeni yatırımcıları, işletmeleri, insanları (sakinleri) ve ziyaretçileri (turistleri) cezbedebilmesi için çeşitli çekici/cezbedici unsurlara sahip olması gerekmektedir. Bir kenti cezbedici yapan unsurlar ise o kentin

sadece fiziki yapısı, lokasyonu, iklimi ya da doğal güzellikleri değil, ayrıca içinde barındırdığı toplumun iradesi, yetkinlikleri, enerjisi, değerleri ve yapısı dikkate alınarak değerlendirilmektedir. Dolayısı ile başarılı bir markalaşma için kenti rakiplerinden ayırt eden veya onu ön plana çıkartan cezbedici unsurları tanımlanmalı ve rekabet avantajları etkili bir şekilde müşterilerine iletilmelidir. Bu süreç kolay olmamakla birlikte imkânsız değildir. Zira Kotler'e göre başarılı bir kent markasının oluşturulması 10 ile 15 yıl arasında bir zaman alabilmektedir (Kotler, vd., 1999).

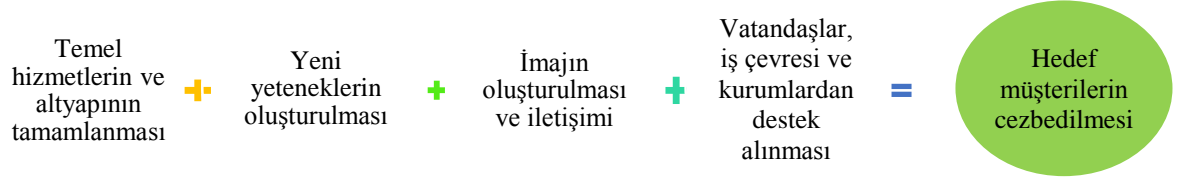
Greenberg (2008)'den aktaran (Eshuis, Klijn, Braun, 2014)'e göre mekân pazarlama/markalaşma faaliyetleri yalnızca yeni yatırımcıları, ziyaretçileri ya da vatandaşları cezbetme amaçlı değil aynı zamanda kentlerin geliştirilmesi, paydaşlar arasında işbirliğinin sağlanması ve yerel sorunlara çözüm bulunması (Busch, Anderberg, 2015) amaçlı olarak da kullanılmaktadır. Bu faaliyetler genellikle sadece yerel yönetimlerin dahil olduğu ve bazı istisnai durumlarda iş çevresinin de katıldığı yukarıdan-aşağı yaklaşımla “top-down manner” gerçekleştirilmektedir (Kotler, vd., 1999; Moilanen, Rainisto, 2009; Järvisalo, 2012; Eshuis, Klijn, Braun, 2014). Hâlbuki söz konusu faaliyetler, o kentte yaşayan vatandaşları doğrudan etkilemektedir çünkü bu faaliyetler kent ile ilgili yaşam ve iş olanaklarının geliştirilmesini kapsamakta (Eshuis, Klijn, Braun, 2014) ve kentin yapısal çevresini ve sistemini değiştirmektedir. Vatandaşlar ise bu değişen çevrede yaşamlarını sürdürmektedir (Moilanen, Rainisto, 2009; Järvisalo, 2012). Sonuç olarak mekân pazarlama bir kentin tüm bileşenlerini etkileyen bir yönetim sürecidir “governance process” (Eshuis, Klijn, Braun, 2014) ve başarılı bir marka, tüm paydaşların katılımı ile yaratılabilmektedir. Çünkü insan çalışır, üretir, kazanır ve vergisini öder; vergiler ile kent beslenir, canlanır ve hizmet sunar; insanlar bu hizmetlerden faydalanır, mutlu olur, sosyalleşir, daha çok çalışır, harcar, vergi öder ve kente daha çok insan ve yatırımcı çeker. Döngünün sürdürülebilirliği insana bağlıdır.

İlgili uygulamalar incelendiğinde, kent markalaşması kapsamında genellikle logo ve slogan oluşturma ya da reklam kampanyaları gibi tanıtıcı (promosyonel) faaliyetlerin gerçekleştirildiği, diğer bir deyiş ile faaliyetlerin markalaşmanın sadece görsel elementleri üzerine yoğunlaştığı görülmektedir (Kotler, vd., 1999; Virgo, de Chernatony, 2006; Kavaratzis, 2009). Bu yaklaşım, markalaşma faaliyetlerinin başarısızlık ile sonuçlanmasına neden olmaktadır. Çünkü tanıtıcı faaliyetler eğer pazarlanan ürün bir

kent/mekan ise tüketicilerin o kent hakkındaki düşüncelerini bir nebze olsa bile değiştiremez. Mekân tüketicilerinin düşünceleri ancak ve ancak mekânın nitelikleri, durumu ve tüketicilere sunduğu fırsatlar iyileştiği zaman değişebilmektedir. Farklı bir ifade ile tanıtıcı faaliyetler tüketici zihnindeki kötü bir mekânı iyi yapmamaktadır (Anholt, 2006). Dolayısı ile kent markalaşmasındaki kilit nokta, markayı kentin kültürünü, değerlerini ve gerçeklerini dikkate alarak oluşturmak ve markanın tüketicilere sahte vaatler sunmamasını sağlamaktır (Crockett, Wood, 2004). Ayrıca, kentlerin sadece doğal güzelliklerine ve doğal kaynaklarına odaklanan reklam harcamaları yaparak da bir kent markasının oluşturulması imkânsızdır. Mekanlar ya da kentler de ürünler ve hizmetler gibi sofistike bir şekilde pazarlanmalıdır (Kotler, vd., 1999). Hatta Kavartzis (2009) ve Marin-Aguilar, Vila-López (2014)'e göre kent markalaşması, kurumsal markalaşma çalışmaları ile yüksek derecede benzerlik göstermektedir. Çünkü her ikisi de çok sayıda paydaşa hitap etmektedir, sosyal sorumluluğu da dikkate almalıdır ve karmaşık bir yapıya sahiptir. Aslında bu yüzden hem kurumsal bir markanın oluşumu hem de bir kent markasının oluşumu Kotler'in de ifade ettiği üzere uzun bir sürece sahiptir (Anholt, 2002). Booth (2005) ve Manzo (2005)'ten aktaran Trueman, Cornelius, Killingbeck-Widdup (2007)'ye göre ise kent markaları kurumsal markalardan daha da karmaşık bir yapıya sahiptir.

Kotler, mekân pazarlama sürecini Şekil 4.'teki gibi dört temel aşama ile tanımlamıştır. İlk aşama vatandaşların, iş çevresinin ve ziyaretçilerin ihtiyaçlarını karşılayacak ve onları tatmin edecek şekilde temel hizmetleri sağlanmak ve kent altyapısının inşa etmek/tamamlamaktır. İkinci aşama, potansiyel vatandaşlar ve iş çevresini kente çekmek için var olan cezbedici ürün ya da hizmetleri geliştirmek veya yenilerini tasarlayıp sunmaktır. Üçüncü aşama, kentin niteliklerinin ve kentte yaşamının sağlayacağı faydaların güçlü bir imaj ve iletişim programı ile potansiyel vatandaşlara, iş çevresine ve ziyaretçilere aktarmak. Son aşama ise kenti yeni yatırımlara, şirketlere ve ziyaretçilere daha misafirperver ve cezbedici hale getirmek için vatandaşlardan, liderlerden ve kurumlardan destek almaktır. Kotler, tüm sürecin başarılı bir şekilde uygulanması ile sürdürülebilir bir toplum yaratılarak kentin hedef müşterileri (yeni vatandaşlar, iş dünyası&yabancı yatırımcılar ve turistler) cezbedebileceğini belirtmiştir.

Fakat kent sakinlerinin katılımı ile oluşacak vatandaş-sektör-yerel yönetim işbirliğini tüm süreç için bir ön şart olarak koşmuştur (Kotler, et al., 1999).



Şekil 4. Kotler'in mekân pazarlama süreci

Bir mekânın imajı, insanların orası ile ilgili sahip olduğu inanç, fikir ve izlenimlerin bütünüdür. Orada yaşayan vatandaşların, işletmelerin ve ziyaretçilerin oraya cevap verme şeklini tanımlar ve insanların nasıl yaşadıklarını ve kentin ne için bilindiğini yansıtır. Dolayısı ile imajın misyonu bir kenti kişiselleştirmek, diğerlerinden ayırmak ve ona kendi kimliğini vermektir. Fakat bir mekânın imajı insanlar arası farklılaşabilmektedir. Aynı mekân farklı insanlar için farklı anlamlara sahip olabilmektedir. Bu yüzden mekânlara ait imajlar iyi yönetilmelidir. Bunun için cezbedilmesi arzu edilen hedef kitle<sup>13</sup> doğru bir şekilde tanımlanmalı, bölümlenmeli (Kotler, vd., 1999) ve imaj kapsamında oluşturulan ayırt edici özellikler hedef kitleye iletilebilir, hedef kitlenin beklentilerini karşılayabilir, sürdürülebilir, inanılır ve özgün nitelikte olmalıdır (Morgan, Pritchard, 2004). Kimi araştırmacılara göre kent markası, sahip olunması arzulanan ayırt edici niteliklere, kimilerine göre ise yerel ve sahip olunan niteliklere odaklanmalıdır (de Jong, vd., 2018). Kentler için başarılı bir marka oluşturma uzun bir süreç ise ve geleceğin belirsiz olduğu da göz önüne alınırsa markalar, gelecek için bugünden hedef pazarları belirlemeli ve onlar için değer ve kimlik yaratmalıdır. Bu neden ile, markanın dayandırılacağı ayırt edici niteliklerini, kimliğini ve müşterilere ne vaat etmesi gerektiğini belirlemek stratejik bir karardır ve kapsamlı bir bakış açısı gerektirir. Arzulanan kimlik ile var olan kimlik birbiri ile ilişkili olmalı ve var olan kimliğin arzulanan kimliği destekler nitelikte olması gerekmektedir (de Jong, vd., 2018).

<sup>13</sup> Kotler (1999)'a göre bir mekanın en kapsamlı hali ile yedi hedef kitlesi olabilir: Yeni vatandaşlar, ziyaretçiler, yönetimler, yatırımcılar, girişimciler, yabancı alıcılar "foreign purchasers" ve konum uzmanları "location specialists".

Genel itibari ile markalaşma çalışmalarında, örgütün ve çalışanların değerleri ile markanın kimliği arasında sembiyotik bir ilişki var olduğu belirtilmektedir (Aaker, 1996). Kent markalarının kurumsal markalar ile yüksek derecede benzeştiği de göz önüne alındığında, kentin kimliği ile kentte yaşayan vatandaşların ve faaliyet gösteren kurumların/işletmelerin değerleri arasında da sembiyotik bir ilişkinin var olduğu belirtilebilir. Açıklamak gerekirse oluşturulacak olan marka ve kimlik kentin, kentte yaşayan vatandaşların ve iş çevresinin değerlerini yansıtmalı (Chan, Marafa, 2014) ama aynı zamanda vatandaşların ve iş çevresinin değerleri de kente enerji vermeli ve kent yönetimini yönlendirebilmelidir. Söz konusu kimlik uyuşması, kentin diğer kentlerden kolaylık ile ayrışmasını sağlamaktadır (Trueman, Cornelius, Killingbeck-Widdup, 2007). Fakat uyumu yakalamak çok kolay değildir. Çünkü vatandaş, yerel yönetim, iş çevresi ve sosyal çevreden oluşan bileşenlerin ihtiyaçları, değerleri ve tercihleri birbirleri ile uyumlu olmayabilir/çelişebilir. Aslında bu uyumsuzluk yüzünden mekân pazarlama bir yönetim sürecidir ve aynı zamanda yeni politikalar oluşturmayı da içermektedir. Bu durum, kent sakinlerini yerel yönetimlerin sunduğu kamusal hizmetleri tüketmekten ziyade aktif vatandaşlara dönüştürmeyi gerektirir ve demokratik katılıma dayalı müzakere biçimlerini şart kılar. Bu yüzden, yerel yönetim ile kent sakinleri arasındaki çift yönlü iletişim (diyalog) çok önemlidir ve çift yönlü iletişimin kalkması kent sakinlerini bir paydaştan ve markayı oluşturan bir bileşenden ziyade sadece tanıtıcı faaliyetler ile ilgilenen bir hedef kitleye dönüştürür (Walsh, 1994'ten aktaran Eshuis, Klijn, Braun, 2014).

Kent sakinlerinin bir paydaş olarak sürece dâhil edilmesi onların duygularının “emotions”<sup>14</sup> markaya aktarılmasını sağlamaktadır. Tüketici gözünde markanın doğru bir duygu ile bağdaşması önem teşkil etmektedir. Çünkü markanın tüketicide oluşturacağı duygu, tüketicinin pazarlanan ürüne (kente) değer atfetmesini ve atfedilen değer sonucunda tüketicinin tutumunu, satın alma ya da almama kararını vermesini sağlamaktadır (O’Shaughnessy, O’Shaughnessy, 2003). Örneğin, (Eshuis, Klijn, Braun, 2014) çalışmasında, Hollanda’daki bir kentin yeniden markalaşma “restructuring”

---

<sup>14</sup> Duygular, hislerden farklı olarak meydana gelen fiziksel ve bilişsel durumlar olup, evrensel olarak kabul görmüş beş duygu durumu (sevinç, üzüntü, korku, şaşkınlık ve kızgınlık) bulunmaktadır. Fakat günümüze kadar gerçekleştirilmiş bilimsel araştırmalar ile insanın daha fazla duyguya sahip olduğu belirtilmiş ve en son 2017 yılında yapılmış bir deneysel çalışmada, insanın 27 farklı duygusu olduğu ileri sürülmüştür.



sürecine kent sakinlerinin katılımının etkisi araştırılmıştır. Çalışmanın bulgularına göre tasarlanacak markanın kimliği ile kent sakinlerinin kimliklerinin uyuşması, markalaşma sürecinde kent sakinlerinde meydana gelebilecek kaygıların yönetilmesinde ve önlenmesinde olumlu bir etkiye sahiptir. Böylelikle marka, kenti oluşturan toplumu dolayısı ile toplumun kimliğini tamamı ile değiştirmek ve ortadan kaldırmak yerine, onu yeniden şekillendirmelidir. Dolayısı ile sembiyotik ilişkinin tarafları birbirine güç katan nitelikte olmalıdır. Bu katılımcı yaklaşım ve sembiyotik ilişki aslında Kotler'in kent markalaşması sürecinin ön şartı olan vatandaş-sektör-yerel yönetim işbirliğini de desteklemektedir. Kent ve vatandaş kimliklerinin karşılıklı etkileşimlerine ek olarak, vatandaşların davranışlarının da marka üzerinde etkisi olabilmektedir (Anholt, 2006).

Mekan pazarlamada, tüketiciler aslında bir ürünü değil de ürünün yarattığı değeri satın almaktadır (Eshuis, Klijn, Braun, 2014). Diğer bir deyiş ile kenti değil, kentin yarattığı değeri satın almaktadır. Kent ise vatandaşlar, iş çevresi, yerel yönetim ve fiziki çevre bileşenlerinden oluşmakta olup satın alınan değeri bu bileşenler oluşturmaktadır. Bu yüzden mekân pazarlama/markalaşma faaliyetlerinde satın alınan “şey”in bir bileşeni olarak vatandaşlar göz ardı edilmemelidir. Vatandaşlar ve yaşam tarzları kentin atmosferini oluşturmakta ve dolayısı ile marka, gücünü sadece kentin fiziki olanaklarından değil aynı zamanda vatandaşlarından da almaktadır (Järvisalo, 2012). Bu özelliği ile mekân pazarlama, katılımcı ve interaktif bir özelliğe sahiptir (Bennett, Savani, 2003; Hatch, Schultz, 2010, Hankinson, 2004; O'Guinn, Muniz, 2004'ten aktaran Eshuis, Klijn, Braun, 2014).

İlgili yazınında, Kotler (1999)'un mekân pazarlama modeli esas teşkil etmek üzere çeşitli mekân pazarlama ve kent markalaşması modelleri/yaklaşımları bulunmaktadır. Kavaratzis (2009) çalışmasında Anholt (2006), Hankinson (2004), Hankinson (2007), Rainisto (2003) ve Trueman & Cornelius (2006) tarafından önerilen kent markalaşması yaklaşımlarını incelemiş ve benzerliklerini tespit edip birbiri ile sentezleyerek bütünlük bir sistem önermiştir. Söz konusu sistem, vizyon ve strateji, iç kültür, yerel topluluklar, sinerji, altyapı, kentsel peyzaj, fırsatlar ve iletişimden oluşmaktadır. Bu sistemin yerel topluluklar kategorisi yerel halk, girişimciler ve iş çevresini ve onların ihtiyaçlarının önceliklendirilmesini kapsamaktadır. Sinerji kategorisi ise yerel halk, girişimciler ve iş çevresinin adil ve dengeli bir şekilde markalaşma sürecine dâhil edilmesini ve süreci

desteklemelerinin sağlanmasını kapsamaktadır. Keza (Yoon, 2010) çalışmasında stratejik kent markalaşması için kavramsal bir model geliştirilmeye çalışılmıştır. Söz konusu modelin merkezinde insan, çevre ve kültür yer almaktadır. Kentin fiziksel tasarımı, imajı ve kimliğinin, merkezde yer alan bu üçlü yapıya bağlı olarak tasarlandığı zaman markalaşmanın başarılı olacağı belirtilmiştir. Çünkü davranışlar fiziksel çevreden etkilenmektedir ve mekânsal yapılanmada gerçekleştirilecek değişimler, insanların rutinlerini ve hatta birbirleri ile etkileşimlerini bile etkilemektedir (Milligan, 1998'den aktaran Borer, 2006).

Gustavsson, Elander (2012)'den aktaran Busch, Anderberg (2015)'e göre kentler, markalaşma faaliyetlerinde kendilerini ya (yatırım yapmak için) karlı bir kent olarak ya da yaşamak için iyi bir kent olarak satmaya odaklanmaktadır. Sürdürülebilir ya da yeşil kent markalaşması ise söz konusu bu iki odak noktasını birbiri ile sentezleyebilmektedir. Curitiba, Portland, Freiburg ve Malmö kentleri ekolojik kentlerin (eko-kent) öncüleri olarak değerlendirilmektedir. Örneğin; Malmö kentinde uygulanan politikalar ve gerçekleştirilen faaliyetlerin sadece insani ve çevresel nedenler dolayısı ile değil aynı zamanda kent ekonomisinin yeniden canlanması için yeni yatırımları cezbetmek amacı ile yürütüldüğü tespit edilmiştir (Holgensen, Malm 2015'ten aktaran Busch, Anderberg, 2015). Hatta son yıllarda ekolojik ya da sürdürülebilir kentlerin popülerlik kazanması ile Hacienda, Auroville ve Masdar City gibi sonradan planlı bir şekilde yeni eko-kentler inşa edilmeye başlanmıştır. Fakat Checker, (2011) çalışmasında, kent markalaşması amacı ile gerçekleştirilen bu ve benzeri sürdürülebilirlik faaliyetlerinin toplum için sadece olumlu değil aynı zamanda olumsuz sonuçlar doğurduğu tartışılmaktadır. Yazar, bu faaliyetlerin çevresel soylulaştırmayı<sup>15</sup> “gentrification” destekler nitelikte olduğunu, çevresel sürdürülebilirlik açısından son derece anlamlı fakat aynı zamanda kar amaçlı faaliyetler olduğunu, sosyal adaleti tehlikeye attığını ve fakir kent halkının kentin daha az gelişmiş bölgelerine taşınmasına neden olduğunu belirtmektedir. Dolayısı ile tasarlanan marka ve kent ile vatandaşlar, kendi kentlerine yabancılaşmamalıdır. Bu durum Yoon (2010)'un da açıkladığı üzere, kent ve içinde yaşayan vatandaşlar arasında kesintisiz bir iletişimi gerektirmektedir. Çünkü kentte gerçekleşen her faaliyet zamanla alışkanlığa, alışkanlıklar

---

<sup>15</sup> Kentsel dönüşüm alan yazınından detaylı bilgi edinilebilir.

geleneklere, gelenekler k lt re, k lt r kentin imajına ve imaj ise kimliđine d n şmektedir.

Chan, Marafa (2014) alıřmasında, yeřil kent markalařması, kentin sadece yeřil kaynakları bađlamında incelenmiřtir. Bu alıřmada her bir kenarı, markalařmanın boyutlarını temsil eden bir markalařma altıgeni, Anholt'un Kent Markaları Endeksi "City Brands Index" referans alınarak tasarlanmıř ve boyutlardan bir tanesi yeřil vatandaşlık olarak belirlenmiřtir.

Marin-Aguilar, Vila-L pez (2014)'e g re ise kent markalařmasında iki yaklařım  nem kazanmaktadır. Bu yaklařımlardan ilki deneyim pazarlama, diđeri ise yeřil pazarlama. Deneyim pazarlama, kentte eřitli b y k etkinlikler organize ederek hedef kitleye unutulmaz anlar yařatma faaliyetlerine odaklanırken, yeřil pazarlama kentin karakteristiklerine ve vizyonuna has ekolojik politikalar oluřturmaya odaklanmaktadır. Kentlerin yeřil pazarlama kapsamında yeřil bir kent kimliđine ve markasına sahip olması ise s rd r lebilir kentsel kalkınmayı beraberinde getirmektedir. Bu bařarım hem kent ekonomisi hem de kent yařamı ve kentliler iin b y k bir fırsattır. Kentlerin yeřil veya s rd r lebilir markalarını pazarlama amacı ile kullanabilmelerinin birka yolu bulunmaktadır: kentte yařanabilirliđe odaklanmak, yeřil teknoloji ve politikaya odaklanmak ve kentin evre  zerinde d ř k bir etkiye sahip olmasına "low impact city" odaklanmak.   nc  seenek, diđer seeneklere nispeten biraz daha zor olarak kabul edilse de bu yaklařım  zerinden bir marka yaratmak ok da imk nsız deđildir.  nk  bu farklı yaklařımlar birbirine zıt deđil, birbirini tamamlayıcı niteliktedir. Sıralı olarak gerekleřtirilmesi durumunda, ilk iki seenek iin yapılan faaliyetler ve onların ıktıları, tam anlamı ile olmasa da, kenti b y k  l de d ř k etkili bir kente d n řt rerek,   nc  seeneđin gerekleřmesine katkıda bulunacaktır.

Chiesura (2004) ve Tzoulas, vd. (2007)'ye g re kentlerdeki yeřil alanlar ve insanın iyiliđi<sup>16</sup> "human well-being" arasında pozitif iliřki bulunmaktadır ve birok kent, fiziksel olarak "yeřil" olmayı bir markalařma aracı olarak kullanmakta ve kendini "yařanılabilir" bir kent olarak pazarlamaktadır.  nk  yařanılabilir kentler, yeni kent sakinleri, iřletmeler ve ziyaretiler iin cezbedici niteliktedir. Bu yaklařım, kentin evresel

---

<sup>16</sup> Refah, iyilik, sađlık, mutluluk ve esenlik anlamındadır.

etkisinden ziyade kent sakinlerinin ve ziyaretçilerinin mutluluğuna odaklanmaktadır. Örneğin yaşanılabilir bir kent markası yaratmak için kentteki yeşil alanların, çatı bahçelerinin ve dikey bahçelerin genişlemesi ve kentin ve çevresinin sahip olduğu ekosistemlerin restorasyonu gibi faaliyetler önerilebilir veya kent hayatında arabadan bisiklete geçişi teşvik edecek politikalar oluşturularak teknolojik yatırımlar yapılabilir (Dinnie, 2011). Bu faaliyetler, hem kentte yaşayan insanların fiziki ve psikolojik sağlığını bir nebze iyileştirerek onların mutluluğunun artmasını hem de söz konusu kentin çevresel etkisini azaltarak, düşük etkili kent olmasını sağlayabilir. Bir başka ifade ile, gerçekleştirilecek olan markalaşma faaliyetleri aslında kenti hem teknolojik ve politik (aslında yönetsel) olarak yeşil yapmakta hem de çevresel etki bağlamında kentin, düşük etkili/yaşanılabilir kent markasına sahip olmasını sağlamaktadır.

#### **4. ÇEVRE DOSTU DAVRANIŞ**

Ekolojik kimlik, bireyin kendisi, insanoğlu, insan dışı varlıklar ve tüm ekosistemden oluşan dörtlü bir yapıya sahiptir. Ekolojik kimliğe sahip olan birey, bu dörtlü yapıdan birisi zarar gördüğü zaman söz konusu zararı kendisine karşı yapılmış olarak algılar ve harekete geçer. Çünkü bu bireyin doğa ile arasında yakın bir ilişki vardır ve kendisini doğanın bir parçası olarak görür. Bu yaklaşım ekopsikolojik yaklaşım olarak adlandırılmaktadır. Ekopsikolojik yaklaşım, içgüdüsel biyofilik eğilimlerin ve insan ile doğa arasındaki deneyimsel bağlantının (fiziksel deneyimin) artması ile insanların çevreye duyarlı davranışlarda bulunacağını savunmaktadır. Dolayısı ile insan ve doğa arasında kuvvetli bir ilişki sağlandığı zaman birey ekolojik bir kimliğe sahip olur, çevreye duyarlı davranışlar sergiler, ekolojik toplum oluşur ve sürdürülebilirlik başarılabılır. Böylelikle vatandaşı yeşil olan, ruhu ve markası yeşil olan sürdürülebilir bir kent yaratılabilir.

Bu bağlamda, sürdürülebilir bir kent yaratabilmek için öncelikle vatandaşların çevreye ne kadar duyarlı bir yaşam sürdürdükleri tespit edilmelidir. Çevreye duyarlı davranışta bulunma, çevre psikolojisinin ilgi alanına girmektedir. Aslında çevre psikolojisi 1960 öncesinde insanın fiziksel çevre (iş-ev ortamı) ve yapı fiziksel çevre (mimari çevre) ile etkileşimine odaklanmaktaydı. 1960'ların sonlarına doğru çevresel problemlerin farkındalığı ile insan faaliyetlerinin biyolojik çevre üzerindeki ve biyolojik çevrenin ise insan sağlığı ve refahı "well being" üzerindeki olumsuz etkisi araştırılmaya

başlanmıştır. 1970'lerden sonra ise enerji teknolojilerinin gelişmesi ile birlikte enerji arzı ve talebi ve risk algısı çevre psikolojisinin odak noktası olmuştur. 1980'lere gelindiğinde, tüketicinin çevre ile ilgili tutumu ve tüketim davranışları arasındaki ilişkiler incelenmeye başlanmıştır. Günümüzde ise iklim değişikliğinin tüm dünyada bariz bir şekilde hissedilmesi, yeryüzündeki bazı ekosistemlerin kaybolmaya başlaması ve bu durumun ana nedenlerinden birisi olarak insan davranışlarının kabul edilmesi ile çevre psikolojisi "Sürdürülebilirlik Psikolojisi" ne evrilmiştir. Çevresel, toplumsal ve ekonomik boyutları bulunan bu genişletilmiş kapsamın ilgi alanı, insan refahı ve yaşam kalitesini koruyarak, insan davranışlarının çevresel problemleri geriye döndürecek şekilde nasıl değiştirilebileceği hakkındadır (Steg, Berg, de Groot, 2013).

Amerika'nın Minnesota eyaletine bağlı Kirlilik Kontrol Ajansı (Pollution Control Agency), sürdürülebilirlik psikolojisi hakkında yayınlamış olduğu raporda Thaler ve Sunstein'in "Nudge" isimli kitabına atıf verilerek insanların, birbiri ile paralel çalışan iki çeşit karar alma sistemine sahip olduğunu belirtmiştir (Manning, 2009). İlki Star Trek'in Mr. Spock karakterine benzeyen bilinçli, rasyonel ve planlı karar almayı sağlayan kural tabanlı sistemdir. Bu sistem aslında Freud'un zihin yapılarından süper egoyu temsil etmektedir. İkinci sistem ise The Simpsons'ın Homer Simpson karakteri gibi bilinçsiz, içgüdüsel/dürtüsel ve duygular ile karar almayı sağlayan çağrışımsal sistemdir. Bu sistem, zihnin id yapısını temsil etmektedir. Kitaba göre kural tabanlı sistem, gerçekleştirilecek davranışı dikkatlice ve kapsamlı bir şekilde inceler ve yavaş karar alır. Çağrışımsal sistem ise bilinçdışı çalışır, karar almada duygular ve deneyimler baskın rol oynar ve hızlı karar alır. Gün içinde aldığımız kararların ve sergilediğimiz davranışların çok büyük bir kısmı çağrışımsal sistem tarafından gerçekleştirilir. Rapora göre ise ÇDD'ler, içimizdeki Homer Simpson'a (çağrışımsal sisteme) çekici gelmemektedir. Dolayısı ile insan günlük hayatta rasyonel davranmadığı için ve duygular ve deneyimler insandan insana değiştiği için eğitim, statü, coğrafi konum, gelir gibi faktörler bağlamında benzer insan gruplarındaki bireylerin ÇDD'leri farklılık gösterebilmektedir.<sup>17</sup> Thaler ve Sunstein'in tespiti ile paralel olarak Gatersleben, Steg (2013) de insanların doğaya olan duyuşsal yakınlığının, çevre ile ilgili bilişsel inançlarından daha iyi bir şekilde ÇDD'yi

---

<sup>17</sup> Rapora göre ÇDD'nin tetiklenebilmesi için bireylere/toplumlara önerilen davranışlar öncelikle çağrışımsal sisteme hitap edecek ve kural tabanlı sistemle de uyumlu olacak şekilde sunulmalıdır.

öngördüğünü alan yazınındaki çeşitli çalışmalara atıf vererek belirtmektedir. Bunun nedeni, yukarıda da açıklandığı üzere, zihnin id yapısının bireyin arzularına bir an evvel doyum arayıp bireyin davranışlarını bu yönde şekillendirmesidir (Koç, 2016). Dolayısı ile, ÇDD araştırmalarında duyuşsal boyutun dikkate alınması önemlidir.

Günümüze kadar yapılan birçok çalışmada, neden bazı bireylerin ÇDD sergilerken diğerlerinin sergilemediği, tutum, bilgi, değer, motivasyon, kişilik, yetenekler ve inançlar gibi değişkenler kullanılarak geliştirilen çeşitli davranış modelleri ile açıklanmaya çalışılmıştır (Nordlund, Garvill, 2002).<sup>18</sup> Fakat hangi etmenlerin ÇDD'yi engellediği veya hangilerinin teşvik ettiği gibi sorulara hala net bir cevap verilememektedir (Kollmus, Agyeman, 2002'den aktaran Çakır, et al., 2015). İnsan-doğa ilişkisi ise ÇDD'yi açıklamaya çalışan güncel yaklaşımlardan bir tanesidir. İnsan-doğa ilişkisi, birey ile çevre arasındaki duyuşsal , bilişsel ve fiziksel ilişki ağını incelemektedir (W. P. Schultz, 2002). Wilson (1993)'e göre insanoğlunun yeryüzünde yaşayan diğer tüm varlıklar ile ilişki halinde olma ihtiyacı bulunmaktadır ve bu ihtiyaç içgüdüselidir. İnsan, her ne kadar yerleşik yaşama geçerek kentlerde yaşıyor olsa da genlerimizde kodlanmış olan bu ihtiyaç biyolojik olarak silinebilir değildir. Alan yazınında bu durum "Biyofili Hipotezi" olarak adlandırılmaktadır. McVay (1993)'ten aktarıldığı üzere Edward Wilson biyofiliyi "İnsanın yaşama ve yaşamın tüm süreçlerine duyduğu sevgi/eğilim" olarak tanımlamakta ve "insan, ancak diğer tüm canlıları (olduğu gibi) kabullendiği zaman, onlara ve kendisine (ve diğer insanlara) değer verir" şeklinde belirtmektedir. Başka bir ifade ile doğa ile daha çok ilişki halinde bulunan insanlar, bitkilere, hayvanlara ve diğer insanlara karşı olumlu bir eğilime sahiptir (P Wesley Schultz, 2000; Stern, Dietz, 1994). Wilson tarafından geliştirilen bu hipotez, daha sonra Kellert (1997) tarafından genişletilmiştir (Nisbet, Zelenski, Murphy, 2009). Kellert (2005)'e göre biyofili/biyofilik eğilim sadece ÇDD'yi desteklemek ile kalmayıp, bireylerin duyuşsal ve psikolojik gelişimini de desteklemektedir.<sup>19</sup> Biyofilik eğilim aslında "zayıf" bir genetik eğilim olup, birey doğayı

---

<sup>18</sup> Steg, Nordlund (2013) çalışmasından yaygın olarak kullanılan ÇDD modelleri hakkında bilgi alınabilir.

<sup>19</sup> Alan yazınında kanıt olarak doğayla daha çok ilişki halinde olan bireylerin daha sağlıklı ve daha mutlu olması, pratik zekâlarının, yaratıcılıklarının, problem çözme, kendini ifade etme ve sosyalleşme yeteneklerinin gelişmesi ve gün geçtikçe kentlerde yaşayan bireylerin doğaya yönelmesi örnek verilmektedir (Kellert, 2005; Nisbet, Zelenski, Murphy, 2009).

deneyimledikçe, öğrendikçe ve toplumsal çevre ve kültür biyofiliyi destekledikçe aktifleşmekte, güçlenmekte, davranışa dönüşmekte ve bireyin fiziksel ve zihinsel sağlığına katkıda bulunmaktadır. Dolayısı ile insanlar doğa ile olan bağlantılarını tam olarak anlarlar ise, tüm canlılar ve gezegen için daha fazla empati yapar, saygı duyar ve onları koruyacak şekilde davranır (Feral, 1998; P Wesley Schultz, 2000).

İnsan-doğa ilişkisi alanında geliştirilen en önemli üç ölçme aracından (ölçek) ilki W. P. Schultz (2002) çalışmasındaki Inclusion of Nature in Self ölçeğidir. Ölçek tek ifadeden oluşup, birey ile doğa arasındaki ilişkiyi bilişsel olarak ölçmektedir. İkinci ölçek Mayer, McPherson Frantz (2004) çalışmasında yer alan ve 14 ifadeden oluşan Connectedness to Nature ölçeğidir. Bu ölçek, birey ile doğa arasındaki duyuşsal ilişkiyi ölçmek için kullanılmaktadır. Üçüncü ölçek ise Nisbet, Zelenski, Murphy (2009) çalışmasında yer alan DİÖ'dür. Söz konusu ölçek bu alanda geliştirilmiş olan en son ölçek olup,<sup>20</sup> insan-doğa ilişkisinin Schultz'un belirttiği üç boyutta incelenmesine olanak sağlamaktadır. Nisbet, Zelenski, Murphy (2009)'a göre insanın doğa ile ilişkisi, bireyin doğaya verdiği değer ve doğanın tüm bileşenleri ile sahip olduğu deneyimler bütünüdür ve doğayı bütünü ile kabullenmek ve kavramaktır. P. W. Schultz (2000) 'e göre, doğa hakkında olumlu duyguları olan bu bireyler doğayı önemserler, doğaya saygılı davranırlar ve onu korurlar. Bu bağlamda, insanın doğa ile ilişkisi, onun çevreye duyarlı (veya duyarsız) davranışlarını etkileyebilmektedir. Örneğin, Dutcher, vd. (2007) çalışmasında gerçekleştirilen çoklu regresyon analizinde, doğa ile ilişkinin çevresel endişe ve ÇDD üzerindeki tahmin ve açıklayıcı gücünün çok yüksek olduğu bulgusuna varılmıştır. Tam (2013) çalışmasında, doğa ile ilişki alanında geliştirilmiş olan yedi farklı ölçek Amerika ve Çin örneklemelerinde birbirleri ile karşılaştırmalı olarak analiz edilmiştir. Çalışmanın bulgularına göre DİÖ değerler, tutum, insan refahı ve ÇDD'yi açıklayıcı gücü en yüksek olan iki ölçekten birisi olarak tespit edilmiştir.<sup>21</sup> Strickland (2014) çalışmasında ise doğa ile ilişkinin balıkçıların çevreye duyarlı avlanma alışkanlıkları üzerindeki etkisi çoklu

---

<sup>20</sup> Bu alandaki diğer ölçeklerin karşılaştırmalı analizleri (Tam, 2013; Nisbet, Zelenski, Murphy, 2009) çalışmalarından incelenebilir.

<sup>21</sup> Diğer ölçek Clayton (2003) çalışmasında yer alan Environmental Identity (EID) ölçeğidir. Bu tez çalışmasında NR ölçeği kullanılmış olup, EID'nin kullanılmama gerekçesi bu ölçeğin, bireyin doğanın diğer bileşenleri hakkındaki bakış açısını (bilişselliği) içermemesi ve NR ölçeğinin duyuşsal, bilişsel ve fiziksel boyutların hepsini içermesidir.

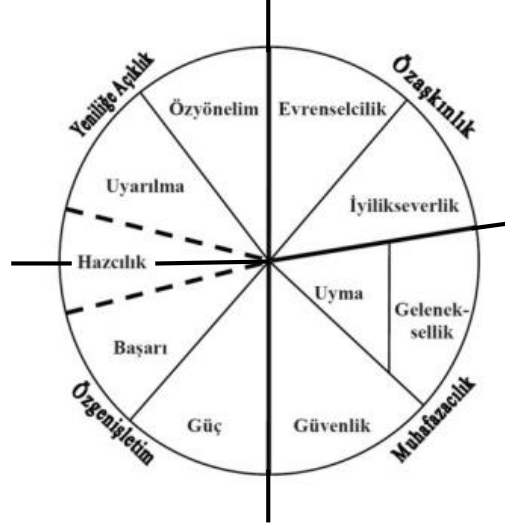
regresyon ile analiz edilmiştir. Çalışmanın bulgularına göre doğa ile ilişkinin davranış değişikliği modeline dâhil edilmesi ile modelin tahmin gücü %6,1'den %67,7'ye yükselmiştir, her ne kadar çevreye duyarlı avlanma yapan ve yapmayan balıkçıların DİÖ'den aldıkları puanlar birbirinden çok farklılaşmasa da.<sup>22</sup> Son olarak, (Bahar, Sahin, 2017) çalışmasında ise DİÖ, ÇEÖ ve ilköğretim öğrencileri için geliştirilmiş olan bir ÇDD ölçeği 7. ve 8. sınıf öğrencilerine uygulanmıştır. Araştırma verileri çoklu regresyon analizine tabi tutulmuş ve sonuç olarak önceki çalışmaların bulguları ile paralel olarak doğa ile ilişkisi yüksek ve çevresel endişe bağlamında hümanist (altruistik) ve doğa dostu (biyosferik) öğrencilerin, egoist ve doğa ile ilişkisi düşük olan öğrencilere nazaran daha fazla ÇDD sergiledikleri belirtilmiştir.

ÇDD araştırmalarındaki bir diğer önemli nokta ise değerlerdir. Değerler, tutum ve davranışları öngördüğü için (W. P. Schultz, Zelezny, 1999; Tam, 2013) ÇDD araştırmaları için çok önemlidir. Değerler, bir önceliklendirme sistemine sahiptir ve önemleri birbirinden farklıdır. Başka bir ifade ile, birbirleri ile çakışan değerler eş anlamlı olarak aktive olduğunda, seçimler daha önemli değere göre yapılmaktadır. Bu özelliği ile değerler, bireyleri, grupları, ülkeleri ve kültürler arası benzerlik ve farklılıkları açıklamak ve tanımlamak için sıklıkla kullanılırlar. Keza, davranış değişikliğini tetiklemek için de bir başlangıç noktası olarak kabul edilmektedir. Belirli değerleri aktive ederek, hedeflenen davranışı etkilemek mümkündür (de Groot, Thogersten, 2013). Alan yazınındaki en tanınmış değer teorilerinden birisi Schwartz'a ait olup, 82 ülkeden toplanan veriler ile 56 ifadeden oluşan 10 çeşit değer ve dört boyut tanımlanmış ve iki boyutlu düzlemde Şekil 5.'te yer alan çembersel model oluşturulmuştur. Aynı boyut altında yer alan değerler veya birbirlerine komşu olan değerler arasında geçişkenlik olduğu belirtilirken, birbirlerinden uzak veya birbirlerine zıt olarak konumlanmış boyutlar veya değerlerin birbirleri ile çatıştığı belirtilmektedir (Schwartz, 2013).

---

<sup>22</sup> Çalışmadaki bu beklenmeyen sonuç, balık tutma ödülleri (prize catch), yarışmalar ve balıkçıların yaşadıkları çevredeki sosyal kurallar ile açıklanmıştır.





Şekil 5. Schwartz'ın çembersel değer modeli (Schwartz (2013)'ten akt. Demirutku, Güngör (2017))

ÇDD alan yazınında ise değerlere bağlı olarak üç çeşit tutumun, bireylerin çevre ile ilgili davranışlarını açıklayabileceği belirtilmektedir: Egoistik, altruistik ve biyosferik<sup>23</sup>. Egoizm, bireyin kişisel çıkarlarına dayandırılmıştır. Birey için iyi-doğru olan her şey toplum için faydalı olarak kabul edilmekte ve bireyin kişisel faydası toplumunkinden önce gelmektedir. Çevreye duyarlılık bağlamında egoizm, bireyin tüketim yaparken sadece kendi konforunu dikkate alarak davranması şeklinde yorumlanabilir. Diğer bir deyiş ile birey, fayda-maliyet analizi yapar ve söz konusu davranışın çıktıları bireye menfaat sağlar ise söz konusu davranışı sergiler. Örneğin, bir kişinin yakıt maliyeti dolayısı ile şahsi arabasını daha az kullanması gibi. Altruizm ise toplumun çıkarlarına dayandırılmıştır. Bu görüşe göre bireylerin ve toplumun, insanlığın faydası için birlikte hareket etmesi gerekmektedir. Diğer bir deyiş ile egoizmin nerede ise tersi bir tutumla bireysel faydayı toplumsal faydanın arkasına atmaktadır. Örneğin kişinin, diğer insanların sağlığını düşünerek daha az araba kullanması gibi. Fakat bu görüş aynı zamanda çevrenin ve doğal kaynakların toplumsal fayda için sonuna kadar kullanılabilmesini de savunmaktadır. Örneğin elektrik üretebilmek için akarsuların üzerine plansız bir şekilde inşa edilen baraj ve hidroelektrik santraller gibi. Özetle, hem egoizm hem de altruizm gerek bireylerin gerek ise toplumların faydası için gerçekleştirilen faaliyetlerin neden olduğu ekolojik değişim ve bu değişimin uzun vadedeki etkilerini göz ardı etmektedir.

<sup>23</sup> Egoistik, altruistik ve biyosferik terimleri ilgili yazınında sırası ile egosantrik, homosantrik/antroposantrik ve ekosantrik/biyosantrik olarak ifade edilmektedir (W. P. Schultz, Zelezny, 1999).

Biyosferizm ise kâinatı/evreni temel almaktadır. Bu görüşe göre tüm canlı ve cansız bileşenleri ile birlikte çevre, bir değerler bütünüdür ve insani faaliyetler bu değerler arasındaki dengeyi bozmayacak ve onu koruyacak şekilde planlanmalıdır. Sonuç olarak biyosferik bireylerin, egoistik ve altruistik bireylere nazaran çevreye daha duyarlı yaşadıkları enerji tüketimi bağlamında ise daha az tüketimde bulunma eğiliminde oldukları belirtilmektedir (Merchant, 2005).

Alan yazınında biyosferik ve altruistik tutumların Schwartz'ın modelindeki evrensellik ve iyilik değerlerini yansıttığı (W. P. Schultz, Zelezny, 1999), biyosferizm ile altruizm arasında pozitif bir ilişki olduğu ve altruizmin de ÇDD'yi öngördüğü belirtilmektedir. Fakat biyosferizmin, ÇDD'yi altruizmden daha iyi bir şekilde öngördüğünü ve altruistik değerlere sahip bireylerin bazı durumlarda, insani menfaatler söz konusu olduğunda, çevreci davranış sergilemekten kaçındığını belirten çalışmalar da bulunmaktadır (de Groot, Thogersten, 2013). Örneğin (W. P. Schultz, Zelezny, 1999) çalışmasında, egoizm ve altruizm tek bir boyut olarak ve biyosferizm ayrı bir boyut olarak ele alınıp Schwartz'ın değer modeli ile ilişkisi incelenmiştir.<sup>24</sup> Çalışmanın bulgularına göre, Şekil 5'te yer alan değer çemberindeki öz-aşkınlığın evrensellik değeri, biyosferizmin pozitif yönlü ve güçlü bir tahminleyicisi iken, tutuculuğun gelenek değeri ile kendini genişletmenin güç değeri biyosferizmin negatif tahminleyicisidir. Bu bulgu teori ile paralel olup, alandaki diğer çalışmaların bulguları ile örtüşmektedir. Diğer taraftan yine aynı çalışmada, öz aşkınlığın evrensellik değeri, Schwartz'ın değer teorisi ile çelişerek, altruizmin bir tahminleyicisi olarak tespit edilememiştir. Teorinin aksine altruizm, özaşkınlığın iyilik ve evrensellik değerleri ile negatif ve kendini genişletmenin güç değeri ile pozitif ilişkili olarak tespit edilmiştir. Ayrıca, tutuculuğun gelenek, uyum ve güvenlik değerleri de altruizm ile pozitif ilişkili olarak bulunmuştur. Bu çalışmanın akabinde W. P. Schultz (2001) çalışmasında ise egoizm, altruizm ve biyosferizmin her birisi ayrı birer boyut olarak incelenmiş ve altruizm, beklenenin aksine, öz aşkınlıkla ve öz genişletme ile biyosferizm kadar güçlü bir (pozitif ve negatif) korelasyona sahip olamamıştır. Söz konusu çalışma beş ayrı ülkenin büyük kentlerinde gerçekleştirilmiş ve kent insanının doğadan kopuk yaşamasından dolayı doğal yaşam ile kendileri arasında bir

---

<sup>24</sup> Çalışma, Güney Amerika ülkeleri, İspanya ve Amerika Birleşik Devletleri (ABD) olmak üzere 14 farklı ülkede gerçekleştirilmiştir.

bağlantı kuramadıkları ve bu yüzden egoistik tutuma sahip oldukları belirtilmiştir. P. W. Schultz, Zelezny (2007)'den aktarıldığı üzere Stern, vd. (1995) çalışmasında da aynı üçlü yapının Schwartzın değer teorisi ve ÇDD ile arasındaki ilişki, Amerika'daki bir eyalette telefon ile görüşme yöntemi ile toplanan veriler kullanılarak araştırılmıştır. Çalışmanın sonuçlarına göre sadece biyosferizm ile ÇDD arasında anlamlı bir ilişki tespit edilebilmiştir. Diğer bir deyiş ile altruizm ile ÇDD arasında anlamlı bir ilişki elde edilememiştir.<sup>25</sup> Ayrıca yazarlar, çevre ile ilgili konularda gerçekleştirilen psikolojik araştırmaların her ne kadar uluslararası geçerliliği olan ölçekler geliştirilmeye çalışılsa da nerede ise hepsinin istisnasız bir şekilde kültüre özel olduğunu belirtmektedir .

ÇDD'nin ölçümünün nasıl yapılması gerektiği, alan yazınında tartışılmalı bir konudur. Uygulamadaki sorunsal, davranışın mı (Oskamp, 2000) yoksa etkisinin mi (Kaiser, Wilson, 2004) ölçülmesi gerektiğidir. Genel kanı ise çalışmanın odak noktasına göre ölçümün şekillendirilmesidir. Davranışın etkisinin veya gerçek çevresel etkinin ölçülmesi çok yönlü bir bakış açısını gerektirmektedir. Çünkü bir davranışın doğuracağı çevresel problemler çeşitlidir ve hatta söz konusu davranış, bir açıdan çevreyi olumlu etkilerken bir başka açıdan olumsuz bir etkiye neden olabilmektedir. Örneğin, organik ürünler satın alarak toprak, hava ve su kirliliğine neden olan zararlı ilaç veya kimyasalların kullanımı azalmış olurken, söz konusu ürün eğer uzak mesafeden nakliye edilmiş ise nakliyesi dolayısı ile meydana gelen karbon salınımı hava kirliliğine neden olmaktadır (Gatersleben, 2013). Davranışın kendisinin ölçümü için ise alan yazınında birisi baskın olmak üzere iki yaklaşım bulunmaktadır. İlk yaklaşım alan yazınındaki genel kanı olup, ÇDD'nin çok boyutlu olarak ele alınmasıdır. Bu yaklaşıma göre, farklı davranışlar farklı faktörler tarafından tetiklenmekte veya spesifik bir faktör farklı davranışların sergilenmesini tetikleyebilmektedir. Başka bir ifade ile çevreci sayılan bir davranış, başka bir çevreci davranışa her zaman neden olmayabilir.<sup>26</sup> İkinci yaklaşım ise

---

<sup>25</sup> Doğal kaynakların kullanımı ve ÇDD bağlamında egoistik ve altruistik (ve yarı altruistik) bireylerin davranışları iktisadi bir kavram olan Ortak Kaynakların Trajedisi-OKT (Tragedy of the Commons) ile açıklanabilir. Keza, altruizmin öz-aşkınlık yerine, kendini genişletme ile yüksek derecede ilişkili çıkması semi-altruistic tutumdan kaynaklanabilir. Söz konusu ilişki bu tez kapsamında incelenmeyecek olup, devam niteliğindeki çalışmalarda değerlendirilebilir (OKT kavramı, 1968 yılında Garrett Hardin tarafından alan yazına kazandırılmış ve teorik alt yapısı Lloyd'un 1833 yılında ortak kaynakların (commons) kullanımı hakkında vermiş olduğu bir örneğe dayanmaktadır).

<sup>26</sup> Çevre dostu bir davranışın başka bir ÇDD'ye neden olması yayılım etkisi "spillover effect" olarak adlandırılmaktadır.

ÇDD'nin tek boyutlu olarak ele alınmasıdır.<sup>27</sup> Bu yaklaşıma göre ise farklı çevresel davranışlar tek bir motivasyonel faktör altında birleşebilir ve bütün ÇDD'ler tek bir amaca hizmet edebilir. Örneğin bu yaklaşıma göre, çevreyi koruma ya da çevre için iyi bir şeyler yapma amacı ile evsel atıkları geri dönüştüren bir kişi, aynı zamanda eşyalarını eskiyene/bozulana kadar kullanır, işe giderken şahsi arabası yerine toplu taşımayı ya da bisiklet kullanır ve elektrikli eşyaları bekleme konumunda bırakmaz. Dolayısı ile bu yaklaşım, çevresel problemleri ve bu problemlerin olası çözümlerinin çok boyutlu olduğunu ve farklı davranışların farklı öncüllerinin olduğunu reddetmekte ve çevre ile ilgili bir davranışın, diğer davranışı tetiklediğini kabul etmektedir.

---

<sup>27</sup> Alan yazınında Campbell paradigması olarak bahsedilmekte olup, matematiksel açıklaması Rasch modeli ile ifade edilmiştir.

## İKİNCİ BÖLÜM

### (VERİ MADENCİLİĞİ ve BİLGİ KEŞFİ)

#### 1. VERİ MADENCİLİĞİ

Veri “Data”, üzerinde herhangi bir işlem yapılmadan bir veya birden fazla kaynaktan elde edilen değerdir (Şeker, 2013). Daha detaylı bir tanımlama ile tecrübe, gözlem, ölçümlene ve deneyler sonucunda elde edilen sayı, kelime ya da görsellerden oluşan değerler kümesidir (Akpınar, 2018). Enformasyon “Information”, verinin işlenmiş halidir (Şeker, 2013). Bilgi “Knowledge”, elde edilen enformasyonun anlamlandırılmasıdır. Bilginin içselleştirilerek tecrübelerle ilişkilendirilmesi, bilinmesi ve bir fayda elde etme amacı ile kullanılabilir durumda olması ise bilgelik “Wisdom” olarak tanımlanmaktadır. Veri-Enformasyon-Bilgi-Bilgellik arasındaki bu ilişki Ackoff (1989) çalışması ile literatüre geçmiştir (Köse, 2018).

VM “Data Mining” çeşitli kaynaklardan çeşitli yöntemlerle toplanan verilerin üzerinde işlem yaparak/bir dizi analiz yöntemini tekrarlı uygulayarak anlamlı ve faydalı bilgi elde etmektir. VM, bilgiye ulaşmak için benimsenmiş olan sistematik bir keşif sürecinin analiz aşamasıdır (Fayyad, Piatetsky-Shapiro, Smyth, 1996a; Tan, Steinbach, Kumar, 2014; Han, Kamber, Pei, 2012). Bir örnekleme kullanarak ana kütleyle temsil eden bir model oluşturmak ya da örüntüleri<sup>28</sup> tespit etmek/keşfetmektir. Söz konusu keşif süreci alan yazınında Veri Tabanlarında Bilgi Keşfi (VTBK) “Knowledge Discovery in Databases-KDD” süreci<sup>29</sup> olarak adlandırılmaktadır. Her ne kadar VM, VTBK’nın sadece bir aşaması olsa da söz konusu sürecin tamamı genellikle VM süreci olarak adlandırılmaktadır.

---

<sup>28</sup> Sürekli devam eden ve tekrar eden yapılardır “pattern” (Şimşek Gürsoy, 2012)

<sup>29</sup> KDD sürecine alternatif olarak IBM firması CRISP-DM ve SAS firması SEMMA süreçlerini aynı amaçla tasarlamıştır. Fakat her iki işletmenin süreç tasarımları KDD ile yüksek derecede benzeşmektedir. KDD, CRISP-DM ve SEMMA süreçlerinin karşılaştırmalı analizine (Azevedo, dos Santos, 2008) çalışmasından erişilebilir.

[https://www.ibm.com/support/knowledgecenter/en/SS3RA7\\_15.0.0/com.ibm.spss.crispdm.help/crisp\\_overview.htm](https://www.ibm.com/support/knowledgecenter/en/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm)  
<http://documentation.sas.com/?docsetId=emref&docsetTarget=n061bzurmej4j3n1jn18bbj1a2.htm&docsetVersion=14.3&locale=en> sitelerinden söz konusu firmaların süreçleri incelenebilir.

Diğer taraftan, VM bir Sorgulama “Query” işlemi değildir. Örneğin, gömlek üreten bir firma, gelecek yıl için kumaş satın alacak olsun ve hangi renk kumaştan ne miktarda sipariş vermesi gerektiğine karar vermek istesin. Bu durumda firma örneğin, yakın geçmişte hangi renk gömlekten kaçar adet satıldığını ve satışlara bağlı olarak her renk kumaştan kaç metre kullanıldığını basit bir filtreleme işlemi yaparak tespit edip bir sonraki yıl için kumaş siparişi verebilir. Bu işlem bir sorgulama işlemidir. Fakat, pazarlama alanında sıklıkla örnek verilen Wallmart market basketi analizi VM’nin başarılı bir şekilde uygulandığı bir bilgi keşfi sürecidir. Söz konusu analizde, cuma günü akşamı bira alan erkek müşterilerin aynı zamanda bebek bezi de aldığı tespit edilmiştir. Başka bir ifade ile alkol alan müşterilerin, alkolün yanında cips, çerez ve türevleri gibi gıdaları alması tahmin edilebilir bir davranış iken, “cuma-akşam-erkek-bira-bebek bezi” ilişkisinin tespiti VM yöntemlerinin<sup>30</sup> kullanıldığı bir keşif sürecidir.

Teknik veya amaç açısından VM ne değildir ya da nedir şeklinde yüzeysel bir değerlendirme yapmak gerekirse veri düzenleme, arama, seçme, gruplama gibi sorgulama işlemleri, veriyi listeleme gibi raporlama işlemleri, maksimum, minimum, standart sapma ve benzeri gibi temel istatistikleri hesaplama ve grafiklerle görselleştirme işlemleri başlı başına birer VM uygulaması değildir. Aksine, verileri kullanarak birbirine benzer kaç küme olduğunun ve hangi gözlemlerin hangi kümelerde bulunduğu belirlenmesi, bir dizi örüntüsüne sahip olan gözlemler arasındaki benzerliklerin tespiti, belirli bir sıklıkta birlikte gözlemlenen belirli nitelikteki olayların hangi oranda birlikte olduklarının tespiti, önbilgi ile türü belirli olan verilerden yola çıkarak yeni verilerin türünün belirlenmesi, iki değişken arasındaki ilişkinin tespit edilerek değişkenlerden birisinin gelecekte varsayılan bir değerine karşılık diğer değişkenin alacağı değer tahmin edilmesi ve belirli niteliklere sahip olan gözlemler arasındaki sıra dışılıkların (anomalilerin) tespit edilmesi gibi işlemler ise birer VM uygulamasıdır (Köse, 2018).

Zhao & Luan (2006)’dan akt. Ersöz (2019)’a göre VM ile istatistik biliminin ortak yönleri olmasına rağmen VM, başlı başına bir istatistiksel analiz süreci değildir. Örneğin, veriyi bilgiye dönüştürmek, veriden öğrenmek, belirsizlikler ile ilgilenmek, bir olayı etkileyen önemli faktörleri belirlemek ve tasarlanan modeller ile gelecekteki olayları

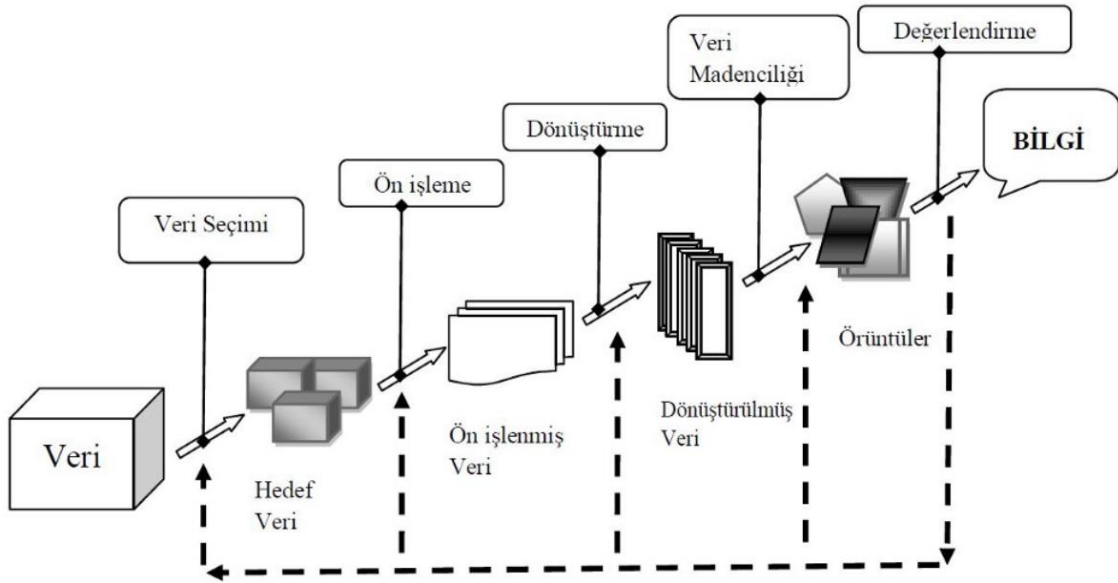
---

<sup>30</sup> VM yöntemi olarak kural çıkarma algoritması kullanılmıştır.

öngörmek hem VM'nin hem de istatistik biliminin ortak yönleri arasında sayılabilir. Diğer taraftan istatistik, bilgiye dayalı teori ile başlar (apriori) ve yokluk hipotezinin reddi/reddedilememesi hakkında kanıt arar. İstatistik keşfediciden çok doğrulayıcı bir süreçtir. VM esas ilgi alanı ise doğrulamak değil keşfetmektir. İstatistiğin amacı aynı anakütle hakkında çıkarım yapmaktır. VM'nin amacı ise yerel "local" veri toplayarak veriye ait özel bilgiyi keşfetmektir. İstatistiksel analizlerde veri kümesinin büyüklüğü birkaç yüz veya bin gözlem iken VM'de, özellikle büyük veri çalışmalarında, milyonlar veya milyarlar olabilmektedir. Fakat verinin temsil gücünün yüksek olduğu durumlarda birkaç bin veya birkaç yüz gözlemlik veri setleri de VM için yeterli olmaktadır.

## 2. BİLGİ KEŞFİ SÜRECİ

Bilgi keşfi süreci beş ana aşamadan oluşmaktadır, Şekil 6. Bu aşamalar, veri seçimi (toplama), veri ön işleme, veri dönüştürme, VM (analiz etme/işleme) ve sonuçları yorumlama şeklindedir (Fayyad, Piatetsky-Shapiro, Smyth, 1996b'den akt.Şeker, 2013; Akpınar, 2014).



Şekil 6. Bilgi keşfi süreci (Savaş, 2020)

### 2.1. Veri Seçimi

Çok sayıda ve çeşitli özellikteki verilerin depolandığı veri tabanından (Savaş, 2020), sadece araştırılan konu ya da üzerinde çalışılan problemle ilgili olduğu düşünülen

verilerin seçildiği ya da verilerin birden fazla veri tabanından derlenip birbirine entegre edildiği/ilişkilendirildiği aşamadır (Akpınar, 2018; Şeker, 2013).

Veri birleştirme işlemi bazen, bir çalışanın personel departmanındaki sigorta, üretim departmanındaki performans ve insan kaynakları departmanındaki eğitim bilgilerinin birleştirilmesi işlemi gibi basit bir işlem olabilirken, bazen de şirket birleşmelerinde iki işletmenin veri tabanlarının birleştirilmesi ya da büyük bir veri setinden uygun bir örnekleme yöntemi ile yeterli miktarda örneklem çekilmesi gibi karmaşık bir işlem de olabilmektedir.

Veri tabanında veriler, genellikle Veri Tabloları veya Örüntü/Öznitelik Matrisleri (ÖM) “Pattern Matrix” şeklinde saklanmaktadır. Gözlemler, kayıtlar ya da nesnelere bu matrislerin satırlarını oluştururken, onlara ait öznitelikler ise sütunlarda yer almaktadır. Öznitelik, bir gözlemi diğerlerinden ayırtan, farklı kılan, onun tanınmasını veya diğer gözlemler ile karşılaştırılmasını sağlayan özelliklerdir.  $N_i$ , bir veri seti ve  $A_j$ , bu veri setini niteleyen öznitelikler olmak üzere  $x_{ij}$ , veri setindeki  $i$ 'inci gözlemin  $j$ 'inci özniteliklerine karşılık gelen değerdir. ( $i = 1, \dots, m$  ve  $j = 1, \dots, n$ ) olmak üzere tüm gözlemler ve gözlemlerin öznitelik değerleri aşağıda yer alan  $(m \times n)$  boyutlu ÖM, Eş.(1)'deki gibi gösterilmektedir.

$$\begin{matrix} & A_1 & \cdots & A_n \\ N_1 & [x_{11} & \cdots & x_{1n}] \\ \vdots & \vdots & \ddots & \vdots \\ N_m & [x_{m1} & \cdots & x_{mn}] \end{matrix} \quad (1)$$

Bir öznitelik değeri bir makinenin yıllık arıza kayıtları, borsada işlem gören bir hisse senedinin gün bazında yıllık fiyat hareketleri veya satış-pazarlama departmanında çalışan bir temsilcinin satış performansı gibi hem sayılar ile temsil edilebilir hem de makinenin ve hisse senedinin türü ya da satış temsilcisinin deneyimli olup olmaması nitelikleri gibi ifadeler ile de sembolize edilebilir. Öznitelik değerinin ne ile sembolize edileceğine, öznitelik doğasına uygun bir ölçme düzeyinin kullanılması ile karar verilmektedir. Ölçme düzeyi, bir nesneyi tanımlayan bir öznitelik değerini sayısal ya da sembolik değerler ile ilişkilendiren bir kuraldır (fonksiyon) (Tan, Steinbach, Kumar, 2014)<sup>31</sup>.

---

<sup>31</sup> Ölçme düzeyleri hakkında ikinci bölüm 4. kısımda detaylı bilgi verilmiştir.



Örnekleme işlemi, veri seçimi aşamasında ele alınan işlemlerdendir (Han, Kamber, Pei, 2012; Tan, Steinbach, Kumar, 2014). Örnekleme, bir anakütleden rassal olarak çekilmiş ve anakütleyi temsil etme yeteneğine sahip daha az sayıdaki gözlemden oluşan bir grup örneği inceleyerek, anakütle hakkında genelleme yapma işlemi olarak tanımlanmaktadır. Kalite kontrolde, kamuoyu araştırmalarında, gelir dağılımı, tüketim ve pazarlama araştırmalarında, seçim tahminlemede ve psikolojik ve demografik birçok araştırmada (Serper, 2014; Şenesen, 2013) tüm anakütleyi<sup>32</sup> incelemek ya da tamsayım<sup>33</sup> yapmak yerine örnekleme yönteminden faydalanmak, daha kısa sürede, daha az maliyet ile ve daha kolay bir şekilde araştırılan konu hakkında bilgi edinmeyi sağlar.

Basit Rassal Örnekleme “Simple Random Sampling”, Tabakalı Örnekleme “Stratified Sampling”, Sistemik Örnekleme “Systematic Sampling”, Kümelere Göre Örnekleme “Cluster Random Sampling” ve Çok Kademeli Örnekleme “Multi-Stage Sampling” olmak üzere çeşitli örnekleme yöntemleri bulunmaktadır.

Basit rassal örneklemede her gözlem şansa bağlı olarak çekilir. Bu yöntemde her gözlemin örnekleme girme şansı birbirine eşittir.

Tabakalı örnekleme, anakütlerdeki gözlemler öznelikleri açısından farklılaştığı zaman (homojen olmadığı durumlarda) kullanılmaktadır. Bu durumda anakütle, birbirine benzer tabakalara (katmanlara) ayrılmakta ve gözlemler her bir tabakadan rassal şekilde seçilerek örneklem kümesi oluşturulmaktadır. Örneğin tüketim harcamaları araştırmalarında alt-orta-üst düzey gelir gruplarından (tabakalarından) belirli sayıda gözlem seçilerek alt örneklemler oluşturulmalıdır.

Sistemik örneklemede, ilk gözlem hariç örnekleme girecek diğer gözlemler bir kurala bağlı olarak (rassal olmayacak şekilde) seçilmektedir.  $N$  anakütle hacmi olmak üzere, anakütle bir  $k$  sayısına bölünür. Daha sonra 1 ile  $k$  arasında bir sayı belirlenir ve bu sayıya karşılık gelen gözlem, örnekleme girecek ilk gözlem olur. Akabinde ilk sayıya here seferinde  $k$  eklenerek örnekleme girecek olan diğer gözlemler tespit edilir. Dolayısı

---

<sup>32</sup> Hakkında bilgi elde edilmek istenen ve homojen birimlerden oluştuğu varsayılan topluluktur. Örneklem ise bu topluluğun bir alt kümesidir.

<sup>33</sup> Örneğin nüfus sayımları ya da motorlu taşıt kayıtları sosyal bilimler alanında tamsayım işlemine verilebilecek örneklerdendir. Fen bilimlerinde ise sıklıkla gerçekleştirilen numune alma işlemi bir örnekleme işlemidir.

ile anakütle hem  $m$  adet kümeye ayrılmış hem de sıra numaraları  $k$ 'nın katları olan gözlemlerden toplamda  $m$  adetlik örnekleme işlemi tamamlanmış olur.

Kümelere göre örneklemede, tabakalı örneklemedeki gibi, anakütle kendi içinde homojen olan kümelere (tabakalara) ayrılır ve gözlemler, tabakalı örneklemenin aksine, her kümeden seçilmez. Yerine, kümeler arası seçim gerçekleştirilir. Örneğin anakütle kendi içinde benzeşen dokuz kümeye ayrılıyor ise bu kümelere dört ya da beş tanesinde (veya daha azında) yer alan gözlemlerin ya tamamı örnekleme alınır ya da söz konusu alt kümelere rassal bir şekilde gözlemler seçilerek örneklem oluşturulur. Bu yaklaşım iki aşamalı kümelere göre örnekleme olarak da adlandırılır. Kümelere göre örneklemede sadece belirli nitelikteki kümelere yer alan gözlemler örnekleme alındığı için elde edilen örneklemin anakütleyi temsil etmeme riski ve örnekleme hatasının büyük olma olasılığı bulunmaktadır.

Çok kademeli örnekleme ise yukarıda sayılan örnekleme yöntemlerinin bir veya birkaçının birlikte kullanılması durumudur. Örneğin sadece sanal mağazası bulunan bir işletme Türk halkının internette alışveriş yapma alışkanlıkları üzerine bir araştırma gerçekleştirmek istesin. İşletme bu durumda örneklemini kademeli olarak belirleyebilir. Tüketiciler öncelikle coğrafi bölgelere göre, daha sonra her bir bölge kendi içinde gelir gruplarına göre ve en sonunda her bir gelir grubu kendi içinde kuşaklara göre tabakalanabilir. Akabinde, her bir tabakadan belirli miktardaki gözlem rassal olarak seçilebilir.

Sayılan örnekleme yöntemleri rassallığa/olasılığa dayalı yöntemlerdir. Örneklemede rassallığa dayalı olmayan kolayda örnekleme, amaca dayalı örnekleme, kota örnekleme ve kartopu örnekleme yöntemleri de bulunmaktadır. Kolayda örnekleme araştırmacının kendi çevresini (ve belki bu çevrenin de çevresini) kapsayan ve anakütleyi temsil ettiği düşünülen gözlem birimlerine ulaşarak gerçekleştirdiği örnekleme yöntemidir. Bu yöntem ile elde edilecek olan verilerin yanlı olma ihtimali çok yüksektir ve oluşturulan örneklemin tüm anakütleyi temsil etme olasılığı da çok düşüktür. Bu yüzden, kolayda örnekleme kullanılarak oluşturulan örneklemler üzerinde yapılacak analizler sonucunda elde edilecek bilgilerin anakütleyle genellenmesi tavsiye edilmemektedir. Amaca dayalı örneklemede gözlem birimleri araştırmacının amacı ile paralel olarak tespit edilmektedir. Örneğin 0-24 ay bebek bakım malzemeleri üzerine

yapılan bir araştırmaya bebeği/torunu olmayan ya da bebek bakımıyla ilgilenmeyen kişilerin dahil edilmemesi gibi. Kota örneklemede belirli bir standart referans alınarak örneklem oluşturulur. Örneğin bir işyerinde çalışanların %45'i erkek ve %55'i kadın ise oluşturulan örneklemin cinsiyet dağılımının da bu orana sahip olması durumu gibi. Son olarak kartopu örnekleme ise genellikle anakütle hacminin belirsiz olduğu durumlarda kullanılmaktadır. Bu yöntemde araştırmaya katılarak örnekleme dahil olan ilk kişinin/birimin yönlendirmesi ile diğer kişilere/birimlere ve bu kişi/birimlerden diğerlerine olacak şekilde bir kartopu gibi kitle büyütülerek örnekleme işlemi gerçekleştirilir. Örneğin toplumdan kendini soyutlamış olan AIDS hastaları, uyuşturucu bağımlıları ya da bir gruba/çeteye ait olan bireyler hakkında yapılacak olan araştırmalarda ya da sosyal ağ analizinde kartopu örnekleme sıklıkla kullanılmaktadır.

Herhangi bir anakütleden herhangi bir yöntem ile örnekleme seçerken iadesiz ve iadeli seçim olmak üzere iki farklı yaklaşımdan faydalanılmaktadır (Serper, 2014; Tan, Steinbach, Kumar, 2014). Bu iki yaklaşım arasındaki fark, seçilmiş olan bir gözlem biriminin örnekleme işlemine sırası ile sadece bir defa veya daha fazla defa dahil edilip edilmemesi durumudur. Örnekleme işleminde gözlemlerin birbirinden bağımsız olarak seçilmesi gerekmektedir. Başka bir ifade ile bir gözlemin örnekleme dahil olma olasılığı diğer gözlemlerin örnekleme dahil olma olasılığını etkilememesi gerekmektedir. Bu yüzden örneklemede ideal olarak iadeli seçim yapılması gerekmektedir. Fakat anakütle hacminin çok büyük ve örneklem hacminin çok küçük olduğu durumlarda seçim iadesiz yapılmış olsa bile seçim işleminin bağımsız olduğu/seçimin iadeli yapılmış olduğu varsayılır (Serper, 2014).

## **2.2. Veri Ön işleme**

Han, Kamber, Pei (2012)'ye göre kaliteli veri doğru, tam ve tutarlı veridir. Veri boyutunun artması, veri kaynaklarının ve veri toplama yöntemlerinin çeşitlenmesi verinin kalitesizleşmesine (yanlış, eksik ve tutarsız olmasına) neden olabilmektedir. Veri analizinde kullanılan yöntemler ne kadar gelişmiş olursa olsun veri setinin kalitesi düşük olduğu sürece elde edilecek olan sonuçlar yanıltıcı olmaktadır. Bu durum veri biliminde GIGO "Garbage in, garbage out" olarak çağırılmaktadır. Bu yüzden ya kaliteli nitelikte veri elde edilmelidir ya da eksik, tutarsız ya da yanlış bilgi içeren veri seti, çeşitli yöntemler kullanılarak VM aşamasında uygulanacak analizlere hazırlanmalıdır. Söz

konusu hazırlık aşaması önışleme aşaması olarak adlandırılmaktadır. Amaç, verileri istatistiksel olarak sağlıklı hale getirmektir. VM'nin en önemli aşaması olup bütün VM sürecinin neredeyse %60-75'ine tekabül etmektedir (Şentürk, 2006).

### 2.2.1. Tekrarlı Veri

Veri seti içinde birbirinin aynısı birden fazla gözlemin bulunması durumudur (Akpınar, 2018; Tan, Steinbach, Kumar, 2014). Bir bankanın veri tabanında aynı müşterinin birden fazla ikametgah adresi olması (Tan, Steinbach, Kumar, 2014) ya da özellikle kadın müşterilerin kayıtlarında sıklıkla yaşanan evlilik öncesi ve sonrası soyadı değişikliği dolayısı ile aynı müşterinin sanki iki farklı müşteriymiş gibi kayıtlanması tekrarlı veri problemlerindedir.

### 2.2.2. Eksik Veri

Eksik veri, veri setinde bazı gözlemlerin bir veya birkaç özniteliğine ait değerlerinin var olmaması durumudur. Örneğin, bir imalat işletmesinin ürün gamındaki parçaların bazılarının malzeme türü bilgisinin ya da ürün müşteri kodunun var olmaması gibi. Bu durum Eksik Veri “Missing Data” olarak adlandırılır.

Eksik veri olması durumunda, çeşitli Veri Tamamlama “Data Imputation” yöntemleri kullanılarak eksik olan değerler yerine yeni değerler üretilebilmektedir (Han, Kamber, Pei, 2012; Şeker, 2013; Tan, Steinbach, Kumar, 2014). Alan yazınında çeşitli veri tamamlama yöntemleri bulunmaktadır.

Liste Boyunca Silme “List Wise Deletion” yönteminde, eksik verisi bulunan gözlem bütün satır boyunca silinir.

		<i>Ciro</i>	<i>Adet</i>	<i>Br. Maliyet</i>
<b>Örnek 1:</b>	<i>A</i> ürünü	50,000	50	500
	<i>B</i> ürünü	65,000	$x_{22}$	450
	$M = C$ ürünü	100,000	$x_{32}$	$x_{33}$
	<i>D</i> ürünü	250,000	100	1000

$M$  matrisi, üç nitelikle tanımlanmış olan dört gözlemden oluşan bir veri setine ait ÖM'dir Bu matrise liste boyunca silme yöntemi uygulanırsa veri setindeki ikinci ve üçüncü gözlem, diğer bir ifade ile B ve C ürünlerine ait gözlemler silinmelidir.

Eşlerin Silinmesi “Pairwise Deletion” yönteminde üzerinde işlem yapılacak olan özniteliğin hangi gözlemlerdeki değeri eksik ise sadece o gözlemler silinir. Örneğin bir analist,  $M$  matrisini kullanarak A, B, C ve D ürünlerini üreten firmanın ortalama birim maliyetini hesaplamak isterse ve ES yöntemini uygulamaya karar verirse sadece üçüncü gözlemi (C ürününe ait gözlemi) silmesi gerekmektedir.

Global Bir Sabit ile Tamamlama “Imputation with a Global Constant” yönteminde ise eksik gözlemi bulunan öznitelik dizilerinde bütün eksik gözlemler yerine sabit bir değer girilir. Örnek 1’de yer alan “adet” özniteliğindeki eksik gözlemlerin hepsine “bilinmiyor” değerinin verilmesi gibi. Bu yöntem uygulanabilirlik açısından kolay olmasına rağmen güvenilir sonuçlar sağlamayacağı için tercih edilmesi önerilmemektedir.

Ortalama ile Tamamlama “Mean Imputation” yönteminde eksik olan veri, ait olduğu öznitelik değerlerinin ortalama değeri hesaplanarak tamamlanır. Örnek 1’de yer alan  $x_{33}$  eksik verisi,  $ortalama\ maliyet = (500 + 450 + 1000)/3 = 650$  şeklinde hesaplanarak tamamlanır.

Medyan ile Tamamlama “Median Imputation” yönteminde eksik olan veri, ait olduğu öznitelik değerlerinin medyan değeri ile tamamlanır. Örnek 1’de yer alan  $x_{33}=500$  olacak şekilde tamamlanır.

Son Gözlemin Taşınması “Last Observation Carried Forward” yönteminde eksik veriler, kendilerinden bir önceki verinin aynısı ile tamamlanır. Örnek 1’deki  $M$  matrisinde yer alan  $x_{33}$  eksik verisinin, kendisinden bir önceki veri olan 450 ile tamamlanması gibi.

Han, Kamber, Pei (2012)’ye göre yukarıda sayılan yöntemler veri setinde yanlılığa “bias” neden olabilmektedir. Yazarlar, sayılan bu yöntemler yerine, mevcut veri seti kullanılarak Regresyon “Regression Imputation”<sup>34</sup>, Bayesçi Çıkarım “Bayesian Inference” ve Karar Ağaçları “Decision Trees” gibi yöntemler ile eksik veriler için değer üretilmesini önermektedir. Çünkü bu yöntemler, yukarıda sayılan yöntemlere kıyasla, mevcut veri setindeki enformasyonun büyük bir kısmından faydalanarak değer

---

<sup>34</sup> İkinci bölüm 2.2.4 numaralı kısımda aykırı olarak tespit edilip ayıklanan gözlemler için regresyon ile değer üretimi hakkında küçük bir uygulama örneği verilmiştir.

üretmesini sağlamaktadır. Zira, eksik verilerin tamamlanması işlemi VM’de başlı başına bir çalışma alanıdır.

### **2.2.3. Hatalı Veri**

Bazı verilerin hatalı olarak kayıtlanmış olması durumu gürültü olarak değerlendirilmektedir. Gürültü, ölçülen bir değişkende gözlemlenen rassal hata veya varyans olarak tanımlanmaktadır (Han, Kamber, Pei, 2012; Tan, Steinbach, Kumar, 2014). Örneğin personelin maaşının sistemde negatif değerli olarak kayıt edilmesi gibi veriyi oluşturan ya da toplayan kişilerden (insan) kaynaklı bir hata olabilir. Kullanılan ölçme aracı uygun olmayabilir ve dolayısı ile araştırılan konu ile alakalı olmayan veriler toplanmış olabilir. Veri dönüştürme aşamasında yanlış boyutlar seçilmiş ya da çıkarılmış olabilir. Eğer gözlem birimi insan ise, araştırmaya katılan bireyler yanlış/yanıltıcı veri sağlamış olabilir. Sayılan bu durumlarda elde edilmiş verilerin hepsi Hatalı Veri “Erroneous Data” olarak adlandırılır. Genellikle tesadüfi hataları temsil eden insan kaynaklı hatalar ve genellikle sistematik hataları temsil eden ölçme aracı kaynaklı hatalar (Akpınar, 2014) birçok veri setinde gözlemlenebilmektedir. Bu tür hatalar istatistik biliminde Parazit/Beyaz Gürültü “White Noise” olarak adlandırılmaktadır (Akpınar, 2018; Tan, Steinbach, Kumar, 2014).

### **2.2.4. Aykırı Veri**

Veri setindeki bazı gözlemlerin bir veya birden fazla özniteliklere ait verilerinde uyuma problemi varsa veya bazı veriler hatalı ise söz konusu veriler Aykırı Veri/Uç Değer, “Outlier” olarak nitelendirilmektedir (Han, Kamber, Pei, 2012; Şeker, 2013; Tan, Steinbach, Kumar, 2014). Aykırı veri, ait olduğu veri setindeki diğer gözlemlerden önemli ölçüde farklı olan veri noktasıdır. Örneğin bir Serpilme Diyagramı “Scatter Plot” çizildiğinde negatif işaretli bir maaş gözleminin, diğer gözlemlerden çok uzak bir konumda belirmesi ya da bir küme analizi yapıldığında rahim ameliyatı geçirmiş ve erkek cinsiyetli bir hastaya ait gözlemin, diğer gözlemlerden farklı bir konumda ve tek başına bir küme oluşturması gibi. Fakat aykırı veriler her zaman hatalı veriler olmayabilir. Örneğin kredi kartı kullanarak yurtiçi online alışveriş yapan bir banka müşterisi olsun. Bu müşterinin bireysel kredi kartı hesabı hareketlerinde yabancı internet sitelerinden aynı anda yapılan çok sayıda harcama tespit edilsin. Söz konusu harcama verileri banka için aykırı veri niteliğindedir. Çünkü bu veriler, müşterinin harcama alışkanlıkları ile

uyuşmayan bir niteliktedir. Diđer bir ifade ile örneđin, müşterinin kredi kartı harcama verileri kullanılarak bir küme analizi yapıldığında, yurtdışı sitelerden aynı anda yapılmış olan söz konusu harcamalar diđer olađan harcamalardan uzak (fakat kendi içinde yoğun) ayrı bir küme oluşturacaktır. Bu durumda banka, bu sıra dışı kredi kartı kullanım hareketlerinin nedenini tespit etmelidir.<sup>35</sup>

Hatalı ve aykırı verilerin tespit edilmesinde ise tanımlayıcı istatistiksel yöntemlerin ve veri görselleştirme yöntemlerinin kullanılması önerilmektedir (Han, Kamber, Pei, 2012). Örneđin, Beş Deđerli Özet “Five-Number Summary” ile veriye ait merkezi dağılım ölçütlerinden medyan (ikinci kartil), birinci ve üçüncü kartil ve minimum ve maksimum gözlem deđerleri kullanılarak Kutu-Bıyık Grafiđi “Box-Whisker/Boxplot” veya Kartil Grafiđi “Quantile-Quantile Plot” çizilebilir.

Kutu-bıyık grafiđi, aykırı verilere duyarlı bir grafikdir. Bu yüzden, grafiđin çiziminde aykırı gözlemleri tespit edebilmek için bıyıkların başlangıç ve bitişlerini temsil eden minimum ve maksimum deđerlerin veri setindeki gerçek minimum ve maksimum gözlemler olmaması tavsiye edilmektedir. Yerine, bıyıkların uç noktalarının, kutuların başlangıç ve bitiş noktalarından ( $Q1$  birinci ve  $Q3$  üçüncü kartil deđerlerinden) 1,5 kartiller arası genişliğe/mesafeye “Interquartile Range” (IQR) kadar uzatılması önerilmektedir Eş. (2, 3).

$$Q1 - 1.5 \times IQR \quad (2)$$

$$Q3 + 1.5 \times IQR \quad (3)$$

Bu şekildeki bir uygulama ile veri setindeki uç gözlemler birinci veya üçüncü kartilden, IQR’ın 1,5 katı kadarlık mesafe içinde kalır ise söz konusu veri setinin aykırı deđer içermediđi ifade edilebilir. Bir başka uygulama ise Eş. (-) ile tanımlı olan aralıklardan  $1.5 \times IQR$  sonraya kadar olan aralıktaki deđerlerin orta derecede aykırı deđer ve üstünün şiddetli aykırı deđer olarak deđerlendirilmesi şeklindedir (Akküçük, 2011). Bu uygulamaya göre şiddetli aykırı deđerler Eş. (4, 5) ile tanımlanan aralıkların sırası ile altında ve üstünde kalan deđerlerdir.

---

<sup>35</sup> VM alan yazımında dolandırıcılık/sahtekarlık tespiti “Fraud Detection” olarak adlandırılmaktadır.

$$Q1 - 3xIQR \quad (4)$$

$$Q3 + 3xIQR \quad (5)$$

**Örnek 2:** Bir grup satış temsilcisinin yıllık toplam satış performansları 1000 TL cinsinden 20, 28, 40, 60, 76, 84, 84, 88, 88, 92, 92, 92, 92, 92, 96, 96, 96, 96, 100, 100, 105, 115, 120, 120 şeklinde olsun. Satış verilerinin ortalaması (86,33333), standart sapması (25,63229), histogram ve yoğunluk fonksiyonları incelendiğinde gözlemlerin normal dağılmadığı, sola çarpık dağılıma sahip olduğu ve aykırı veriye sahip olabileceği Şekil 7’den de öngörülebilmektedir. Şekil 8 (a)’dan kutu-bıyık grafiği incelendiğinde ise üç gözlemin grafikte aykırı veri olarak yer aldığı ve ikinci ve üçüncü kartilerin birbirlerine çok yakın olduğu (verinin sola çarpık olduğu) rahatlıkla gözlemlenebilmektedir. Bıyıkların uç noktalarının tespitinde IQR’ın 1,5 katı kuralı uygulanmadığında ise Şekil 8(b)’den de inceleneceği üzere satış dizisinin aykırı değere sahip değilmiş gibi dağıldığı gözlemlenmektedir.

```
#Aykırı değer içeren gözlem dizisi
satis=c(20, 28, 40, 60, 76, 84, 84, 88, 88, 92, 92, 92, 92, 92, 96, 96, 96, 96, 100,
100, 105, 115, 120, 120)

#Ortalama ve standart sapma
mean(satis)
sd(satis)

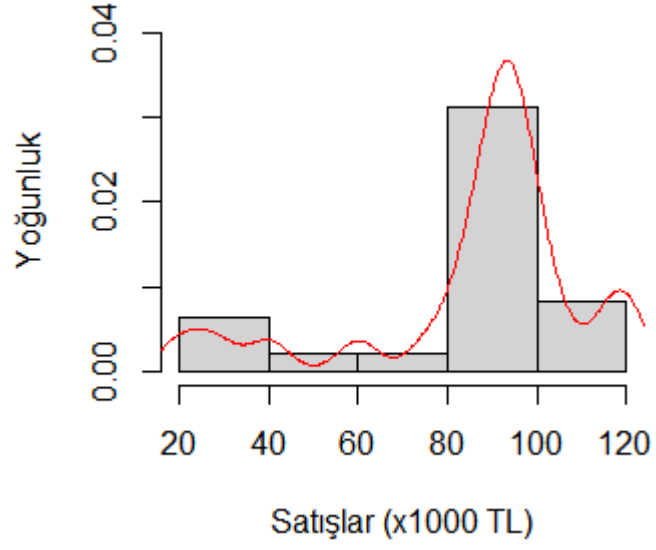
#Histogram ve yoğunluk fonksiyonu
hist(satis, prob=TRUE, col="lightgray", ylim=c(0,.04), xlab="Satışlar (x1000 TL)",
ylab="Yoğunluk")
lines(density(satis), col="red")
#Kutu-Bıyık grafiği - range=1,5*IQR default olarak belirlenmiş
boxplot(satis)
boxplot.stats(satis)
#Kutu-Bıyık grafiği range=0 olur ise
boxplot(satis,range = 0)
boxplot.stats(satis)
```

Kartil grafikleri incelendiğinde de orijinal veri dizisinin normal bir dağılıma sahip olmadığı Şekil 9.’dan de anlaşılabilir. Şekil 9.’dan de anlaşılabilir.

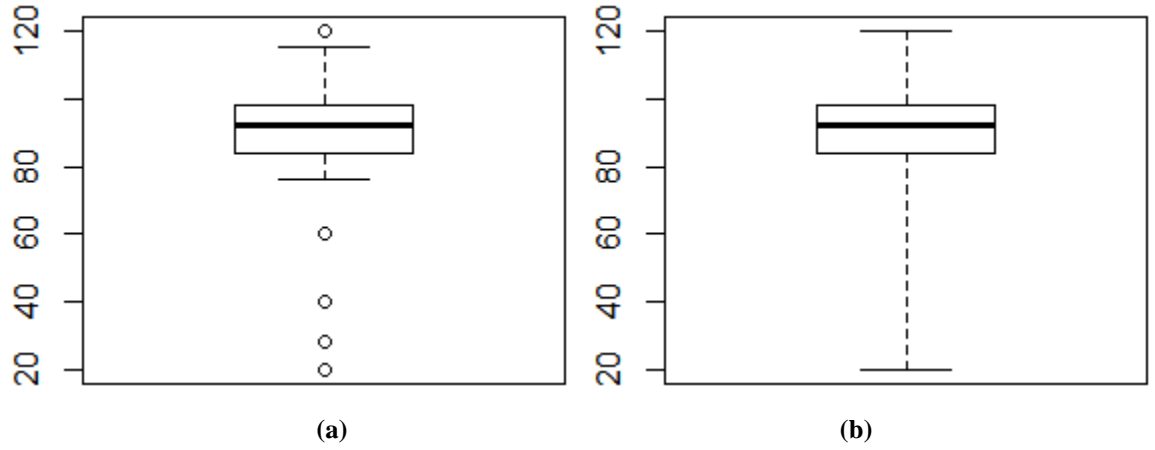
```
#q-q grafiği
qqnorm(satis, xlab="Teorik Kartiller", ylab = "Satışlar (x1000 TL)")
```



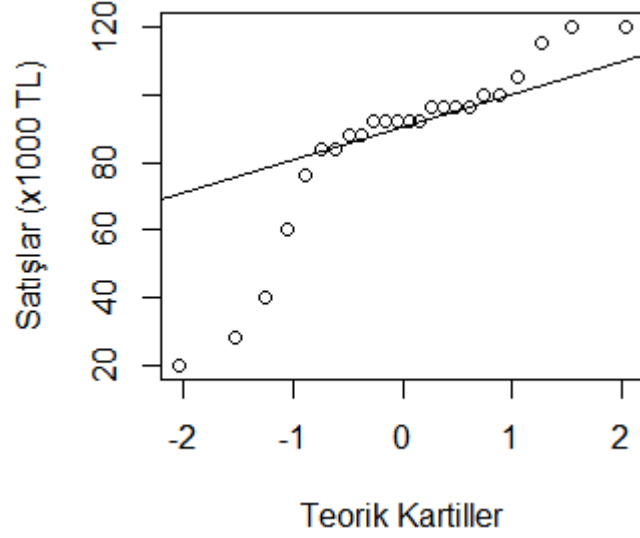
qqline(satis)



Şekil 7. Örnek 2'nin histogram ve yoğunluk grafiği



Şekil 8. Örnek 2'nin kutu-bıyık diyagramı bıyık uzunluklarının IQR'nın 1.5 katı kuralı (a) dikkate alınarak ve (b) alınmayarak tespit edilmesi



**Şekil 9.** Örnek 2'nin teorik kartiller dikkate alınarak tespit edilen aykırı veriler grafiği

Aykırı olarak tespit edilmiş olan gözlemler diziden çıkartılarak temizlenmiş olan satış dizisi grafiksel olarak tekrar incelendiğinde gözlemlerin aritmetik ortalamasının yükseldiği (92,29412), standart sapmasının azaldığı (6,962082), normale daha yakın bir dağılıma sahip olduğu Şekil (10-11)'den incelenebilmektedir.<sup>36</sup>

#Aykırı değer içermeyen gözlem dizisi

```
satis2=c(76, 84, 84, 88, 88, 92, 92, 92, 92, 92, 96, 96, 96, 96, 100, 100, 105)
```

#Ortalama ve standart sapma

```
mean(satis2)
```

```
sd(satis2)
```

#Histogram ve yoğunluk fonksiyonu

```
hist(satis2, prob=TRUE, col="lightgray", xlab="Satışlar (x1000 TL)", ylab="Yoğunluk")
```

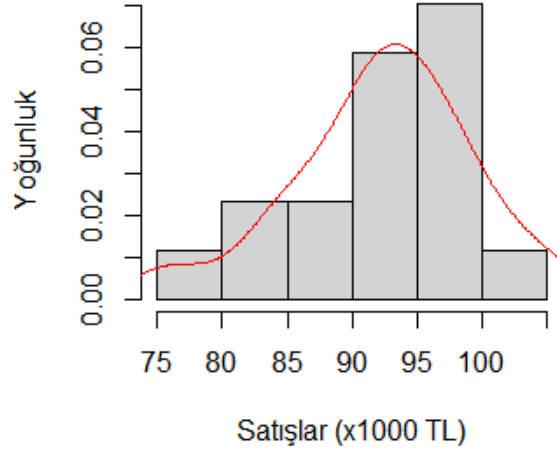
```
lines(density(satis2), col="red")
```

#Kutu-Bıyık grafiği - range=1,5\*IQR default olarak belirlenmiş

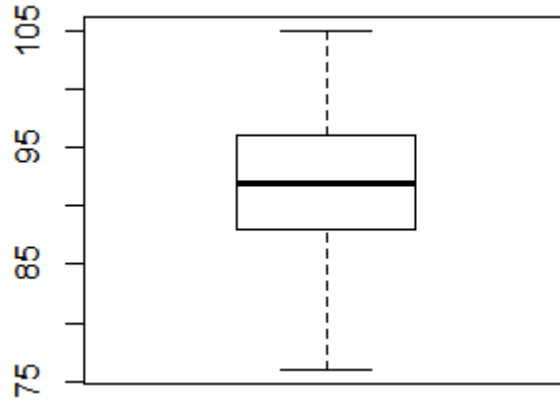
```
boxplot(satis2)
```

```
boxplot.stats(satis2)
```

<sup>36</sup> Aykırı değer içeren yıllık toplam satış dizisinden 20, 28 ve 40 değerli gözlemler (ilk üç gözlem) çıkartılmış, histogram, yoğunluk ve kutu-bıyık grafikleri tekrar incelenmiştir. Bu incelemeler sonucunda 60 değerli olan dördüncü gözlem de aykırı olarak tespit edildiği için bu gözlem de veri setinden çıkartıldıktan sonra elde edilen yeni diziyeye ait grafikler tekrar oluşturulmuş ve Şekil (10-11) elde edilmiştir.



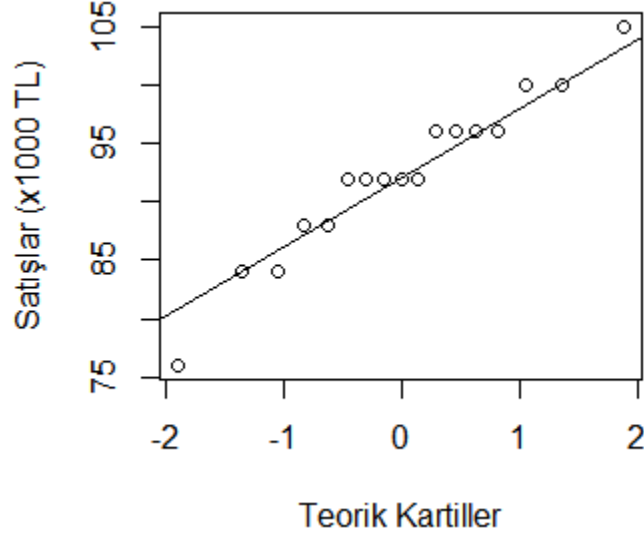
**Şekil 10.** Örnek 2'nin veri setinden aykırı gözlemler çıkartıldıktan sonra elde edilen histogram ve yoğunluk grafiği



**Şekil 11.** Örnek 2'nin veri setinden aykırı gözlemler çıkartıldıktan sonra elde edilen kutu-bıyık diyagramı

Aykırı veriler temizlendikten sonra elde edilen kartil grafiği incelendiğinde, gözlemlerin normallik doğrusu etrafında ve doğruya yakın bir alanda saçılmış olmasından dolayı temizlenmiş veri dizisinin normale daha yakın bir dağılıma sahip olduğu Şekil 12.'den söylenebilir.

```
#q-q grafiği
qqnorm(satis2, xlab="Teorik Kartiller", ylab = "Satışlar (x1000 TL)")
qqline(satis2)
```



**Şekil 12.** Örnek 2'nin veri setinden aykırı gözlemler çıkartıldıktan sonra teorik kartiller dikkate alınarak elde edilen dağılım grafiği

Çok değişkenli veri setlerinde ise Konveks Kabuk (2 ve 3 boyutlu setlerde kullanılmakta) özellikle mekânsal veri setleri için önerilen bir yöntemdir. Kabuğun dış katmanlarındaki gözlemler aykırı veri olarak nitelendirilmektedir. Ayrıca K-En Yakın Komşuluk gibi uzaklık tabanlı, maksimum olabilirlik gibi olasılık tabanlı ve DBSCAN ve OPTICS gibi yoğunluk temelli yaklaşımlar da yaygın olarak kullanılmaktadır. Bu yaklaşımlarda ise, veri setindeki boyut/değişken sayısı arttıkça arama uzayı seyrekleşmekte ve uzaydaki her bir gözlem neredeyse aykırı değer olarak tespit edilme olasılığı bulunmaktadır (Akpınar, 2018; Tan, Steinbach, Kumar, 2014). Özetle, çok boyutlu veri setlerinde aykırı değer tespiti başlı başına bir çalışma alanı olup, bu alanda gerçekleştirilen analizlerde kullanılan yöntemler aslında küme analizlerinde kullanılan yöntemler ve onların aykırı değer tespiti için geliştirilmiş versiyonlarıdır.

Hatalı veya aykırı veri içeren veri setlerinin düzenlenmesi için çeşitli yöntemler bulunmaktadır (Şeker, 2013). Bu yöntemlerin ilki Veri Kutulama “Data Binning” yöntemidir. Bu yöntem, elektrik, elektronik ve telekomünikasyon gibi alanlarda kullanılan Veri Nicemleme “Data Quantization” ya da Veri Ayrıklaştırma “Data Discretization” yöntemine benzemektedir. Veri nicemleme yönteminde, sürekli nitelikte olan sinyal verisi kesikli niteliğe dönüştürülmektedir. Veri kutulamada<sup>37</sup> “Data Binning”

<sup>37</sup> Demetleme olarak da isimlendirilmektedir (Şentürk, 2006).

ise sıralanan veriler eşit boyutlarda gruplar elde edecek şekilde kutulanır ve her bir kutu için tek bir değer atanır. Başka bir ifade ile her bir kutu içinde yer alan her bir gözlemin değeri birbirine eşitlenir/aynılanır. Böylelikle bir kutu içinde yer alan gözlemlere ait veriler, tek bir veriye indirgenmiş olur ve bir kutu içinde kalan örneğin hatalı, çelişkili, aykırı ya da hatta eksik verinin tüm veri setini etkilemesi önlenmiş olur. Çünkü söz konusu problematik verinin değeri, kutulama sonrasında ya kutudaki verilerin medyanı, ya da kutu sınır değerlerinin ortalaması ile değiştirilmiş olur.

İkinci yöntem, kümeleme “Clustering” yöntemidir. Bu yöntemde veriler, veri setine uygun<sup>38</sup> bir kümeleme yöntemi ile kümelenecek, hiçbir kümeye ait olmayan gözlemler elenir. Böylelikle genel bir ifade ile gürültülü olarak tanımlanan aykırı, eksik, hatalı ve uyumsuz verilerin, tüm veri setini ve devamında gerçekleştirilecek olan analizleri etkilemesi engellenmiş olur.

**Örnek 3:** Örnek 2.’deki satış verisi  $k = 2$  için k-Ortalamlar yöntemi kullanılarak kutulandığında, kutulama işlemi sonucunda gözlem değerleri (başka bir ifade ile her bir satış temsilcisine ait satış değişkeni/özniteliği) sürekli nitelikten sıralı niteliğe dönüştürülmüştür. Bu dönüşüm satış temsilcilerinin aynı zamanda iki kümeye ayrılmasını sağlamıştır.

Aşağıda yer alan R-Studio Console çıktısından da inceleneceği üzere  $k = 2$  için ilk dört gözlem birinci kümeye ve devamındaki gözlemler ikinci kümeye atanmıştır. Kutulama sonucunda elde edilen enformasyon, kutu-bıyık, kartil ya da histogram grafiklerinden elde edilen enformasyon ile birleştirildiğinde birinci kümedeki gözlemlerin aykırı değerlere sahip olan bir grubu temsil ettiği, satış ekibinin genel başarısını temsil etmediği ve bu yüzden söz konusu satış temsilcilerinin ayrıca değerlendirilmesi veya söz konusu temsilcilerin satış portföyündeki ürünlerin ya da sorumlu oldukları pazarların incelenmesi ve belki de yarı yapılandırılmış görüşmeler ile satış temsilcilerinin başarısızlıklarının nedenlerinin araştırılması gerektiği bilgisine ulaşılabilmektedir. Bu ve benzeri gibi temel bilgilere kutu sayısı arttırılarak ulaşılabılır.

---

<sup>38</sup> Uygunluk ifadesi ile kümelenecek verideki değişken türleri dikkate alınarak bir kümeleme algoritmasının benimsenmesi kastedilmiştir. Çünkü sürekli, ordinal ve nominal türdeki değişkenlere uygulanabilecek kümeleme algoritmaları ve uzaklık ölçütleri birbirinden farklıdır.



Call:

```
lm(formula = satis ~ yıl)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.521	-3.106	-0.394	3.248	8.449

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3496.8669	341.2546	-10.25	6.13e-09 ***
yıl	1.8060	0.1715	10.53	4.02e-09 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.423 on 18 degrees of freedom

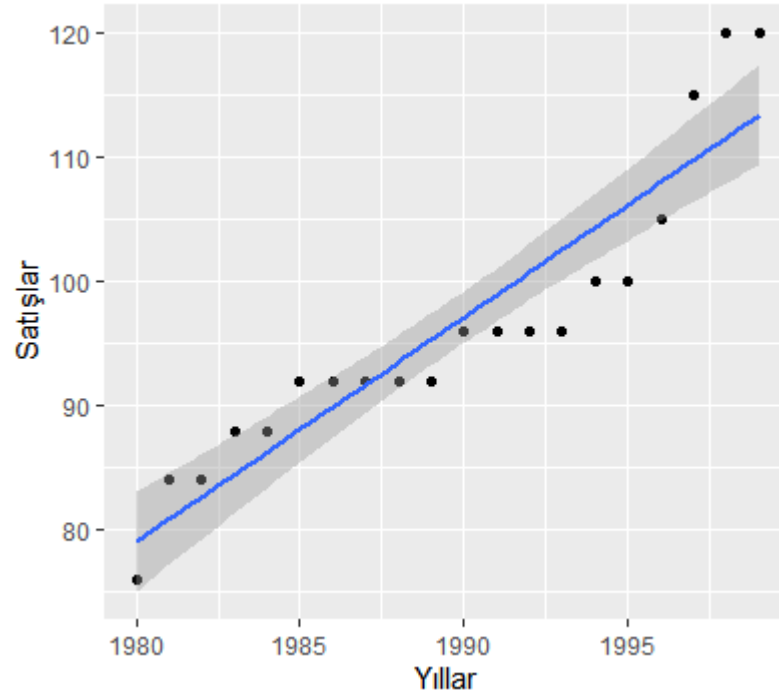
Multiple R-squared: 0.8603, Adjusted R-squared: 0.8526

F-statistic: 110.9 on 1 and 18 DF, p-value: 4.017e-09

#Fonksiyonun grafiği çizilir Şekil 13

```
p=ggplot(m, aes(x=yıl, y=satis))
```

```
p+geom_point()+stat_smooth(method = lm)+labs(x="Yıllar", y="Satışlar")
```



Şekil 13. Örnek 3'ün saçılım grafiği

```
#Tahmin edilecek yıllar oluşturulur
```

```
yeni=data.frame(yıl=1976:1979)
```

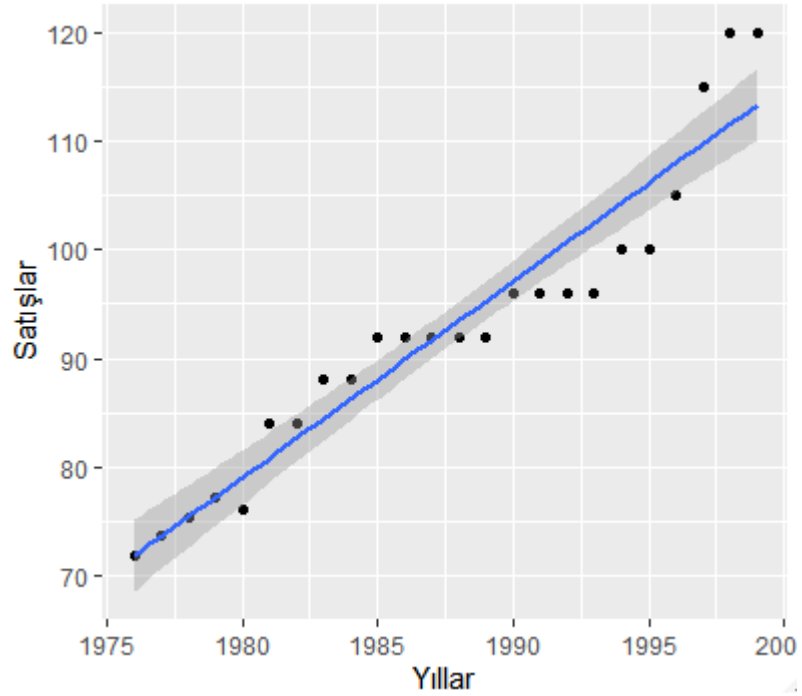
```

#Dört yıl geriye ekstrapolasyon yapılır
pr=as.data.frame(predict(lm(satis~yil),yeni))
#Tahmin edilen satış değerleri, yıllar ile birleştirilir
birlesik2=cbind(yeni,pr)
birlesik2
      yıl   satis
1 1976 71.81880
2 1977 73.62481
3 1978 75.43083
4 1979 77.23684

#Elde edilen data frame'in sütun isimleri orijinal veri setinin sütun isimleri ile
aynı hale getirilir (Aksi halde, sütun isimleri farklı olduğunda, birleşmez)
names(birlesik2)[1]="yil"
names(birlesik2)[2]="satis"
#Ayıklanan aykırı veriler için oluşturulan yeni veri seti, orijinal veri seti ile
birleştirilir
son=rbind(birlesik2,birlesik)

#Yeni grafik çizilir Şekil 14
p1=ggplot(son, aes(x=yil, y=satis))
p1+geom_point()+stat_smooth(method = lm) +labs(x="Yıllar", y="Satışlar")

```



**Şekil 14.** Örnek 3.'ün veri setinden aykırı gözlemler çıkartıldıktan sonra, çıkartılan gözlemler yerine regresyon ile yeni verinin elde edilmesi



### 2.3. Veri Dönüştürme

Önişlenmiş verinin, çeşitli dönüştürme yöntemlerine tabi tutularak VM'ye (işlemeye/analize) hazır hale getirilmesi aşamasıdır. Veri Düzleştirme/Perdahlama “Data Smoothing”, Birleştirme “Data Aggregation”, Normalleştirme “Data Normalization” (Şeker, 2013), Veri Ayırıklaştırma “Data Discretization” ve Veri İndirgeme “Data Reduction”, bu aşamada gerçekleştirilebilecek işlemlerdendir.

#### 2.3.1. Veri Düzleştirme

Bu işlemin amacı verideki, özellikle ekonomi ve finans verileri ile görüntü ve sinyal verilerindeki, gizli yapıları bir yakınsama fonksiyonu oluşturarak yakalamaktır. Rassal Yürüme “Random Walk”, Eğri Uydurma “Curve Fitting” ve Hareketli Ortalamalar “Moving Average” en sık kullanılan düzleştirme algoritmalarındandır (Akpınar, 2018).<sup>39</sup>

#### 2.3.2. Veri Normalleştirme

Bu işlem, bir veri setinde bulunan veriler arasındaki farklılığı azaltarak onları tek düzene sahip olacak şekilde, örneğin  $[-1, +1]$  veya  $[0, +1]$  aralığında yer alacak şekilde, dönüştürülmesi işlemidir. Amaç farklı ölçeklendirme sistemine ait verileri birbirleri ile karşılaştırılabilir kılmaktır. Örnek vermek gerekirse, AB ülkelerine ait bir veri setinde nüfus ve harcanabilir gelir değişkenlerinin tanım aralıkları veya ölçekleri birbirinden farklıdır. Bu durumda bu değişkenlere ait ham veriler kullanılarak gerçekleştirilecek analizler, yanıltıcı sonuçlar doğurabilir. Verilerin belirli bir aralığa çekilmesi hem söz konusu değişkenlerin birbirleri ile karşılaştırılabilirliğini kolaylaştırır hem de analizlerin daha sağlıklı sonuçlar vermesini sağlar. Alan yazınında çeşitli veri normalleştirme yöntemleri bulunmaktadır (Akpınar, 2018). Bunlardan bazıları Minimum-Maksimum Normalleştirme “Min-Max Normalization”,  $[-1, +1]$  Normalleştirme, Z-Skor Normalleştirme “z-Score Normalization” ve 10 Tabanında Logaritma Alma “Log Base 10” şeklindedir. R dilinde kodlanmış olan bir fonksiyon ile kullanılan ve alan yazınında yer alan çeşitli normalleştirme yöntemleri Ek 1.'de tablo şeklinde özetlenmiştir.

---

<sup>39</sup> Zaman serileri analizi kitaplarından sayılan algoritmalar hakkında detaylı bilgi edinilebilir.

### 2.3.2.1. Minimum-Maksimum normalleştirme

Bu yöntem, [0, +1] aralığında normalleştirme olarak da adlandırılmaktadır. Bu yöntemde, normalleştirilecek veri setindeki en büyük ve en küçük değerler dikkate alınır ve diğer veriler bu değerlere göre normalleştirilir. Normalleştirme ile birlikte, veri setinin en küçük değeri 0 ve en büyük değeri 1 olur. Bu yüzden, aykırı değerlere karşı hassas bir yöntemdir (Akküçük, 2011).

$$x_{ij}' = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (6)$$

#### Örnek 4:

```
satis=c(20, 28, 40, 60, 76, 84, 84, 88, 88, 92, 92, 92, 92, 92, 96, 96, 96, 96, 100, 100, 105, 115, 120, 120)
```

```
#Min-Max normalleştirme  
n1=normalize(satis)  
n1
```

### 2.3.2.2. [-1, +1] normalleştirme

Bu yöntemde, veri dizisindeki en büyük ve en küçük değerlerin ortalaması 0 olarak dönüştürülüp, diğer bütün değerler sıfır noktasının, sağında (0, +1] ve solunda [-1, 0) olacak şekilde sıralanır.

$$x_{ij}' = \frac{x_{ij} - \frac{\min(x_j) + \max(x_j)}{2}}{\frac{\max(x_j) - \min(x_j)}{2}} \quad (7)$$

#### Örnek 5:

```
satis=c(20, 28, 40, 60, 76, 84, 84, 88, 88, 92, 92, 92, 92, 92, 96, 96, 96, 96, 100, 100, 105, 115, 120, 120)
```

```
#[-1,+1] normalleştirme  
n2=normalize(satis,-1,1)  
n2
```

### 2.3.2.3. z-Skor normalleştirme

Bu yöntem standart skor normalleştirme ya da standardizasyon olarak da adlandırılmaktadır. Veriler, ait oldukları veri dizisinin ortalaması ve standart sapması dikkate alınarak standardize edilir. Bu yöntem ile elde edilen “-“ işaretli standardize değerler, veri setinin ortalamasının altında kalan gözlemlere, “+” işaretli değerler ise

ortalamadan yüksek gözlemlere aittir (Akküçük, 2011). Elde edilen standardize değerler  $[0, +1]$  normalleştirme ya da  $[1, +1]$  normalleştirme yöntemlerindeki gibi bir alt ve üst sınırı yoktur. Bu yöntem kullanılarak elde edilen değerler, dönüştürülen gözlemin, ortalamadan kaç standart sapmalık mesafede uzak olduğunu ifade etmektedir. Başka bir ifade ile dönüştürülmüş değer, sayıların değişim hızını temsil etmektedir denilebilir.

$$x_{ij}' = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \quad (8)$$

### Örnek 6:

```
satis=c(20, 28, 40, 60, 76, 84, 84, 88, 88, 92, 92, 92, 92, 92, 96, 96, 96, 96, 100, 100, 105, 115, 120, 120)
```

```
#z-skor normalleştirme (standardizasyon)
n3=scale(satis)
n3
#VEYA
data.Normalization(satis,type="n1")
```

#### 2.3.2.4. Logaritma alma

Bu yöntem genellikle çarpık “Skewed” veri setlerinin normal dağılıma sahip bir veri setine dönüştürülmesi amacı ile kullanılmaktadır. Özellikle tarihsel veri “historical data” türlerinde bu dönüştürme yöntemine sıklıkla başvurulmaktadır. Fakat bu yöntem, veri setinin yapısını değiştirdiği için eleştirilmektedir. Çarpık veri setlerinde, logaritmik dönüşüm yerine, normallik varsayımı gerektirmeyen istatistiksel analiz yöntemlerinin uygulanması önerilmektedir (Feng, vd., 2014).

### Örnek 7:

```
satis=c(20, 28, 40, 60, 76, 84, 84, 88, 88, 92, 92, 92, 92, 92, 96, 96, 96, 96, 100, 100, 105, 115, 120, 120)
```

```
#Logaritma alma
n4=log10(satis)
n4
```

#### 2.3.3. Veri Ayrıklaştırma

Bölüm 2.2.4.’te açıklanan Veri Kutulama/Nicemleme yönteminin aynısıdır. Bu dönüşüm yönteminin amacı, sürekli nitelikteki verileri kesikli veriye dönüştürmektir. Çünkü örneğin Karar Ağacı “Decision Tree” gibi bazı VM algoritmaları kesikli veri ile

çalışmaktadır. Bu dönüşüm yöntemi, aynı zamanda veri sayısının azaltılmasını da sağlamaktadır. Ayırıklaştırma işlemi farklı yaklaşımlar benimsenerek gerçekleştirilebilir. Örneğin, ayırıklaştırılacak veri dizisi yaş olsun ve gözlemler genç, orta yaşlı ve yaşlı şeklinde üç kategoriye ayrılacak istensin. Bu durumda, istatistik kurumunun kabul ettiği yaş kategorilerinin kesim noktaları referans alınabilir. Bu yaklaşım var olan bir standardı referans olarak ayırıklaştırma olarak nitelendirilebilir. Veri dizisi ayrıca, karar vericinin belirlediği bir frekans değeri referans alınarak da eşit aralıklı ayırıklaştırılabilir ve akabinde her bir dizideki değerlerin sınır değerleri veya ortalaması referans alınarak tekrar ayırıklaştırma işlemine tabi tutulabilir

**Örnek 8:** 4, 8, 15, 21, 21, 24, 25, 28 ve 34 değerlerinden oluşan bir veri dizisine sayılan kutulama yaklaşımları uygulandığında aşağıdaki şekli alır (Han, Kamber, Pei, 2012; Tan, Steinbach, Kumar, 2014).

Eşit frekanslı ayırıklaştırma →	4, 8, 15 21, 21, 24 25, 28, 34
Sınır değerler referans alınarak ayırıklaştırma →	4, 4, 15 21, 21, 24 25, 25, 34
Dizi ortalaması referans alınarak ayırıklaştırma →	9, 9, 9 22, 22, 22 29, 29, 29

Bu basit olarak nitelendirilebilecek yaklaşımlara ek olarak kümeleme, karar ağacı, k-ortalamlar ve histogram ve korelasyon analizi ile ayırıklaştırma gibi gelişmiş yaklaşımlar da bulunmaktadır (Akpınar, 2018; Gama, Pinto, 2006; Han, Kamber, Pei, 2012; Şeker, 2013).

#### **2.3.4. Veri İndirgeme**

Forbes Mayıs 2018 sayısında, dünya günlük veri üretiminin 2,5 kentilyon bayt olduğu, Google arama motorunda bir saniyede ortalama 3,7 milyar arama yapıldığı, Facebook'ta bir saniyede ortalama beş yeni profilin oluşturulduğu, Instagram'a günde ortalama 95 milyon fotoğraf ve video yüklendiği, bir dakikada 103,447,520 adet spam nitelikte elektronik posta yollandığı ve Internet of Things (IoT) özellikli akıllı araç

“device”<sup>40</sup> sayısının 2019 yılı sonunda 200 milyara ulaşacağı ifade edilmektedir. International Data Corporation büyük veri ve iş analitiği forumunun verilerine göre ise 2020 yılında dünya çapında oluşacak veri 35 zb’dir (Ersöz, 2019).<sup>41</sup> İnsanlık olarak ürettiğimiz verinin boyutunun hızlı bir şekilde artması (enine ve boyuna büyümesi) ile paralel olarak, veri toplama ve depolama araçlarının da kapasiteleri artmıştır. Fakat veri boyutunun artması ile birlikte veri üzerinde gerçekleştirilen işlemlerin istatistiki açıdan güvenilirliği azalmaktadır. Bu durum boyut laneti “curse of dimensionality” olarak adlandırılmaktadır (Akpınar, 2018; Tan, Steinbach, Kumar, 2014). Bu nedenle Büyük Veri’nin “Big Data” boyutunun azaltılması gerekmektedir. Bu işlem veri indirgeme olarak adlandırılmaktadır. Veri indirgemedede iki temel yaklaşım bulunmaktadır. Bunlar, öznelik/boyut azaltma “Feature/Dimension Reduction” ve nesne/gözlem sayısını azaltma “Volume Reduction” şeklindedir (Akpınar, 2014).

#### **2.3.4.1. Öznelik/boyut azaltma**

Bir veri setinde, gerçekleştirilecek analiz ile alakasız “irrelevant”, birbirini tekrarlayan/fazlalık “redundant” ya da birbirleri ile yüksek korelasyonlu öznelikler olabilir. Bu türdeki özneliklerin analize dahil edilmesi ile elde edilecek sonuçlar yanıltıcı olabilir. Bu durumda, söz konusu özneliklerin veriden çıkartılması veya başka bir ifade ile gerçekleştirilecek analiz ile alakalı ve önemli özneliklere karar verilmesi önem teşkil etmektedir.

Analize dahil edilecek özneliklerin karar verilmesinde iki temel yaklaşım bulunmaktadır. İlki, Alt Öznelik Seçimi “Feature/Variable/Attribute Subset Selection” yaklaşımıdır. Bu yaklaşımda çeşitli yöntemler kullanılmaktadır.

Adım Adım Regresyon “Stepwise Regression” yönteminde bağımsız değişkenler, yani öznelikler, modele birer birer eklenir (ileriye doğru seçme “forward”) veya modelden birer birer çıkartılır (geriye doğru çıkarma “backward”). Her bir ekleme ve çıkarma işlemi bir adım olarak nitelendirilir ve her bir adımda yeni bir regresyon modeli oluşturulmuş olduğu için her bir modelin, bağımlı değişkeni açıklama gücü hesaplanır.

---

<sup>40</sup> Bu araçlar, insan hayatını kolaylaştırırken, insanlardan aldıkları sesli/text mesajlı komutları kullanarak bizlere (insanlara) ait çeşitli bilgileri de toplamaktadır.

<sup>41</sup> 1 Zeta byte =  $10^{21}$  byte

Amaç, olası en az sayıda öznitelik ile gerçeğe yakın sonuçlar elde etmektir. Örneğin ileriye doğru seçme yaklaşımı benimsenmiş ise modele yeni eklenmiş bir özniteliği, modelde tutmak (modelden çıkarmak) için Çok Değişkenli Modellerde “Multivariate Models”, Uyum İyiliğini “Goodness of Fit” temsil eden  $R^2$  değerinin bir önceki adımda elde edilmiş olan  $R^2$  değerinden büyük (küçük) olması veya Ortalama Hata Kare (OHK) “Mean Squared Error” değerinin bir önceki adımda elde edilmiş olan değerden küçük (büyük) olması gerekmektedir. Bu yöntem özellikle öznitelik sayısının yüksek olduğu veri setlerinde pek de uygulanabilir değildir. Örneğin, 100 öznitelikten oluşan bir veri seti, sadece 15 özniteliğe indirgenmek istensin.  $C(100,15) = 2,53338E + 17$  farklı alt öznitelik kümesi oluşturulabilir. Her bir öznitelik kümesine ait  $R^2$  ve OHK değerlerinin tespit edilmesi ve bu değerlere bağlı olarak öznitelik seçiminin yapılması işlemi maliyetlidir. Bu yüzden adım adım regresyon yöntemi yerine, makine öğrenmesindeki “Machine Learning” alt küme seçimi algoritmalarının kullanılması önerilmektedir. Bu algoritmalar, Paketleyici “Wrapper”, Filtreleyici “Filters” ve Gömülü “Embedded” olmak üzere üç gruba ayrılmaktadır<sup>42</sup>. Söz konusu algoritmalar hakkında detaylı bilgi için Solorio-Fernández, Carrasco-Ochoa, Martínez-Trinidad (2019), Stańczyk, Jain (2015) ve Khalid, Khalil, Nasreen (2014) çalışmaları incelenebilir.

İkinci yaklaşım ise Öznitelik Çıkarımı/Dönüşümü “Feature Extraction/Transformation” yaklaşımıdır. Bu yaklaşımda, veri setini oluşturan öznitelikler<sup>43</sup> kullanılarak, çeşitli yöntemler ile<sup>44-45</sup> yeni öznitelikler<sup>46</sup> oluşturulmaktadır.

---

<sup>42</sup> Aslında Adım Adım Regresyon yöntemi de Paketleyici algoritmalarından birisi olarak sayılmaktadır.

<sup>43</sup> İstatistiki olarak bağımsız/gözlemlenebilir değişkenler şeklinde ifade edilmektedir.

<sup>44</sup> *Ateşböceği Algoritması* “Firefly Algorithm”, *Karınca Kolonisi Algoritması* “Ant Colony Algorithm”, *Genetik Algoritma* “Genetic Algorithm”, *İkili Örüntü* “Local Binary Patterns”, IsoMap, *Kernel Temel Bileşenler Analizi* ve *Semidefinite Embedding* gibi. Öznitelik çıkarımında kullanılan gelişmiş yöntemlerin bazıları için Micheli-Tzanakou (2000) ve Huan Liu, Motoda (1998) çalışmaları incelenebilir.

<sup>45</sup> Öznitelik çıkarımında kullanılması gereken çıkarım yöntemi, işlemin yapılacağı veri setine bağlı olarak değişebilmektedir. Örneğin görsel verilerde, zamana bağlı değişim gösteren verilerde, sürekli, kategorik ya da karma nitelikte öznitelikler içeren veri setlerinde kullanılan çıkarım yöntemleri birbirinden farklıdır.

<sup>46</sup> İstatistiki olarak gizil/yapay/latent değişkenler şeklinde ifade edilmektedir.

Faktör Analizi (FA)<sup>47</sup> ve Temel Bileşenler Analizi (TBA) öznitelik çıkarımında sıklıkla kullanılan yöntemlerdendir. Her iki yöntemin de amacı birbirleri ile korelasyonu olan özniteliklerin bir dönüştürme işlemine tabi tutularak kendi aralarında korelasyon olmayan<sup>48</sup> ve tamamen yeni ve daha az sayıda niteliklerle ki, bu nitelikler bileşen<sup>49</sup> olarak adlandırılmaktadır, temsil etmesini sağlamaktır (Köse, 2018).

Genel anlamda faktör analizi iki amaçla kullanılmaktadır, yapıları keşfetmek (açımlamak/açıklamak) ve yapılar arası ilişkileri doğrulamak. Açımlayıcı faktör analizinde (AFA), öznitelik çıkarımı/boyut indirgeme işlemi yapılmaktadır. Diğer bir ifade ile çok boyutlu gözlemlenebilir değişken seti kullanılarak, daha az boyutlu gizil değişkenler oluşturulmaktadır. Doğrulayıcı faktör analizinde (DFA) ise AFA sonucunda elde edilen gizil değişkenler arasındaki ilişkilerin yapısı incelenmektedir.

FA, parametrik olmayan “non-parametric”, açımlayıcı/keşfedici “exploratory” bir analiz yöntemidir. Dolayısıyla FA’yı uygulayabilmek için, parametrik analiz yöntemlerinin gerektirdiği normallik varsayımının sağlanması gerekmemektedir (Işığışık, 2011; Civelek, 2018). Bu yöntem, fazlalık<sup>50</sup> olan öznitelikleri, kovaryans analizi ile elimine ederek, aranan<sup>51</sup> bilgiyi varyans analizi ile elde etmeye çalışır (Cateni, vd., 2013; Khalid, Khalil, Nasreen, 2014). Veri bilimi alan yazınında, FA’nın gürültülü veriyi tolere edip, aşırı uyumu engelleyebildiği de belirtilmektedir (Khalid, Khalil, Nasreen, 2014). TBA’da da aynen FA’da olduğu gibi çok boyutlu gözlemlenebilir değişken seti kullanılarak, daha az boyutlu bir değişken seti tanımlanmaktadır. Fakat TBA’da kovaryans matrisi hakkında bir varsayım yapılmadan boyut indirgeme ya da dönüşüm gerçekleştirilmektedir. FA’da ise verilerin bir modele uyduğu varsayılmaktadır. Bu yüzden TBA’da (bir uyum aranmadığı için) ortak faktör varyansı hesaplanırken hata

---

<sup>47</sup> Söz konusu iki analiz arasında minör farklılıklar bulunmakta olup, gözlem sayısının yüksek olduğu veri setlerinde her iki analiz yönteminin çok benzer sonuçlar sağladığı belirtilmektedir (Joseph F. Hair, vd., 1998).

<sup>48</sup> Dönüştürme işleminde dikey döndürme yöntemlerinden birisi kullanıldığı zaman birbirleri ile korelasyonsuz yeni bileşenler elde edilebilmektedir. Aksi halde eğer eğik döndürme yöntemlerinden birisi ile döndürme işlemi gerçekleştirilirse oluşturulan yeni bileşenler arasında korelasyon var olabilir (olur).

<sup>49</sup> Bileşenler, YEM’lerde gizil/örtük değişken olarak adlandırılmaktadır.

<sup>50</sup> Redundant

<sup>51</sup> Relevant

terimi ihmal edilir. FA'da ise ortak faktörler tarafından açıklanamayan varyans, artık varyans veya residual olarak adlandırılan hata varyansı/terimi dikkate alınmaktadır. Dolayısı ile özellikle hacmen büyük veri setlerinde açıklanamayan varyans, küçük olanlara nazaran daha az olduğu için büyük veri setlerinde TBA ile FA'nın sağladığı sonuçlar birbirine benzemektedir. Ayrıca, değişken sayısının fazla (30 ve üzeri) ve ortak faktör varyansının yüksek (0.60 ve üzeri) olduğu durumlarda da TBA ve FA'nın birbirine çok benzer sonuçlar sağladığı ifade edilmektedir (Kline, 1994; Joseph F. Hair, vd., 1998).

Sonuç olarak, TBA ve FA arasındaki farklılıklar ve benzeşimler dikkate alındığında boyut indirgemedeki amaç çok sayıdaki değişkeni, bu değişkenlerin bir bileşkesi olarak daha az sayıda değişken ile ifade etmek ise TBA'nın kullanılması uygundur. Diğer taraftan amaç boyut indirgeme yaparken aynı zamanda indirgeme işlemi ile elde edilen yeni değişkenlerin bir yapıyı temsil edip etmediğini veya birer gizil değişken olup olmadığını da tanımlamak ise FA'nın kullanılması daha uygundur. Bu tez çalışmasında kullanılacak olan veri toplama aracı ölçeklerden meydana gelen bir soru formu olduğu için ve orijinal ölçekler endişe, perspektif, davranış ve özbenlik gibi gizil yapılar içerdiği için boyut azaltma işlemi FA analizi ile gerçekleştirilecektir. FA analizinin kullanılması aynı zamanda kullanılan ölçeklerin geçerliğinin test edilmesini de sağlayacaktır.

Geçerlik, bir değişkenin diğer değişkenlerle olan kuramsal ilişkisi ile ilgilidir (DeVellis, 2014). Geçerliğin test edilmesindeki amaç, belirli bir yapıyı diğer yapılarla karıştırmadan ölçebilecek gözlemlenebilir değişkenleri seçerek kendi içinde tutarlı bir yapı elde etmek/oluşturmaktır (Tezbaşaran, 2008). Bu amaçla FA yöntemi kullanılacaktır.

FA'nın ön şartı, değişkenler arasında belirli bir oranda korelasyon bulunmasıdır. Dolayısıyla FA gerçekleştirilmeden önce gözlemlenebilir değişkenler arası korelasyonların incelenmesi gerekmektedir. Bunun için birden fazla yöntem bulunmaktadır. İlk yöntem, korelasyon matrisinin gözle kontrol edilmesidir. Gözlemlenebilir değişkenler arasındaki korelasyon katsayıları küçük ise bu veri setine FA'nın uygulanması uygun değildir. Spesifik olarak belirtmek gerekirse korelasyon matrisindeki değerler incelendiğinde 0.30 ve üzeri değere sahip korelasyon sayısı (adedi) çok fazla değilse FA'nın yapılması önerilmemektedir (Joseph F. Hair, vd., 1998). İlgili yazınında ayrıca, FA'dan önce Kaiser-Mayer-Olkin (KMO) ve Barlett küresellik



testlerinin uygulanması da tavsiye edilmektedir. Barlett küresellik testi değişkenler arasında yeterli oranda ilişki (korelasyon) olup olmadığını incelemek için kullanılan başka bir yöntemdir. Bu yöntem korelasyon matrisini bütünsel olarak ele alarak değişkenler arasındaki korelasyonu incelemektedir. Bu ilişki “Korelasyon matrisi birim matristir” sıfır hipotezi ile test edilerek incelenir. Amaç, sıfır hipotezinin reddedilmesidir ki bu durum, birim matrisin köşegen dışı değerlerinin sıfır olmadığı ve dolayısıyla değişkenler arasında ilişki olduğu durumudur. Barlett testi sonucu  $p < 0.05$  anlamlılık değerinden düşük ise değişkenler arasında FA yapmaya yeterli düzeyde bir ilişki olduğu kabul edilmektedir (Kalaycı, 2010; Durmuş, Yurtkoru, Çinko, 2016; Joseph F. Hair, vd., 1998). KMO testi ile elde edilecek KMO değerleri ise ölçeği oluşturan maddeler ve örneklem arasındaki korelasyonun uygunluğu ile ilgili bir büyüklük olarak tanımlanmaktadır. KMO değerinin en az 0.60 olması<sup>52</sup> beklenmektedir. Bir diğer test ise Measure of Sampling Adequacy – MSA’dır. KMO testi ölçekte yer alan bütün maddelerin genel olarak FA’ya uygunluğunu test ederken, MSA analizi her bir maddenin FA’ya uygunluğunu ölçmektedir. Bu değer, bir maddenin diğer maddeler tarafından hatasız bir şekilde ne derece tahmin edilebildiğini temsil etmektedir (Joseph F. Hair, vd., 1998). MSA değerinin 0.50’den küçük olması durumunda ilgili maddenin ölçekten çıkartılması önerilmektedir (Durmuş, Yurtkoru, Çinko, 2016). Diğer bir deyiş ile MSA değeri 0.50’nin altındaki değerler, FA’nın veri seti için uygun olmadığını bir göstergesidir. İlgili yazınında 0.80 ve üzeri değerler yüksek, .70-.79 arası değerler orta, 0.60-0.69 arası değerler vasat ve 0.50-0.59 arası değerler zayıf olarak belirtilmektedir (Coşkun, vd., 2015). MSA değerinin, örneklem büyüklüğü, değişken (madde) sayısı, korelasyon değerleri arttıkça ve faktör sayısı azaldıkça arttığı belirtilmektedir.

Yukarıda sayılan korelasyon incelemeleri gerçekleştirildikten sonra FA analizi hem ölçeklerin geçerliğinin test edilmesi hem de boyut indirgeme amacı ile uygulanır. FA’da büyük bir madde grubundan (gözlemlenebilir değişken setinden) daha küçük bir değişken grubuna (örtük değişkenlere) geçerek orijinal bilgi özetlenir. Analiz sonucunda ise olası en az bilgi kaybıyla kendi aralarında ilişkisiz olan az sayıda faktörle güçlü bir şekilde ilişkili olan az sayıda madde elde edilmesi beklenmektedir (Özdamar, 2013; DeVellis,

---

<sup>52</sup> Gözlemlenebilir değişkenlerin / maddelerin FA’ya orta seviyede uygun olması durumu

2014; Coşkun, vd., 2015; Joseph F. Hair, vd., 1998). FA özellikle tutum ve davranış gibi yapıların bileşenlerinin belirlenmesinde en sık kullanılan yöntemlerdendir (Tezbaşaran, 2008). İlgili yazınında hangi madde gruplarının bir boyut meydana getirdiğini belirlemenin en iyi yolunun FA olduğu belirtilmektedir (DeVellis, 2014).

FA'nın uygulanmasında bazı önemli hususlar bulunmaktadır. İlki, faktör döndürme yönteminin tespitidir. Faktör döndürme, bir örtük değişkeni kuvvetli bir şekilde ifade edebilen gözlemlenebilir bir değişken kümesi tanımlayarak bilginin yorumlanabilmesini sağlamaktadır (DeVellis, 2014). Bu işlem ile her bir faktörde baskın bir şekilde etkili olan gözlemlenebilir değişkenler maddeler belirgin olarak ortaya konabilmektedir (Özdamar, 2013). Döndürme, maddeler arası ilişkileri tanımlayan farklı bakış açılarını tespit etme işlemidir. Maddeler arası ilişkilerden oluşan bilgi seti, incelendiği bakış açısına göre daha az veya çok anlamlı olabilmektedir. Diğer bir deyişle faktör matrisinin satır (gözlemlenebilir değişkenler) ve sütunlarında (faktörler) sadeleştirme yaparak (Joseph F. Hair, vd., 1998), verilerin organizasyon yapısının görünür ve basit olduğu bir bakış açısı elde edilmeye çalışılmaktadır. Bu yüzden hangi döndürme yönteminin seçileceği önem arz etmektedir. İki tür döndürme yöntemi bulunmaktadır, dik "orthogonal" ve eğik "oblique" döndürme. Dik döndürme yöntemleri, örtük değişkenlerin birbirleriyle ilişkili olmadığı durumlarda önerilmektedir. Eğik döndürme ise örtük değişkenlerin (faktörlerin) birbirleriyle ilişkili olduğu, dolayısıyla faktörler arası korelasyon nedeniyle birden fazla faktörün eş anlamlı olarak birden fazla madde ile bağlantılı olduğu durumlarda önerilmektedir. Ayrıca eğik döndürme yöntemi ilgili yazınında daha gerçekçi bir yaklaşım olarak nitelendirilmektedir. Çünkü gerçek hayatta birbirinden tamamen bağımsız olabilecek bir yapının bulunamayacağı veya varsa da çok az yapının kendi içinde bağımsız olabileceği belirtilmektedir (Joseph F. Hair, vd., 1998). Fakat eğik döndürme yöntemi maddeleri tek boyutluluktan uzaklaştırdığı için yani birbirlerinden bağımsız bir yapıya kavuşturmadığı için anlamlandırılmaya çalışılan teorik yapı basitlikten uzaklaşabilmektedir (DeVellis, 2014). Diğer bir deyiş ile örneğin bir maddenin, birbirine yakın faktör yük değerine sahip iki farklı faktörde yer alma durumu ile karşılaştırılması muhtemel olabilmektedir. Bu nedenle, DeVellis (2014) çalışmasında da önerildiği üzere, ölçeğe eğik döndürme yöntemine ek olarak dik döndürme yöntemi de uygulanmalı ve her iki yöntem ile elde edilen sonuçlar karşılaştırılmalıdır. Bu

karşılaştırma ile eğik döndürme yönteminin sağladığı faktörler arası korelasyon büyüklüğü (Tabachnick, Fidell, 2013) ve her iki yöntem ile elde edilen faktör yapısı (basit/karmaşık) dikkate alınarak elde edilen sonuçlar üzerinden FA'ya devam edilmelidir. Faktör döndürme yöntemine karar vermede önemle belirtilen bir diğer nokta ise örneklem hacminin, gözlemlenebilir değişken madde sayısının veya ortak faktör varyansının “communality” yüksek olduğu durumlarda seçilen döndürme yönteminin faktörleşme üzerinde pek farklılık yaratmadığı ve birbirine yakın sonuçlar/faktörleşme/faktör yapısı verdiğidir (Tabachnick, Fidell, 2013). Eğer her iki faktör döndürme yöntemi birbirine yakın yapılar ve değerler sağlar ise faktörler arasında korelasyona (yani doğrusal ilişkiye) izin vermeyen Dik Döndürme yöntemi ile elde edilen sonuçların kullanılması tavsiye edilmektedir.

İkinci önemli nokta faktör sayısının tespitidir. Faktör sayısının tespitinde Kaiser kriteri, Yamaç-Birikinti grafiği “Cattell Scree Test / Scree plot”, açıklanan varyans kriteri, Joliffe kriteri (Özdamar, 2013; Kalaycı, 2010), Paralel Analiz ve En Küçük Ortalamalı Kısmi Korelasyon yöntemleri bulunmaktadır. İlgili yazınında ilk yöntemler en sık kullanılan yöntemler olmasına rağmen olması gerekenden daha fazla faktör sayısı tespitine neden olabilmektedir. Son iki yöntem ise diğer iki yönteme nazaran faktör sayısını tespit etmeye objektif kriterlerle yaklaştığı kabul edilmektedir (Hayton, Allen, Scarpello, 2004; Koçak, Çokluk, Kayri, 2016; DeVellis, 2014; Tabachnick, Fidell, 2013). Bu tez çalışmasında uygun faktör sayısının tespiti için Kaiser kriteri ve Yamaç-Birikinti kriteri birlikte kullanılacaktır.

Üçüncü önemli nokta ise maddelerin faktör yük değerlerinin tespitidir. Faktör yükleri, gözlemlenebilir değişkenlerle faktörler arasındaki korelasyonları temsil etmektedir. Faktör yükünün artması, değişken ve faktörün birbirleriyle yakından ilişkili olduğunun bir göstergesidir (Coşkun, vd., 2015). Bir maddenin yer aldığı faktördeki yük değerinin en az 0.45 (Büyüköztürk, 2005) veya 0.50 (DeVellis, 2014; Durmuş, Yurtkoru, Çinko, 2016; Kalaycı, 2010; Joseph F. Hair, vd., 1998) hatta 0.70 (Durmuş, Yurtkoru, Çinko, 2016) ve üzerinde olması, o maddenin ilgili yapıyı temsil edebilen veya ölçebilen bir madde olarak değerlendirilmesi ve ilgili faktörde yer alması gerektiği anlamına gelmektedir. Faktör yük değerinin 0.70-0.80 arasında olması durumunda varyansın %50'si, 0.50 olması durumunda ise %25'i açıklanmış olmaktadır. Fakat 0.80 ve üzeri

faktör yükleri aşırı olarak kabul edilip, anormal olarak karşılanmaktadır (Joseph F. Hair, vd., 1998). Faktör yük değerlerindeki bir diğer önemli nokta ise bu değer ile örneklem büyüklüğünün ilişkisidir. Örneklem büyüklüğünün 100 olması durumunda faktör yükü 0.55 ve üzerinde olursa eğer ilgili faktör yükü istatistiksel olarak anlamlı kabul edilmektedir (Joseph F. Hair, vd., 1998).

Dördüncü önemli nokta ise yukarıda da bahsedildiği üzere maddelerin tek bir faktörde yüksek yük değerine ve diğer faktörlerde olabildiğince düşük yük değerine sahip olması gerekliliğidir. İlgili yazınında, bir maddenin en yüksek iki yük değeri arasındaki farkın en az 0.10 olması şeklinde önerilmektedir (Büyüköztürk, 2005).

Beşinci önemli nokta, FA sürecinde ölçekten çıkartılan her bir maddeden sonra analizin tekrar edilme gerekliliğidir. Ayrıca, eğer bir faktör altında sadece tek bir madde yer alıyorsa ya da bir madde birden fazla faktör altında birbirine yakın değerlerde yüklenmiş ise bu maddenin ölçekten çıkartılması ve analizin tekrarlanması gerekmektedir (Durmuş, Yurtkoru, Çinko, 2016).

Son önemli nokta ise faktörlerin herhangi bir maddede birlikte açıkladıkları ortak faktör varyansının en az 0.66'nın üzerinde olması gerekliliğidir. Bu değer yüksek olması modele ilişkin açıklanan toplam varyansın da yüksek olması anlamına gelmektedir. Karar verilen faktör sayısına bağlı olarak analize dâhil edilen değişkenlerle ilgili açıklanan varyansın da keza 0.66 ve üzeri olması önemlidir. Hatta çalışmanın fen ya da sosyal bilimler alanında olması durumuna göre kabul edilebilir varyansın değiştiği görülmektedir. Hair vd. (1998) çalışmasında belirttiği üzere fen bilimleri için toplam varyansın en az 0.95 olması beklenmektedir. Bu değer sosyal bilimler uygulama alanları için ise en az 0.60 ve üzerinde ve hatta daha da düşük olduğu bazı durumlarda bile olurlu olarak kabul edilebilmektedir. Açıklanan varyansın yüksek olması ölçülmeye çalışılan yapının iyi bir şekilde temsil edildiği anlamına gelmektedir (Büyüköztürk, 2005; Coşkun, vd., 2015).

Sonuç olarak FA'da, kuram ile paralel olarak bir ölçekte yer alan örtük değişken sayısı kadar faktör elde etmek ve her bir örtük değişkenin temsil etmiş olduğu boyuta ait olan gözlemlenebilir değişkenlerin ilgili boyut altında toplanmasını sağlamak hedeflenmektedir. Diğer yandan gözlemlenebilir değişkenler ilgili boyut altında toplanmıyorsa ya da kuramla uyumlu bir faktör yapısı elde edilemezse ilgili değişkenler

ölçekten çıkartılarak veya iyileştirilerek pilot uygulama tekrarlanıp söz konusu analizler yinelenir.

Bu tez çalışmasında FA, yalnızca ölçeğin geçerliliğini sağlamakla kalmayıp, gözlemlenebilir değişkenleri gizil değişkenlere dönüştürerek verinin indirgenmesini de sağlamaktadır. Başka bir ifade ile SÖÖH'lerin çıktığı katmanındaki nöronlar, ölçek verisine ait gözlemlenebilir değişkenler olmayıp, ölçeklerin boyutlarını temsil eden faktörlerdir. Böylelikle SÖÖH'ler ile yapılacak olan küme analizinde daha küçük ağ yapıları ile çalışılması mümkün kılınmıştır. Bu şekildeki bir uygulama, FA sonuçlarının bağımlı ya da bağımsız değişken olarak çeşitli istatistiksel tekniklerde kullanılması (Joseph F. Hair, vd., 1998) ile benzeşmektedir.

Boyut indirgemedede kullanılan diğer iki yöntem ise Tekil Değer Ayırışımı “Singular Value Decomposition” ve Çok Boyutlu Ölçeklemedir “Multidimensional Scaling” (Tan, Steinbach, Kumar, 2014). Tekil değer ayrışımında, TBA ve FA'nın aksine çok boyutlu veri, bileşenlerine, özdeğerler yerine tekil değerler esas alınarak indirgenmektedir. TBA ve FA'da temel bileşenler yönündeki varyanslar hesaplanırken, tekil değer ayrışımında standart sapmalar hesaplanmaktadır. Çünkü  $(m \times n)$  boyutlu  $A$  matrisinin tekil değerleri,  $A^T A$  matrisinin özdeğerlerinin kareköküne eşittir. Çok boyutlu ölçeklemede ise gözlemler arası mesafeler dikkate alınarak veri seti daha az boyuta sahip olacak şekilde indirgenmektedir. Amaç, indirgeme işlemi öncesi gözlemler arası mesafe ile indirgeme sonrası mesafeyi bir nevi minimize eden ve stres olarak adlandırılan benzeşmezlik değerini minimize etmektir. Özniteliklerin/değişkenlerin kategorik (sıralayıcı ölçme düzeyleri için) veya nümerik (hem eşit aralıklı hem de oranlı ölçme düzeyleri için) olması durumuna göre çok boyutlu ölçeklemede metrik ve metrik olmayan teknikler geliştirilmiştir.

#### **2.3.4.2. Gözlem sayısı azaltma**

Gözlem sayısı azaltma, istatistikte kullanılan örnekleme yöntemlerinin uygulanması yolu ile ana kütleyi olabildiğince yüksek düzeyde temsil edebilen çok daha küçük örnek bir kütlenin seçilmesi işlemidir. Detaylı bilgi için Yamane (2009) ve Serper, Aytaç, Bayram (2016) çalışmaları incelenebilir. Ayrıca ikinci bölüm Kısım 2.1'de örnekleme yöntemleri hakkında kısaca açıklama yapılmıştır.

## 2.4. Veri Madenciliği

Verinin, çeşitli yöntemler ile işlenerek/analiz edilerek, ondan anlamlı, yorumlanabilir ve anlaşılabilir çıktılar elde etme aşamasıdır. VM’de gerçekleştirilen işlemler Türkçe alan yazınında sınıflandırma, kümeleme ve birliktelik kuralları şeklinde üç temel gruba ayrılmış olsa da aslında yabancı yazınında VM’de gerçekleştirilen tüm işlemler amaçlarına göre iki gruba ayrılmaktadır (Tan, Steinbach, Kumar, 2014). İlk grup, verinin yapısını keşfetme amaçlı gerçekleştirilen tanımlayıcı yöntemleri “Descriptive Tasks” içermektedir. Örneğin, korelasyon, trend, kümeleme, ilişki ve aykırı değer analizi gibi. Bu analizler verideki gizli ilişkilerin gün yüzüne çıkartılmasını sağlamaktadır. İkinci grup ise veri setindeki bir veya daha fazla özneliğin değerini (hedef/bağımlı değişken), veri setindeki diğer özneliklerin değerlerini (açımlayıcı/bağımsız değişkenler) kullanarak tahmin etmeyi amaçlayan sınıflandırma ve regresyon analizi gibi yöntemleri “predictive tasks” içermektedir.

### 2.4.1. Sınıflandırma

Bir veri setindeki gözlemlerin istatistiksel yöntemler ya da makine öğrenmesinde kullanılan algoritmalar ile önceden belirli olan sınıflara atanması ya da etiketlenmesi işlemidir. Lineer, Kuadratik ve Lojistik Diskriminant Analizleri “Linear, Quadratic and Logistic Discriminant Analysis”, BAYES Sınıflandırıcıları “BAYES Classifiers” ve k-En Yakın Komşuluk “k-Nearest Neighborhood” algoritması en sık kullanılan istatistiksel sınıflandırma yöntemlerindedir. Karar Ağaçları “Decision Trees”, Destek Vektör Makineleri “Support Vector Machines” ve Yapay Sinir Ağları (YSA) “Artificial Neural Networks” ise makine öğrenmesi kapsamında yer alan sınıflandırma amaçlı kullanılan yöntemlerdendir (Michie, vd., 1994; İşçi Güneri, Aydın, 2017; Kiang, 2003). Sayılan yöntemler sınıf atama işlemini, bir veri kümesinden öğrenerek gerçekleştirirler. Söz konusu veri kümesi eğitim kümesi “training set” olarak adlandırılmaktadır. Örneğin bir banka şubesi yeni oluşturduğu bireysel kredi paketini telefonla görüşme yöntemi ile potansiyel müşterilere ulaşılarak pazarlama kararı almış olsun ve bankanın veri tabanında bir tanesi mevcut müşterilerine ait ve diğeri potansiyel müşterilere ait VS-1 ve VS-2 şeklinde iki veri seti bulunsun. Her iki veri seti de meslek, yaş, medeni hal, eğitim durumu, ev sahibi olma durumu, gelir, yatırım aracı kullanıp kullanmadığı, daha önce kredi çekip çekmediği, kredi ödeme durumu, kredi kartı sahibi olup olmadığı ve kredi

kartı borcu ödeyebilme durumu şeklinde 11 özneliğe sahip olsun. VS-1’de VS-2’den farklı olarak daha önce telefon aracılığı ile bir promosyondan faydalanıp faydalanmama durumunu temsil eden 12’nci öznelik yer alsın. En son öznelik müşterinin daha önce bu yöntemle bir ürün satın alma/almama durumunu temsil ettiği için bu öznelik hedef çıktı olup, sınıfı “class” temsil etmektedir ve iki adet sınıf etiketi “class label” bulunmaktadır, telefonla görüşme yöntemi ile bir ürünü alan müşterileri temsil eden sınıf (1) ve telefonla görüşme yöntemi ile bir ürün almayan müşterileri temsil eden sınıf (0).

Şubenin amacı telefonla görüşme yöntemi ile potansiyel müşterilere bir ürün satmak olduğu için, şube VS-1’i kullanarak seçtiği bir sınıflandırma algoritmasını eğitip test edecektir. Daha sonra, eğitilmiş algoritmayı VS-2’yi kullanarak çalıştırıp, söz konusu veri setindeki her bir gözlemi 0-1 şeklinde etiketleyecek ve 1 etiketine sahip olan gözlemleri telefon ile arayarak onlara söz konusu bireysel kredi paketini satmaya çalışacaktır, Tablo 1.

**Tablo 1.** Mevcut müşterilere ait veri seti

No	Ö1	Ö2	Ö3	Ö4	Ö5	Ö6	Ö7	Ö8	Ö9	Ö10	Ö11	Sınıf
1	Öğretmen	35	B	Lisans	0	5000	0	0	0	1	1	1
2	İşçi	50	E	İlköğretim	1	5500	1	1	1	1	1	1
3	Emekli	63	E	Lis.Üstü	1	3000	1	1	-1	1	-1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

İşletmecilik alanında sınıflandırma problemleri pazarlama alanı dışında üretim-kalite departmanının kusurlu üretilen parçaları görsel veri işleme algoritmaları ile üretim hattında belirleyerek, onları kusurlu olarak sınıflandırması ve montaja sevk edilmeden hattan ayıklanması, yatırım departmanının geçmiş verileri kullanarak finans piyasalarının trendlerini sınıflandırması ve sonuçlara göre yatırım kararı alması, satın alma departmanının tedarikçileri belirli performans kriterleri ışığında iyi-orta-kötü olarak sınıflandırması ve gelecek siparişlerini bu bilgiyi kullanarak gerçekleştirilmesi, insan kaynakları departmanının firma çalışanlarını performans, eğitim ve iletişim kriterlerini dikkate alarak sınıflandırması ve yıl sonu primlerini bu sınıflandırmaya göre dağıtması ya da muhasebe departmanının mevcut müşterilerin geçmiş bilgilerini kullanarak onları borç ödeyebilme durumlarına göre sınıflandırması ve bu bilgiyi satış departmanı ile

paylaşarak riskli gruptaki müşterilere hiç ya da küçük hacimli satış işlemi yapılması şeklinde örneklenebilir.

#### **2.4.2. Kümeleme**

Kümeleme veya küme analizi, genel bir ifade ile bir veri setinde bulunan benzer gözlemleri bir araya getirerek veri setindeki tüm gözlemlerin gruplanması şeklinde tanımlanabilir. Küme analizi sonucunda aynı grupta yer alan gözlemler birbirine oldukça benzer ve homojen olmalıdır. Ayrı kümelerde yer alan gözlemler ise birbirinden farklı, birbirine benzemeyen ve heterojen olmalıdır. Bu tezde karma veri küme analizi gerçekleştirilmiş olduğu için, küme analizi ile ilgili açıklamalar tezin dördüncü bölümünde verilmiştir.

#### **2.4.3. Birliktelik Kuralları**

Birliktelik kuralları, eş anlı gerçekleşen olaylar/ardışık zamanlı örüntüler arasındaki ilişkileri analiz ederek geleceğe ilişkin tahmin yapılması olarak tanımlanabilir (Özkan, 2008; Akpınar, 2018). Birliktelik kuralları özellikle pazarlama alanında perakende işletmelerinin uyguladığı Pazar Sepeti Analizinde “Market Basket Analysis” kullanılan analiz yöntemidir. Amaç, müşterilerin alışveriş sırasında satın aldıkları ürün (ya da hizmetler) arasındaki ilişkileri tespit ederek, müşterilerin satın alma eğilimlerini/alışkanlıklarını karakterize etmektir. Örneğin, bira satın alan müşterilerin cips veya yer fıstığı satın alması, günlük süt alan müşterilerin organik yumurta satın alması, tamir seti satın alan müşterilerin çivi, cıvata ve somun satın alması veya mühendislik matematiği kitapları satın alan müşterilerin yazılım ve programlama kitaplarını satın alması gibi. İşletmeler birliktelik kurallarını kullanarak benzer örneklerdeki gibi bilgileri keşfetmekte ve raf ve katalog tasarımlarını, promosyonlarını ve ayrıca elektronik posta bilgilendirmelerinin içeriklerini bu tespitler ışığında şekillendirmektedir. Birliktelik kuralları pazarlama alanı dışında sağlık bilimlerinde de kullanılmaktadır. Örneğin ameliyat edilen bir hastanın ameliyat sonrasındaki ilk 15 gün içinde enfeksiyon geçirme olasılığının tespit edilmesi, diyabet hastası olan bir insanın hangi olasılıkla böbrek hastası da olabileceği ya da yüksek tansiyon hastası olan bir insanın kalp hastası da olma ihtimalinin tespit edilmesi gibi (Altunkaynak, 2017; Akpınar, 2018).



Alan yazınındaki en sık kullanılan birliktelik kuralı algoritmaları Apriori, Equivalence Class Transformation (ECLAT), Frequent Pattern Growth (Fp-Growth) algoritmaları ve bu algoritmaların türevleridir (Supriyamenon, Rajarajeswari, 2017; Zhang, Zhang, 2002; Akpınar, 2014).

#### **2.4.4. Tahminleme**

Tahminleme, regresyon analizi, zaman serileri analizi ve YSA gibi yöntemler ile geçmiş verileri kullanarak geleceğe yönelik çıkarımlarda bulunmaktır. Tahminleme yöntemleri fen, sağlık ve sosyal bilimlerde geniş bir uygulama alanına sahiptir. İşletmecilik özelinde ise satış ve talep miktarı, pazar payı, üretim, işçilik, bakım-onarım ve yatırım maliyetleri, döviz kuru ve finansal piyasa hareketleri ve verimlilik alanlarında sıklıkla kullanılmaktadır.

#### **2.5. Sonuçların Yorumlanması**

VM aşamasında elde edilen örüntünün “pattern” yorumlanması ve anlamının sorgulanması aşamasıdır. Elde edilen örüntüler beklendiği kadar bilgi içermiyorsa Şekil 6’dan da inceleneceği üzere, VM sürecindeki önceki aşamaların tekrarlanması gerekmektedir.

### **3. KEŞİF SÜRECİNİ OLUMSUZ ETKİLEYEN FAKTÖRLER**

Bilgi keşfi süreci veriden ve insandan kaynaklı çeşitli faktörlerden olumsuz etkilenabilmektedir (Fayyad, Piatetsky-Shapiro, Smyth, 1996b; Fayyad, Stolorz, 1997; Tan, Steinbach, Kumar, 2014; Verma, 2015; Wang, Storey, Firth, 1995). İkinci bölüm Kısım 2.3.4’te de ifade edildiği üzere dünya günlük veri üretiminin 2,5 kentilyon bayt olduğu dikkate alındığında büyük ve çok boyutlu veri, bilgi keşfi sürecindeki belki de en önemli problemlerden bir tanesi olabilir. Büyük veri tabanları yüzlerce alan, tablo ve milyarlarca kayıt içerebilmektedir. Verinin niteliksel ve niceliksel olarak büyük olmasından ötürü arama uzayı büyümekte, gözlemler seyrekleşmektedir. Dolayısı ile üzerinde çalışılan/araştırılan konu ile ilgili gerçekçi bir çıkarım yapmak veya model kurmak için konu ile ilgili daha fazla gözlem elde etmek gerekebilmektedir. Daha verimli algoritmalar geliştirmek/kullanmak, örnekleme yapmak ve paralel işleme “parallel processing” yapmak verinin niceliksel büyüklüğünün oluşturduğu dezavantajları giderebilir. Boyut indirgeme yöntemlerini kullanmak ve üzerinde çalışılan konu ile ilgili

bir ön bilgiye sahip olmak örneğin, önceki çalışmalardan elde edilmiş bilgiler, konu ile ilgili değişkenler, değişkenler arasındaki ilişkiler vb. gibi, verinin niteliksel büyüklüğünün oluşturduğu dezavantajları giderebilir.

Dünyanın hızlı değişimi/dönüşümü ile birlikte hakkında araştırma yapılan bir olay/olguyu tanımlayan 20 yıl önceki öznitelik sayısı günümüzdekinden çok daha azdır. Örneğin; sanal mağazacılık ve mobil bankacılık kavramları hayatımıza ve alan yazına girmiş bulunmaktadır. Verinin artan eğilimli çok boyutluluğu önceden keşfedilen yapıları/kalıpları geçersiz kılabilir. Bu değişimlerden olumsuz etkilenmemek için değişim ile birlikte güncelleme yapmak şarttır.

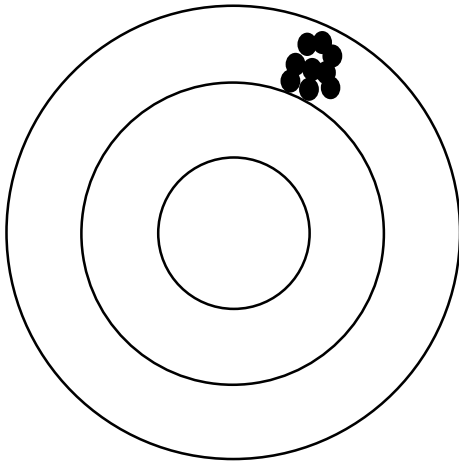
Nitelik sayısının artışı veya verinin enlemsel olarak büyümesi araştırılan konu ile alakalı özniteliklerin tespitini, tespit edilen öznitelikler arası ( karmaşık) ilişkilerin keşfedilmesini ve keşif sonuçlarının yorumlanabilirliğini/anlaşılabilirliğini zorlaştırmaktadır. Gelişmiş algoritmalar ve veri görselleştirme yöntemleri ile söz konusu problemin negatif etkisi azaltılabilir.

Aykırı, eksik veya daha genel bir ifade ile gürültülü veri, keşif sürecini olumsuz etkileyen bir diğer faktördür. Arama uzayı büyüdükçe üzerinde araştırma yapılan konu ile alakasız gözlemlerin ve boyutların kapsama dahil olma olasılığı artmaktadır. Dolayısı ile bilgi keşfi sürecinin sonunda anlamlı bir bilgi elde edememe, yanıltıcı bilgi elde etme ya da VM yöntemleri ile oluşturulan modelin veri setine aşırı uyumlu olması gibi sonuçlarla karşılaşılabilir. Örneğin aşırı uyum problemi tespit edilirse oluşturulan model verideki gerçek örüntüyü öğrenmiş değil, gürültülü veriyi ezberlemiş olur ve bu da kullanılan modelin performansını düşürür. Veri setini ön işleme tabi tutarak ya da konu ile ilgili daha önce yapılmış çalışmaları, farklı niteliklerdeki (nitel/nicel) veri setlerini veya farklı analiz yöntemlerini kullanmak sureti ile çapraz kontrol yaparak (aslında triangule ederek) söz konusu olumsuz etkiler bertaraf edilebilir.

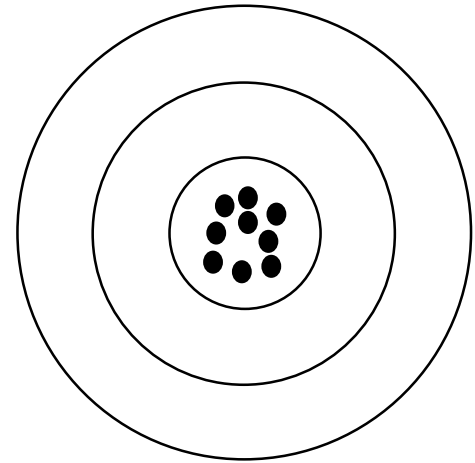
Yukarıda sayılan faktörlere ek olarak, bazen verilerin tek bir kaynak yerine birden fazla kaynaktan sağlanması gerekmektedir. Bu durum, farklı veri kaynaklarından derlenen verilerin birbiri ile entegrasyonunu gerektirmekte ve bazen de veri güvenliğini riske atabilmektedir. Bazı durumlarda ise veri çeşitlilikleri söz konusu olup geleneksel analiz yöntemleri yerine problemin doğasına has gelişmiş yöntemlerin tasarlanması gerekebilmektedir. Ayrıca bu süreç doğru ve kullanılabilir bilgiye ulaşabilmek adına

disiplinler arası çalışmayı da gerektirmektedir (Tan, Steinbach, Kumar, 2014; Han, Kamber, Pei, 2012).

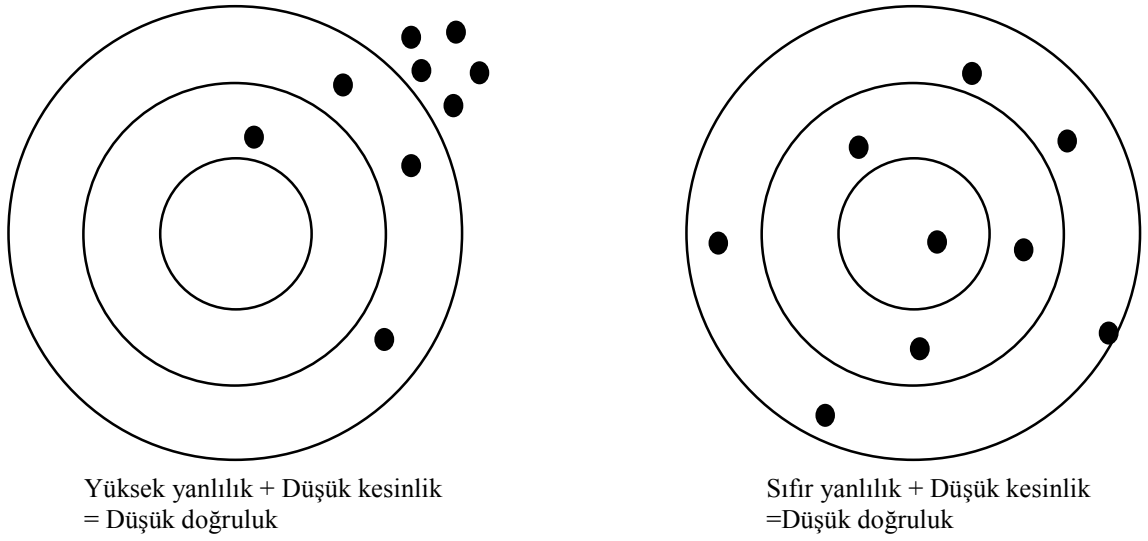
Son olarak, kesinlik, yanlılık ve doğruluk ile tanımlanan veri kalitesinin azalması da keşif sürecini etkilemektedir. Kesinlik “precision” tekrarlı ölçümlerde elde edilen ölçüm verilerinin birbirine benzerliği veya sonuçların tekrarlanabilirliği, yanlılık “bias” ölçüm değerlerinde meydana gelen sistematik varyans ve doğruluk “accuracy” ölçülen değerlerin gerçeğe ne kadar yakın olduğu şeklinde tanımlanmaktadır. Kesinlik ölçülen değerlerin standart sapması ile tespit edilirken, yanlılık, bu değerlerin ortalaması ile gerçek değerler arasındaki fark alınarak tespit edilir. Dolayısı ile yanlılık yalnızca, bilinen nesnelere için belirlenebilir. Tan, Steinbach, Kumar (2014) kesinlik ve yanlılığı 1 gr kütleli olduğu bilinen bir ağırlığın ölçümü ile ilgili bir deney ile açıklamaktadır. Söz konusu ağırlık beş defa ölçülmüş ve ölçüm sonuçları gr cinsinden {1.015, 0.990, 1.013, 1.001, 0.986} şeklinde elde edilmiş. Ölçümlerin ortalaması 1.001 gr ve ölçülen ağırlığın gerçek kütlesi 1 gr olduğu için ölçümün yanlılığı bu iki değer arasındaki fark olup, 0.001’dir. Ölçümün kesinliği ise ölçümlerin standart sapması olup 0.013’tür.



Yüksek yanlılık + Yüksek kesinlik  
= Düşük doğruluk



Sıfır yanlılık + Yüksek kesinlik  
= Yüksek doğruluk



**Şekil 15.** Yanlılık, doğruluk ve kesinliğin grafiksel gösterimi

Doğruluk ise yanlılık ve kesinliğe bağlı genel bir kavram olup, genel bir formülü ya da hesaplama yöntemi bulunmamaktadır. Doğruluğu temsil etmede/tanımlamada en sık başvurulan yöntem, ölçüm değerlerinin ondalık gösterimli olarak kullanılmasıdır. Örneğin bir cismin bir ayrıtı, en küçük birimi milimetre olan, bir cetvel ile ölçüldüğünde, ölçümün doğruluğu  $\pm 0.5$  mm olarak ifade edilmektedir. Kesinlik, yanlılık ve doğruluğun daha iyi anlaşılabilmesi için VM ve istatistikte sıklıkla kullanılan hedef tahtası örneği Şekil 15'ten incelenebilir.

Psikometride ise doğruluk yerine, geçerlik “validity” terimi, kesinlik yerine, güvenilirlik “reliability” terimi kullanılmaktadır.

#### 4. ÖLÇME DÜZEYLERİ

Ölçme, bir değişkene (bu tez çalışması bağlamında her bir özneliliğe) ait gözlem sonuçlarının simgelerle gösterilmesi olarak tanımlanmaktadır. Ölçme işleminde simge olarak genellikle sayılar kullanılsa da bazı durumlarda sıfatlar da kullanılabilir. Fakat gözlem sonuçlarının örneğin Çok iyi-İyi-Orta-Zayıf gibi sıfatlar ile ifade edildiği durumlar sayılarla ifade edildiği durumlara nazaran daha az bilgi içerir ve dolayısı ile daha az kesinlik taşır (Serper, 2014).

Bir ölçme işleminde amaç, mesafe, zaman, ağırlık vb gibi özneliliklerin değerlerini tespit etmek ise kullanılması gereken ölçme aracına karar vermede ve ölçüm değerlerini anlamlandırmada veya söz konusu değerleri kullanarak matematiksel işlemler yapmada

çok fazla problemle karşılaşılmayacağı neredeyse aşikardır. Fakat amaç, zeka, yetenek, refah, mutluluk ya da statü gibi olguları ölçmek ise hem bu olguları ölçmek ve anlamlandırmak hem de ölçüm sonuçları ile matematiksel işlemler yapmak zor olabilmektedir (Şenesen, 2013). (Belki de) bu yüzden çeşitli Ölçme Düzeyleri “Scales/Levels of Measurement”<sup>53</sup> ve her bir düzeyde gerçekleştirilebilecek matematiksel işlemler geliştirilmiştir.

Stanley Smith Stevens psikofizik alanında uzmanlaşmış ünlü bir deneysel psikologdur. Stevens 1946 yılında bilimsel araştırmalarda kullanılan veri türlerini dikkate alarak günümüzde hala geçerliliğini koruyan sınıflayıcı “nominal”<sup>54</sup>, sıralayıcı “ordinal”, eşit aralıklı “interval” ve oranlı “ratio” olmak üzere dört temel ölçme düzeyini geliştirmiştir (Stevens, 1946; Teghtsoonian, 2001).<sup>55</sup>

Sınıflayıcı veya sıralayıcı ölçme düzeyleri, nitel/kategorik/kalitatif değişkenlere ait gözlemleri ölçmek için kullanılırken, eşit aralıklı veya oranlı ölçme düzeyleri, nicel/nümerik/nicel değişkenlere ait gözlemleri ölçmek için kullanılmaktadır (Serper, 2014; Esin, Ekni, Gamgam, 2006).

Bir nicel değişkenin en büyük ve en küçük gözlem değerleri arasındaki fark ise değer aralığı olarak adlandırılmaktadır. Ayrıca, nicel değişkenler aldıkları sayı değerlerine bağlı olarak sürekli  $[-\infty, +\infty]$  veya kesikli  $[-\infty, \dots, -1, 0, 1, \dots, +\infty]$  nitelikte olabilmektedir (Akpınar, 2018; Şenesen, 2013).

#### 4.1. Sınıflayıcı Ölçme Düzeyi

Sınıflayıcı ölçme düzeyinde birbirine benzer olan gözlemleri diğer gözlemlerden ayırt etmek için, söz konusu birimlere aynı simge atanır. Örneğin müşteri numarası, cinsiyet, tercih edilen marka, yaşanılan kent ve meslek gibi. Sınıflayıcı düzeyde ölçülen bir özneliliğin matematiksel özelliği eşitliktir. Başka bir ifade ile bir sınıftaki gözlemler, diğer öznelilikleri açısından farklı olsa da sınıflandıkları öznelilik açısından eşittir.

---

<sup>53</sup> VM’de ölçme düzeyi, Özellik Türleri “Attribute Types” olarak adlandırılmaktadır.

<sup>54</sup> Latince “nomen” yani isim-ad “name” kelimesinden gelmektedir (Köse, 2018).

<sup>55</sup> Stevens’in dört temel ölçme düzeyi dışında, bu ölçme düzeyi sisteminin yetersiz olduğu/kapsayıcı olmadığı savunularak 7’li ve 10’lu ölçme düzeyleri sırası ile Mosteller & Tukey (1977) ve Chrisman (1998) tarafından geliştirilmiştir (McIver, 2004).

Örneğin bir markanın sadık olarak sınıflandırdığı müşteri grubu her ay alışveriş yapan müşterileri temsil etsin. Söz konusu sadık müşterilerin %50'sini gençler, %30'unu orta yaşlılar ve %20'sini de yaşlılar oluştursun. Yaşları açısından farklı düzeylerde olan söz konusu genç, orta yaşlı ve yaşlı müşteriler sadakat açısından eşit sayılmaktadır. Sınıflamada gözlem sonuçları bazen sayılarla ifade edilse de örneğin, Bekar:1, Evli:2, Eşi Vefat Etmiş:3 ve Ayrılmış:4 vb. gibi, söz konusu sayılar sayısal bir değer taşımamaktadır (Serper, 2014).

Sınıflayıcı ölçme düzeyi gözlemleri, belirli niteliklerden aldıkları değerlere göre farklı kümelere ayırmaktadır. Amaç, aynı kümedeki gözlemlerin birbirinden farkının, kümeler arası farklara nazaran daha az olmasını sağlamaktır. Diğer bir ifade ile amaç, küme içi türdeşliği elde etmektir. Kümeleme, ölçümlemenin en düşük düzeyi olarak nitelendirilmektedir. Elde edilen kümeler hem kapsayıcılık hem de örtüşmezlik koşullarını sağlamalıdır. Eğer elde edilen kümeler bütün gözlemleri içeriyorsa/dışarıda hiçbir gözlem kalmıyorsa kapsayıcıdır ve aynı zamanda birbirleri ile örtüşmüyorsa/kesişmiyorsa/her bir gözlem sadece bir kümeye aitse örtüşmezdir ve bazı istatistiki işlemlerin yapılması uygundur (Şenesen, 2013).

Sınıflayıcı ölçme düzeyinde değişkenler çoklu değere sahip olabildiği gibi ikili değere de sahip olabilmektedir. İkili değere sahip değişkenler kendi içinde iki gruba ayrılmaktadır, ikili-dikotom ve ikili-binary (Akpınar, 2018). Dikotom değişkenler kadın-erkek ve hasta-sağlıklı gibi simetrik iki değere sahip olan değişkenlerdir. Değişkenin her iki durumda da sahip olduğu değer eşit derecede önemlidir. Binary değişkenler ise 0-1 türündeki varlık ve yokluğu, evet ve hayır temsil eden asimetric değerli değişkenlerdir. Örneğin, elektronik ya da elektrikli cihazlardaki aç-kapa simgesi binary çalışan bir komuttur. Aç:1 komutunda cihaza enerji akışı sağlanır ve Kapa:0 komutunda cihazın enerjisi kesilir (Akpınar, 2018; Han, Kamber, Pei, 2012; Tan, Steinbach, Kumar, 2014).

Sınıflayıcı ölçme düzeyinde simge olarak kullanılan sayılar ile sadece sınıflama yapılmaktadır, sıralama, toplama, çıkarma, çarpma ve bölme işlemleri yapılamamaktadır (Serper, 2014; Şenesen, 2013). Sınıflayıcı ölçme düzeyi ile elde edilen her bir kümede yer alan toplam gözlem sayısı/sıklık/frekans hesaplanabilir. Maksimum frekans/mod ve minimum frekanslar tespit edilebilir. Her kümedeki gözlem sayısı toplam gözlem sayısına bölünerek paylar/yüzdeler bulunabilir veya kümelerin frekansları birbirine bölünerek

oranlanabilir (Han, Kamber, Pei, 2012; Şenesen, 2013), Ki-kare testi uygulanabilir ve entropi hesaplanabilir (Tan, Steinbach, Kumar, 2014).

#### 4.2.Sıralayıcı Ölçme Düzeyi

Sıralayıcı ölçme düzeyinde gözlemlere atanan sayılar hem gözlemleri birbirinden ayırt eder, hem de söz konusu öznelik açısından gözlemlerin önem sıralarını belirler. Bu ölçme düzeyindeki sayılar arasında sadece küçüklük, eşitlik ve büyüklüğü temsil eden mantıksal ilişkiler kurulur (Esin, Ekni, Gamgam, 2006). Başka bir ifade ile gözlemler, aldıkları değerlere göre bir eksen üzerinde sıralanır. Eksen üzerinde aynı noktaya düşen gözlemler birbirinden ayrıştırılmaz ve aynı küme içinde gruplanır. Bir kümeye giren gözlemler, diğer kümelere giren gözlemlerden daha düşük ya da daha yüksek değerlidir çıkarımında bulunulabilir. Fakat ne kadar yüksek olduğu tespit edilemez (Şenesen, 2013). Örneğin, hanelerin refah seviyesi ve öğrencilerin başarı durumu sıralayıcı ölçme düzeyindeki özneliklere örnek olarak verilebilir. Sıralayıcı ölçme düzeyi, sınıflayıcı ölçme düzeyine göre daha fazla bilgi içerdiği için daha hassas bir ölçme düzeyi olarak kabul edilir (Serper, 2014). Sıralayıcı ölçme düzeyi nesnel olarak ölçülemeyen niteliklerin öznel değerlerini tespit etmede sıklıkla kullanılmaktadır (Han, Kamber, Pei, 2012). Örneğin bir tüketici anketinde katılımcının tercih ettiği ilk üç çamaşır deterjanı markası sorulduğunda, tüketicinin markalardan en çok tercih ettiğini 1, bir sonrakini 2 ve en sonuncusunu 3 olarak sıralaması, söz konusu markaların tüketici açısından öznel bir değere sahip olmasını sağlamaktadır. Ölçümlenen bu değer, gözlem sonuçlarından hangisinin hangisinden küçük ya da büyük olduğu bilgisini içermektedir. Fakat söz konusu küçüklük ya da büyüklüğün miktarı hakkında herhangi bir bilgi içermemektedir.

Sıralayıcı ölçme düzeylerinde sayıların sadece sırasının anlamı bulunmaktadır; sayılar arasındaki farkın hiçbir anlamı yoktur. Gözlem sonuçları ile sadece sınıflama ve sıralama yapılmaktadır. Toplama, çıkarma, çarpma ve bölme işlemleri kesinlikle yapılamamaktadır (Serper, 2014; Şenesen, 2013).

Sıralayıcı ölçme düzeyi ile gözlemler kümelenebilir; elde edilen her bir kümede yer alan toplam gözlem sayısı/sıklık/frekans hesaplanabilir. Maksimum frekans/mod ve minimum frekanslar tespit edilebilir. Her kümedeki gözlem sayısı toplam gözlem sayısına bölünerek paylar/yüzdeler bulunabilir; kümelerin frekansları birbirine bölünerek oranlanabilir. Ortanca değer/medyan tespit edilebilir ve bu değer bir anlam taşıyabilir.

Sıralanan gözlemler kartillere bölünebilir (Han, Kamber, Pei, 2012; Şenesen, 2013), sıra korelasyonu hesaplanabilir (Tan, Steinbach, Kumar, 2014).

### 4.3.Eşit Aralıklı Ölçme Düzeyi

Eşit aralıklı ölçme düzeyinde gözlemlere atanan sayılar kısmi de olsa matematiksel açıdan anlam taşımaktadır. Çünkü bu ölçme düzeyinde sıfır/yokluk anlamı taşıyan gerçek bir sayı bulunmamaktadır. Başlangıç noktası olarak herhangi bir nokta belirlenebilir ve belirlenen bu nokta bir anlam ifade etmez. Bu ölçme düzeyi için sıklıkla verilen örnek Celcius ölçeğidir (Han, Kamber, Pei, 2012; Serper, 2014).  $0^{\circ}C$  suyun donma noktası olup Fahrenheit ölçeğinde  $32^{\circ}F$  değerine karşılık gelmektedir. Bu ölçme düzeyinde hem sayıların sırası hem de sayılar arası farklılık anlam taşımaktadır. Örneğin bugünkü hava sıcaklığı  $10^{\circ}C$  ve bir önceki günü hava sıcaklığı  $5^{\circ}C$  olsun. Bu durumda bugün, dünden daha sıcaktır ve bugün dünden  $5^{\circ}C$  daha sıcaktır sonucuna varılabilir. Fakat bugün dünden iki kat daha sıcaktır sonucu çıkarılamaz. Likert tipi ölçme araçları da eşit aralıklı ölçme düzeyine sahiptir. Örneğin bir araştırmaya katılan iki denekten (katılımcı) birincisi soru formundaki bir ifadeye 5'li Likert ölçeğinin<sup>56</sup> üst uç değeri olan Tamamen Katılıyorum:5 şeklinde cevaplamış olsun. İkinci denek ise aynı ifadeye ölçeğin orta değeri olan Ne Katılıyorum Ne Katılmıyorum:3 cevabını versin. Bu durumda birinci denek, ikinci denekten söz konusu ifadeye iki kat daha fazla katılıyor ya da söz konusu ifadeyi iki kat daha fazla onaylıyor çıkarımında bulunulması imkansızdır (Serper, 2014). Zira tutum, değer, fayda, zeka ve refah da mutlak sıfır noktasına ait olmayan (veya isteğe bağlı bir sıfır noktasına sahip olabilen) ve eşit aralıklı ölçme düzeyleri ile ölçülen olgulardandır.

Eşit aralıklı ölçme düzeyi ile elde edilen gözlem sonuçları ile sınıflama, sıralama, toplama ve çıkarma işlemleri yapılabilirken, çarpma ve bölme işlemleri kesinlikle yapılamaz. Sayılar arasındaki farklar sadece sıralama, eşitlik, büyüklük ve küçüklüğü ifade eder ve sayılar birbirleri ile oranlanamaz (Serper, 2014; Şenesen, 2013). Aritmetik ortalama, standart sapma, kök alma, üs alma, logaritma alma gibi dönüşümler

---

<sup>56</sup> 5'li Likert tipi ölçek 5 düzeyden oluşup dört eşit aralığa sahiptir. Orta düzey olan 3, alt düzey 1 ve üst düzey 5'e eşit mesafededir. Bu yüzden 1-3 arası mesafe ile 1-5 arası mesafe eşit olup miktar açısından eşittir çıkarımında bulunulamaz.



uygulanabilir, geometrik ve harmonik ortalama alınabilir ve iki deęişken arasında korelasyon katsayısı hesaplanabilir (Han, Kamber, Pei, 2012; Şenesen, 2013) ve  $t$  ve  $F$  testleri uygulanabilir (Tan, Steinbach, Kumar, 2014).

#### **4.4.Oranlı Ölçme Düzeyi**

Oranlı ölçme düzeyinde ise gözlemlere atanan sayılar sınıfları, sıralamayı, farkları ve aynı zamanda varlık ve yokluğu da temsil edebilmektedir. Başka bir ifade ile bir başlangıç noktası vardır ve sıfır noktası bir anlam ifade etmektedir<sup>57</sup>. Bu yüzden en hassas ölçme düzeyi oranlı ölçme düzeyidir. Satış miktarı, gelir, mesafe, zaman, ürün gamındaki toplam ürün sayısı, tüketilen elektrik miktarı vb. gibi tüm öznitelikler oranlı ölçme düzeyinde ölçülebilmektedir. Oranlı ölçme düzeyi ile elde edilen gözlem sonuçları ile sınıflama, sıralama, toplama, çıkarma, çarpma ve bölme işlemlerinin hepsi yapılabilir. Gözlemlerin aldığı deęerler birbirleri ile oranlanabilir ve dönüşüm yapılabilir. Örneğin para birimleri birbirlerine çevrilebilir (Han, Kamber, Pei, 2012; Serper, 2014; Şenesen, 2013). Varyans, çarpıklık, basıklık, regresyon ve korelasyon analizleri gerçekleştirilebilir (Şenesen, 2013; Tan, Steinbach, Kumar, 2014).

### **5. BENZERLİK VE BENZEŞMEZLİK (YAKINLIK VE UZAKLIK) ÖLÇÜTLERİ**

İki gözlemin birbirine yakınlıkları/uzaklıkları “proximity/distance” veya bu iki nesnenin birbirine benzerlikleri/benzeşmezlikleri “similarity/dissimilarity” bu nesnelere öznitelik deęerleri kullanılarak hesaplanır. İkinci bölüm Kısım 4’te açıklandığı üzere Stevens’in oluşturduğu dördü ölçme düzeyi dikkate alınırsa özniteliklerin ölçüm deęerleri sınıflayıcı, sıralayıcı, eşit aralıklı ve oranlı olmak üzere dört farklı türde olabilir. Söz konusu ölçme düzeylerinde gerçekleştirilebilecek matematiksel işlemler farklılaştığı için farklı ölçme düzeyleri ile ölçülen özniteliklere sahip olan gözlemlerin birbirleri ile benzerliklerinin hesaplanmasında farklı formülizasyonlar kullanılmaktadır. VM’de kullanılan kümeleme ve en yakın komşu sınıflandırma algoritmalarında ve aykırı deęer tespitinde bu ölçütler kullanılmaktadır. Başka bir ifade ile söz konusu algoritmalar (ham veya önışlenmiş) veri matrisi yerine benzerlik/benzeşmezlik matrisini kullanarak bilgiye

---

<sup>57</sup> Kelvin ölçeğinde mutlak sıfır noktası bulunmaktadır.



hesaplanıyor ise  $d(A, B) = 0$  olur. Bu durumda fonksiyon, birinci pozitiflik özelliğini sağlamaktadır. Fakat diğer taraftan, ikinci pozitiflik özelliğini sağlayamamaktadır. Çünkü ikinci pozitiflik özelliğine göre  $d(A, B) = 0$  ise  $A = B$  olmalıdır. Başka bir ifade ile bu iki kümenin eşit kümeler olması gerekmektedir. Fonksiyon ayrıca simetri özelliğini de sağlamamaktadır;  $d(A, B) = 0$  ve  $d(B, A) = 4$  olduğu için  $d(A, B) \neq d(B, A)$ 'dır;  $A$  ile  $B$  arasındaki mesafe,  $B$  ile  $A$  arasındaki mesafeye eşit değildir. Günlük hayattan bir örnek olarak ise günün saatleri arasındaki mesafenin hesaplanmasında kullanılan fonksiyon verilebilir. Eğer  $t_1 \leq t_2$  ise  $t_1$  ile  $t_2$  zamanı arasındaki mesafe  $d(t_1, t_2) = t_2 - t_1$  şeklinde,  $t_1 \geq t_2$  ise  $d(t_1, t_2) = 24 + (t_2 - t_1)$  şeklinde hesaplanmaktadır. Bu mesafe fonksiyonuna göre 13:00 ile 14:00 arasındaki mesafe  $d(13:00, 14:00) = 1$  saat iken, 14:00 ile 13:00 arasındaki mesafe  $d(14:00, 13:00) = 23$  saattir. Dolayısı ile bu fonksiyon da metrik değildir.

Benzerlik ve benzeşmezlik ölçütleri her zaman pozitif değerli olup, birbirlerine dönüştürülebilen ölçütlerdir. Dönüşümde en sık kullanılan yöntemler,  $s$  benzerlik ve  $d$  benzeşmezlik değerlerini temsil etmek üzere Eş. (10)'daki gibidir.

$$s = 1 - d, 0 \leq \{s, d\} \leq 1 \quad (10)$$

Bu kısımda anlatılan benzeşmezlik ölçütlerine ek olarak özellikle yazılı dokümanların ve multimedya nesnelerinin benzerliklerinin tespit edilmesinde, metin madenciliğinde, intihal tespitinde ve arama motorlarında Kosinüs Benzerlik “Cosine Similarity”, Genişletilmiş Jaccard Endeksi “Extended Jaccard Coefficient”, Bregman Diverjansı “Bregman Divergence” ölçütleri ve ayrıca Ki-kare “Chi-square”, Pearson Korelasyonu, Hamming mesafesi, Levenstein mesafesi, Canberra Mesafesi, Bray-Curtis mesafesi, Tanimoto Benzerliği ve Kord Mesafesi “Chord Distance” ölçütleri de bulunmaktadır. Detaylı bilgi için Han, Kamber, Pei (2012), McCune, Grace (2002), Shirkhorshidi, Aghabozorgi, Wah (2015) ve Tan, Steinbach, Kumar (2014) çalışmaları incelenebilir

### 5.1. Tek Özniteliğe Sahip Gözlemlerin Benzerlik-Benzeşmezlik Ölçütleri

Sınıflayıcı ölçme düzeyi ile ölçülen nominal öznitelikler, niteledikleri gözlemlerin birbirlerinden farklı olma durumları hakkında bilgi taşımaktadırlar. Başka bir ifade ile nominal öznitelikli gözlemler ya birbirinin tamamen aynıdır ya da birbirlerinden

tamamen farklıdır. Dolayısı ile birbirinin tamamen aynısı olan  $N_1$  ve  $N_2$  gibi iki gözlemin benzerlik değeri  $s(N_1, N_2) = 1$ , tamamen farklı olan  $N_1$  ve  $N_3$  gibi iki gözlemin benzerlik değeri  $s(N_1, N_3) = 0$ 'dır.

Sıralayıcı ölçme düzeyi ile ölçülen ordinal öznitelikler ise niteledikleri gözlemlerin sıralamaları hakkında bilgi taşımaktadırlar. Örneğin bir işletme dört farklı kalitede {*düşük kalite* = 1, *orta kalite* = 2, *iyi kalite* = 3 ve *yüksek kalite* = 4} ürün üretiyor olsun. Dolayısı ile düşük kaliteli bir ürün ile yüksek kaliteli bir ürün arasındaki benzeşmezlik  $d(\bar{U}_1, \bar{U}_2) = 4 - 1 = 3$  olarak hesaplanır. Hesaplanan benzeşmezlik değeri  $[0,1]$  aralığına çekilirse  $d(\bar{U}_1, \bar{U}_2) = \frac{4-1}{4} = 0.75$  olarak hesaplanır. Bu hesaplamada düşük ile orta kalite, orta ile iyi kalite ve iyi ile yüksek kalite arasındaki mesafelerin eşit olduğu varsayılmış olmaktadır. Fakat gerçek hayatta söz konusu öznel mesafelerin tam anlamı ile eşit olması imkânsız olabilmektedir.

Eşit aralıklı ve oranlı ölçme düzeyleri ile ölçülen nümerik özniteliklere sahip gözlemlerin birbirleri ile benzerlik/benzeşmezlikleri, ölçümlenen değerler arasındaki farkın mutlak değeri hesaplanarak tespit edilir. Elde edilen değer,  $\mathbb{R}^+$ 'in bir elemanı olacağı için benzerlik/benzeşmezlik ölçütleri dönüşümünde Eş.(10)'daki formülizasyonun kullanılması mümkün değildir. Alan yazınında nümerik özniteliklere ait benzerlik/benzeşmezlik ölçütlerinin dönüşümünde Tablo 2.'de yer alan eşitliklerin kullanılması önerilmektedir (Tan, Steinbach, Kumar, 2014). Tek özniteliğe sahip gözlemlerin benzerlik/benzeşmezliklerinin hesaplanması durumunda veya gözlemlerin sadece tek bir öznitelikleri dikkate alınarak birbirleri ile ne kadar benzer/benzeşmez oldukları tespit edilmek istendiğinde yukarıda verilen açıklamalar dikkate alınarak aşağıdaki gibi bir hesaplama ve dönüşüm tablosu oluşturulabilir.

**Tablo 2.** Tek öznitelikli gözlemler için benzerlik/benzeşmezlik ölçütleri

Öznitelik türü	Benzeşmezlik	Benzerlik
Nominal	$d = \begin{cases} 0, & x = y \\ 1, & x \neq y \end{cases}$	$s = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}$
Ordinal	$d =  x - y  / (n - 1)$	$s = 1 - d$
Nümerik (Eşit aralıklı ve Oranlı)	$d =  x - y $	$s = -d, s = 1 / (1 + d), s = e^{-d}$ $s^* = 1 - \frac{d - \min_d}{\max_d - \min_d}$

\*Bu eşitliğin kullanılabilmesi için öncelikle tüm benzeşmezlik değerleri hesaplanarak en küçük benzeşmezlik  $\min_d$  ve en büyük benzeşmezlik  $\max_d$  değerleri tespit edilmelidir.

## 5.2. Tüm Öznitelikleri Nominal Olan Gözlemler İçin Benzerlik- Benzeşmezlik Ölçütleri

Nominal öznitelikler, sınıflayıcı ölçme düzeyleri ile ölçülmektedir. Ölçülen nesnelerin ölçüm konusu ile ilgili durumları hakkında bilgi verirler. Örneğin, yaşam memnuniyetine katılan deneklerin, çocuk, genç, orta yaşlı ya da yaşlı olması ya da ihracat araştırmalarında, araştırmaya katılan işletmelerin otomotiv, hazır giyim ve konfeksiyon, demir-çelik, kimya ve elektronik gibi faaliyet alanlarına göre sınıflandırılması gibi. Bu ve benzeri özniteliklere sahip nesneler, benzeşmezlik oranı dikkate alınarak karşılaştırılmaktadır.  $i$  ve  $j$  gibi iki gözlem için benzeşmezlik ölçütü  $d(i, j)$ ,  $n$  toplam öznitelik sayısını,  $t$   $i$ 'inci ve  $j$ 'inci gözlemlerin toplam kaç öznitelik için aynı değere sahip olduğunu ve  $p$   $i$ 'inci ve  $j$ 'inci gözlemlerin toplam kaç öznitelik için farklı değere sahip olduğunu temsil etmek üzere sırası ile Eş. (11) ve Eş. (12)'deki gibidir.

$$d(i, j) = \frac{n-t}{n} \quad (11)$$

$$s(i, j) = 1 - d(i, j) = \frac{p}{n} \quad (12)$$

**Örnek 9:** Bir işletme üç tane nominal özniteliği dikkate alarak tedarikçilerini değerlendiriyor olsun. İlk beş tedarikçisi hakkındaki bilgiler aşağıdaki gibidir.

**Tablo 3.** Örnek 9'a ait veri seti

Tedarikçiler	Firma ölçeği	Ürün	Ulaştırma
1	Küçük	B	Kara
2	Küçük	C	Tren
3	Büyük	A	Tren
4	Orta	A	Kara
5	Büyük	B	Kara

Bu bilgiler ışığında söz konusu beş tedarikçiye ait her bir nominal öznitelik için benzeşmezlik matrisleri Tablo 3 dikkate alınarak aşağıdaki gibi oluşturulur.

$$\text{Firma ölçeği özniteliği için: } d(i, j) = \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 1 & 1 & 0 & & \\ 1 & 1 & 1 & 0 & \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

$$\text{Ürün özniteliği için: } d(i, j) = \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 1 & 1 & 0 & & \\ 1 & 1 & 1 & 0 & \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

$$\text{Ulaştırma özniteliği için: } d(i, j) = \begin{bmatrix} 0 & & & & \\ 0 & 0 & & & \\ 1 & 1 & 0 & & \\ 1 & 1 & 1 & 0 & \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Eş. (11) uygulandığında final benzeşmezlik matrisi aşağıdaki gibi hesaplanır.

$$d(i, j) = \begin{bmatrix} 0 & & & & \\ 2/3 & 0 & & & \\ 1 & 2/3 & 0 & & \\ 2/3 & 1 & 2/3 & 0 & \\ 1/3 & 1 & 2/3 & 2/3 & 0 \end{bmatrix}$$

Final benzeşmezlik matrisi incelendiğinde  $d(5,1) = 1/3$  ile birinci ve beşinci tedarikçilerin en düşük benzeşmezlik değerine sahip olduğu, dolayısı ile bu tedarikçilerin birbirleri ile (sadece söz konusu öznitelikler bağlamında) yüksek derecede benzeştiği ve aslında aralarındaki uzaklığın diğer tedarikçilerle olan uzaklıkları ile karşılaştırıldığında nispeten daha küçük olduğu sonucuna varılabilmektedir.

Nominal türdeki özniteliklere sahip gözlemler arası benzeşmezliklerin hesaplanmasında kullanılan başka bir yaklaşım, nominal özniteliklerin simetrik veya asimetrik ikili değişkenlere dönüştürülmesi sureti ile gerçekleştirilmektedir. Örnek 9'daki problem ele alınırsa, firma ölçeği ve ürün özniteliklerinin üç değere (duruma) ve ulaştırma özniteliğinin iki değere (duruma) sahip olduğu görülmektedir. İkili değişkenlere dönüştürme yaklaşımı benimsendiğinde örneğin firma ölçeği özniteliği üç durum ile temsil edildiği için bu öznitelikten her bir durum bir yeni öznitelik olmak üzere üç yeni öznitelik oluşturulur, küçük, orta ve büyük öznitelikleri gibi. Dolayısı ile her bir yeni öznitelik alabileceği değerler  $Evet = 1$  ve  $Hayır = 0$  olarak düzenlenir. Bu dönüşümden sonra, alabileceği değerler sadece  $\{0, 1\}$  olan, yeni öznitelik uzayı tanımlanmış olur<sup>58</sup>. Örneğin bu örnek için sekiz tane yeni öznitelik uzayı oluşturulmalıdır. Fakat bu durumda öznitelik uzayı büyümekte veya enlemsel olarak genişlemekte ve benzeşmezlik matrisi seyrekleşmektedir. Bu durum ikinci bölüm kısım 3'te de açıklandığı üzere hem öznitelikler arası ( karmaşık) ilişkilerin keşfedilmesini hem de keşif sonuçlarının yorumlanabilirliğini/anlaşılabilirliğini zorlaştırmaktadır

---

<sup>58</sup> Bu durumda nominal olan öznitelik, ikinci bölüm kısım 5.3'te açıklanan asimetrik ikili türde bir öznitelğe dönüştürülmüş olur.

### 5.3. Tüm Öznitelikleri İkili Türde Olan Gözlemler İçin Benzerlik- Benzeşmezlik Ölçütleri

İkili türdeki özniteliklere sahip gözlemler arası benzeşmezlik ölçütünün hesaplanmasında Kontenjans Tablosundan “Contingency Table” faydalanılmaktadır, Tablo 4.  $a$  ve  $b$  sırası ile her iki gözlemin değerinin 1 ve 0 olduğu öznitelik sayısını,  $c$  birinci gözlemin değerinin 1 ve ikinci gözlemin değerinin 0 olduğu öznitelik sayısını ve son olarak  $d$  birinci gözlemin değerinin 0 ve ikinci gözlemin değerinin 1 olduğu öznitelik sayılarını temsil etmek üzere toplam öznitelik sayısı  $n = a + b + c + d$ 'dir.

Öznitelik, simetrik ikili türde dikotom bir değişken (örneğin kadın ve erkek ya da hasta ve sağlıklı gibi  $\{1,0\}$  değerleri eşit derecede öneme sahip) ise Eş. (13)'teki Simetrik İkili Benzeşmezlik “Symmetric Binary Dissimilarity” formülü kullanılmalıdır.

$$d(i, j) = \frac{c+d}{n} \quad (13)$$

**Tablo 4.** İkili türdeki öznitelikler için kontenjans tablosu

		Gözlem $j$		Toplam
		1	0	
Gözlem $i$	1	$a$	$c$	$a+c$
	0	$d$	$b$	$d+b$
Toplam		$a+d$	$c+b$	$n$

Öznitelik, asimetrik ikili türde bir değişken (örneğin varlık ve yokluk ya da evet ve hayır gibi  $\{1,0\}$  değerleri eşit derecede öneme sahip değil) ise Eş. (14)'deki Asimetrik İkili Benzeşmezlik “Asymmetric Binary Dissimilarity” formülü kullanılmalıdır. Dikkat edileceği üzere bu eşitlikte, her iki gözlemde de yokluğu temsil eden  $b$  bir öneme sahip değildir. Çünkü bu tür özniteliklere sahip olan veri setlerinde amaç, araştırılan konunun varlığını tespit etmektir.

$$d(i, j) = \frac{c+d}{a+c+d} \quad (14)$$

Asimetrik ikili benzerlik hesabında ise Eş. (15) ile ifade edilen Basit Eşleşme Katsayısı “Simple Matching Coefficient” ve Eş. (16) ile ifade edilen Jaccard İndeksi “Jaccard Index/Coefficient” olmak üzere iki ölçüt sıklıkla kullanılmaktadır (Han, Kamber, Pei, 2012; Tan, Steinbach, Kumar, 2014).

$$s(i, j) = \frac{a+b}{n} = 1 - d(i, j) \quad (15)$$

$$s(i,j) = \frac{a}{a+c+d} = 1 - d(i,j) \quad (16)$$

**Örnek 10:** Örnek 9’da yer alan problemdeki öznitelikler (özniteliklerin alabileceği durumlar dikkate alındığında) simetrik ikili özniteliklere dönüştürüldüğünde yeni veri seti (dönüştürülmüş öznitelik matrisi) Tablo 5’teki şekli alır.

**Tablo 5.** Örnek 10’a ait veri

Tedarikçiler	Küçük ölçek	Orta ölçek	Büyük ölçek	A ürünü	B ürünü	C ürünü	Kara	Tren
1	1	0	0	0	1	0	1	0
2	1	0	0	0	0	1	0	1
3	0	0	1	1	0	0	0	1
4	0	1	0	1	0	0	1	0
5	0	0	1	0	1	0	1	0

Örneğin birinci ve beşinci tedarikçilerin kontenjans tablosu Tablo 6’daki gibi oluşturulur.

**Tablo 6.** Örnek 10’a ait örnek bir çözüm

		Gözlem 5		
		1	0	Toplam
Gözlem 1	1	2	1	3
	0	1	4	5
Toplam		3	5	8

Birinci ve beşinci gözlemler arası asimetrik ikili benzeşmezlik  $d(1,5) = \frac{1+1}{2+1+1} = 0.500$  olarak hesaplanır. İnceleneceği üzere aynı gözlemler için Örnek 9’da hesaplanan benzeşmezlik değeri  $d(1,5) = 0.333$  ile bu örnekteki yaklaşım kullanılarak elde edilen benzeşmezlik değerleri (iki gözlem arasındaki uzaklıklar) farklıdır. Söz konusu farklılık, hem kontenjans tablosundaki  $b$  (yokluk) değerinin hesaplamada dikkate alınmamasından hem de öznitelik uzayının büyümesinden kaynaklanmaktadır. Uzayın büyümesi ile elde edilen dönüştürülmüş öznitelik matrisi seyrekleşmiş ve seyrekleşmenin bir sonucu olarak da iki gözlem arasındaki mesafe artmıştır.

Aynı gözlemler için basit eşleşme katsayısı  $d(1,5) = \frac{2+4}{8} = 0.750$  ve Jaccard indeksi  $d(1,5) = \frac{2}{2+1+1} = 0.500$  olarak hesaplanmaktadır. Görüleceği üzere söz konusu gözlemler basit eşleşme katsayısına göre 0.750 seviyesinde birbirlerine benzer olarak değerlendirilirken, Jaccard indeksine göre 0.500 seviyesinde benzer çıkmıştır. Başka bir ifade ile aynı gözlemler için birbirinden çok farklı iki benzeşmezlik değeri elde edilmiştir. Buradaki farklılık ise esas olarak Jaccard indeksinin sadece  $a$  (varlık) değerini dikkate almasından kaynaklanmaktadır. Bu yüzden özellikle varlık ve yokluk arasındaki farkın



önemli olduğu asimetrik ikili değişkenlere sahip problemlerde Jaccard indeksinin kullanılması önerilmektedir.

Veri setinde hem simetrik hem de asimetrik ikili türde öznitelikler olması durumunda ise ikinci bölüm kısım 5.6’da yer alan yaklaşım kullanılarak benzeşmezlik hesabı yapılmalıdır.

#### **5.4. Tüm Öznitelikleri Ordinal Türde Olan Gözlemler İçin Benzerlik- Benzeşmezlik Ölçütleri**

Ordinal türde olan veya nümerik olup, kutulama yöntemleri ile ayrıklaştırılarak ordinal hale getirilen özniteliklere sahip gözlemler arası benzeşmezliklerin ölçümünde, nümerik türdeki özniteliklere sahip gözlemler için kullanılan benzeşmezlik ölçütleri, kullanılmaktadır. Fakat gözlemler arası uzaklık hesaplanmadan önce, ordinal türdeki verilerin Eş. (17) kullanılarak  $[0, 1]$  aralığına çekilmesi (bir nevi normalize edilmesi) gerekmektedir.  $r_{if}$ ,  $i$ 'inci gözlemin  $f$ 'inci öznitelik için ordinal sıra değeri ve  $M_f$ ,  $f$ 'inci özniteliğin alabileceği ordinal değer sayısı olmak üzere normalize değer  $z_{if}$ ,

$$z_{if} = \frac{r_{if}-1}{M_f-1} \quad (17)$$

Örneğin, ordinal bir öznitelik kötü, iyi ve orta olmak üzere sadece üç değere sahip ise  $M_f = 3$  olur. Eğer bir gözlemin söz konusu öznitelik için değeri orta ise  $r_{if} = 2$  olur. Bu durumda  $z_{if} = \frac{2-1}{3-1} = 0.5$  olarak  $i$ 'inci gözlemin  $f$ 'inci ordinal öznitelik için normalize değeri 0.5 hesaplanmış olur. Bu işlem gerçekleştirildikten sonra, örneğin Öklit uzaklığı veya başka bir uzaklık fonksiyonu ile gözlemler arası mesafe hesaplanabilir. Örneğin bir başka gözlemin aynı öznitelik için normalize değeri 1 ise bu iki gözlem arasındaki Öklit mesafesi ise  $d(i, j) = \sqrt{(x_{i1} - x_{j1})^2} = \sqrt{(1 - 0.5)^2} = 0.5$  olarak hesaplanır.

#### **5.5. Tüm Öznitelikleri Nümerik Türde Olan Gözlemler İçin Benzerlik- Benzeşmezlik Ölçütleri**

Öklit “Euclidean”, Manhattan ve Minkowski benzeşmezlik ölçütleri eşit aralıklı ve oranlı türde özniteliklere sahip gözlemler arası uzaklığın ölçümünde kullanılmaktadır.  $n$  öznitelik sayısını temsil etmek üzere, gözlemler arası Öklit ve Manhattan uzaklıkları, aşağıda yer alan Eş. (18, 19) ile hesaplanmaktadır. Her iki mesafe de pozitiflik, simetriklik

ve üçgen eşitsizliği koşullarının hepsini sağladığı için metrik nitelikte olan uzaklık fonksiyonlarındandır.

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{in} - x_{jn})^2} = \sqrt{(x_{in} - x_{jn})^T (x_{in} - x_{jn})} \quad (18)$$

$$d(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{in} - x_{jn}| \quad (19)$$

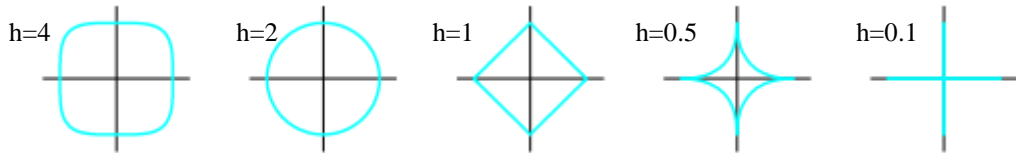
$h \geq 1$  olmak üzere Eş. (20) ile ifade edilen Minkowski mesafesi ( $p$  öznitelik sayısını temsil etmek üzere alan yazınında  $L_p$  norm olarak da adlandırılmaktadır), Öklit ve Manhattan mesafelerinin genelleştirilmiş halidir. Örneğin  $h = 1$  ve  $h = 2$  için hesaplanan Minkowski uzaklığı, sırası ile Manhattan ( $L_1$  norm, city block, taxicab) ve Öklit ( $L_2$  norm) uzaklıklarına karşılık gelmektedir.

$$d(i, j) = \sqrt[h]{|x_{i1} - x_{j1}|^h + |x_{i2} - x_{j2}|^h + \dots + |x_{in} - x_{jn}|^h} \quad (20)$$

Supremum<sup>59</sup> Mesafe “Supremum Distance” ise Minkowski mesafesinin genelleştirilmiş halidir ve alan yazınında  $L_{max}$ ,  $L_{\infty}$  norm ve Chebyshev uzaklığı olarak da adlandırılmaktadır.  $f$  iki gözlem arasındaki maksimum uzaklığı sağlayan öznitelik olmak üzere Chebyshev uzaklığı Eş. (21)’de verildiği gibidir. Uzaklığın, farklı  $h$  değerleri için geometrik anlamı Şekil 16’dan incelenebilir.

$$d(i, j) = \lim_{h \rightarrow \infty} \left( \sum_{f=1}^n |x_{if} - x_{jf}|^h \right)^{\frac{1}{h}} = \max_f |x_{if} - x_{jf}| \quad (21)$$

Örneğin iki gözlem  $A(3, 5)$  ve  $B(7, 6)$  şeklinde boyutlu uzayda tanımlı olsun. Gözlemler arası mesafe, birinci öznitelik dikkate alındığında  $|3-7|=4$  birim ve ikinci öznitelik dikkate alındığında  $|5-6|=1$  birimdir. Dolayısı ile söz konusu iki gözlem arasındaki Chebyshev mesafesi  $d(A, B) = 4$  birim olarak tespit edilmiş olur.



Şekil 16. Farklı  $h$  değerlerine ait uzaklıkların geometrik anlamı (Friedman, Hastie, Tibshirani, 2008)

<sup>59</sup> TÜBA bilim terimleri sözlüğünde en küçük üst sınır olarak çevrilmiştir.

Son olarak, özneliklerin bir önem derecesine sahip olma durumunda benzeşmezlik ölçütlerine  $w$  gibi bir ağırlık katsayısı dahil edilebilmektedir. Bu işlem tüm benzeşmezlik ölçütlerine uygulanabilir. Örneğin, Öklit mesafesi, ağırlıklandırma işleminden sonra aşağıda yer alan Eş. (22)'ye dönüşmektedir.

$$d(i, j) = \sqrt{w_1(x_{i1} - x_{j1})^2 + w_2(x_{i2} - x_{j2})^2 + \dots + w_q(x_{in} - x_{jn})^2} \quad (22)$$

### 5.6.Farklı Türde Özneliklere Sahip Gözlemler İçin Benzerlik-Benzeşmezlik Ölçütleri

Farklı türde özneliklere sahip veri setlerinde gözlemler arası mesafenin hesaplanmasında Eş. (23)'te yer alan benzeşmezlik ölçütü kullanılmaktadır. Bu ölçüt bazı kaynaklarda Gower Uzaklığı "Gower Distance" olarak da adlandırılmaktadır.

$$d(i, j) = \frac{\sum_{f=1}^n \delta_{ij}^{(f)} a_{ij}^{(f)}}{\sum_{f=1}^n \delta_{ij}^{(f)}} \quad (23)$$

Eğer  $i$  ve  $j$  gözlemlerinin  $f$  özneliği için gözlem değerleri  $x_{if}$  veya  $x_{jf}$  eksik gözlemler ise veya  $x_{if} = x_{jf} = 0$  ve  $f$  asimetrik ikili bir türde bir öznelik ise Eş.(23)'te yer alan  $\delta_{ij}^{(f)} = 0$ 'dır. Diğer tüm durumlarda  $\delta_{ij}^{(f)} = 1$  olarak kabul edilir. Başka bir ifade ile iki gözlem arasında hesaplanacak bir uzaklık olduğu tüm durumlarda  $\delta_{ij}^{(f)} = 1$ , diğer tüm durumlar için  $\delta_{ij}^{(f)} = 0$  olarak kabul edilmektedir. Bu şekilde tüm gözlemler için hesaplanan mesafeler toplamı, gözlemleri birbirleri ile karşılaştırılabilir kılan toplam öznelik sayısına bölünmüş ve böylelikle bir nevi normalize edilmiş olur. Diğer taraftan  $f$  özneliği nümerik ise gözlemler arası uzaklık Eş. (24) ile, nominal veya ikili ise Eş. (25) ve ordinal ise Eş. (26) kullanılarak hesaplanır.

$$d_{ij}^{(f)} = \frac{|x_{if} - x_{jf}|}{\max_{hf} x_{hf} - \min_{hf} x_{hf}} \quad (24)$$

$$d_{ij}^{(f)} = \begin{cases} 0, & x_{if} = x_{jf} \\ 1, & x_{if} \neq x_{jf} \end{cases} \quad (25)$$

$$z_{if} = \frac{r_{if} - 1}{M_f - 1} \quad (26)$$

**Örnek 11:** Dört farklı plastik ham madde kullanarak ticari araçlar için trim parça üreten bir işletme bu malzemeleri kendi kalite laboratuvarında bazı testlere tabi

tutmaktadır. İşletme, bu test sonuçlarını kullanarak malzemeleri kalitelerine göre gruplamak istemektedir. Bu ve benzeri gibi problemlerde farklı türde öznitelikler ile tanımlanan gözlemlerin gruplanmasında her bir öznitelik türü için öncelikle, ilgili uzaklık eşitliği kullanılarak gözlemler arası mesafe hesaplanmalıdır ve hesaplanan uzaklıklar Eş. (23) kullanılarak birbirleri ile entegre edilmelidir.

**Tablo 7.** Örnek 11'e ait veri

Malzeme	Test 1 Nominal	Test 2 Ordinal	Test 3 Sürekli
1	Kırmızı	İyi	45
2	Mavi	Yeterli	22
3	Sarı	Orta	64
4	Kırmızı	İyi	28

Test 1 sonuçlarına göre gözlemler arası mesafe Eş.(25) kullanılarak aşağıdaki gibi hesaplanır.

$$\text{Test 1 için gözlemler arası benzeşmezlik: } d(i, j) = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ 1 & 1 & 0 & \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

Test 2 sonuçlarına göre gözlemler arası mesafenin hesaplanabilmesi için öncelikle ordinal değerlere nümerik sıra değerleri atanmalı ve daha sonra Eş. (26) kullanılarak uzaklıklar hesaplanmalıdır.

**Tablo 8.** Örnek 11'e ait örnek bir çözüm

Malzeme	Test 2 Ordinal	Test 2 Normalize
1	3	1
2	1	0
3	2	0.5
4	3	1

$$\text{Test 2 için gözlemler arası benzeşmezlik: } d(i, j) = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ 0.5 & 0.5 & 0 & \\ 0 & 1 & 0.5 & 0 \end{bmatrix}$$

$\max_{hf} x_{hf} = 64$  ve  $\min_{hf} x_{hf} = 22$  olmak üzere Test 3 sonuçlarına göre gözlemler arası uzaklık matrisi Eş. (24) kullanılarak aşağıdaki gibi hesaplanır.

$$\text{Test 3 için gözlemler arası benzeşmezlik: } d(i, j) = \begin{bmatrix} 0 & & & \\ 0.55 & 0 & & \\ 0.45 & 1 & 0 & \\ 0.40 & 0.14 & 0.85 & 0 \end{bmatrix}$$

Tüm öznitelikler dikkate alındığında Eş. (23) kullanılarak örneğin birinci ve ikinci gözlemler arası benzeşmezlik  $d(i, j) = \frac{1(1)+1(1)+1(0.55)}{3} = 0.85$  olarak hesaplanır.

### 5.7.Yüksek Korelasyonlu Özniteliklere Sahip Gözlemler İçin Benzerlik-Benzeşmezlik Ölçütleri

Aralarında yüksek korelasyon bulunan veya tanım aralıkları yaş ve maaş gibi birbirinden çok farklı genişlikte olan özniteliklere ait gözlemler arası uzaklığın belirlenmesinde Mahalanobis Uzaklığı<sup>60</sup> “Mahalanobis Distance” kullanılmaktadır (Tan, Steinbach, Kumar, 2014).  $n$ , öznitelik sayısını temsil etmek üzere  $x_{in}$  ve  $x_{jn}$  gibi iki gözlem (vektör) arasındaki uzaklık Eş. (18) ile tanımlanan Öklit uzaklığı ile hesaplanmak istendiğinde, söz konusu iki gözlem dışındaki diğer tüm gözlemlere ait bilgiler göz ardı edilir. Fakat uzaklık matrisleri kullanılarak gerçekleştirilen sınıflandırma, kümeleme ya da aykırı değer analizlerinde, veri setinin tamamına daha bütünsel bir bakış açısı ile yaklaşan bir uzaklık ölçütünün kullanılması daha uygun olabilmektedir.  $x_j$  veri setinin merkezi ve  $x_i$  veri setindeki bir gözlem olmak üzere Eş. (27) ile tanımlanmış olan Mahalanobis uzaklığı, bir gözlemin, ait olduğu veri setinin merkez noktasına uzaklığının hesaplanmasında, tüm veri setine ait kovaryans matrisini (C) kullanarak, diğer gözlemlerin davranışlarını da dikkate almaktadır.

$$d(i, j) = \sqrt{(x_i - x_j)^T C^{-1} (x_i - x_j)} \quad (27)^{61}$$

Bölüm 2 Kısım 2.3.2.3’de açıklanan z-Skor normalleştirme bir gözlemin, ait olduğu veri setinin ortalamasından kaç standart sapma uzaklıkta olduğu bilgisini vermektedir. Mahalanobis mesafesi ise z-Skor’un çok boyutlu genelleştirilmiş hali olarak düşünülebilir. Açıklamak gerekirse Eş. (27)’teki kovaryans matrisinin tersi<sup>62</sup> ile çarpım

---

<sup>60</sup> Prasanta Chandra Mahalanobis, “Analysis of race-mixture in Bengal” isimli çalışmasında, kafataslarına ait çeşitli antropometrik ölçütleri kullanarak Bengaldeki ırk popülasyonları arasındaki benzerlik/benzeşmezlikleri (mesafeleri) tespit etmeye çalışmıştır. Bu süreçte Karl Pearson’ın geliştirdiği Irk Benzerlik Katsayısı “Coefficient of Racial Likeness” olarak adlandırılan parametrenin yetersiz olduğunu fark etmiştir. Mahalanobis (1927) çalışmasının devamındaki yıllarında ırklar arası diverjansın büyüklüğünü ölçebilmek amacı ile bugün Mahalanobis mesafesi olarak anılan  $D^2$  istatistiğinin son halini Mahalanobis (1936) çalışmasında alan yazına kazandırmıştır (Ghosh, Majumder, 1994).

<sup>61</sup> Tek değişkenli ( $i$ ’inci öznitelige ait) aykırı değer analizinde  $x_j$  değişkeni yerine  $\bar{x}_i$  değişkeni kullanılmaktadır. Çünkü amaç, gözlemlerin kendi ortalamasından sapmasını tespit etmektir.

<sup>62</sup> Kovaryans matrisinin tersi, *kesinlik matrisi* “precision matrix” veya *konsantrasyon matrisi* “concentration matrix” olarak da adlandırılmaktadır.

işlemi gözlemler arası uzaklığın varyansa bölünerek, veri setinin standardize edilmesini sağlar. Böylelikle gözlemler arası uzaklık standart sapma cinsinden ifade edilmiş olur. Kovaryans matrisinin birim matris olması durumunda ise Mahalanobis mesafesi, Öklit mesafesine eşitlenir. Bu durum, veri setini tanımlayan öznelikler arasında kovaryans olmadığı ve özneliklerin birbirine lineer bağımlı olmadığı anlamına gelmektedir.

Fakat diğer taraftan, kovaryans matrisinin hesaplanması bazı durumlarda zorlaşabilir. Örneğin veri seti enlemsel olarak çok geniş ise veya başka bir ifade ile veri, çok fazla öznelik kullanılarak tanımlanıyor ise öznelikler arası korelasyon olma ihtimali yükselir (çoklu doğrusallık) ve kovaryans matrisinin tekil matris olmasına neden olarak matrisin tersinin alınmasını imkansızlaştırır. Bir diğer problem ise veri setlerinin hem enlemsel hem de boylamsal olarak büyük olduğu durumlarda araştırmacıların öznelik indirgeme işlemi ile veri setinin boyutlarını küçülttükten sonra elde ettikleri çıktılar ile Mahalanobis mesafesini hesaplamasıdır. Bu durumda ise mesafe hesaplanırken kullanılan değerler (yeni özneliklerin değerleri) zaten ortogonal (korelasyonsuz) olduğu için gözlemler arası mesafelerin kovaryans ile düzeltilmesine gerek yoktur. Bu durumda kalıntı olan temel bileşenlerin, başka bir ifade ile varyansı en fazla açıklayan bileşenler dışında kalan diğer bileşenlerin, ağırlıklandırılması gerekmektedir (De Maesschalck, Jouan-Rimbaud, Massart, 2000).

**Örnek 12:** İki öznelik ile tanımlanmış olan bir veri seti için Mahalanobis uzaklıklarının hesaplanması ve görselleştirilmesi aşağıdaki gibi yapılabilir.

```
x1 <- c(100, 90, 150, 60, 110, 70, 55, 85, 140, 140, 100, 90, 120, 70, 130, 80, 30,
10, 175, 165, 115)
x2 <- c(5.2, 7.5, 8, 4, 6, 5, 2.5, 5, 7, 5.5, 4, 4, 4.5, 2.5, 6.5, 4, 2, 2, 8, 7, 2)
x <- cbind(x1, x2)
x_mean <- apply(x,2,mean)#x matrisindeki kolonların ortalamasını hesapla
x_cov <- var(scale(x))#x veri matrisini normalize et
x_cov <- var(x)
x_mah <- mahalanobis(x, x_mean, x_cov)
x_mah
-----
[1] 0.06005593 5.61807058 2.48279178 1.04065530 0.41505102 1.36806810 1.44881189
0.37865249
[9] 1.18784547 1.35377442 0.50371177 0.22571336 1.02143382 1.55355287 0.69151166
0.22253782
[17] 2.72019410 4.45986496 3.24873549 2.40840405 7.59056310
```

```

----
qchisq(.95, df=2)
----
[1] 5.991465
----
#Örneğin hesaplanan Mahalanobis mesafelerinden %95 ve s.d.=2 için ki-Kare değeri
5.991'den büyük olan en son gözlem, Şekil...’da yer alan A (115, 2) noktası, bu veri
seti ile benzeşmemektedir ve dolayısı ile bir aykırı değerdir. Aynı zamanda 5.61’lik
mMahalanobis mesafesi ile iki numaralı gözlem de aykırı değer olma sınırındadır.
plot(x, xlab = 'x1', ylab = 'x2',xlim=c(-60,260), ylim=c(-3,14))
ellipse(center = x_mean, shape = x_cov, radius = sqrt(qchisq(.90, df=2)),col =
'blue')
ellipse(center = x_mean, shape = x_cov, radius = sqrt(qchisq(.95, df=2)),col =
'blue')
ellipse(center = x_mean, shape = x_cov, radius = sqrt(qchisq(.99, df=2)),col =
'blue')
points(115, 2, col="red", pch=16)
text(115, 2, adj=2, labels="A")
points(175, 8, col="red", pch=16)
text(175, 8, adj=2, labels="B")

```

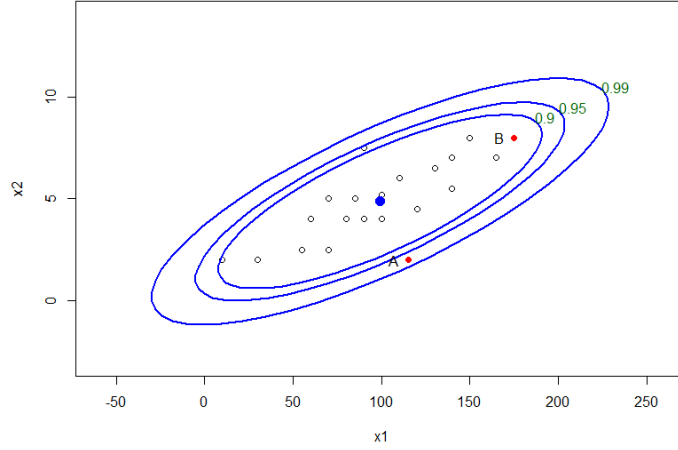
Şekil 17.’de yer alan elipsin merkezi, örnekleme oluşturan veri setinin ortalamasına/orta noktasına, şekli ise veri setinin kovaryansına eşittir. Dolayısı ile Mahalanobis mesafesinde bir gözlemin merkezden uzaklığı, elipsin gözlem doğrultusundaki genişliğine bölünür. Böylelikle uzaklık, veri setinin dağılımı ile göreceli olarak ilişkilendirilmiş olur. Ayrıca, bir elipsin üzerinde yer alan noktalar elipsin merkezine eşit uzaklıkta Mahalanobis mesafesine sahiptir.

Öklit uzaklığına göre A gözlemi merkeze daha yakın ve B ise merkezden daha uzaktır. Mahalanobis uzaklığına göre ise durum tam tersidir. B noktası, serbestlik derecesi (s.d.) “degrees of freedom”  $n = 2$  iken 0.95’lik güven aralığının<sup>63</sup> içinde kalmıştır ve aslında A noktası ile kıyaslandığında merkeze daha yakındır. 0.95’lik güven aralığı, Mahalanobis mesafesinde kesme değeri “cut-off value” olarak adlandırılmaktadır. Dolayısı ile A gözlemi, söz konusu veri setindeki gözlemlerin dağılımları dikkate alındığında, bir aykırı değer olarak nitelendirilir. Çünkü 0.95 aralığının dışında kalarak diğer gözlemlerden uzak bir mesafede konumlanmış olup, onlarla benzeşmemektedir.

---

<sup>63</sup> İstatistikte  $\alpha$  güvenirlilik seviyesi sosyal bilimler için genellikle 0.95 ( $p = 0.05$ ) olarak kabul görmektedir.

Sonuç olarak Öklit uzaklığı, iki gözlem arasındaki çizgisel mesafeyi hesaplarken Mahalanobis uzaklığı, gözlemlerin varyans (birinci ana bileşen) yönündeki dağılımını dikkate alarak hesaplama yapmaktadır.



Şekil 17. Mahalanobis uzaklığının geometrik gösterimi

Özniteliklerin<sup>64</sup> normal dağılıma sahip olduğu durumlarda, Eş. (27) ile tanımlanmış olan Mahalanobis mesafesinin karesi Ki-kare dağılımı özelliği taşımaktadır (Filzmoser, Garrett, Reimann, 2005). Fakat sosyal bilimlerde yer alan öznitelikler çoğu zaman normal dağılıma uymamaktadır ve dolayısı ile hesaplanan Mahalanobis mesafesinin anlamlandırılması zorlaşabilmektedir. Çünkü Mahalanobis mesafesinin bir üst sınırı yoktur. Bu nedenle mesafe, genellikle Ki-kare  $p$  değerine “Chi-square  $p$  values” dönüştürülür. Böylelikle mesafe  $[0,1]$  aralığına çekilmiş olur ve farklı güvenirlilik seviyeleri için anlamlandırılabilir.

$p$  değeri, veri setindeki (örneklemdeki) bir gözleme ait hesaplanan Mahalanobis mesafesinin, gerçek Mahalanobis mesafesinden büyük olma olasılığını temsil etmektedir. Sıfıra yakın  $p$  değerleri (kabul edilen güvenirlilik seviyesine tekabül eden  $p$  değerinden küçük olan değerler), söz konusu gözleme ait Mahalanobis mesafesinin çok büyük olduğu ve bu gözlemin, ait olduğu veri setini temsil etmediği/veri seti ile benzeşmediği anlamına gelmektedir. Bu yüzden Mahalanobis mesafesi çok değişkenli veri setlerinde aykırı değerlerin tespiti için de kullanılmaktadır.

Alan yazınında yakın geçmişte gerçekleştirilen çalışmalar sonunda, Mahalanobis mesafesinin aykırı değerlerden etkilenebildiği tespit edilmiştir. Bu durum maskeleye

<sup>64</sup> İstatistikte bağımsız ya da egzogen değişken olarak tanımlanmaktadır.

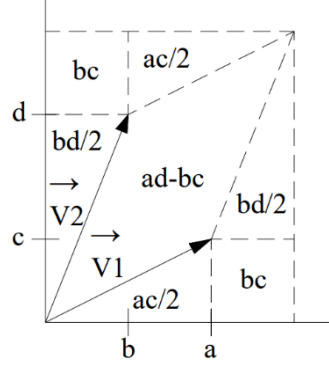


etkisi “masking effect” ve batırma etkisi “swamping effect” olarak meydana gelmektedir. Maskeleye etkisi, veri setindeki bir veya bir grup aykırı gözlemin, kendilerine yakın ve fakat merkezden daha uzak konumlanmış olan başka bir veya bir grup aykırı gözlemin tespitini, veri setinin ortalamasını arttırarak merkezini kaydırmak sureti ile engellemesi/maskeleye durumu. Batırma etkisi ise bu durumun tam tersidir. Bir veya bir grup aykırı gözleme yakın konumlanmış olan ve aslında aykırı olmayan gözlemlerin, aykırı olarak tespit edilmesine neden olmasıdır (Acuna, Rodriguez, 2004). Her iki tür etkinin giderilmesi amacı ile Sağlam Mahalanobis Mesafesi (SMM) “Robust Mahalanobis Distance” Rousseeuw tarafından geliştirilmiştir (P. J. Rousseeuw, 1984; P. Rousseeuw, 1985). Fakat bu ölçüt, algoritması P. J. Rousseeuw, Driessen (1999) çalışması ile iyileştirilene kadar, alan yazınında pek yer bulamamıştır (M. Hubert, Debruyne, 2010). SMM ölçütüne ait algoritma,  $m$  adet’lik gözlem içeren veri setinde  $h$ ’ar adetlik alt örneklem oluşturarak her bir alt örneklemin kovaryans matrisinin determinantını hesaplar ve en küçük determinanta sahip olan veri setini referans olarak aykırı değerleri veya gözlemler arası mesafeleri tespit eder. Bu yüzden SMM mesafesinin bir diğer adı minimum kovaryans mesafesidir(MKM) “minimum covariance distance”.

MKM’nin anlaşılabilmesi için, determinant kavramının anlaşılması önemlidir. Bir matrisin determinantı, o matrisin büyüklüğünü temsil etmektedir. Örneğin  $(2 \times 2)$  veya  $(3 \times 3)$  boyutlu bir matrisin determinantı, matrisi oluşturan satır vektörlerinin iki veya üç boyutlu düzlemde meydana getirdikleri alanı (paralelkenar “parallelogram”) veya hacmi (paralel-yüzlü “parallelepiped”) temsil etmektedir (Strang, 2016).

$(2 \times 2)$  boyutlu bir matrisin determinantının koordinat düzlemindeki gösterimi Şekil 18.’deki gibi ise, matrisin determinantı aşağıda yer alan Eş. (28) ile tanımlanır.

$$\begin{aligned}
 A &= \begin{vmatrix} a & c \\ b & d \end{vmatrix} = (a + b)(c + d) - \left( \frac{bd}{2} + bc + \frac{ac}{2} + \frac{bd}{2} + bc + \frac{ac}{2} \right) \\
 &= ac + ad + bc + bd - bd - 2bc - ac \\
 &= ad - bc \\
 &= \det A
 \end{aligned} \tag{28}$$



Şekil 18.  $(2 \times 2)$  bir matrisin determinantının geometrik gösterimi<sup>65</sup>

Bu alanın sıfıra yaklaşması, paralelkenarın,  $\vec{v}_1$  ve  $\vec{v}_2$  vektörleri arasındaki açının daralması ve vektörlerin birbirine yaklaşması ile gerçekleşir. Başka bir ifade ile vektörlerin bitiş noktalarına denk gelen paralelkenarın köşelerindeki  $bc$  alanları büyüyerek vektörleri daha dar bir alana sıkıştırıp, söz konusu vektörlerin birbirleri ile lineer bağımlılığını arttırmalıdır.

Eş. (29) ile tanımlanan  $(2 \times 2)$  boyutlu bir kovaryans matrisi bu bağlamda incelenirse  $bc$  alanının büyümesi ile  $\vec{v}_1$  ve  $\vec{v}_2$  vektörlerinin bağımlılığının artarak bu vektörleri tanımlayan iki öz niteliğin kovaryanslarının artması ve bu artışa nispeten varyanslarının daha düşük olması anlamına gelmektedir. Bu durum, iki öz niteliğin ölçümlendiği tüm gözlemlerin söz konusu vektörler etrafındaki salınımlarının azalırken bağımlılıklarının artması olarak da yorumlanabilir.

$$C = \begin{bmatrix} Var_1 & Cov_{12} \\ Cov_{21} & Var_2 \end{bmatrix} \quad (29)$$

SMM ölçütü bu bağlamda tekrar değerlendirilirse SMM,  $m$  gözlemlili  $n$  öz nitelikli bir veri setini  $h$  adetlik alt örneklemelere bölerek her bir alt örneklemin kovaryans matrisinin determinantını hesaplar ve öz nitelik vektörleri arası en küçük kovaryans determinantını sağlayan alt örneklemin ortalama ve varyans parametrelerini tespit eder. Amaç, veri setindeki en merkezi veri alt kümesini tespit edip bu veri setine ait istatistikleri kullanarak veri setinde geri kalan tüm gözlemleri söz konusu istatistikler ışığında değerlendirmektir.

<sup>65</sup> Blinn (2003) çalışması referans alınarak görselleştirilmiştir.

Alan yazınında  $h$  değeri için farklı oranlar yer almaktadır, sırası ile Eş. (30) (P. J. Rousseeuw, Driessen, 1999; M. Hubert, Debruyne, 2010), Eş. (31) (Filzmoser, Garrett, Reimann, 2005) ve Eş. (32) (Leys, vd., 2018).

$$\frac{m+n+1}{2} \leq h \leq m \quad (30)$$

$$\frac{m}{2} \leq h \leq m \quad (31)$$

$$h = 0.75m \quad (32)$$

SMM ölçütü, veri setinin merkezinin konumunu ve yayılımını sırası ile alt örneklemin ortalaması ve kovaryansı ile tahminleyen bir ölçüttür. Bu tür tahminleyicilerin tahmin performansının, veri setinin belirli bir miktarda aykırı değer içermesi durumunda bozulacağı ifade edilmektedir. Bozulmanın gerçekleştiği veri setindeki aykırı değer oranı, Bozulma Noktası “breakdown point” olarak adlandırılmaktadır. Bu sebeple, bir tahminleyicinin bozulma noktası ne kadar yüksek ise o tahminleyicinin sağlam olduğu ifade edilmektedir ve alan yazınında kabul edilen en yüksek bozulma noktası 0.5’tir. Bozulma noktası 0.5 olan sağlam bir tahminleyici, neredeyse yarısı aykırı değerden oluşan bir veri setinin ortalama, medyan, varyans ve benzeri gibi istatistiklerini başarılı bir şekilde tahminleyecektir.

**Örnek 13:** Örnek 12’deki veri seti SMM ölçütü ile değerlendirilir ise.

```
#library(mvoutlier)
x1 <- c(100, 90, 150, 60, 110, 70, 55, 85, 140, 140, 100, 90, 120, 70, 130, 80, 30,
10, 175, 165, 115)
x2 <- c(5.2, 7.5, 8, 4, 6, 5, 2.5, 5, 7, 5.5, 4, 4, 4.5, 2.5, 6.5, 4, 2, 2, 8, 7, 2)
x <- cbind(x1, x2)

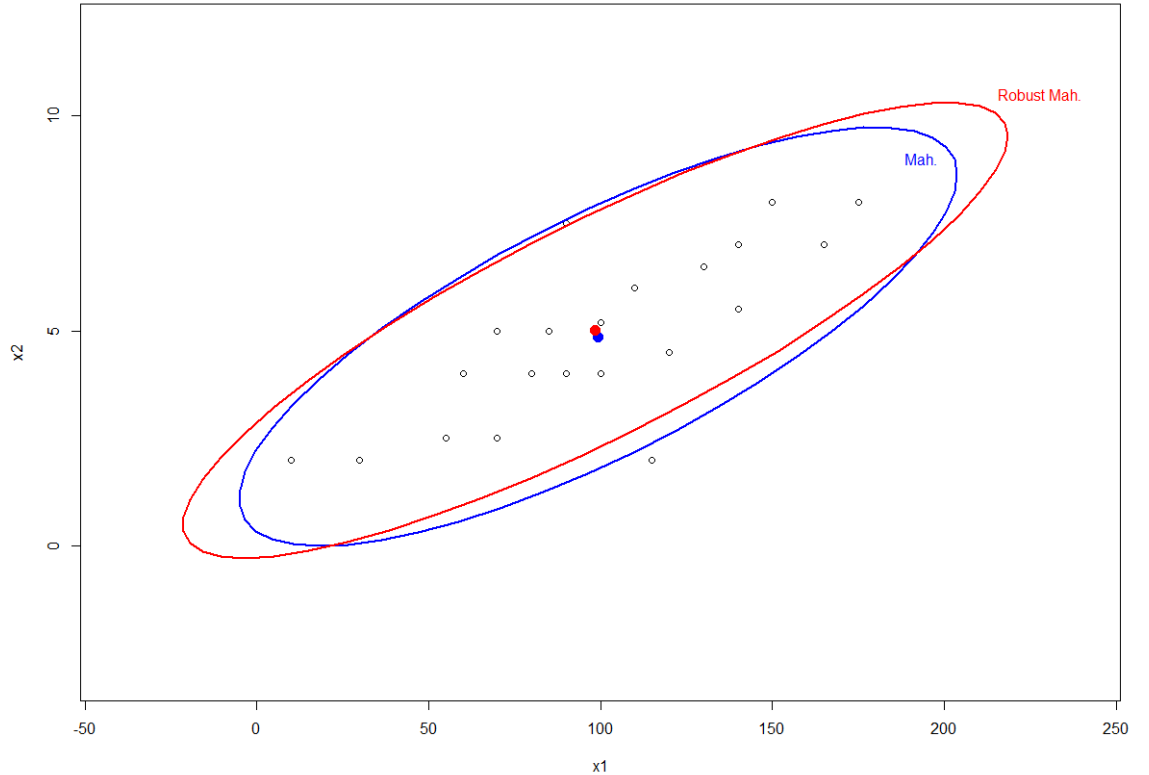
x_mean <- apply(x,2,mean)#x matrisindeki kolonların ortalamasını hesapla
x_cov <- var(scale(x))#x verisini normalize et
x_cov <- var(x)
x_mah <- mahalnobis(x, x_mean, x_cov)
qchisq(.95, df=2)

#library(robustbase)
#library(car)
#library(MASS)
x.mcd <- covMcd(x)
x.mcd$center
```

```
x.mcd$cov
```

```
plot(x1, x2, xlim=c(-40,240), ylim=c(-3,12))  
ellipse(center = x_mean, shape = x_cov, radius = sqrt(qchisq(.95, df=2)),col =  
'blue')  
ellipse(x.mcd$center, x.mcd$cov, radius = sqrt(qchisq(.95, df=2)), add=TRUE, col =  
'red')  
text(198, 9, adj=1, labels="Mah.", col="blue")  
text(240, 10.5, adj=1, labels="Robust Mah.", col="red")
```

Elde edilen sonuca göre veri setindeki bir gözlem daha aykırı değer olarak belirlenmiştir. Veri setinin merkezi değişmiştir. Elipsin eni daralmıştır. Elipsin şeklini Mahalanobis uzaklık ölçütü tanımından da hatırlanacağı üzere veri setinin kovaryansı belirliyordu. Bu veri seti için, veri setindeki tüm gözlemlerin kovaryansının determinantı, SMM ile elde edilen kovaryans matrisinin determinantından daha küçüktür, Şekil 19. Dolayısı ile orijinal veri seti aykırı değer içeren bir veri setidir.



Şekil 19. Mahalanobis mesafesi ve SMM'nin geometrik gösterimi

## ÜÇÜNCÜ BÖLÜM

### (KÜME ANALİZİ)

#### 1. KÜME ANALİZİ

Küme analizi, birbirinden çok farklı disiplinlerde yaygın bir şekilde kullanılan veri analizi yöntemlerinden bir tanesidir. Örneğin, sağlık bilimciler çeşitli parametreleri dikkate alarak tümörleri kümelemekte, uzay bilimciler şekillerine göre galaksileri gruplamakta, işletmeler tüketim davranışlarına göre tüketicileri gruplamakta, videoları sahnelere ya da kitapları ve metinleri içeriklerine göre kümelemektedir.

Küme analizi nicel bir analiz yöntemi olmasına rağmen, kümeleme veya gruplama ihtiyacının doğuşu (enteresan bir şekilde) antropoloji bilimi ile ilişkilidir.<sup>66</sup> Etimolojik yapısı “anthropos” insan ve “logos” bilim olan antropoloji, insanın fiziki yapısının ve davranışlarının evrimini, geçmiş ve günümüz insan topluluklarının biyolojik özelliklerini, toplumları ve kültürleri araştıran bir bilim dalıdır. Bilimsel nitelikte sayılabilecek ilk antropolojik araştırmalar 18-19. yüzyılda ve yoğunluklu olarak Amerika, Avustralya ve Afrika’da yürütülmüştür. Çağdaş nitelikte sayılabilecek günümüz antropoloji araştırmaları ise 20. yüzyılda başlamıştır (Beals, Hoijer, Çev. Erginer, 1972). Bu yıllarda özellikle phrenology (kafatasçılık) ve eugenics (kusursuz ırkın yaratılması)<sup>67</sup> araştırmaları ön plana çıkmıştır (Akpınar, 2018). Bu çalışmalar kapsamında kafatasının fiziksel yapısı incelenerek insanın kişiliği, karakteri ve suça yatkınlığı öngörülmeğe çalışılmış, saf, kusursuz, zeki, statü sahibi ve sağlıklı olan beyaz ırklı insanların çoğaltılması, beyaz olmayan tüm insanların, engellilerin ve sosyal tabakanın alt sınıfından olan “beyazların” toplumdaki ayrıştırılarak (veya kısırlaştırılarak veya hayatları sonlandırılarak) toplumların saflaştırılması çalışmaları da yapılmıştır.<sup>68</sup> İşte bu çalışmalar kapsamında korelasyon, Ki-kare istatistiği, TBA ve Mahalanobis mesafeleri geliştirilmiştir. Küme analizinin sosyal

---

<sup>66</sup> Aristo’nun canlıları sınıflandırması, bilinen ilk (hiyerarşik) kümeleme çalışmalarıdır (Hennig, Meila, 2016).

<sup>67</sup> Bu “bilim” dalları artık günümüzde sözde bilim dalı olarak değerlendirilmektedir, pseudoscience.

<sup>68</sup> Galton ve öğrencisi Pearson, eugenics çalışmalarının öncülerindedir. Galton’un en bilinen çalışmaları siyah ve beyaz bebeklerin ayrımı, zekanın kalıtsallığı ve korelasyondur. Pearson ise temel bileşenler analizi ve ki-kare istatistiği ile tanınmaktadır (Ki-kare istatistiği her ne kadar Pearson tarafından geliştirilmiş olsa da teorisi, Galton tarafından oluşturulmuştur). Mahalanobis de bu dönemin bilim insanlarından olup ünlü araştırması hakkında ikinci bölüm 5.7 numaralı kısımdan bilgi edinilebilir.

bilimlere yansması ise öncelikle kültürel özellikler “traits” dikkate alınarak kabilelerin tespiti, kabileler arası veya kültürler arası etkileşim ve davranışlar dikkate alınarak insanların gruplanması/tipolojilerinin çıkartılması konularında olmuştur (Köse, 2018).

Kümeleme “clustering” olarak da adlandırılan<sup>69</sup> küme analizi “cluster analysis” genel bir ifade ile bir veri seti içindeki grupların bulunması (Hennig, Meila, 2016) ya da bir veri setinin altbölümlere bölünmesi (Han, Kamber, Pei, 2012) işlemi şeklinde tanımlanabilir. Amaç, aralarında ortak özellikler bulunan birbirleri ile ilişkili nesnelere, gözlemleri veya olayları tanımlamak ve onları bu ortak özelliklere göre önceden belirlenmemiş kategorilere ayırmak ya da gruplamaktır (Duda, Hart, Stork, 2000; Waske, Benediktsson, 2014; Haykin, 2009). Ayırma ya da gruplama işlemi, kümeleme algoritması ve benzerlik/benzeşmezlik ölçütleri kullanılarak gerçekleştirilmekte; fakat farklı kümeleme algoritmaları aynı veri setini farklı kümelerle bölebilmektedir (Han, Kamber, Pei, 2012).

Küme analizinde en önemli noktalardan birisi, analiz sonucunda elde edilen kümelerde, aynı kümeyle ait olan gözlemlerin birbirine benzer/homojen veya geometrik açıdan yakın olması, farklı kümelerde yer alan gözlemlerin ise birbirinden farklı/heterojen veya geometrik açıdan ayrık olmasıdır.<sup>70</sup> Bu bağlamda küme analizinde hangi özneliklerin ayırt edici nitelik olarak seçilip, analize dahil edileceği, seçilen özneliklerin birbirlerine kıyasla göreceli önem dereceleri var ise bunların nasıl ağırlıklandırılacağı, hangi kümeleme algoritmasının kullanılacağı, gözlemler arası benzerliklerin hangi benzerlik/benzeşmezlik ölçütleri ile tespit edileceği, gözlemlerin doğru kümelerle atanıp atanmadığı, hangi gözlemlerin kümeler arasında kararsız kaldığı ve analiz sonucunda kaç kümenin olması gerektiğinin tespiti yüksek derecede önem taşımaktadır (P. J. Rousseeuw, 1987; Han, Kamber, Pei, 2012).

---

<sup>69</sup> Alan yazınında numerical taxonomy, grouping, clumping, typology ve q-analysis olarak da adlandırılmaktadır. Kümeleme analizindeki ilk kitap Robert Tryon tarafından 1939 yılında yazılmıştır (Wilmink, Uytterschaut, 1984; Murtagh, 2016).

<sup>70</sup> Küme içi homojenliğin, kümeler arası heterojenliğin ve kümeleme çalışmalarında birden fazla algoritma kullanılarak analizlerin karşılaştırmalı yapılmasının önemi ilk defa Cormack (1971) tarafından ifade edilmiştir (Wilmink, Uytterschaut, 1984).

## 1.1.Kümeleme Algoritmaları

Küme analizi, belirli bir algoritmanın adı olmayıp, çeşitli algoritmalar yardımı ile kümeleme işleminin gerçekleştirilmesi sürecidir. Bu süreçte genellikle istatistik, makine öğrenimi ve veri tabanı yönetimi alanlarında geliştirilen yöntemler/algoritmalar kullanılmaktadır. Kümeleme algoritmaları, sınıflama algoritmaları gibi veri setini öğrenerek tahminleme yapamazlar. Sadece, mevcut veri setini kullanarak incelenen popülasyonun tanınmasını sağlarlar. Bu yüzden gerçekleştirilen işlemler, veri setindeki gizli örüntüleri tarama amaçlı yapılır ve denetimsiz/danışmansız öğrenme “unsupervised learning” araçlarından birisi olarak kabul edilirler (Akpınar, 2018; Köse, 2018).

Alan yazınında çok fazla sayıda kümeleme algoritması bulunduğu için küme analizinde kullanılan algoritmaların net bir şekilde gruplanması veya kümelenmesi pek mümkün değildir. Örneğin Roger K. Blashfield 1976 yılında nerede ise küme analizi yapan kullanıcı sayısı kadar kümeleme algoritması olduğunu belirtmiştir (Wilmink, Uyterschaut, 1984). Günümüzde ise net bir sayı verilememek ile birlikte kümeleme algoritmalarının sayısının 100’ün üzerinde olduğu ifade edilmektedir (Akpınar, 2018). Bu neden ile kümeleme algoritmaları farklı kaynaklarda farklı şekillerde gruplandırılmaktadır.

Friedman, Hastie, Tibshirani (2008) çalışmasında kümeleme algoritmaları üç gruba ayrılmıştır: kombinatoryal algoritmalar “combinatorial algorithms”, karma modeller “mixture models” ve mod arama “mode seekers-bump hunters” algoritmaları. Kombinatoryal algoritmalar, herhangi bir olasılık modeli ile ilişkilendirilmeden doğrudan gözlemlenen veriler üzerinde çalışır. Mesafe ya da maliyet tabanlı algoritmalar olarak da adlandırılabilir. Amaç, küme içi mesafeleri minimize etmektir. Küçük veri setleri için uygun bir yaklaşım olarak kabul edilmektedir çünkü 19 gözlemin 4 kümeye atanması bile 34,105 farklı şekilde yapılabilmektedir.<sup>71</sup> Ayrıca, kombinatoryal algoritmaların lokal optimuma yakın bir sonuç verdiği de belirtilmektedir. Bu durumda sezgisel algoritmalar<sup>72</sup>

---

<sup>71</sup>  $N$  gözlem ve  $K$  küme sayısı olmak üzere, birbirinden farklı atama sayısı  $S(N, K) = \frac{1}{K!} \sum_{k=1}^K (-1)^{K-k} \binom{K}{k} k^N$  şeklinde hesaplanır.

<sup>72</sup> Genetik, arı kolonisi, kedi, karınca kolonisi, ateşböceği, parçacık, balık, bakteri, kurt, tabu arama, benzetmeli tavlama, yapay sinir ağları, destek vektör makineleri gibi.

kullanılarak global optimuma ulaşılabilir. Karma modellerde ise veri seti birden fazla olasılık yoğunluk fonksiyonu ile karakterize edilir. Her bir fonksiyon, bileşen yoğunluk fonksiyonu “component density function” olarak adlandırılır ve birer kümeyi temsil eder. Oluşturulan karma model daha sonra grafik yöntem, maksimum olabilirlik, beklenti maksimizasyonu veya Bayesçi yaklaşımlar ile tahminlenir.<sup>73</sup> Mod arama algoritmaları ise olasılık yoğunluğu fonksiyonunun farklı modlarını doğrudan tahmin etmeye çalışan, parametrik olmayan bir yaklaşıma sahiptir. Her bir moda en yakın gözlemler, kümeleri temsil etmektedir.<sup>74</sup>

Everitt, vd. (2011) çalışmasında kümeleme algoritmaları hiyerarşik, olasılıklı, model esaslı, yoğunluk esaslı, kısıtlı, bulanık ve Kohonen ağları şeklinde ve Han, Kamber, Pei (2012) çalışmasında ise bölümlenmeli, hiyerarşik, yoğunluk temelli, ızgara temelli ve gelişmiş şeklinde ayrılmıştır. Gelişmiş kümeleme algoritmaları altında ise bulanık, olasılıklı, çok boyutlu, ve kısıtlı kümeleme algoritmaları ile büyük boyutlu veri setlerinin ve ağ tipindeki veri setlerinin “graph/network data”<sup>75</sup> kümelenmesi incelenmiştir.

Hennig, Meila (2016) çalışmasında ise kümeleme algoritmaları dikotomik olarak incelenmiştir;

- Eğer her bir gözlem sadece bir kümenin elemanı ise kullanılan algoritma bir “hard” kümeleme algoritmasıdır, örneğin k-ortalama. Eğer bir gözlem aynı anda fakat farklı derecelerde birden fazla kümenin elemanı olabiliyorsa kullanılan algoritma “soft” kümeleme algoritmasıdır, örneğin bulanık kümeleme.

- Eğer oluşturulan kümeler alt kümelere ayrılabilir ise kullanılan algoritma hiyerarşik kümeleme algoritmasıdır, örneğin BIRCH “Balanced Reducing and Clustering Using Hierarchies”, CURE “Clustering Using Representatives”, ROCK “A Robust Clustering Algorithm for Categorical Attributes”, Chamelon; eğer değil ise yatay

---

<sup>73</sup> Alan yazınında “fitting mixture distributions” veya “fitting models” olarak yer almaktadır.

<sup>74</sup> Yazarlar PRIM algoritmasını örnek olarak sunmuşlardır. Söz konusu algoritma, öznitelik uzayının tamamını kapsayacak şekilde bir kutu ile uzayı taramaya başlar ve kutu içinde kalan gözlemlerin ortalaması maksimize olana kadar her iterasyonda kutu küçültülür.

<sup>75</sup> Sosyal ağlar ve tüketici-ürün eşleştirme grafikleri örnek verilebilir. Bu tip verilerin analizinde jeodezik mesafe ölçütleri kullanılmaktadır.



kümeleme algoritmasıdır, örneğin k-ortalamalar, k-medoid, k-prototip, k-mode, k-medyan ve Canopy gibi.

- Veri setindeki gözlemler, her bir özneliğe karşılık gelen değerler vektörü olabilir. Bu durumda algoritma ÖM ile çalışır. Bazı durumlarda ise ÖM yerine gözlemlerin birbirleri ile olan ilişki matrisi kullanılır ve algoritma benzerlik/benzeşmezlik matrisi ile çalışmış olur. Kullanılan algoritma ilişki kümeleme algoritması olarak kabul edilir “relational clustering”.

- Veri setindeki gözlemler örneğin mod, medyan, ortalama veya medoid gibi bir küme merkezi ile ilişkilendirilip, her bir gözlem kendisine en yakın merkezin olduğu kümeye atanabilir. Atama sonrası küme merkezleri güncellenir ve bu süreç iteratif bir şekilde küme içi mesafeler minimize edilene kadar devam eder. Amaç elde edilen kümelerin toplam biçim bozukluğunun “distortion” minimize edilmesidir. Biçim bozukluğu, küme içi mesafelerin karelerinin toplamı ile ölçülmektedir ve bu ölçüm maliyet “cost” olarak adlandırılmaktadır. Bu tip kümeleme algoritmaları maliyet tabanlı algoritmalarıdır (Ben-Israel, vd., 2007). Diğer taraftan olasılık tabanlı algoritmalar da bulunmaktadır. Olasılıklı kümelemede her bir gözlem belirli olasılık değerleri ile birden fazla kümenin elemanı olabiliyor olsa da esas amaç, her bir gözlemi gerçek fakat bilinmeyen tek bir kümeye atamaktır. Veri seti bu yüzden her bir gözlemin sadece bir kümenin elemanı olma olasılığını maksimize edecek şekilde bölünür veya kümelere ayrılır. Bulanık kümeleme de olasılıklı kümeleme algoritmaları kapsamında incelenen kümeleme algoritmalarındandır.

Türkçe alan yazınında ise en kapsamlı sayılabilecek gruplandırma Akpınar (2018) çalışmasında yer almaktadır. Bu çalışmada, kümeleme algoritmalarını klasikler ve çağdaşlar olmak üzere ikiye ayırmıştır. Klasik algoritmalar Tekli Bağlantı/En Yakın Komşu “Single Linkage/Nearest Neighbor”, Tam Bağlantı/En Uzak Komşu “Complete Linkage/Furthest Neighbor”, Aritmetik Ortalamalı Bağlantı/Gruplararası Bağlantı “Average Linkage/Between-Groups Linkage”, Merkezi Bağlantı “Centroid Linkage”, Medyan Bağlantısı “Median Linkage” ve Ward Bağlantısı “Ward’s Linkage” şeklinde listelenmiştir. Çağdaş algoritmalar ise kendi içinde beş gruba ayrılmıştır. Söz konusu beş grup BIRCH, CURE, ROCK ve Chamelon’u kapsayan Hiyerarşik Kümeleme

algoritmaları<sup>76</sup>, k-means, k-medodis ve uzantıları (PAM “Partition Around Medodis”, CLARA “Clustering Large Applications” ve CLARANS “Clustering Large Applications Based Upon Randomized Search”), k-prototypes, k-modes, k-medians, k-means++ ve Canopy algoritmalarını kapsayan Bölümleyici Kümeleme Algoritmaları,<sup>77</sup> DBSCAN “Density-Based Spatial Clustering of Applications with Noise”, OPTICS “Ordering Points to Identify the Clustering Structure” ve DENCLUE “Density-Based Clustering” algoritmalarını kapsayan Yoğunluk Temelli Kümeleme Algoritmaları,<sup>78</sup> STING “Statistical Information Grid” ve CLIQUE algoritmalarını kapsayan Izgara Temelli Kümeleme Algoritmaları<sup>79</sup> ve son olarak PROCLUS “Projected Clustering” ve SUBCLU “Density-Based Subspace Clustering” algoritmalarını kapsayan Alt Uzaklık Arama Algoritmaları şeklindedir.<sup>80</sup>

Türkçe alan yazınında kapsamlı sayılabilecek bir diğer gruplandırma Köse (2018) çalışmasında yer almaktadır. Bu çalışma, Akpınar (2018)’ ek olarak dağılım/olasılık temelli kümeleme algoritmalarını da dikkate almıştır. Bu kapsamda Beklenti Maksimizasyonu “Expectation Maximization-EM” ve  $\alpha$ -Beklenti Maksimizasyonu “ $\alpha$ -EM” algoritmalarından bahsedilmiştir.<sup>81</sup>

Yukarıda sayılan algoritma grupları hakkında kısaca bilgi vermek gerekir ise hiyerarşik kümeleme algoritmaları bağlantı temelli ve gözlemler arası uzaklığı dikkate alan bir yaklaşıma sahiptir. Bu yüzden, eşit aralıklı ve oranlı ölçme düzeyleri ile ölçülmüş olan veri setlerinin kümelenebilmesinde kullanılabilir. Gözlemler arası uzaklık hesabında, veri setini tanımlayan özniteliklerin türlerine uygun bir şekilde Bölüm 2 Kısım

---

<sup>76</sup> Sayılan algoritmalar ek olarak, DIANA “Divisive Anaysis” ve AGNES “Agglomerative Nesting” algoritmaları da bu grupta yer almaktadır.

<sup>77</sup> Sayılan algoritmalar ek olarak, MONA “Monothetic Analysis” algoritması da bu grupta yer almaktadır.

<sup>78</sup> Sayılan algoritmalar ek olarak, DBCLASD “Application Based Clustering Algorithms for Mining in Large Spatial Databases” algoritmaları da bu grupta yer almaktadır.

<sup>79</sup> Sayılan algoritmalar ek olarak, MAFIA “Merging of Adaptive Intervals Approach to Spatial Data Mining”, Wave Cluster ve O-CLUSTER “Orthogonal Partitioning Clustering” algoritmaları da bu grupta ek olarak yer almaktadır.

<sup>80</sup> Sayılan algoritmalar ek olarak, ORCLUS “Arbitrarily Oriented Projected Cluster Generation “, LAC “Locally Adaptive Clustering” ve Biclustering algoritmaları da bu grupta yer almaktadır.

<sup>81</sup> Sayılan algoritmalar ek olarak, PDQ “Probabilistic Distance Clustering” ve Gaussian Mixture Models algoritmaları da bu grupta yer almaktadır.

5'te açıklanan benzerlik/benzeşmezlik ölçütleri kullanılmaktadır. Kümeleme, aşağıdan yukarı (birleştirici)<sup>82</sup> veya yukarıdan aşağı (bölümleyici)<sup>83</sup> olmak üzere iki şekilde gerçekleştirilebilir. Bu yüzden elde edilen kümeler, sıklıkla dendrogramlardan<sup>84</sup> faydalanılarak görselleştirilir. Aşağıdan yukarı yaklaşımda, her bir gözlem bir kümeyi temsil edecek şekilde kümelenirken, yukarıdan aşağı yaklaşımda durum tam tersidir. Dolayısı ile bu grupta yer alan algoritmalar veri setindeki tüm durumları (veri setindeki tüm gözlemler arasındaki uzaklıkları) hesapladığı için durum uzayını arama “state space search” algoritmaları olarak da adlandırılmaktadır. Küme sayısı ise karar vericinin belirlemiş olduğu bir eşik değerine göre belirlenir. Dolayısı ile göreceli olarak değişkendir.

Bölümleyici kümeleme algoritmalarında  $m$  adet gözlemden oluşan bir veri seti, karar verici tarafından belirlenmiş  $k$  tane kümeye bölünmektedir. Küme adedi  $k$ 'nın algoritmanın başında tanımlanmış olması bölümleyici kümeleme algoritmalarının en çok eleştirilen özelliğidir. Bu gruptaki algoritmalarda aynı zamanda maksimum iterasyon sayısı ve yakınsama kriteri de<sup>85</sup> “convergence criterion” önceden tanımlanmalıdır.

Izgara temelli kümeleme algoritmalarında, veri seti sonlu sayıda hücrelere bölünerek çok boyutlu bir ızgara yapısı oluşturulur. Bu yaklaşım veri setini hiyerarşik bir şekilde katmanlara (hücrelere) bölmektedir. Izgaranın granülleşme düzeyi karar verici tarafından belirlenmektedir. Her bir hücre, alt katmanlardaki hücrelerin aritmetik ortalaması, minimumu, maksimumu, dağılımı, standart sapması ve benzeri gibi istatistiki bilgisini içermektedir. Bu gruptaki algoritmalar genellikle coğrafi bilgi sistemlerinde kullanılmaktadır.

Hem hiyerarşik hem de bölümleyici kümeleme algoritmaları, gözlemlerin koordinat düzlemi üzerindeki dağılımı öbek (küresel veya konveks) şeklinde olan kümelerin tespitinde etkilidir. Fakat gözlemler her zaman bir öbek şeklinde dağılmayıp

---

<sup>82</sup> Bottom up (agglomerative)

<sup>83</sup> Top down (divisive)

<sup>84</sup> Latince “dendron” ağaç, “gramma” çizim anlamına gelmektedir.

<sup>85</sup> Her iterasyonda yeniden hesaplanan küme merkezlerinin ardışık iki iterasyon arasındaki değişim miktarıdır (küme merkezlerindeki hareketlilik). Küme merkezlerindeki hareketlilik yakınsama kriterinden daha küçük bir değere sahip olduğunda algoritma sonlanmaktadır.

bir desen boyunca ve bir yoğunluğa bağlı olarak da dağılabilmektedir. Bu durumda, belirli bir yoğunlukta sürekliliği devam eden gözlem gruplarının birer küme olarak tespit edilebilmesine imkan sağlayan yoğunluk temelli kümeleme algoritmaları kullanılmaktadır. Dolayısı ile bu grupta yer alan algoritmalarda gözlemler arası mesafeler dikkate alınmamaktadır. Yerine, yoğunluk esas alınmaktadır. Yoğunluk, belirli bir yarıçapta ( $\epsilon$ ) kalan minimum nokta (gözlem) sayısı (*MinPts*) olarak ifade edilmektedir. Örneğin rassal olarak seçilen bir noktanın  $\epsilon$  etrafındaki gözlem sayısı, minimum gözlem sayısında küçük ise söz konusu nokta gürültü veya aykırı değer olarak adlandırılmaktadır. Bu gruptaki algoritmaların en büyük darboğazı  $\epsilon$  değerinin karar verici tarafından belirlenmesidir. Fakat çekirdek yoğunluk fonksiyonu “kernel based density function” veya sezgisel algoritmalar ile bu darboğaza çözümler getirilebilmektedir.

Yoğunluk temelli kümelemede gözlemlerin belirli bir desen boyunca sabit bir yoğunluğa sahip olması gerekmektedir. Fakat gerçek hayattaki veri setleri her zaman sabit bir dağılıma sahip olmayabilir. Yoğunluk bazen değişken olabilir. Bu gibi durumlarda ise olasılıklı kümeleme algoritmaları kullanılır. Bu algoritmalar, her bir olasılık dağılımı bir kümeyi oluşturacak şekilde veri setini kümeler ayırmaktadır. Amaç ise her bir gözlemin hangi kümeye dahil olduğu durumda hatanın minimize edilmesi<sup>86</sup> ya da olabilirliğin maksimize edilmesidir.<sup>87</sup> Karma modeller “mixture models” olarak da adlandırılmaktadır.

Bulanık kümeleme, gözlemler ile kümeler arasındaki bağlantının üyelik derecesi ile temsil edilmesi esasına dayanmaktadır. Bu gruptaki algoritmalar, üyelik derecesi  $[0, 1]$  aralığında bir değer olmak üzere, bir gözlemin birden fazla kümenin üyesi olmasına imkan verir. Kümeleme algoritması iteratif olarak çalışıp algoritmanın durma kriteri ya iki iterasyon arasındaki farkın karar verici tarafından önceden belirlenmiş bir  $\epsilon$  değerinden küçük olması ya da iterasyon sayısının azami bir değere ulaşmasıdır. Alan yazınında sıklıkla kullanılan bulanık kümeleme algoritması Bulanık c-ortalamalar “Fuzzy c-means” olup diğer bulanık kümeleme algoritmaları Soft k-means, Flame, Gustafson-Kessel ve Gath-Geva’dır (Akpınar, 2018).

---

<sup>86</sup> Regresyon Analizi ve En Küçük Kareler yöntemleri için

<sup>87</sup> Maksimum Olabilirlik ve Naive Bayes yöntemleri için.

İkinci bölüm 2.3.4 numaralı kısımda da bahsedildiği üzere büyük boyutun laneti, VM sürecini olumsuz etkileyen faktörlerden bir tanesidir. Bu durum kümeleme analizi için de geçerlidir çünkü çok boyutlu verinin anlaşılması ve görselleştirilmesi nerede ise imkansızdır. Boyut sayısı arttıkça ÖM'deki gözlemler arası mesafe yakınsanacak “converge” ve uzaklık kavramı anlamını yitirecektir. Ayrıca gerçekleştirilmek istenen analizin amacı ile alakasız veya gereksiz olan öznitelikler yanıltıcı sonuçlar doğurabilmektedir<sup>88</sup>, örneğin alakasız özniteliklerin bazı kümelerin oluşmasını engellemesi ve gerçekte var olmayan bazı kümelerin de oluşmasına neden olması gibi. Bu yüzden alakasız ve gereksiz özniteliklerin tespit edilerek analiz kapsamından çıkartılması gerekmektedir. Bu amaç ile alan yazınında alt uzay kümeleme algoritmaları (Akpınar, 2018)<sup>89</sup> ve boyut indirgeme algoritmaları geliştirilmiştir. Alt uzay kümeleme algoritmaları veri setini alt uzaylara bölerek hem her alt uzayı tanımlayan öznitelikleri hem de bu alt uzaylardaki kümeleri tespit edebilmektedir. Bu gruptaki algoritmalar Alt Uzay Arama “Subspace Search Methods”,<sup>90</sup> Korelasyon Temelli Kümeleme “Correlation-Based Clustering Methods”<sup>91</sup> ve “Biclustering Methods”<sup>92</sup> olmak üzere üç alt gruba ayrılmaktadır. Boyut Indirgeme algoritmaları ise orijinal veri setinde yer alan öznitelikler kullanılarak yeni öznitelikler (boyutlar) oluşturulması esasına dayanmaktadır. (Han, Kamber, Pei, 2012).

Kısıtlı kümeleme algoritması, karar vericinin bazı ön bilgiye sahip olduğu veya problemin doğasına has bazı koşulların analize dahil edilmek istendiği durumlarda tercih edilmektedir. Bu algoritma iki aşamaya sahiptir. İlk aşamada kısıtlar gruplanmaktadır.

---

<sup>88</sup> Alakasızlık, söz konusu özniteliğin konu ile tamamı ile alakası olmaması, gereksizlik ise veri setindeki birbirleri ile çok yüksek derecede korelasyon bulunan iki özniteliğin aynı anda analize dahil edilmesi durumudur.

<sup>89</sup> Öznitelik seçimi veya boyut indirgeme algoritmaları tüm uzayı kullanmaktadır. Başka bir ifade ile bu algoritmalar global düzeyde çalışmaktadır. Alt uzay kümeleme algoritmaları ise global uzayı alt parçalara bölerek yerelde çalışmaktadır.

<sup>90</sup> CLIQUE ve PROCLUS algoritmaları.

<sup>91</sup> Örneğin TBA ile birbirleri ile korelasyonsuz yeni boyutlar oluşturularak, yeni bir uzay veya onun alt uzaylarında kümeleme analizi yapılması gibi.

<sup>92</sup> Optimizasyon temelli metodları ve sayımlama temelli metodları kapsamaktadır.

Kısıt gruplama; gözlem kısıtları<sup>93</sup>, küme kısıtları<sup>94</sup> ve benzerlik ölçütü kısıtları olmak üzere üç farklı şekilde gerçekleştirilmektedir. İkinci aşamada ise kümeleme algoritması seçilmektedir. Seçim işlemi, kısıtların “hard”<sup>95</sup> ve “soft”<sup>96</sup> olması durumuna göre değişmektedir.

YSA’lar kapsamında yer alan denetimsiz öğrenme algoritmaları<sup>97</sup> genellikle kümeleme amaçlı kullanılmaktadır (Hagan, vd., 1996). Öz-Örgütlemeli (Kendini Düzenleyen) Haritalar (ÖÖH) "Self-Organizing Maps" ise alan yazınında en sık kullanılan algoritmalarındandır (Akpınar, 2018). Teuvo Kohonen tarafından geliştirildiği için Kohonen Ağları olarak da adlandırılmaktadır. Söz konusu tez çalışmasının kümeleme uygulaması ÖÖH’ler kullanılarak gerçekleştirildiği için, Bölüm 3 Kısım 3’te ÖÖH’ler hakkında detaylı bilgi verilmiştir.

Yukarıda kısaca özetlenmiş olan kümeleme algoritmaları, bu algoritmaların karşılaştırılması ve küme geçerliliklerinin (kalitesinin) test edilmesi hakkında Halkidi, Batistakis, Vazirgiannis (2001) ve Xu, Tian (2015) çalışmalarından bütünsel bir şekilde özet bilgi edinilebilir.

## 1.2.Uygun Küme Sayısının Tespiti

En uygun küme sayısının tespitinde genel kabul görmüş bir yaklaşım ya da ölçüt maalesef bulunmamaktadır. Bu ölçüt ve yaklaşımlardan bazıları aşağıda açıklanmaya çalışılmıştır. Alan yazınında bu amaçla kullanılan 30 ölçüt Charrad, vd. (2014) çalışmasından incelenebilir.<sup>98</sup>

Küme sayısının bir parametre olarak ele alındığı algoritmalar için, küme sayısı gözlem sayısının yarısının karekökü olarak belirlenebilir, Eş. (33). Bir diğer yaklaşım ise

---

<sup>93</sup> Birbirleri ile kesinlikle bağlantılı olması ya da olmaması gerekli gözlemler.

<sup>94</sup> Küme sayısı, kümedeki minimum gözlem sayısı veya kümenin şekli.

<sup>95</sup> COP-k-means algoritması

<sup>96</sup> CVQE algoritması

<sup>97</sup> İlişkisel Öğrenme “Associative Learning”, Rekabetçi Ağlar “Competitive Networks”, Hamming Ağları “Hamming Networks”, Öğrenen Vektör Nicemlemesi “Learning Vector Quantization”, Radyal Temelli Ağlar “Radial Basis Networks”, Grossberg Ağı “Grossberg Network” ve Adaptif Rezonans Teorisi “Adaptive Resonance Theory”.

<sup>98</sup> Küme sayısının tespitinde kullanılan ölçütlerden bazıları aynı zamanda küme kalitesinin incelenmesinde de kullanılmaktadır.

dirsek yöntemidir “elbow method”. Bu yöntemde farklı küme sayısı  $k$ 'lar için küme içi Kareler Toplamı (KT) “Sum of Squares-SS”<sup>99</sup> hesaplanmakta ve  $(k, KT)$  grafiği çizilmektedir. Grafiğin bükülme noktasındaki küme sayısı en uygun sayı olarak belirlenmektedir.

$$k = \sqrt{\frac{m}{2}} \quad (33)$$

Aşağıda algoritması açıklanan çapraz doğrulama “cross validation” yaklaşımı, yeniden örnekleme “resampling” yöntemlerinden olup küme sayısının tespitinde sıklıkla kullanılmaktadır (Han, Kamber, Pei, 2012; Kawamoto, Kabashima, 2017).<sup>100</sup>

1.  $t - 1$  parça eğitim seti  $D_t^E$  ve 1 parça test seti  $D^T$  olarak kullanılmak üzere veri seti  $D$ ,  $t$  adet parçaya ayrılır  $D_i$ ,  $i = 1, \dots, t$ .
2. Birinci eğitim seti  $D_1^E$  küme analizine tabi tutulur ve kümeler elde edilir.
3.  $D^T$  küme analizine tabi tutulur.
4. İkinci aşamanın çıktıları kullanılarak,  $D^T$ 'deki gözlemler,  $D_1^E$ 'deki kendilerine en yakın kümelere atanır. Böylelikle test seti, eğitim setinin özelliklerine dayalı olarak kümelenebilir olur. Atama işlemi ise gözlemler ile ikinci aşamada tespit edilmiş olan küme merkezleri arasındaki mesafeler dikkate alınarak yapılır. Dolayısı ile uzaklıkların karelerinin toplamı KT de tespit edilmiş olur.
5. Bu süreç, her bir parça test seti olarak kullanılabildiği kadar  $t$  defa tekrarlanır.
6.  $t$  tane hesaplanmış olan KT'lerin toplamının ortalaması alınır ve genel bir kümeleme kalitesi ölçütü hesaplanır.
7. Tüm algoritma farklı  $k$ 'lar için tekrarlanarak en küçük kümeleme kalitesi değerine sahip olan  $k$ , uygun küme sayısı olarak tespit edilir.

Bu yaklaşımdaki en önemli husus  $k$  sayısının tespitidir.  $k$  sayısı çok küçük belirlenir ise yanlılık artabilir ve varyans azalabilir. Tam tersi durumda ise varyans<sup>101</sup>

---

<sup>99</sup> Değişim “variation” olarak da adlandırılmaktadır. Kareler toplamının gözlem sayısına bölümü varyansa, varyansın karekökü ise standart sapmaya eşittir.

<sup>100</sup> Çapraz doğrulama aynı zamanda küme analizi sonuçlarının değerlendirilmesinde ve kümeleme algoritmalarının performanslarının karşılaştırılmasında da kullanılmaktadır.

<sup>101</sup> Test setindeki gözlemlerin eğitim seti küme merkezlerine olan uzaklıkları esas alınarak hesaplanmaktadır.

artabilir ve yanlışlık azalabilir (Diamantidis, Karlis, Giakoumakis, 2000). Ayrıca veri setinin genellikle %10 – %90 ya da %20 – %80 test-eğitim setleri şeklinde bölünmesi önerilmektedir. Dolayısı ile  $k = 10$  ya da 5 olarak algoritma çalıştırılmalıdır (Friedman, Hastie, Tibshirani, 2008).

Yukarıda açıklanan kriterlere ek olarak Ortalama Karesel Standart Sapma (OKSS) “Root-Mean-Square Standard Deviation”, Yarı Kısmi R-kare (YKR) “Semi-Partial R-Squared”, R-kare (RK) “R-squared” ve kümeler arası mesafeler de (KAM) küme sayısının tespitinde sıklıkla kullanılan ölçütlerdendir. Halkidi, Batistakis, Vazirgiannis (2001)’e göre, özellikle hiyerarşik kümelemede, küme sayısının doğru tespit edilebilmesi için bu dört ölçütün birlikte değerlendirilmesi gereklidir. Ayrıca dendrogramlar da hiyerarşik kümeleme algoritmaları için uygun küme sayısının tespitinde kullanılmaktadır (Kassambara, 2015). OKSS ölçütü kümelerin homojenliğinin bir göstergesidir ve bu değer olabildiğince küçük olması gereklidir. Başka bir ifade ile en küçük OKSS değerini sağlayan  $k$ , olası en uygun küme sayısıdır.  $KT_{Iç}$  küme içi KT,  $KT_{Ara}$  kümeler arası KT ve  $KT_{Veri}$  tüm veri setine ait KT,  $m$  gözlem sayısı ve  $n$  öznitelik sayısı olmak üzere KT, OKSS, YKR ve RK sırası ile Eş. (34-38) ile ifade edilmektedir.

$$KT = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_j)^2 \quad (34)$$

$$KT_{Veri} = KT_{Iç} + KT_{Ara} \quad (35)$$

$$OKSS = \sqrt{\frac{KT}{n(m-1)}} \quad (\text{öznitelikler eşit sayıda gözleme sahip}) \quad (36)$$

$$YKR = \frac{KT_{Iç}^{(yeni)} - (KT_{Iç}^{(eski)1} + KT_{Iç}^{(eski)2})}{KT_{Veri}} \quad (37)$$

$$RK = \frac{KT_{Ara}}{KT_{Veri}} \quad (38)$$

$KT_{Ara}$  kümeler arası farklılığı temsil etmektedir ve Eş. (35)’den de inceleneceği üzere,  $KT_{Ara}$ ’nın değeri arttıkça,  $KT_{Iç}$ ’in değeri azalacaktır ve bu nedenle elde edilen kümeler homojen nitelikli olacaktır. Dolayısı ile Eş. (38) ile tanımlanmış olan RK hem kümeler arası farklılığı hem de küme içi homojenliği nitelendiren bir ölçüttür. Bu ölçütün değeri sıfıra yaklaştıkça kümeler arası benzerlik yükselirken küme içi homojenlik



azalmakta, 1'e yaklaştıkça ise kümeler arası benzerlik azalırken küme içi homojenlik artmaktadır.

Eş. (37)'den de inceleneceği üzere YKR, birleştirilen iki kümenin birleşme öncesi ve sonrası KT değerlerinin değişim oranını vermektedir. Bu oran homojenlik kaybı “loss of homogeneity” olarak da adlandırılmaktadır. Eğer KT ölçütü sıfır ise birleştirilen iki kümenin KT değerleri birbirine eşittir ve dolayısı ile iki “homojen ve benzer” kümeden tek bir küme elde edilmiş olur. Diğer taraftan, bu ölçütün değerinin sıfırdan uzaklaşması ise yeni kümenin aslında iki heterojen kümenin birleşiminden oluştuğu anlamına gelmektedir.

KAM ise 2. Bölüm 5.5 numaralı Kısımda açıklanan benzerlik/benzeşmezlik ölçütlerinden bir tanesi kullanılarak hesaplanabilir. OKSS, YKR, RK ve KAM ölçütleri hesaplandıktan sonra  $(k, OKSS - YKR - RK - KAM)$  grafiği çizilerek kırınım noktası uygun küme adedi olacak şekilde tespit edilir.

### 1.3.Küme Kalitesinin İncelenmesi

Küme analizi sonuçlarının uygunluğunun başka bir ifade ile kümelerin geçerliliklerinin “validity” doğrulanması, uygun olan en az küme sayısına karar verilmesi kadar önemli bir konudur. Küme analizi sonucunda elde edilen kümelerin birbirinden ayrık olması, küme içindeki gözlemlerin birbirlerine olabildiğince yakın olması, kümelerin kendi içlerinde birbirine benzer nitelikli gözlemleri içermesi (homojenlik/uyum/saflık) ve birbirine benzer olan gözlemlerin tek bir kümede toplanarak diğer kümelerde yer almaması (bütünlük) (Amigó, vd., 2009; Han, Kamber, Pei, 2012)<sup>102</sup>

---

<sup>102</sup> Ayrıca, “rag bag” de kümeleme analizinde dikkat edilen bir diğer kriterdir. “rag bag” kelimesi kümeleri bozmak olarak tanımlanabilir. Şöyle ki, aynı veri seti iki farklı kümeleme algoritması ile kümelensin. İlk algoritma sonucunda elde edilen  $A$  kümesi neredeyse kusursuz bir şekilde homojen olsun.  $B$  kümesi ise heterojen gözlemler ve bir aykırı değer içeriyor olsun. İkinci algoritma ile elde edilen  $C$  kümesi homojen gözlemler ve bir aykırı değer içersin.  $D$  kümesi ise sadece heterojen gözlemlerden oluşsun. Söz konusu aykırı değer, kümelerin yapısını bozan bir elemandır ve bu elemanın, küme kalitesi zaten düşük olan  $B$  kümesinde yer alması, küme kalitesi yüksek olan  $C$  kümesinde yer almasına her zaman tercih edilir. Dolayısı ile kümeleme algoritmalarının performansları karşılaştırılmak istendiğinde mevcut veri setine aykırı değerler veya (aynı ana küleden olduğu düşünülen) farklı bir sınıf etiketine ait gözlemler karıştırılır ve sonuçlar karşılaştırılır.

Kümelemedeki bir diğer önemli kriter az sayıda gözlem içeren küçük grupların (sınıfların) korunmasıdır “cluster preservation”. Eğer küçük gruplar daha küçük kümelere ayrılır ise bu kümelere ait olan gözlemlerin gürültü/aykırı değer olarak değerlendirilmesi kaçınılmazdır. Dolayısı ile bu kritere göre, az sayıda gözlem içeren kümelerin (bir sınıfı temsil ettiği bilinen gözlemlerin) daha küçük kümelere

ve belki de en önemlisi, kümelerin kararlı “stable” olması gereklidir. Başka bir ifade ile ana küleden rassal olarak yeni bir gözlem (aykırı değer olmadığı sürece) çekildiğinde bu gözlemin analize dahil edilmesi halinde bile benzer küme analizi sonucu elde edilebilmelidir (Mucha, Bartel, 2017). Bu bağlamda, küme kalitesinin incelenmesi içsel “internal” ve dışsal “external” olmak üzere iki farklı yaklaşımla yapılabilmektedir (Tan, Steinbach, Kumar, 2014; Han, Kamber, Pei, 2012).<sup>103</sup> Meila (2016) çalışmasından bu bölümde Kısım 1.3.1 ve 1.3.2’de açıklanan değerlendirme kriterleri ve daha fazlası hakkında detaylı bilgi edinilebilir.

### 1.3.1. İçsel Değerlendirme Kriterleri

Bu yaklaşımda kümeler, kendilerini oluşturan veri setinden yararlanılarak değerlendirilmektedir. Amaç, bir kümedeki gözlemlerin diğer kümelerden ne kadar uzak olduğunu ve birbirleri ile uyumlarını “cohesion” kontrol etmektir (Amigó, vd., 2009). Davies-Bouldin, Dunn, Calinski-Harabasz ve Silhouette endeksleri bu kapsamda sıklıkla kullanılan geçerlilik endekslerindedir (Akpınar, 2018). Ayrıca, bu bölümde Kısım 1.2’de açıklanan KT, OKSS, YKR, RK ve KAM ölçütleri de içsel değerlendirme kriteri olarak kullanılmaktadır.

$k$  küme sayısı,  $M_k$   $k$ ’ncü kümenin merkezi,  $\mu_k$   $k$  kümesinde bulunan tüm noktaların  $M_k$ ’ya olan uzaklıklarının ortalaması ve  $d(M_i, M_j)$   $i$ ’nci ve  $j$ ’nci küme merkezleri arasındaki uzaklık olmak üzere Davies-Bouldin endeksi Eş. (39) ile tanımlanmaktadır. Küme içi uzaklıkların düşük, kümeler arası uzaklıkların yüksek olması durumunda düşük bir Davies-Bouldin endeksi elde edilir ve dolayısı ile kümeleme kalitesi yüksek olarak değerlendirilir.

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \left( \frac{\mu_i + \mu_j}{d(M_i, M_j)} \right) \quad (39)$$

---

bölünmesi, kümeleme sonuçlarını miktarsal açıdan daha fazla gözlem içeren kümelerin daha küçük kümelere bölünmesinden daha kötü etkilemektedir.

<sup>103</sup> Ayrıca göreceli “relative” değerlendirme de yapılmaktadır. Bu kapsamda, küme analizi birden fazla kümeleme algoritması ile gerçekleştirilip, her bir algoritmanın çıktısı içsel ve dışsal değerlendirme kriterleri ile karşılaştırılmaktadır.

$d(i, j)$   $i$ 'inci ve  $j$ 'inci kümeler arasındaki uzaklık<sup>104</sup> ve  $d'(k)$   $k$ 'inci kümedeki iki nokta arasındaki uzaklık<sup>105</sup> olmak üzere Dunn endeksi Eş. (40) ile tanımlanmaktadır. Dunn endeksi küme içi uzaklıkların düşük ve kümeler arası uzaklıkların yüksek olması durumunda büyük bir değer almaktadır ve bu durum kümelemenin kalitesinin iyi olduğuna işaret etmektedir. Bu endeks aynı zamanda birden fazla kümeleme algoritması ile kümeleme yapıldığı zaman, algoritmaların karşılaştırılmasında da kullanılabilir. Bu durumda, en yüksek endeks değerini sağlayan kümeleme algoritmasının diğer algoritmalara göreceli olarak daha iyi performans sergilediği kabul edilmektedir.

$$D = \frac{\min_{1 \leq i < j \leq k} d(i, j)}{\max_{1 \leq l \leq k} d'(k)} \quad (40)$$

Calinski-Harabasz endeksi kümeler arası varyansın, küme içi varyansa oranı olarak tanımlanmaktadır, Eş. (41). Dolayısı ile büyük endeks değerleri, yüksek küme kalitesi anlamına gelmektedir. Çünkü büyük endeks değeri, değerlendirilen küme bölümlenmesinin kümeler arası yüksek uzaklığa, küme içi düşük uzaklığa sahip olduğu anlamına gelmektedir.

$$CH = \frac{KT_{Ara}/(k-1)}{KT_{Iç}/(n-k)} \quad (41)$$

$a(i)$ , içsel benzeşmezlik “within dissimilarity”  $i$ 'inci gözlem ile bu gözlemin bulunduğu kümede yer alan diğer gözlemler arasındaki mesafelerin ortalamasını,  $b(i)$   $i$ 'inci gözlemin (kendi kümesi hariç) diğer tüm kümeler ile arasındaki ortalama mesafelerin<sup>106</sup> en küçüğü olmak üzere Silhouette endeksi Eş. (42) ile tanımlanır. Bu

---

<sup>104</sup>  $d(i, j)$  iki farklı küme arasındaki uzaklığın hesaplanmasında iki kümenin merkezleri arasındaki mesafe ya da iki kümenin birbirine en yakın veya en uzak elemanları arasındaki mesafelerden herhangi birisi olarak hesaplanabilir.

<sup>105</sup>  $d'(k)$  bir küme içindeki iki nokta arasındaki uzaklığın hesaplanmasında, küme içinde birbirine en uzak iki nokta arasındaki mesafe, küme içindeki tüm noktalar arası mesafenin ortalaması veya küme içindeki tüm noktaların merkeze olan mesafelerinin ortalaması şeklinde hesaplanabilir.

<sup>106</sup> Örneğin veri seti üç kümeye ayrılmış olsun.  $A$  kümesindeki bir  $i$  gözleminin,  $B$  kümesine ortalama uzaklığı  $d(i, B)$ ,  $i$ 'nin  $B$  kümesindeki tüm gözlemlere olan uzaklıklarının ortalamasıdır ve  $C$  kümesine ortalama uzaklığı  $d(i, C)$  ise  $i$ 'nin  $C$  kümesindeki tüm gözlemlere olan uzaklıklarının ortalamasıdır. Dolayısı ile  $b(i) = \min(d(i, B), d(i, C))$  şeklinde hesaplanmaktadır.

endeks, her bir gözlemin ait olduğu küme dışında ona en yakın konumlanmış olan komşu küme bilgisini de sağlamaktadır.<sup>107</sup>

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (42)$$

Bir kümedeki her bir gözlem için endeks değerleri hesaplandıktan sonra bu değerlerin aritmetik ortalaması alınarak, söz konusu kümenin homojenlik değeri tespit edilmiş olur. Bu değer 1'e yakın olması, içsel benzeşmezliğin kümeler arası benzeşmezlikten çok küçük olduğu ve söz konusu gözlemin neredeyse kusursuz bir şekilde en uygun kümeyle atandığı anlamına gelmektedir. Diğer taraftan bu değer -1'e yakın olması, içsel benzeşmezliğin kümeler arası benzeşmezlikten çok büyük olduğu ve söz konusu gözlemin yanlış kümeyle atandığı anlamına gelmektedir. Bu durumda, söz konusu gözlemin  $b(i)$  değerini sağlayan komşu kümeyle atanması gereklidir çünkü bu kümeyle daha yakın bir mesafede konumlanmışır. Son olarak, endeks değerinin 0'a yakın olması ise hem içsel hem de kümeler arası benzeşmezlik değerlerinin birbirine çok yakın olduğu, söz konusu gözlemin aslında hiçbir kümenin elemanı olmadığı ve kümeler arasında kararsız bir şekilde konumlandığı anlamına gelmektedir. Bu gözlem, "intermediate case" olarak değerlendirilmektedir. Silhouette endeksi en uygun küme sayısının tespiti amaçlı da kullanılmaktadır. Bu durumda her farklı  $k$  sayıda küme için birer homojenlik değeri hesaplanır ve en büyük homojenlik değerine sahip olan küme adedi uygun küme sayısı olarak seçilir (Han, Kamber, Pei, 2012; P. J. Rousseeuw, 1987).

### ***1.3.2. Dışsal Değerlendirme Kriterleri***

Dışsal yaklaşımda ise kümeler, bir kıyaslama bilgisi "gold standart/prior information" esas alınarak değerlendirilmektedir (Amigó, vd., 2009). Bu yüzden eğer veri etiketleri veya gözlemlerin sınıfları hakkında bir ön bilgi yok ise dışsal değerlendirme yapılması zorlaşmaktadır (S. Huang, vd., 2014). Böyle bir durumda araştırmacıların farklı seçenekleri bulunmaktadır.

İlk seçenek, kümeleme sonuçlarını birden fazla içsel değerlendirme kriteri ile değerlendirmek, kümeler arası farklılığı varyans analizine tabi tutmak (t-testi veya F-testi) mümkün ise veri setini başka algoritmalar ile de kümeleyip, kümeleme sonuçlarını

---

<sup>107</sup>  $b(i)$  değerini sağlayan küme,  $i$  gözleminin komşu kümesidir.

karşılaştırmak veya veri setini test ve eğitim seti olmak üzere ikiye bölüp, test setini eğitim seti üzerinden kümeleyerek dışsal değerlendirme kriterlerini hesaplamaktır.<sup>108</sup> Fakat Diamantidis, Karlis, Giakoumakis (2000) çalışmasında da belirtildiği üzere bu ve benzeri yaklaşımlar (algoritmanın ikinci aşamasında veri setinin bir kısmı analiz dışında tutulduğu için) beklenenden daha kötü sonuçlar sağlayabilmektedir. Başka bir ifade ile bu algoritma ile hesaplanacak olan dışsal değerlendirme kriterleri olası muhtemel kötü değerlere sahip olacaktır. Bu da aslında (belki de) iyi gerçekleştirilmiş bir küme analizi çalışmasının, kötü değerlendirilmesine neden olacaktır

İkinci seçenek (Costa, De Carvalho, De Souto, 2004, 2002) k-parçalı çapraz doğrulamanın<sup>109</sup> bir versiyonu olup, aynı zamanda 3. Bölüm 1.2 numaralı Kısımda uygun küme sayısının tespit edilmesi için açıklanan algoritmaya benzemektedir:

1. Veri seti  $D$ , küme analizine tabi tutulur.
2. Veri seti  $D$ ,  $k - 1$  parça eğitim seti  $D^E$  ve 1 parça test seti  $D^T$  olarak kullanılmak üzere  $k$  adet parçaya ayrılır.
3.  $D^E$ , birinci aşamada kullanılmış olan kümeleme algoritması ve benzerlik/benzeşmezlik ölçütü ile küme analizine tabi tutulur ve kümeler elde edilir.
4. Üçüncü aşamanın çıktıları kullanılarak,  $D^T$ 'deki gözlemler,  $D^E$ 'de kendilerine en yakın olan kümelere atanır. Böylelikle, test seti eğitim setinin özelliklerine dayalı olarak kümelenebilir olur. Atama işlemi, üçüncü aşamada kullanılmış olan benzerlik/benzeşmezlik ölçütü ile mesafeler hesaplanarak gerçekleştirilir.
5. Birinci ve dördüncü aşamaların çıktıları kullanılarak, dışsal değerlendirme kriterleri hesaplanır.
6. Bu süreç, her bir parça test seti olarak kullanılabilecek kadar  $k$  defa tekrarlanır.
7.  $k$  tane dışsal değerlendirme kriterinin ortalaması alınır ve genel bir kümeleme kalitesi ölçütü hesaplanır.

---

<sup>108</sup> Aynı zamanda kümeleme algoritmalarının performanslarının ölçümü için de kullanılmaktadır ve “holdout with random sampling” olarak adlandırılmaktadır.

<sup>109</sup> Cross validation, stratified cross-validation, leave one out cross validation, k-fold cross validation, bootstrapping (empirical, parametric, random, jittering) gibi tekniklerin hepsi yeniden örnekleme yaklaşımlarındandır. Detay bilgi için Efron, Tibshirani (1994), Diamantidis, Karlis, Giakoumakis (2000), Chernick, LaBudde (2011), Leisch (2016) ve Mucha, Bartel (2017) çalışmaları incelenebilir.

Yazarlar aynı zamanda, bootstrap algoritması ile de mevcut veri setinden geri iadeli yeni örneklem oluşturmak üzere dışsal değerlendirme kriterlerinin hesaplanabileceğini ve aynı zamanda kümeler arası varyans analizi de yapılabileceğini belirtmişlerdir.<sup>110</sup>

Üçüncü seçenek (Halkidi, Batistakis, Vazirgiannis, 2001):

1.  $(m \times n)$  boyutlu orijinal veri seti  $D$ , birbirlerine neredeyse eşit sayıda gözlem içerecek şekilde rassal olarak bölünür “partition”.
2.  $D$ , bölümleyici kümeleme algoritmaları(ndan bir tanesi) ile kümelendir.
3. Birinci ve ikinci aşamanın çıktıları kullanılarak orijinal veri setine ait dışsal değerlendirme kriterleri hesaplanır.
4.  $D$  kullanılarak Monte Carlo simülasyonu ile  $k$  tane sentetik veri seti  $D'_i$  oluşturulur,  $i = 1, \dots, k$ ,  $m' = m$  ve  $n' = n$ .
5.  $D'_i$  veri setindeki sentetik gözlemler birinci aşamada oluşturulan kümelere kendilerine en yakın olanlara atanır.
6. İkinci aşamada kullanılan kümeleme algoritmasının aynısı kullanılarak  $D'_i$  küme analizine tabi tutulur.
7. Beşinci ve altıncı aşamanın çıktıları kullanılarak sentetik veri setleri için dışsal değerlendirme kriter(ler)i hesaplanır.
8. Üçüncü ve yedinci aşamalarda elde edilen dışsal değerlendirme kriter(ler)inin olasılık yoğunluk fonksiyonları çizilir ve 0.95 anlamlılık seviyesinde dağılımların benzerlikleri (veya her iki fonksiyonun da aynı dağılımdan olup olmadığı) test edilir.

Saflık “purity”, kesinlik “precision”, Hubert  $\Gamma$  istatistiği, Rand endeksi, F-ölçütü “F-measure”, Jaccard endeksi, entropi ve Folwkes-Mallows endeksi bu amaçla en sık kullanılan dışsal değerlendirme kriterlerindedir. Söz konusu kriterlerin formülasyonları

---

<sup>110</sup> Kümeler arası varyans analizinde bootstrap algoritması geri iadeli olarak her bir küme için ayrı ayrı  $k$  defa çalıştırılır ve bootstrap örneklem oluşturulur. Oluşturulan örneklem seti varyans analizine tabi tutulur ve bir  $t$  (iki küme olması durumunda) veya  $F$  (üç ve üzeri küme olması durumunda) istatistiği hesaplanır. Bu süreç tüm  $k$ 'lar için tekrarlanır ve  $k$  tane  $t_{(B)}/F_{(B)}$  istatistiği elde edilir. Orijinal veri seti varyans analizine tabi tutulur ve  $t_{(O)}/F_{(O)}$  istatistiği hesaplanır.  $t_{(B)}/F_{(B)} \geq t_{(O)}/F_{(O)}$  şartını sağlayanların toplam adedi (#) belirlenir ve toplam,  $k$  sayısına bölünerek elde edilen anlamlılık seviyesinin bootstrap tahmini hesaplanır “the bootstrap estimate of the achieved significance level-ASL” (Efron, Tibshirani, 1994).

ve kullanımı, Halkidi, Batistakis, Vazirgiannis, (2001), Amigó, vd. (2009) ve Tan, Steinbach, Kumar (2014) çalışmalarından faydalanılarak aşağıda açıklanmıştır.

$m$  gözlem sayısı,  $k$  küme sayısı,  $l$  sınıf sayısı,  $C_i$   $i$ 'inci kümeyi ve  $T_j$   $j$ 'inci sınıfı temsil etmek üzere saflık Eş. (43) ile temsil edilmektedir. Bu eşitliğe göre, her bir kümede baskın olan sınıfa ait gözlem sayılarının toplamının, toplam gözlem sayısına bölümü ile küme analizinin saflığı derecelendirilmiş olur. Dolayısı ile saflık, bir kümede yer alan gözlemlerin aynı sınıftan olma/olmama durumunu derecelendirerek, yanlış kümelenen gözlemleri cezalandırmakta fakat birlikte kümelenmiş olan aynı sınıfa ait gözlemleri ödüllendirmemektedir. Örneğin, Eş. (43)'e göre aynı sınıftan olan gözlemlerin aynı kümelerde gruplanması ile saflık derecesi 1'e yaklaşacaktır. Fakat diğer taraftan, her bir gözlem bir küme olacak şekilde kümelenirse de, ( $m = k$ ) durumunda, saflık yine 1'e eşit olacaktır. Bu, uygun bir kümeleme değildir.

$$Saflık = \frac{1}{m} \sum_{i=1}^k \max_j |C_i \cap T_j| \quad j = 1, \dots, l \quad (43)$$

Bu durumda  $i$ 'inci küme için hesaplanan kesinlik ise Eş. (44) ile ifade edilmektedir.

$$Kesinlik(C_i, T_j) = \frac{|C_i \cap T_j|}{|C_i|} \quad (44)$$

Hubert  $\Gamma$  istatistiği, matrislerden bir tanesi dikotom olmak üzere,  $X$  ve  $Y$  gibi iki matris arasındaki nokta seri korelasyonudur Eş. (46). Küme analizindeki kullanımında ise  $X$  benzeşmezlik matrisi,  $Y$  dikotom değerlere sahip sınıf matrisi (ön bilgi/golden standard),  $S$  standart sapma ve  $m$  gözlem sayısı olma üzere Eş. (47) ile ifade edilmektedir. Hubert  $\Gamma$  istatistiği,  $X$  ve  $Y$  gibi iki girdi arasındaki lineer uyumu ölçen bir kriterdir. Hubert  $\Gamma$  istatistiğinin aldığı değer ne kadar büyük olursa, söz konusu iki matrisin bir biri ile uyumlu olduğu belirtilmektedir (Jain, Dubes, 1988).

$$Y(i, j) = \begin{cases} 0 & \text{eğer } i \text{ ve } j \text{ gözlemleri aynı kümede ise} \\ 1 & \text{değil ise} \end{cases} \quad (45)$$

$$\Gamma = \sum_{i=1}^{n-1} \sum_{j=i+1}^n X(i, j)Y(i, j) \quad (46)$$

$$\Gamma = \left\{ (1/m) \sum_{i=1}^{n-1} \sum_{j=i+1}^n [X(i, j) - \bar{X}][Y(i, j) - \bar{Y}] \right\} / S_x S_y \quad (47)$$

İki farklı küme analizi sonucunda elde edilen kümeler dikkate alındığında<sup>111</sup>  $a$  her iki küme analizinde aynı kümede yer alan gözlem çiftlerinin sayısını (doğru pozitif),  $b$  her iki küme analizinde ayrı kümelerde yer alan gözlem çiftlerinin sayısını (doğru negatif),  $c$  ilk küme analizinde ayrı kümede fakat ikinci küme analizinde aynı kümelerde yer alan gözlem çiftlerinin sayısını (yanlış pozitif) ve  $d$  ise ilk küme analizinde aynı kümelerde fakat ikinci küme analizinde ayrı kümede yer alan gözlem çifti sayısını (yanlış negatif) temsil etmek üzere Rand endeksi Eş. (48) ile ifade edilmektedir. Dolayısı ile Rand endeksi doğru alınmış kararların bir yüzdesi olarak düşünülebilir çünkü hem doğru kümelemeyi ödüllendirmekte hem de yanlış kümelemeyi cezalandırmaktadır.

$$RE = \frac{a+b}{a+b+c+d} \quad (48)$$

Alan yazınında, Rand endeksi değerleri kullanılarak, bir hata matrisi “confusion matrix” oluşturulmuştur, Şekil 20’de. Matris, kontenjans tablosu olarak da kullanılmaktadır.<sup>112</sup>

		Birinci kümeleme analizi (Sınıf etiketi/Ön bilgi)	
		Aynı küme	Farklı küme
İkinci kümeleme analizi (Tahmin edilen etiket/Sınıf/Küme)	Aynı küme	a	c
	Farklı küme	d	b

Şekil 20. Hata matrisi

Fakat Rand endeksi  $c$  ve  $d$  değerlerini eşit değerlendirmektedir. Başka bir ifade ile farklı etiketlere sahip olan gözlemlerin aynı kümeye atanması durumunu  $c$ , cezalandırmamaktadır. Bu amaçla  $F$  ölçütü geliştirilmiştir. Bu ölçüt “recall” veya “sensitivity” olarak adlandırılan hassasiyetin, kesinlikten “precision” bir  $\beta$  katsayısı kadar önemli derecelendirilmesine imkan vermektedir. Bu katsayının artması ile  $c$

<sup>111</sup> İlk küme analizine ait atamalar, sınıf etiketleri ya da “gold standard” olarak adlandırılan bir ön bilgi de olabilir.

<sup>112</sup> İkinci bölüm 5.3 numaralı kısımdaki Tablo 4 incelenebilir.



değerlerinin  $d$  değerlerine karşı göreceli önemi azalacaktır. Bu yüzden, gerçekte farklı sınıflara atanmış olup da kümeleme sonucunda aynı kümeye atanmış olan gözlemlerin önemi azalmış olacaktır. Hata matrisi dikkate alınarak kesinlik oranı Eş. (49) ile yeniden yapılandırıldığında ve hassasiyet Eş. (50) ile ifade edildiğinde F ölçütü Eş. (51) ile gösterilir. Dolayısı ile örneğin hassasiyet ile kesinlik eşit öneme,  $\beta = 1$ , sahip ise F ölçütü hassasiyet ve kesinliğin harmonik ortalamasına tekabül etmektedir.

$$Kesinlik = \frac{a}{a+c} \quad (49)$$

$$Hassasiyet = \frac{a}{a+d} \quad (50)$$

$$F_\beta = (1 + \beta^2) \frac{Kesinlik * Hassasiyet}{(\beta^2 * Kesinlik) + Hassasiyet} \quad (51)$$

Dolayısı ile kesinlik, aynı sınıfta kümelenen (aynı sınıfta olduğu tahmin edilen) gözlem çiftlerinin gerçekte ne kadarının aynı sınıfa ait olduğunun oranını temsil ederken hassasiyet, gerçekte aynı sınıfta olan gözlem çiftlerinin kümeleme (tahminleme) sonucu hangi oranda yine aynı sınıfta yer aldığıнын oranını temsil etmektedir.

Bölüm 2 Kısım 5.3.'te Jaccard endeksi hakkında bilgi verilmiş olup bu kısımda her herhangi bir tekrar yapılmayacaktır.

Fowlkes-Mallows endeksi Jaccard ve Rand endeksleri gibi kümeler arası benzer ve benzer olmayan gözlem çifti sayılarını dikkate alan bir dışsal değerlendirme kriteridir. Eş. (52) ile ifade edilen endeks kesinlik ve hassasiyetin geometrik ortalamasıdır. Bu yüzden G-ölçütü "G-measure" olarak da adlandırılmaktadır.

$$FM = \sqrt{Kesinlik * Hassasiyet} = \sqrt{\frac{a}{a+c} * \frac{a}{a+d}} \quad (52)$$

$m$  gözlem sayısı,  $i = 1, \dots, k$  küme sayısı,  $j = 1, \dots, l$  sınıf (etiket) sayısı,  $p(C_j)$  bir gözlemin  $i$ 'inci kümede  $j$ 'inci sınıftan olma olasılığı ise Eş. (53) ile tanımlanan entropi, bir kümeye ait gözlemlerin ait oldukları sınıfların küme içindeki dağılımına dair bir ölçüttür. Bir küme ne kadar çok aynı sınıftan olan gözlemlerden oluşuyorsa o kümenin entropisi sifira yaklaşacaktır çünkü bu durumda eşitlikte yer alan oran 1'e yaklaşacak ve 1'e yakın bir değerın logaritması ise 0'a yakın bir değer olacaktır.

$$E(C_i) = - \sum_{j=1}^l p(C_j) \log p(C_j) = - \sum_{j=1}^l \frac{|C_j|}{m_i} \log \frac{|C_j|}{m_i} \quad (53)$$

Küme analizinin genel entropisi ise  $m_i$   $i$ 'inci kümedeki gözlem sayısı ve  $m$  toplam gözlem sayısı olmak üzere, her bir küme için Eş. (53) ile hesaplanmış olan entropilerin ağırlıklı ortalamasıdır, Eş. (54).

$$Entropi = \sum_{i=1}^k E(C_i) \frac{m_i}{m} \quad (54)$$

Alan yazınında ayrıca Cophenetic matris, Goodman-Kruskal  $\gamma$  istatistiği, Hubert-Arabie, Wallace endeksleri ile Mirkin, Van Dongen metrikleri, BCubed ve V-ölçütü “V-measure” de bulunmaktadır. Cophenetic matris ve Goodman-Kruskal  $\gamma$  istatistiği için Jain, Dubes (1988), Hubert-Arabie ve Wallace endeksi için (Jain, Dubes, 1988) Hubert, Arabie (1985), Mirkin ve Van Dongen metrikleri için Meila (2005), V-ölçütü için Rosenberg, Hirschberg (2007) ve BCubed ölçütü için ise Amigó, vd. (2009) ve Han, Kamber, Pei, (2012) çalışmalarından detaylı bilgi edinilebilir.

#### **1.4.Küme Analizini Olumsuz Etkileyen Faktörler**

Küme Analizi sürecini ve dolayısı ile analiz sonuçlarını olumsuz etkileyebilecek çeşitli faktörler bulunmaktadır (Akpınar, 2018; Tan, Steinbach, Kumar, 2014):

- Küme analizi, konveks olmayan, bir tür kesikli optimizasyon “discrete optimization” problemidir. Lokal minimum sayısının çok fazla olmasından ötürü tüm mümkün küme bölünmeleri denenmeden global minimumun bulunması nerede ise imkansızdır. Global minimum ancak gözlem ve küme sayısının çok küçük “extremely small” olduğu durumlarda tespit edilebilir (Tian Zhang, Ramakrishnan, Livny, 1997).

- Veri setlerinin boylamsal olarak genişlemesinden ötürü analizler genellikle ana kütleliyi temsil ettiği varsayılan örneklem üzerinde gerçekleştirilmekte ve yerel sonuçlar globale genellenmektedir. Bu yaklaşım çoğunlukla olumlu sonuçlar sağlamasına rağmen söz konusu örneklemin ana kütleliyi tamamen temsil etme garantisi bulunmamaktadır.

- Büyük veri setlerinin neden olduğu bir diğer problem, boyutlar arttıkça yoğunluğun sifıra yaklaşması ve ÖM'nin seyrekleşmesidir. Bu gibi bir durumda kümeleme algoritmalarının performansları düşmektedir ve araştırmacılar boyut indirgeme teknikleri ile veri setini daha kompakt bir hale getirerek küme analizi yapmaktadır.

- Kümeleme algoritmalarının büyük bir kısmı parametrik olarak çalışmaktadır. Parametre değerlerinin özellikle çok boyutlu (öznitelikli) veri setlerinde (karar vericiler

tarafından) doğru bir şekilde tanımlanması önemlidir. Çünkü bu algoritmalar, parametre değerlerine karşı oldukça hassastır. Değerlerdeki en küçük bir farklılık, veri setinin farklı bir şekilde bölümlenerek kümelenemesine neden olabilmektedir.

- Veri setinin normal dağılmaması, birden fazla dağılıma sahip olması, farklı yoğunlukta, büyüklükte ve birbirine yakın ya da iç içe geçmiş kümelere sahip olması ve aşırı miktarda aykırı değer içermesi de küme analizini olumsuz etkileyen faktörlerdendir.

## 2. KARMA VERİ KÜMELEME ALGORİTMALARI

Alan yazınındaki kümeleme algoritmalarının ve benzerlik/benzeşmezlik ölçütlerinin birçoğu ya sadece sınıflayıcı ve sıralayıcı ölçme düzeyleri ile ölçülen nitel öznitelikler içeren veri setlerine ya da sadece eşit aralıklı ve oranlı ölçme düzeyleri ile ölçülen nicel öznitelikler içeren veri setlerine uygundur. Fakat gerçek veri setleri genellikle hem nitel hem de nicel öznitelikler içermektedir. Örneğin riskli tedarikçilerin tespitinde fire oranı, ppm<sup>113</sup> oranı, teslimat süresi, çalışanların ve işletmenin yıl cinsinden deneyim süresi ve benzeri gibi nicel özniteliklere ek olarak işletmenin türü (küçük-orta-büyük), işletme stratejisi, esnekliği, işbirlikçiliği ve firma için stratejik önemi gibi sıralayıcı ya da sınıflayıcı ölçme düzeyleri ile ölçülen nitel öznitelikler de önemlidir. Bu ve benzeri gibi veri setleri karma veri setleri olarak adlandırılmaktadır.

Karma veri setlerinin kümeleneşmesi genellikle “zor” olarak nitelendirilmektedir. Çünkü bu tür veri setlerine toplama, çıkarma, çarpma, bölme ya da ortalama veya standart sapma alma gibi matematiksel işlemler yapılamamaktadır (Ahmad, Khan, 2016). Bu yüzden bu tür veri setlerinin kümeleneşmesinde, aslında birer veri dönüştürme işleminden ibaret olan iki yaklaşım yaygın olarak kullanılmaktadır (Shih, Jheng, Lai, 2010; Liang, vd., 2012; Balaji, Lavanya, 2018). Bu yaklaşımların ilki, 2. Bölüm 2.3.3 numaralı Kısımda açıklanan veri ayırıklaştırma yöntemlerinden birisini kullanarak nicel özniteliklerin nitel özniteliklere dönüştürülmesidir. Diğer yaklaşım ise örneğin kategorik özniteliklerin 1, 2, 3, ... gibi değerler kullanarak sayısallaştırılması veya 2. Bölüm 5.3 numaralı Kısımda yer alan Örnek 10'daki gibi, özniteliklerin asimetrik ikili değişken türlerine dönüştürülerek {0, 1} tanım aralığında sayısallaştırılmasıdır. Söz konusu

---

<sup>113</sup>  $ppm = \frac{\text{hatalı parça sayısı}}{\text{toplam parça sayısı}} * 10^6$

dönüştürme işlemlerinden bir tanesi gerçekleştirildikten sonra ise araştırmacılar, veri setine uygun kümeleme algoritması ve benzerlik/benzeşmezlik ölçütlerini kullanarak küme analizi yapmaktadır. Fakat her iki yaklaşımın da küme analizi sonuçlarını olumsuz yönde etkileyecek dezavantajlar bulunmaktadır (Liang, vd., 2012; C. C. Hsu, Chen, 2007b). Örneğin ilk yaklaşımda, sayısallaştırılacak olan kategorik öznitelik renk, şekil ya da cinsiyet gibi bir öznitelik ise ve özneliğin alabileceği nitel değerler {1, 2, 3} şeklinde sayısallaştırılıp, gözlemler arası uzaklık Öklit mesafesi ile hesaplandığında, elde edilecek sonuçların gerçeği yansıtmayacağı aşıkardır. Diğer taraftan, nümerik bir özneliğin kategorik hale getirilmesinde ise örneğin her bir kategorinin alt ve üst sınır değerlerindeki gözlemler eşit değerlendirilmiş olacak ve bu durum bilgi kaybına neden olacaktır ya da tam tersi bir durumda, kategoriler arası geçişlerin bulunduğu bölgelerde yer alan ve değerleri birbirine çok yakın olan iki gözlem ise birbirinden farklı kategorilerde yer alacaktır.

Karma veri setlerinin kümelenmesinde kullanılan en son yaklaşım ise bu bölümde 1 numaralı kısımda özetlenen kümeleme algoritmalarına nazaran çok daha az sayıda olan (Pathak, Pal, 2016) karma veri kümeleme algoritmalarından birisini kullanmaktır.

Karma veri kümeleme algoritmalarının sınıflandırılması, diğer kümeleme algoritmalarının (tek tip verileri kümeleyen algoritmaların) sınıflandırılması gibi biraz zordur. Fakat genel itibari ile bir grupta yapılmak istenir ise bölümlenmeli, hiyerarşik, model temelli, YSA temelli ve diğer algoritmalar şeklinde gruplanabilir.

### **2.1.Bölümlenme Esaslı Karma Veri Kümeleme Algoritmaları**

Bu grupta yer alan birçok algoritmanın eleştirilen ortak özelliği  $k$  küme sayısının bilindiği varsayımı ve  $C_i$  küme merkezlerinin rassal olarak tespit edilmesi gerekliliğidir. Bu yüzden araştırmacılar, bu bölümde Kısım 1.2'de açıklanan ve benzerleri gibi yöntemler ile küme sayısını başka bir algoritma ile tespit ederek, bu bilgiyi algoritmaya verdikten sonra kümeleme yapmaktadır. Küme merkezlerinin tespiti için de alan yazınında çeşitli algoritmalar yer almaktadır. Bu algoritmalarından bazıları için Ji, vd. (2015), Ahmad, Hashmi (2016) ve Jinyin, vd. (2017) çalışmaları incelenebilir.

Z. Huang (1998) çalışmasında yazar, öncelikle veri setini kategorik ve nümerik nitelikte öznitelikler içerecek şekilde ikiye ayırmış ve her bir veri setinde yer alan

gözlemler arası mesafeleri sırası ile Hamming ve Öklit mesafelerini kullanarak ayrı ayrı hesaplamıştır. Daha sonra bir eşitlik yardımı ile bu mesafeleri toplamış ve toplam mesafe matrisini kullanarak K-prototype algoritması ile küme analizini gerçekleştirmiştir. Fakat bu yaklaşım aslında tartışmaya açık bir yaklaşımdır çünkü küme merkezleri kategorik özniteliklerin modu ile temsil edilmiştir. Başka bir ifade ile küme merkezlerinin tespitinde nümerik öznitelikler dikkate alınmamıştır. k-ortalamar algoritmasının karma veri setleri için Ahmad, Dey (2007) tarafından geliştirilmiş olan versiyonunda ise gözlemlerin birlikte görülme sıklıklarını dikkate alan yeni bir uzaklık ölçütü ve dolayısı ile bir maliyet fonksiyonu tanımlanmıştır. Söz konusu fonksiyon, Z. Huang (1998) çalışmasında kullanılan maliyet fonksiyonunun gelişmiş versiyonudur. Eski versiyonda nümerik özniteliklerin hepsi 1 değeri ile ağırlıklandırılırken kategorik öznitelikler karar vericinin tercihine bağlı olarak serbest bir şekilde ağırlıklandırılabilirdi. Yeni versiyonda ise kategorik öznitelikler gözlemlenme sıklıklarına göre ağırlıklandırılırken, nümerik öznitelikler ayrıklaştırıldıktan sonra benzer bir algoritma ile ağırlıklandırılmaktadır. Ayrıca, küme merkezlerinin tespitinde hem kategorik değişkenler hem de nümerik değişkenler dikkate alınmaktadır. Yazarlar önerilen algoritmanın k-prototypes algortmasından daha iyi bir performansa sahip olduğunu ifade etmişlerdir. Cheng, Leu (2009) çalışmasında yer alan Kısıtlı k-prototype karma veri kümeleme algoritması ise k-prototype algoritmasına ait maliyet fonksiyonunun oluşturulan kümeler bazında kısıtlandırılmış<sup>114</sup> versiyonudur. Örneğin söz konusu çalışmada Taipei, Tayvan'daki köprüler bakım çalışmaları için kümelendiği. Küme kısıtlarını ise bakım maliyeti ve köprüler ile ilgili bazı teknik kriterler oluşturmuştur. Bir kısıta sahip olan her bir öznitelik için bir alt ve üst limit değerleri tanımlanmış ve her iterasyonda bir kümedeki kısıtlandırılmış öznitelige ait değer, önceden tespit edilmiş limitlerin içinde değil ise maliyet fonksiyonu, küme içi gözlemler arası ortalama uzaklık artı kısıt maliyeti kadar arttırılmaktadır. Son olarak Ji, vd. (2012)'nin çalışmasında ise Ahmad, Dey (2007) çalışmasında yer alan uzaklık ölçütü  $\alpha$  katsayısı ile bulanıklaştırılmış ve küme

---

<sup>114</sup> Kısıtlı kümeleme problemleri dört gruba ayrılmaktadır: problemin amacı ile alakalı gözlemler hakkındaki kısıtlar (örneğin mali değeri 1 milyon dolar ve üzerinde olan lüks konutların kümelmesi), problemin doğasına özgü kısıtlar, küme sayısı ve benzeri gibi bazı parametreler ile ilgili kısıtlar ve kümelerin kendileri ile alakalı kısıtlar (Tung, vd., 2001).

merkezlerinin hesaplanmasında, hem bulanıklık hem de küme ortalamaları dikkate alınmıştır.

## 2.2.Hiyerarşik Karma Veri Kümeleme Algoritmaları

Bu grupta yer alan algoritmalarda en önemli iki nokta gözlemler arası uzaklıkların hangi benzeşmezlik ölçütü ile ölçüleceği ve kümeler arası bağlantı kriterinin ne olacağıdır.

Hsu, Chen (2007a) çalışmasında karma veriler entropi ve varyans temelli bir algoritma ile kümelenebilir çalışılmıştır. Varyans, nümerik öznitelikler için ve entropi ise kategorik öznitelikler için gözlemler arası uzaklığın ölçümünde kullanılmıştır. Gözlemlerin entropi değerleri mesafe hiyerarşisi<sup>115</sup> ile ağırlıklandırılmıştır. Yaklaşım olarak aşağıdan yukarı hiyerarşik kümeleme yaklaşımı kullanılmıştır. Algoritma, veri setinde birbiri ile en benzeşmeyen gözlemlerden kümeler oluşturup daha sonra ikinci aşamada geri kalan gözlemleri bu kümelere atama yapmak şeklinde çalışmaktadır. İlk aşamanın zaman karmaşıklığı  $n$  gözlem sayısı olmak üzere  $n^2$  olduğu için, büyük veri setlerinin kümelenebilirliği açısından pratik olmadığı ifade edilmektedir. Bu yüzden veri setinin büyük olması durumunda mevcut veri setinden örneklem seçilerek algoritmanın ilk aşamasının gerçekleştirilmesi önerilmiştir. Lim, vd. (2012) çalışmasında da veri seti kategorik ve nümerik olarak ikiye ayrılmış ve kategorik veri seti kullanılarak Entropi ile birinci ön kümeleme işlemi yapılmıştır ve dolayısı ile bir  $k$  tespit edilmiştir. Daha sonra,  $k$  bilgisi bir girdi olmak üzere, sadece nümerik veriler kullanılarak ikinci ön-kümeleme işlemi yapılmıştır. Kümeler, içerdikleri gözlem sayılarına ve türlerine göre ağırlıklandırılmış (kümeler arasında daha dengeli dağılan özniteliklerin ağırlıkları, dengesiz dağılanlara nazaran daha az olacak şekilde) ve bu ağırlıklar kullanılarak yeni bir benzerlik matrisi oluşturulmuş ve bu matris aşağıdan yukarı yaklaşımla hiyerarşik olarak kümelenebilir. Son olarak hiyerarşik kümelemenin özel bir türü olan HCPC<sup>116</sup> de karma veri setleri için geliştirilmiştir. Bu yöntem üç temel aşamadan oluşmaktadır. İlk aşamada,

---

<sup>115</sup> “Distance hierarchy” olarak adlandırılmaktadır. Kavram hiyerarşisinin “concept hierarchy” bir uzantısıdır. Aynen AHP veya karar ağacı biçimselliğine sahiptir. Dolayısı ile alt seviyedeki her bir düğüm belirli bir ağırlık ile bir üst seviyedeki düğüme bağlıdır.

<sup>116</sup> Hierarchical Clustering on Principal Components

veri setindeki deęişkenlerin tamamı Karma Veriler İin Faktör Analizi<sup>117</sup> kullanılarak, daha az sayıda yeni deęişkenlere (temel bileşenlere) dönüştürülür. İkinci aşamada, bir önceki aşamada elde edilmiş olan TBA Ward's mesafe kriteri kullanılarak Hiyerarşik Küme Analizi ile kümelendir. En uygun küme sayısı, küme analizi sonucunda elde edilen dendrogramdan tespit edilir. Üçüncü aşamada ise bölümlenmeli kümeleme yaklaşımlarında birisi uygulanır (Kassambara, 2017).

### **2.3.Model Esaslı Karma Veri Kümeleme Algoritmaları**

Bu gruptaki algoritmaların başarımı veri setini temsil ettiği düşünölen istatistiksel modele ve parametrelerine baęlıdır. Bu grupta yer alan alıřmaların bazıları ařaęıdaki gibidir:

B. S. Everitt (1988) alıřmasında sonlu karma model “finite mixture density” kullanılarak nümerik, ikili ve sıralı türde öznitelikler içeren karma veri setine küme analizi uygulanmıştır. Verideki gözlemlenebilir sıralı ve ikili deęişkenlerin<sup>118</sup> aslında bir eşik deęere “treshhold” baęlı olan gözlemlenemeyen sürekli deęişkenlerden türetildięi belirtilmiştir. Bu nedenle, gözlemlenemeyen her bir sürekli deęişken için bir daęılım fonksiyonu oluşturularak, gözlemlenebilir olan ikili ve kategorik türdeki deęişkenlere ait daęılımlar elde edilmiştir. Bu yaklaşımın zaman karmaşıklığı yüksek olduęu için, alıřmanın yazarının da ifade ettięi üzere, ikili ve sıralı deęişken sayısının fazla olduęu veri setlerinde kullanımı önerilmemektedir. Ayrıca, sıralı ve ikili deęişkenlerin sürekli deęişkenlerden oluşturulduęunu varsayarak her bir deęişken için, o deęişkeni temsil eden sürekli deęişkenler yaratılması pek uygun ve uygulanabilir deęildir. Örneęin cinsiyet ya da renk gibi kategorik deęişkenler için bu şekilde bir dönüşümün yapılması nerede ise imkansızdır. Moustaki, Papageorgiou (2005) alıřmasında ise gizil sınıf modeli genişletilerek nitel verilerin de kümelmesine imkan saęlayacak şekilde gizil sınıf karma modeli “latent class mixture model” ile küme analizi gerçekleştirilmiştir. Kategorik öznitelikler ikili özniteliklere dönüştürölmüş ve multinom (okterimli) daęılım ile temsil edilmiştir. Pathak, Pal (2016) alıřmasında veri seti, nümerik ve kategorik özniteliklerden

---

<sup>117</sup> Bu yöntem, Temel Bileşenler Analizi ve oklu Uyum Analizi (Multiple Corrspondance Analysis) yöntemlerinin karması olan entegre bir yöntemdir.

<sup>118</sup> Özniteliklerin

oluşmak üzere ikiye ayrılmıştır. Öncelikle nümerik veri seti Öklit uzaklığı kullanılarak bulanık c-ortalamlar algoritması ile kümelendirilmiştir. Daha sonra, küme sayısı olarak nümerik veri setinde tespit edilmiş olan  $k$  referans alınarak, kategorik veri seti, (olasılık temelli) karma model ile kümelendirilmiştir. Böylelikle nümerik veri setinin küme analizi çıktıları referans alınarak kategorik veri setine ait kümelendirme tespit edilmiştir. Akabinde aynı işlemler bu sefer önce kategorik, daha sonra nümerik veri seti için tekrarlanarak, kategorik veri setinin küme analizi çıktıları referans alınmış ve nümerik veri setine ait kümelendirme tespit edilmiştir. En son şamada ise söz konusu kümelerdeki ortak yapıların tespiti için işbirlikçi “collaborative” kümeleme algoritması kullanılmıştır. Foss, vd. (2016) çalışmasında karma veri kümeleme için yarı-parametrik bir algoritma<sup>119</sup> geliştirilmiştir. Söz konusu algoritma, k-ortalamlar ile çok terimli Gauss karma modellerini birbirine entegre etmiştir. Amaç, nümerik ve kategorik özneliklerin küme analizi üzerindeki etkisini dengelemektir.

## 2.4.Yapay Sinir Ağları

ÖÖH’ler ve Adaptif Rezonans Teorisi (ART)<sup>120</sup>, kümeleme alan yazınında sıklıkla kullanılan YSA’lardandır. ÖÖH’ler, söz konusu tez çalışmasının uygulama kısmında kullanılan tekniklerden bir tanesi olduğu için ÖÖH’lere bu kısımda değinilmeyecektir.

---

<sup>119</sup> KAMILA

<sup>120</sup> ART ağları, ÖÖH’ler gibi danışmansız öğrenme kapsamında incelenen bir YSA türüdür. Bu ağlar iki katmandan oluşmaktadır. İlk katman girdi katmanı olup, ikinci katman çıktılarıdır (sınıflar/kümelere). Ağa giren ilk vektör (girdi/gözlem/örnek), ağda örnek model olarak saklanır. İkinci vektörün, araştırmacı tarafından tanımlanan, belirli bir benzerlik oranını sağlaması durumunda, ilk vektör ile aynı sınıfta kümelendirilir. Aksi halde ikinci vektör yeni bir kümede saklanır ve ağırlıklar, benzerlik kararına göre güncellenir. Dolayısı ile ağın hafızasına alınan eski bilgiler, yeni veri geldikçe güncellenir. Bu yüzden YSA’lardaki gibi, eski bilgiyi kaybetme tehlikesi bulunmaktadır. Fakat ART ağları, bu problemi aşmak için çift yönlü öğrenecek şekilde tasarlanmıştır. Başka bir ifade ile bu ağlar, bilgiyi sabit ve adaptif olarak saklamaktadır. Açıklamak gerekirse algoritma önce girdi katmanından çıktı katmanına doğru, daha sonra ise çıktı katmanından girdi katmanına doğru çalışmaktadır. Girdi vektörünün, araştırmacı tarafından tanımlanan, belirli bir benzerlik oranını sağlaması durumunda vektör bilgileri girdi-çıkıtı katmanları arasındaki uzun dönemli hafızaya alınır ve çıktı katmanında kümelendirilir (sabit saklama). Aksi halde ise kısa dönemli hafızadan silinir ve bir sonraki girdi vektörü kısa dönemli hafızada işlenmek üzere ağa verilir (adaptif saklama). Ayrıca, kümeleme işlemi ağa en son gözlem girene kadar devam eder. Başka bir ifade ile küme sayısının artma ihtimali yükselir ve bu, ağın gereksiz büyümesine neden olabilir. Ayrıca girdi katmanında kullanılan ağırlıklar, çıktı katmanındaki proses elemanının aktivasyon değeri, eşik değeri, öğrenme katsayısı ve benzerlik katsayısı gibi parametreler araştırmacı tarafından hazır bilgi olarak ağa verilmeli/tanımlanmalıdır (Öztemel, 2012; Aydın Keskin, 2008; Chabane, Islam, Zerr, 2017).



C.-C. Hsu, Huang, Hsiao (2006) ve bir benzeri olarak C. C. Hsu, Huang (2008) çalışmalarında, kategorik veriler için mesafe hiyerarşisi ve nümerik veriler için farkların karelerinin toplamının karekökü kullanılmış ve ART ağları ile küme analizi gerçekleştirilmiştir. Fakat 2008 çalışmasındaki kategorik öznitelikler için kullanılan mesafe hiyerarşisinde, gözlemler arası uzaklıklar hesaplanırken gözlemler, hiyerarşi ağacındaki yaprakları temsil eden kategoriler arasında konumlandırılmıştır. Dolayısı ile yapraklar arası mesafe 1 olmak üzere, gözlemin yapraklar arasındaki bağlantı üzerindeki konumu, uzman görüşlerine başvurulmuş,  $[0,1]$  aralığında sürekli bir değer ile ağırlıklandırılmıştır. Chao, vd. (2011) çalışmasında ise mobil ağ hataları, C. C. Hsu, Huang (2008) çalışmasındaki ART ağı kullanılarak kümelenebilmiştir. Son olarak Lam, Wei, Wunsch (2015) çalışmasında bulanık ART ağları kullanılarak öznitelik seçimi yapılmıştır. Bu çalışmada da kategorik değişkenler ikili değişkenlere dönüştürülmüştür. Öznitelik seçimi gerçekleştirildikten sonra veri seti görselleştirilmiş ve bir  $k$  sayısı görsel inceleme sonucunda öngörülmüştür. Öngörülen  $k$  sayısı kadar küme,  $k$ -ortalamalar algoritması ile elde edilmiş ve küme analizi sonuçları incelenmiştir.

Yukarıda kısaca açıklanan kümeleme algoritmalarına ek olarak alt uzay kümeleme, yoğunluk temelli kümeleme, spektral kümeleme ve topluluk öğrenmesi “ensemble method” algoritmaları da kullanılarak karma veri setleri kümelenebilmektedir.

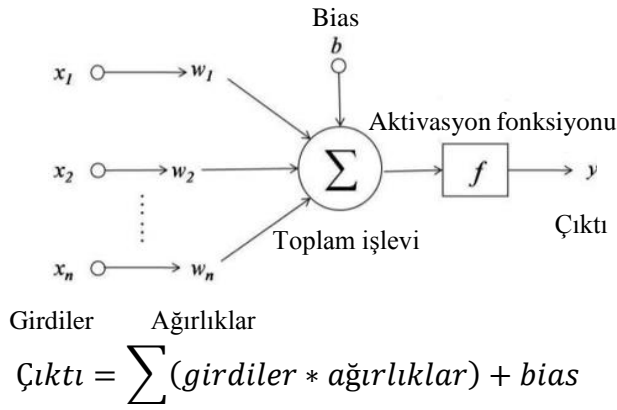
### 3. ÖZ-ÖRGÜTLEMELİ HARİTALAR

#### 3.1. Temel YSA Kavramları

ÖÖH’ler bir danışmana (ön bilgiye) gerek kalmadan yarışmacı öğrenme yaklaşımı ile eğitilen bir tür ileri beslemeli YSA’dır. YSA’lar biyolojik sinir ağlarını taklit eden programlardır. İnsan beyninden esinlenerek tasarlanmıştır. İşlem elemanları düğümlerdir Her bir düğüm aslında yapay bir sinir hücresi olup, biyolojik sinir hücresinin (nöron) basitleştirilmiş halidir ve düğümler birleşerek, öğrenen, hatırlayan ve genelleme yapabilen ağları oluşturmaktadır.

Bütün YSA’lar Şekil 21.’teki yapı esas alınarak geliştirilmiştir. Bu yapıdaki farklılıklar YSA’ların da farklılaşmasına neden olmaktadır. İleri beslemeli çok katmanlı bir YSA’nın gösterimi Şekil 22.’deki gibidir. Girdi katmanı, çevreden aldığı bilgiyi işlenmek üzere gizli katmana iletmektedir. Ağırlıklar, girdilerin nöron üzerindeki etkisini

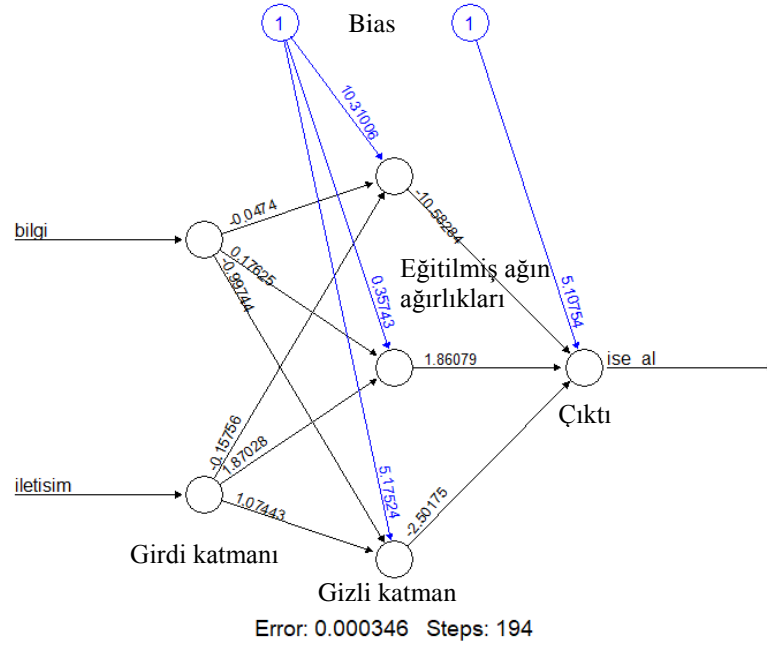
temsil eden katsayılarıdır ve bu katsayılar, tahmin edilen çıktının hatasındaki değişimine, öğrenme oranına ve momentum terimine bağlı olarak güncellenmektedir. Toplam işlevi, nörona gelen girdileri bir fomülizasyon ile birleştirerek aktivasyon fonksiyonuna iletir. Toplam işlevi bazı durumlarda, minimum, maksimum ya da normalleştirme gibi işlevler de olabilmektedir. Aktivasyon fonksiyonunun görevi ise giriş sinyalini çıkış sinyaline dönüştürmektir. YSA'lar doğrusal olmayan yapıya bu fonksiyon ile kavuşmaktadır. Aktivasyon fonksiyonunun olmaması, YSA'nın bir doğrusal regresyon modeline dönüşmesine neden olmaktadır. Çıktı katmanı ise ağın tahmin ettiği değer ile beklenen çıktıyı dikkate alarak bir değerlendirme yapar ve hata belirli bir seviyenin üstünde ise bu bilginin, geriye doğru olmak üzere önce gizli katman(lar)a ve sonra girdi katmanına yayılmasını sağlar.



Şekil 21. Yapay bir nöronun/düğümün yapısı

**Örnek 14:** Bir insan kaynakları departmanı, çalışan veri tabanını kullanarak, aday işgörenleri değerlendirmek istemektedir. Örneğin bu amaç ile üretim birimlerinde deneyimi olan beyaz yakalı lisans mezunu teknik personeller için işe alma/ma kararını otomatize etmek istesin. Bu kararı, adayın başvurduğu pozisyon ile alakalı olmak üzere bilgisi ve iletişim becerileri şeklinde iki kriter kullanarak şekillendirecek olsun. Bu durumda bir YSA, girdi katmanı (iki öznelik olduğu için) iki düğüm, gizli katman üç düğüm ve çıktı katmanı bir düğüm olmak üzere bir YSA ile aşağıdaki gibi görselleştirilebilir ve eğitilebilir.<sup>121</sup>

<sup>121</sup> Söz konusu örnek, bir sınıflandırma problemi olup, ileri beslemeli, geri yayımlı ve delta öğrenme kuralı ile öğrenen bir YSA'dır. Çok katmanlı bir ağ yapısına sahiptir. Girdi katmanı özneliklerden oluşup, çıktı katmanı ise beklenen çıktılarıdır (sınıf etiketleri). Eğitim verisindeki bir vektör (gözlem) girdi katmanı ile ağa giriş yapar, gizli katmanda ağırlıklar da dikkate alınarak işlenir. Çıktı



Şekil 22. Örnek 14'e ait YSA

```
#install package
#install.packages("neuralnet")
#veri seti
bilgi=c(20,10,30,20,80,30)
iletisim=c(90,20,40,50,50,80)
ise_al=c(1,0,0,0,1,1)
df=data.frame(bilgi,iletisim,ise_al)
# load library
#require(neuralnet)

#eğitim
nn=neuralnet(ise_al~bilgi+iletisim,data=df, hidden=3,act.fct = "logistic",
linear.output = FALSE)
plot(nn)

#test veri seti
bilgi=c(30,40,85)
```

katmanına iletilir. Çıktı katmanında, beklenen çıktı ile elde edilen çıktı (tahmin) karşılaştırılır. Tahminin hatası hesaplanır ve hesaplanan hata, belirli formülizasyonlar ile geriye doğru olmak üzere önce gizli katmandaki nöronlara ve daha sonra girdi katmanındaki nöronlara yayılarak nöron ağırlıkları güncellenir. Gizli katman ve her bir gizli katmandaki nöron sayısının tespiti ise genellikle çoklu denemelerden sonra en düşük hata değerini veren kombinasyon şeklinde tespit edilmektedir. Veri seti bu tür ağlarda, eğitim ve test seti olmak üzere ikiye ayrılmaktadır. Ağ, eğitim seti kullanılarak eğitilir. Ağın performansı ise test seti ile kontrol edilir. Bu ağlar aynı zamanda zaman serilerinde de kullanılabilir. Bu durumda girdi ve çıktı katmanları belirli sayıda birer gün dizisidir. Bu tür ağların başarımı gizli katman sayısı, gizli katman(lar)daki nöron sayısı, öğrenme katsayısı, momentum terimi, hata farkındaki değişim ve iterasyon sayısına bağlıdır.

```

iletisim=c(85,50,40)
test=data.frame(bilgi,iletisim)

#tahmin
Predict=compute(nn,test)
Predict$net.result

#[0, 1] aralığındaki tahmin çıktısının threshold=0.5 için {0, 1} olarak
dönüştürülmesi. Bu dönüşüm, aktivasyon fonksiyonu lojistik hiperbolik olduğu için
gereklidir.
prob <- Predict$net.result
pred <- ifelse(prob>0.5, 1, 0)
pred

```

### 3.2.Öz-Örgütleli Haritalar ve Öğrenme Algoritmaları

Pözlbauer (2004)'ün ifadesi ile, veri setleri iki öznelikten oluşmadığı sürece veri setlerindeki gözlemlerin bir kağıt üzerinde görselleştirilmesi imkansızdır. Vektör yansıtma “vector projection”, girdi uzayının boyutlarını kullanarak çıktı uzayını daha az boyuta indirgeme işlemi olarak tanımlanmaktadır. Vektör yansıtma işlemindeki en hassas noktalardan bir tanesi, girdi uzayındaki gözlemlerarası mesafenin çıktı uzayında da korunarak indirgeme işleminin gerçekleştirilmesidir. Bu işlemi gerçekleştiren çeşitli yöntemler<sup>122</sup> bulunmaktadır ve vektör niceme bunlardan bir tanesidir. Bu yöntemde, girdi vektörleri kullanılarak bu vektörleri temsil eden yeni vektörler oluşturulmak sureti ile indirgeme işlemi gerçekleştirilmektedir ve ÖÖH'ler de bir tür vektör niceme yöntemidir. Başka bir ifade ile çok sayıdaki girdi vektörünü kullanarak daha az sayıdaki sinaptik ağırlıkları dönüştürmek sureti ile yeni vektörler oluşturmaktadır (ağırlık vektörleri/sinaptik ağırlıklar).

Örnek 13.'ün aksine ÖÖH'lerin en belirgin özelliği, olayları öğrenmek için herhangi bir danışmana/öğretmene ya da çıktı/etiket bilgisine gerek duymamasıdır. Özellikle, çıktılarının belirlenemediği ya da bilinmediği problemlerde yüksek başarımlı sağladığı için (Öztemel, 2012) kümeleme ve hatta bazen öngörülemede “forecasting”

---

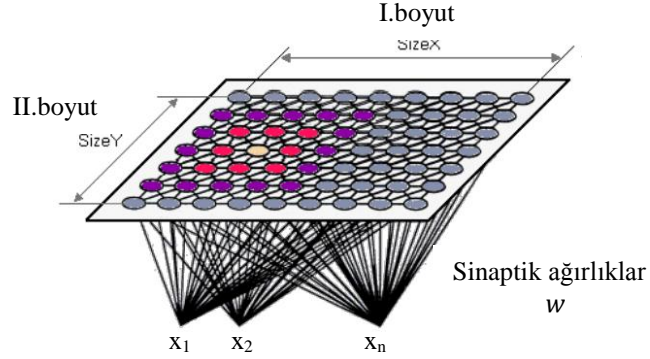
<sup>122</sup> Metrik Çok Boyutlu Ölçekleme “Metric Multidimensional Scaling”, Bağımsız Bileşenler Analizi “Independent Component Analysis”, Temel Bileşenler Analizi, Eğrisel (Eğriçizgisel) Bileşenler Analizi “Curvilinear Component Analysis” t-Dağılımlı Stokastik Komşu Yerleştirme “t-Distributed Stochastic Neighbor Embedding”, ve Sammon'un Doğrusal Olmayan Haritalaması “Sammon's Nonlinear Mapping” (Pözlbauer, 2004; Thrun, 2018)

sıklıkla kullanılmaktadır. ÖÖH'ler yapısal olarak da diğer YSA'lardan çok farklıdır. ÖÖH'lerde gizli katman bulunmamaktadır. ÖÖH'ler sadece girdi ve çıktı katmanından oluşmaktadır ve çıktı katmanı iki boyutlu uzayda bir düzlemi temsil etmektedir (Öztemel, 2012). İşlem elemanları iki boyutlu düzlemde ızgara benzeri yapı üzerindeki uniform dağılan ve birbirleri ile bağlantılı olan nöronlardır (Akpınar, 2018) ve çok katmanlı ağların aksine her bir girdi, çıktı katmanındaki tüm nöronlar ile ilişki halindedir.

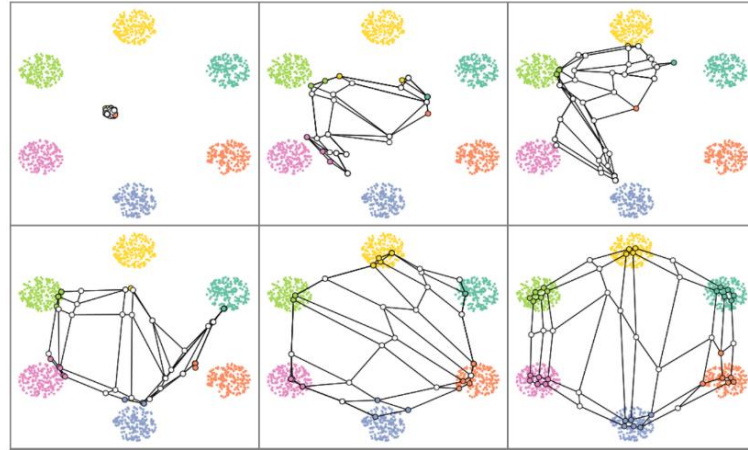
ÖÖH'ler genellikle Şekil 23.'ten de inceleneceği üzere iki boyutlu tasarlanmaktadır. Sıklıkla kullanılan komşuluk alanının geometrik yapısı/topoloji ise kare veya altıgen şeklindedir. İki boyutlu ÖÖH'lere nazaran daha az sıklıkla kullanılıyor olsalar da tek boyutlu ve üç boyutlu ÖÖH'lerde tasarlanabilmektedir. Silindir ve halka "toroid" geometrik yapıları üç boyutlu ÖÖH'lerde sıklıkla kullanılan komşuluk alanlarıdır. Bu topolojilere ek olarak hiyerarşik ÖÖH'ler de tasarlanabilir (Miljkovic, 2017). Hiyerarşik ÖÖH'lerde birden fazla çıktı katmanı vardır ve her bir katmanın çıktısı, kendisinden sonra gelen katmanın girdisidir. Katmanlaşma, en son katmanda tek/dört nöron kalana kadar devam etmektedir (Polanco, François, Lamirel, 2001; Chair, Charrad, 2016). Çok katmanlı ÖÖH'ler uygun küme sayısının tespitinde ve öznelik sayısının gözlem sayısından çok daha fazla olduğu veri setlerinde diğer yöntemlere nazaran çok daha başarılıdır (Lu, Segall, 2013; Khanchouch, Charrad, Limam, 2015).

İki boyutlu bir ÖÖH örneği Şekil 24.'ten de inceleneceği üzere bir  $X$  veri setindeki  $n$  boyutlu (öznelikli)  $x$  girdi vektörlerinden bir tanesi örneğin  $(9 \times 7)$  boyutlu 63 nörondan oluşan bir ağa girer ve eğitim başlar. Öncelikle,  $x$  girdisine en çok benzeyen "kazanan" nöron  $i$  (yavruağzı renkli küme merkezi), daha sonra kazanan nörona belirli bir komşuluk mesafesindeki komşu nöronlar  $j$  (kırmızı ve mor renkli nöronlar) tespit edilir. Son olarak  $i$  ve  $j$  numaralı nöronların  $x$  girdi vektörü arasındaki sinaptik ağırlıkları (çıkı katmanı ağırlıkları) bazı parametreler dikkate alınarak güncellenir ve akabinde ikinci girdi vektörü ağa verilerek eğitim süreci devam eder. Eğitim sürecinin başlangıcında girdi vektörlerine (veri setine) pek benzemeyen (girdilere uzak bir mesafede olan) ağırlık vektörü, ağın eğitimi bittiğinde, veri setini temsil edecek şekilde onlara yakın mesafelerde konumlanmışlardır.

Ağırlık vektörü “weight vectors” alan yazınında prototip “prototypes”, model “models”, kod vektörü “code vectors” ve merkez “centroids” olarak da adlandırılmaktadır (Cottrell, vd., 2016).



Şekil 23. Örnek bir ÖÖH'nin yapısı (Breard, 2017)



Şekil 24 Örnek bir ÖÖH'nin eğitim süreci (Breard, 2017)

ÖÖH'lerin çalışma prensibi rekabetçi öğrenme “competitive learning” olup, yarışmayı kazanma ve kazanan elemanın “best matching unit/neuron” 1, diğerlerinin 0 değerini alma esasına dayanmaktadır. Algoritmanın başlangıcında çıktı katmanındaki nöron sayısı, öğrenme katsayısı  $\eta$  ve komşuluk yarıçapı  $\sigma$  belirlenir; her nöron ve öznitelikler arasındaki sinaptik bağlantılara tesadüfi olarak ağırlık “weights”  $w$  değerleri atanır ve ağ topolojisi tanımlanır. Bir girdi vektörü “input”  $x$  (veri setindeki tek bir gözlem) bir  $t$  anında (iterasyon) ağa verildiğinde algoritma aşağıdaki şekilde ağı eğitir (Haykin, 2009):

1. Girdi vektörü  $x$  ile her bir nörona ait ağırlık vektörleri arasındaki mesafe  $d_{i,x}$  Eş. (55) ile tanımlanmış olan Öklit uzaklığı ile hesaplanır.  $r$  çıktı katmanındaki nöron sayısı

olmak üzere, girdi vektörüne en yakın olan ağırlık vektörünün ait olduğu  $i$ 'inci nöron, yarışmayı kazanan nöron olarak belirlenir. Kazanan nöron 1 ve diğer nöronlar 0 değerini almak üzere, sürekli nitelikli girdi uzayı, kesikli nitelikli çıktı uzayına dönüşmüş olur. Bu aşama rekabetçi süreç “competitive process” olarak adlandırılmaktadır.

$\| \cdot \|$  öklit uzaklığını temsil etmek üzere

$$d_{i,x} = \min_i \|x - w_i\| \quad i = 1, \dots, r \quad (55)$$

2. Kazanan nöron, kendisine  $\sigma$  mesafedeki (yarıçaptaki) komşu nöronların merkezi konumuna gelir ve Eş. (56) ile ifade edilen  $h_{j,i}$  komşuluk fonksiyonunun tanımlanmasında kullanılır.

$$h_{j,i}(t) = \exp\left(\frac{-\|d_i - d_j\|^2}{2\sigma^2(t)}\right) \quad (56)$$

$d_{j,i}$ , kazanan nöron  $i$  ve ona komşu olan bir  $j$  nöronu olmak üzere, bir  $t$  anında  $i$  ve  $j$  arasındaki mesafeyi temsil etmektedir. Bu mesafe sıfıra yaklaştıkça,  $\|d_i - d_j\| \rightarrow 0$ , komşuluk fonksiyonunun değeri maksimuma yaklaşmalıdır,  $h_{j,i}(t) = e^0 = 1$ . Böylelikle, kazanan nöronun etrafındaki nöronlar için komşuluk fonksiyonunun değeri maksimize edilmiş olur. Diğer taraftan, mesafe sonsuza yaklaştıkça,  $\|d_i - d_j\| \rightarrow \infty$ , komşuluk fonksiyonunun değeri sıfıra yaklaşmalıdır/sönümlenmelidir ki yakınsama gerçekleşsin,  $h_{j,i}(t) = e^{-\infty} = 0$ . Yazınında, her iki koşulu da sağlayabilen çeşitli fonksiyonlar bulunmaktadır Meksika şapkası “Mexican hat”, baloncuk “bubble” ve Gauss fonksiyonu gibi (Breard, 2017; Kohonen, 2001). Sayılan fonksiyonlardan ve Eş. (56) ile ifade edilen Gauss fonksiyonu, en sık kullanılan fonksiyondur.<sup>123</sup> Gauss fonksiyonu, haritadaki nöronların sinaptik ağırlıklarının girdi vektörüne yaklaşma miktarını kazanan nöron merkezde olmak üzere, merkezden uzaklık ile ters orantılı olacak şekilde arttırmaktadır. Başka bir ifade ile fonksiyonun çan eğrisi şeklindeki yapısından

---

<sup>123</sup> Meksika şapkası fonksiyonunda fonksiyonun ortası kazanan nöron ve komşularının sinaptik ağırlıklarını girdi vektörüne yaklaştırırken, geri kalan nöronların ağırlıklarını uzaklaştırmaktadır. Başka bir ifade ile onları cezalandırmaktadır (Kohonen, 2001). Baloncuk fonksiyonunda ise fonksiyon sadece kazanan nöron ve kazanan nöron merkezde olmak üzere onun etrafını saran birinci dereceden komşularının ağırlığını güncellemektedir (Breard, 2017).

ötürü, en fazla artış merkez nöronda (kazanan nöronda) gerçekleştirilirken, merkezden uzaklaştıkça (komşuluğun sınırına yaklaştıkça) artış miktarı azalmaktadır.

Ayrıca, komşuluk mesafesi  $\sigma$ 'nın da zamanla azalması gerekmektedir. Bunun için  $\tau_1$ , karar verici tarafından atanan zaman sabiti olmak üzere, Eş. (57)'da yer alan üssel azalma "exponential decay" yaklaşımı sıklıkla kullanılmaktadır. Chabane, Islam, Zerr (2017) çalışmasında ise  $r$  çıktı katmanındaki nöron sayısı,  $\sigma_{max} = r/2$  ve  $\sigma_{min} = 1/2$  olmak üzere bir  $t$  anındaki komşuluk mesafesinin Eş. (58) kullanılarak güncellenmesi önerilmektedir.

$$\sigma(t) = \sigma_0 \exp \frac{t}{\tau_1} \quad (57)$$

$$\sigma(t) = \sigma_{max} \left[ \frac{\sigma_{max}}{\sigma_{min}} \right]^{\frac{t}{T}} \quad (58)$$

Bu aşamanın tamamı işbirlikçi aşama "cooperative process" olarak adlandırılmaktadır.

**3.** Bu aşamada ise güncellenen komşuluk mesafesi ve komşuluk fonksiyonu ile kazanan nörona komşu olan nöronlar tespit edilir ve sadece bu nöronların ağırlıkları güncellenir.

Danışmanlı öğrenme algoritmalarında, bu bölümde Kısım 3.1'de kısaca açıklandığı üzere, beklenen çıktı referans alınarak ağırlık eğitimi (ağırlıkların güncellenmesi) bütünsel olarak gerçekleştirilmektedir. Fakat danışmansız öğrenmede bu şekilde bir ön bilgi bulunmadığı için mevcut veri setindeki ilişki ağı kullanılarak öğrenme işlemi gerçekleştirilmektedir. Örneğin TBA'da da bir tür ilişki analizi yapılarak birbirleri ile benzeşen gözlemler dikkate alınıp faktöriyel yapı ortaya çıkarılmaktadır. Bu tür öğrenme yaklaşımlarında Hebb kuralı esas alınır. Hebb kuralına göre, eğer bir sinaptik bağlantının iki ucunda yer alan iki nöron aynı anda uyarılıyor ise söz konusu sinaptik bağlantı güçlü bir bağlantı olmalıdır (ağırlık güncellenmelidir). Fakat söz konusu iki nöron eş anlı olarak uyarılmıyor ise (farklı durum/şartlarda uyarılıyor ise) nöronlar arasındaki sinaptik bağlantı zayıf olmalı veya hatta mümkünse iptal edilmelidir (Haykin, 2009). Eş. (59)'ten de inceleneceği üzere, Hebb kuralının ilk kısmına göre bir  $t$  anındaki sinaptik ağırlığın yeni değeri, söz konusu bağlantının eski ağırlığı, söz konusu nöronların (presinaptik  $x_j$  ve postsinaptik  $y_k$  nöronların) değerleri ve öğrenme oranına bağlıdır.



$$\Delta w_{kj}(t) = \eta y_k(t) x_j(t) \quad (59)$$

Yukarıda yer alan Eş. (59)'a göre,  $j$ 'inci presinaptik nöronun her aktivasyonunda veya başka bir ifade ile her girdi sinyalinde,  $k$ 'inci postsinaptik nöron uyarılarak bir çıktı değeri oluşacak ve  $\Delta w_{kj}(t)$  değeri üssel bir şekilde arttıracaktır. Bu durum, rekabetçi öğrenme için bazı dezavantajlar meydana getirmektedir (Hagan, vd., 1996). İlki, nöronlar arası ağırlıkların sınırsız bir şekilde sürekli artarak belirli bir süre sonra bir doyum noktasına ulaşması ve dolayısı ile ağa yeni sinyal (bilgi) geldiğinde söz konusu bağlantının bu bilgiyi öğrenememesi. İkinci dezavantaj, tüm ağırlıkların yüksek değerlere sahip olması dolayısı ile ağın seçicilik özelliğinin kaybolması (güçlü olan sinaptik bağlantının tespit edilememesi) (Haykin, 2009). Üçüncü bir diğer dezavantaj ise ağırlıkların büyümesini engelleyecek bir mekanizma olmadığı için, veri seti gürültü içeriyorsa ağın bunu fark edemeyerek düşük oranda bir hızla da olsa sinaptik ağırlıkların büyümeye devam etmesidir (Hagan, vd., 1996). Sayılan bu dezavantajlar dolayısı ile Hebb kuralının ikinci kısmı (zayıf bağlantıların ağırlıklarının azaltılması veya ortadan kaldırılması) sağlanamamaktadır. Bu yüzden Eş. (59) ile tanımlanan Hebb kuralına bir sönümleme veya unutmama terimi “decay/forgetting term”  $g(y_j)w_j$  eklenmiştir. Böylelikle Hebb kuralı,  $w_j$ ,  $j$ 'inci nöronun ağırlık vektörü ve  $g(y_j)$ , pozitif bir fonksiyon ve  $y_j = 0$  için  $g(y_j)=0$  olması durumunda Eş. (60) ile ifade edilmektedir. Sönümleme teriminin değeri 1'e yaklaştıkça ağ öğrendiği eski bilgileri unutmakta ve sadece yakın geçmişte öğrendiği bilgileri hafızada saklamaktadır. Böylelikle, Hebb teriminin limitsiz bir şekilde büyümesi engellenmektedir (Hagan, vd., 1996).

$$\Delta w_j = \overbrace{\eta y_j x}^{\text{Hebb terimi}} - \overbrace{g(y_j)w_j}^{\text{Sönümleme terimi}} \quad (60)$$

$y_j = 0$  için  $g(y_j) = 0$  şartı, Eş. (61)'de yer alan doğrusal fonksiyon ile sağlanmaktadır.

$$g(y_j) = \eta y_j \quad (61)$$

$x$  girdi vektörü olmak üzere, kazanan nöron  $i(x)$  için ise Eş. (62) tanımlanmaktadır.

$$y_j = h_{j,i(x)} \quad (62)$$

Dolayısı ile Eş. (61, 62), Eş. (60)'ta yerine yazıldığında, Hebb kuralı Eş. (63)'teki şeklini alır.

$$\Delta w_j = \overset{\text{Hebb terimi}}{\widehat{\eta y_j x}} - \overset{\text{Sönümlenme terimi}}{\widehat{\eta y_j w_j}} = \eta y_j (x - w_j) = \eta h_{j,i(x)} (x - w_j)$$

$$\Delta w_j = \eta h_{j,i(x)} (x - w_j) \begin{cases} i: \text{kazanan nöron} \\ j: \text{aktive edilen komşu nöron} \end{cases} \quad (63)$$

Dolayısı ile  $(t + 1)$  anındaki ağırlıkların güncellenmesi Eş. (64) ile gerçekleştirilir. Bu eşitlik, kazanan nöron  $i$ 'ye ait  $w$  ağırlık vektörünün, girdi vektörü  $x$ 'e yaklaşmasını sağlamaktadır.

$$w_i(t + 1) = w_i(t) + \eta(t) h_{j,i(x)}(t) (x(t) - w_i(t)) \quad (64)$$

Böylelikle eğitim sürecinde girdiler ağa verildikçe, ağırlık vektörünün dağılımı, girdi vektörünün dağılımına benzeyecek ve komşu nöronların sinaptik ağırlıklarının birbirine benzer olması dolayısı ile de ağ topolojik olarak kendisini güncelleyecektir. Fakat her iterasyonda, öğrenme oranı  $\eta$ 'nın da güncellenmesi gereklidir.  $\tau_2$  karar verici tarafından atanan bir diğer zaman sabiti olmak üzere  $t$  zamanındaki öğrenme oranı Eş. (65) ile güncellenir. Chabane, Islam, Zerr (2017) çalışmasında ise  $T$  toplam iterasyon sayısı olmak üzere, öğrenme oranının Eş. (66) kullanılarak güncellenmesi önerilmektedir.

$$\eta(t) = \eta_0 \exp\left(-\frac{t}{\tau_2}\right) \quad (65)$$

$$\eta(t) = \eta_0 \left(1 - \frac{t}{T}\right) \quad (66)$$

Üç numaralı aşamanın tamamı uyumlama/yakınsama süreci “adaptive/convergence process” olarak adlandırılmaktadır. Algoritmanın devamında ise bir sonraki gözlem (girdi vektörü) işlenmek üzere ağa verilir. Bu süreç haritada neredeyse hiçbir değişiklik olmayana kadar iteratif bir şekilde devam eder.

**Örnek 15:** Dört öznitelik ve dört gözlemden oluşan bir veri seti  $x$ , ÖÖH kullanılarak kümelenecektir. Kümeleme için oluşturulacak olan ÖÖH iki nörona sahip

olsun; sadece iki nörona sahip olduğu için komşuluk mesafesi  $\approx 0$  (yaklaşık sıfır) olsun;<sup>124</sup> girdi katmanı ile çıktı katmanı nöronları arasındaki ağırlıklar rassal olarak  $w_{ij}$  şeklinde belirlensin; öğrenme katsayısı 0.6 ve iterasyon sayısı 100 olarak kabul edilsin. Bu durumda birinci iterasyona ait hesaplamalar ve kümeleme probleminin R-Studio kullanılarak final çözümü aşağıdaki gibi olur:

$$x = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad w = \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \\ 0.5 & 0.7 \\ 0.9 & 0.3 \end{bmatrix}$$

Birinci gözlemin ağa girmesi ile birinci gözlem ve birinci ve ikinci nöronlar arasındaki mesafe Eş. (55) ile hesaplanır.

$\| \cdot \|$  öklit uzaklığını,  $x = 1$  birinci satır vektörünü (birinci gözlemi) ve  $i = 1$  ve 2 nöron numaralarını temsil etmek üzere

$$d_{i,x} = \min_i \|x - w_i\| \quad \text{için} \quad d_{1,1} = \|x - w_1\| = \sqrt{((0.2 - 1)^2 + (0.6 - 1)^2 + (0.5 - 0)^2 + (0.9 - 0)^2)} = 1.364$$

$$d_{2,1} = \|x - w_2\| = \sqrt{((0.8 - 1)^2 + (0.4 - 1)^2 + (0.7 - 0)^2 + (0.3 - 0)^2)} = 0.990$$

$d_{1,1} = \min_i \|x - w_i\| = \min_i \{1.364, 0.990\} = 0.990$  olduğu için ikinci nöron kazanan nöron  $i = 2$  olarak tespit edilir.

Akabinde komşuluk fonksiyonunun değeri Eş. (56) ile hesaplanır. Fakat bu örnekte komşuluk yarıçapı 0 olarak belirlendiği için kazanan nöronun kendisi dışında bir komşusu bulunmamaktadır. Bu yüzden komşuluk fonksiyonu aşağıdaki gibi değerlendirilir.

$$h_{j,i}(t) = \exp\left(\frac{-\|d_i - d_j\|^2}{2\sigma^2(t)}\right) = h_{j,2}(0) = \exp\left(\frac{-\|d_2 - d_j\|^2}{2\sigma^2(0)}\right) = \exp\left(\frac{-\text{sayı}}{0}\right) = e^0 = 1$$

Sadece kazanan nöron olan ikinci nöronun ağırlıkları,  $t = 0$  için, Eş. (64) kullanılarak güncellenir. Eğer kazanan nöronun komşu nöronları bulunsaydı aşağıda yer alan işlemler, komşu nöronlar için de tekrarlanacaktı.

---

<sup>124</sup> Çıktı katmanında sadece iki nöron bulunduğu için ve her bir nöronun birer kümeyi temsil etmesi hedeflendiği için, komşuluk mesafesi (yarıçap) sıfır olarak atanmıştır. Böylelikle Eş. ( )'deki  $h_{j,i(x)}(t)$  çarpanının değeri 1 olacaktır.

$$w_j(t + 1) = w_j(t) + \eta(t)h_{j,i(x)}(t) (x(t) - w_j(t))$$

$$w_2(1) = w_2(0) + \eta(0)h_{j,2(x)}(0) (x(0) - w_j(0)) = 0.8 + 0.6 * 1(1 - 0.8) =$$

0.92 olarak ikinci nöronun birinci vektörün birinci öznitelik değeri için güncellenmiş ağırlığı tespit edilir ve  $t = 1$  için güncellenmiş tüm ağırlık vektörü aşağıdaki gibi hesaplanır.

$$w(1) = \begin{bmatrix} 0.2 & 0.92 \\ 0.6 & 0.76 \\ 0.5 & 0.28 \\ 0.9 & 0.12 \end{bmatrix} \begin{cases} = 0.8 + 0.6 * 1(1 - 0.8) \\ = 0.4 + 0.6 * 1(1 - 0.4) \\ = 0.7 + 0.6 * 1(0 - 0.7) \\ = 0.3 + 0.6 * 1(0 - 0.3) \end{cases}$$

Hatırlanacağı üzere ilk girdi vektörü  $[1,1,0,0]$  ve kazanan nöronun ağırlık vektörü  $[0.8,0.4,0.7,0.3]$  şeklinde idi. Rekabetçi öğrenmeye göre her iterasyonda, kazanan nöronların ağırlık vektörünün girdi vektörüne yaklaşması (ona benzemesi) gerekmektedir. Güncellenmiş ağırlık vektöründen de inceleneceği üzere, girdi vektörünün ilk iki elemanının değeri 1 olduğu için, ağırlık vektörünün ilk iki elemanının değeri artarak 1'e yaklaşmıştır. Diğer taraftan girdi vektörünün son iki elemanının değeri 0 olduğu için ise ağırlık vektörünün son iki elemanının değeri azalarak 0'a yaklaşmıştır.

İkinci nöronun ağırlık güncellemesinden sonra öğrenme oranı Eş. (65) ile güncellenir ve ağa ikinci gözlem vektörü  $[0,0,0,1]$  verilerek eğitime devam edilir.

```
require(kohonen)
b=matrix(rbind(c(1,1,0,0),c(0,0,0,1), c(1,0,0,0), c(0,0,1,1)), nrow = 4, ncol = 4)
rownames(b) <- c("x1", "x2", "x3", "x4")
colnames(b) <- c("o1", "o2", "o3", "o4")
b
data_train <- b

set.seed(100)
som_model <- som(data_train, xdim=1, ydim=2, topo="hexagonal", rlen=1000, alpha =
c(0.6,0.01), radius = 0.0001, neigh = "gaussian" )
summary(som_model)
plot(som_model, type="codes")
colors <- c( "green", "red")
plot(som_model, type = "mapping", col = colors[som_model$unit.classif],pch =
obj.classes, main = "b")
#nesnelerin merkeze olan ortalama mesafelerine göre "quality" grafiği, bu değer
küçüldükçe kümelemenin kalitesi artmaktadır
```

```

plot(som_model, type = "quality")
rapor=cbind(som_model$unit.classif, som_model$distances)
rapor

```

Yukarıda yer alan kod çalıştırıldığında elde edilen çıktılar aşağıdaki gibidir, Şekil 25. Eğitim raporundan da inceleneceği üzere birinci ve üçüncü gözlemler ikinci kümede, ikinci ve dördüncü gözlemler ise birinci kümede yer almıştır. Birinci kümede yer alan gözlemler için birinci ve ikinci öznitelikler (o1, o2) dengeli bir şekilde (birisi diğerinden daha baskın olmayacak şekilde) ayırt edici özellik iken, ikinci kümede yer alan gözlemler için ise üçüncü ve dördüncü öznitelikler (o3, o4) dengeli bir şekilde ayırt edici özellik olarak ön plana çıkmıştır. Ayrıca “quality plot” çıktısından da inceleneceği üzere, aslında aralarında çok fazla fark olmasa da birinci kümedeki gözlemler arası mesafe ikinci kümeye nazaran daha azdır.

```

> summary(som_model)
SOM of size 1x2 with a hexagonal topology and a bubble neighbourhood function.
The number of data layers is 1.
Distance measure(s) used: sumofsquares.
Training data included: 4 objects.
Mean distance to the closest unit in the map: 0.263.

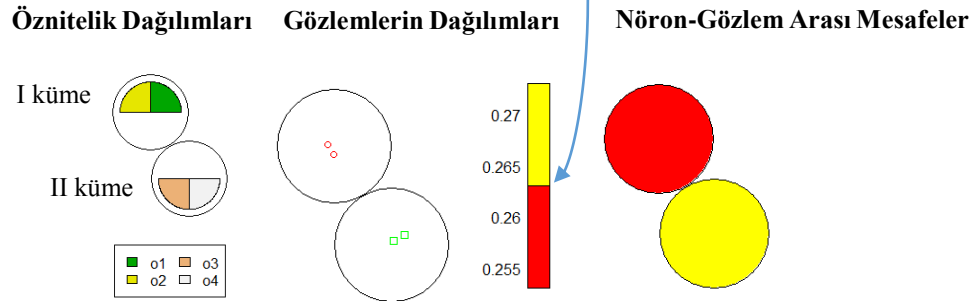
```

```
> rapor
```

```

      [,1] [,2]
[1,]    2 0.3089152
[2,]    1 0.1214125
[3,]    2 0.1973124
[4,]    1 0.4245265

```



Şekil 25. Örnek 14’ün ÖÖH ile eğitiminin R Studio çıktısı

Yukarıda açıklanan algoritma, girdi katmanı ile çıktı katmanı arasındaki ağırlık vektörünü her iterasyon sonrası güncellemektedir. Bu yaklaşım çevrimiçi öğrenme “online learning” olarak adlandırılmaktadır. Çevrimiçi öğrenmenin en büyük dezavantajı, veri setindeki gözlemlerin diziliminden kaynaklanmaktadır. Açıklamak gerekir ise ÖÖH’ler geçici hafıza ile çalıştığı için, ağa ilk giren gözlemlere ait bilgiler iterasyon sayısı ilerledikçe unutulmakta ve ağ güncel ve yakın geçmiş iterasyonlardan elde ettiği bilgiler ile öğrenmektedir. Bu yüzden örneğin sıradışı verilerin çoğunluğunun,  $p > k$

olmak üzere,  $k$  tanesinin en son  $p$  iterasyona kaldığı varsayılır ise  $T - p$  inci iterasyona kadar öğrenme başarımı yüksek bir ağ elde edilmişken, son  $p$  adet iterasyonda ağın yapısı bozularak başarımı düşecektir. Çevrimiçi öğrenmenin eleştirilen bir diğer yönü de öğrenme oranının, ağın öğrenme hızı ile kararlı bir yapıya sahip olması arasında bir ödünleşim yapılmasını gerektirmesidir. Çünkü yüksek öğrenme oranı öğrenme hızını arttırırken, ağın kararlılıktan uzaklaşmasına neden olabilmektedir. Sayılan bu sebeplerden ötürü, çevrimiçi öğrenmeye alternatif olarak yığın öğrenme “batch learning” algoritması geliştirilmiştir. Çevrimiçi öğrenmeden en önemli farkı, ağırlık vektörlerinin ve ağ topolojisinin, ağa bütün girdiler verildikten sonra güncellenmesidir, Şekil 26.

1. Birinci gözleme ait ilk girdi vektörü ağa girer.

2. İlk girdi vektörü  $x_j$  ile her bir nörona ait ağırlık vektörleri arasındaki mesafe  $d_j$  Eş. (67) ile tanımlanmış olan Öklit uzaklığı ile hesaplanır.  $r$  çıktı katmanındaki nöron sayısı olmak üzere, girdi vektörüne en yakın olan ağırlık vektörünün ait olduğu  $i$ 'inci nöron, yarışmayı kazanan nöron olarak belirlenir ve bu bilgi kayıtlanır.

$$d_j = \min_i \|x_j - w_i\| \quad i = 1, \dots, r \quad \text{ve} \quad j = 1, \dots, m \quad (67)$$

3. Veri setindeki tüm gözlemler, ağda hiçbir güncelleme yapılmadan ağa girer ve tüm gözlemler için ikinci aşama tekrarlanır.

4. Tüm veri seti ağda işlem gördükten sonra sinaptik ağırlıklar ve komşuluk fonksiyonu güncellenir. Eş. (68)'e göre  $i$ 'inci nöronun kazanan nöron olduğu tüm  $j$  girdileri için hesaplanan komşuluk fonksiyonu değerlerinin söz konusu girdi vektörleri ile çarpımlarının toplamının, tüm  $i \& j$  çiftleri için hesaplanan komşuluk fonksiyonu değerlerinin toplamları kullanılarak elde edilen ortalama değer,  $i$  nöronunun yeni sinaptik ağırlığıdır.

Sinaptik ağırlıkların güncellenmesinde herhangi bir  $t$  zamanı (iterasyonu) olmadığı için  $w_j(t) = 0$ 'dır. Ayrıca Eş. (68)'den inceleneceği üzere, pay ve paydadaki öğrenme oranı birbiri ile sadeleşmekte ve dolayısı ile yığın öğrenme algoritması ile eğitilen ÖÖH'lerde öğrenme oranına gerek duyulmamaktadır (Kohonen, 2001). Bu özelliği ile ÖÖH'lerin çevrim içi öğrenme algoritmasına kıyas ile daha hızlı bir şekilde eğitildiği ifade edilmektedir (Matsushita, Nishio, 2010; Kohonen, 2001; Hagan, vd., 1996).

$$\begin{aligned}
w_i^{yeni} &= \frac{\sum_{j=1}^m 0 + \eta h_{ji}(x_j - 0)}{\sum_{j=1}^m \eta h_{ji}} \\
&= \frac{\sum_{j=1}^m h_{ji} \bar{x}_j}{\sum_{j=1}^m h_{j,i(x)}}
\end{aligned} \tag{68}$$

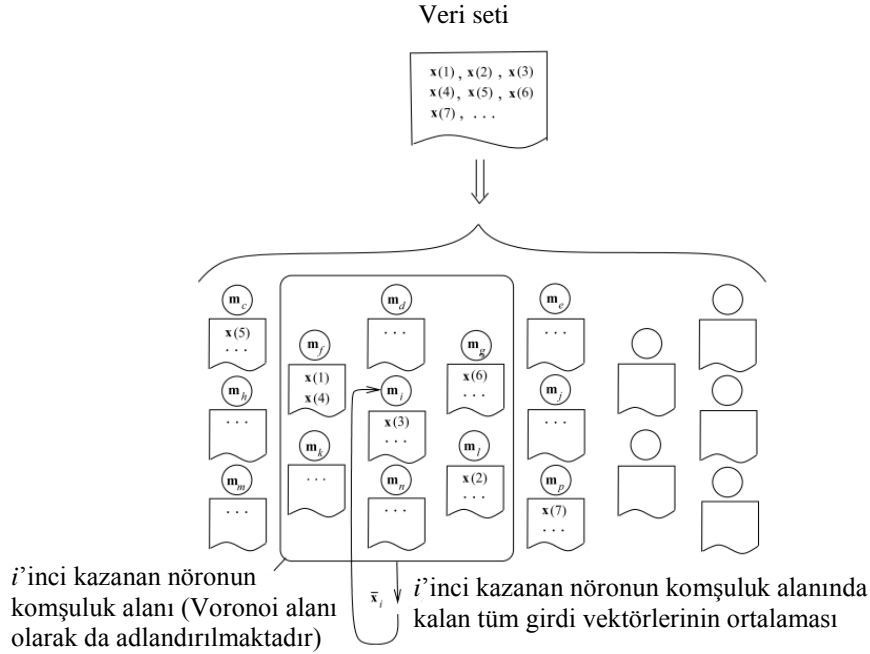
5. Veri setindeki tüm gözlemler ağı verildikten ve ağı güncellendikten sonra tüm algoritmanın birkaç sefer tekrarlanması ile ağı eğitimi tamamlanmış olur. Başka bir ifade ile çevrim içi öğrenmede ağı eğitimi için birkaç bin veya on binlerce iterasyon gerekirken, yığın öğrenmede en fazla 20-30 veya hatta 3-4 iterasyon ile eğitim tamamlanabilmektedir. Algoritma, eğitimin son birkaç iterasyonunda Forgy veya başka bir ifade ile k-ortalama algoritmasına benzer hareket etmektedir. Çünkü üyelik fonksiyonu son iterasyonlara doğru nerede ise ağıdaki her bir nöronu kazanan nöron yapacak şekilde komşuluk alanını daraltmaktadır (Kohonen, 2001; Cottrell, vd., 2016). Örneğin Hagan, vd. (1996) çalışmasında sunulan uygulama örneğinde  $\sigma_0 = 4$  olarak atanmış ve ilk iterasyon sonunda (ikinci iterasyonda kullanılmak üzere)  $\sigma_1 = 1$  olarak ağı tarafından güncellenmiştir. Fakat diğer taraftan Fort, Letremy, Cottrell (2002) çalışmasında çevrim içi ve yığın öğrenme algoritmalarının performansları tek boyutlu ÖÖH'ler için hem gerçek hem de gürültü içeren sentetik veri setleri kullanılarak karşılaştırılmıştır. Çalışmanın bulgularına göre haritanın topografik olarak düzenlenmesinde (nöronların dizilimi)<sup>125</sup> yığın öğrenme algoritmasının performansı çevrimiçi öğrenme algoritmasının performansından daha kötüdür. Yazarlara göre, çevrimiçi öğrenme algoritması yığın öğrenme algoritmasına kıyas ile (ağırlık vektörü, öğrenme oranı ya da komşuluk yarı çapı gibi) ÖÖH başlangıç parametrelerine daha az hassastır.

Yığın öğrenme algoritmasının eleştirilen yönü ise algoritmanın başarımının nöronların başlangıç ağırlıklarına bağlı olmasıdır. Bu dezavantajın giderilmesi için ise (özniteliklerin sadece nümerik nitelikte olması durumunda başka bir ifade ile Öklit mesafesinin hesaplanabilir olduğu durumlarda) doğrusal ağırlıklandırma “linear initialization” yaklaşımının kullanılması önerilmektedir. Diğer tüm durumlarda ağırlık

---

<sup>125</sup> Söz konusu çalışmanın amacı, girdileri sıralamak sureti ile kümelemektir. Örneğin gerçek veri setindeki 50 gözlem 6 kümeye ayrılmıştır.

vektörünün başlangıç değerinin rassal olarak atanması önerilmektedir (Kohonen, 2013, 2001).



**Şekil 26.**  $m_i$  kazanan nöronun sinaptik ağırlığı olmak üzere yığın algoritmasının bir iterasyonu (Kohonen, 2013)

Cottrell, vd. (2016)'ya göre ise yığın öğrenme algoritması, çevrimiçi öğrenme algoritmasına kıyas ile deterministik bir algoritmadır. Çünkü yığın algoritması, algoritma her çalıştırıldığında, aynı başlangıç parametreleri için aynı sonuçları sağlamaktadır ve bu yüzden sıklıkla tercih edilmektedir. Fakat diğer taraftan bu algoritma bazen girdi uzayında birbirinden uzak konumda olan gözlemleri aynı veya komşu nöronlara atayabilmektedir. Çevrimiçi öğrenme algoritması ise aynı başlangıç parametreleri için farklı sonuçlar sağlamaktadır. Bu yüzden stokastik bir algoritma olarak değerlendirilir. Çevrimiçi algoritmanın bu dezavantajı, eğitimin birden fazla defa (bootstrap ile) tekrarlanıp, elde edilen sonuçların karşılaştırılması ile elimine edilebilmektedir. Bu amaç ile yazarlar kararlılık endeksinin “stability index” kullanılmasını önermektedir.

$K$ , haritadaki toplam nöron sayısı,  $L$ , çevrim içi öğrenme algoritmasının tekrarlanma sayısı<sup>126</sup>  $Y_{i,j}$ ,  $x_i$  ve  $x_j$  gözlem çiftlerinin aynı kümede gözlemlenme sayısı

<sup>126</sup> Bu değer iterasyon sayısı değildir. Örneğin iterasyon sayısına 1000 olarak karar verildi ise algoritmanın 1000 iterasyonu kaç defa tekrarlanması gerektiğidir.



olmak üzere  $i$  ve  $j$  gözlem çiftleri için kararlılık endeksi  $M_{i,j}$  %95 anlamlılık seviyesinde Eş. (69) ile hesaplanmakta ve aşağıdaki gibi yorumlanmaktadır.

$$M_{i,j} = \frac{Y_{i,j}}{L} \quad (69)$$

$$A = \frac{9}{K} \text{ ve } B = 1.96 \sqrt{\frac{9}{KL} \left(1 - \frac{9}{K}\right)} \quad (70)$$

- $M_{i,j} > A + B$  ise  $i$  ve  $j$  gözlemlerinin birbirleri ile ilişkisi (birlikte görülme/aynı kümede olma sıklıkları) anlamlıdır.
- $A - B \leq M_{i,j} \leq A + B$  ise  $i$  ve  $j$  gözlemlerinin birbirleri ile ilişkisi (birlikte görülme/aynı kümede olma sıklıkları) rastlantısaldir ve söz konusu gözlem çiftleri değişken/kararsız çift olarak adlandırılmaktadır “fickle pair”.
- $M_{i,j} < A - B$  ise  $i$  ve  $j$  gözlemlerinin birbirleri ile ilişkisizliği (birlikte görülmemeye/aynı kümede olmama sıklıkları) anlamlıdır.

Yığın öğrenme algoritmasında, veri setindeki tüm gözlemler ağa tanıtıldıktan sonra ağ güncellendiği için çevrim dışı öğrenmektedir. Başka bir ifade ile bu algorithmada ağa yeni veri girişi mümkün değildir. Diğer taraftan çevrim içi öğrenme algoritmasında, ağa giren her bir gözlemden sonra ağ güncellenmektedir ve dolayısı ile bu algorithmada ağa yeni veri girişi mümkündür (Miljkovic, 2017).

### 3.3.Öz-Örgütlemeli Haritaların Başarımı

Su, Liu, Chang (2002) çalışmasında da belirtildiği üzere, ÖÖH'lerin başarımında nöron sayısı, iterasyon sayısı, haritanın yapısı, öğrenme oranı, komşuluk fonksiyonu ve başlangıç ağırlık vektörleri değerlerinin tespiti önemli bir role sahiptir. Çünkü söz konusu parametreler kümeleme sonuçlarını etkilemektedir ve ÖÖH'lerde danışmanlı öğrenme yöntemlerindeki gibi genel kabul görmüş bir performans başarı kriteri bulunmamaktadır. Bu yüzden haritanın mimarisi değişik yapılarla oluşturularak çeşitli ağ yapıları denenmelidir.

Örneğin nöron sayısı, ağın genelleme yeteneğini doğrudan etkilemektedir. Verinin boyutu büyüdükçe haritada yer alması gereken nöron sayısı da artmalıdır. Nöron sayısının artması ile elde edilecek çıktıların hassaslaşması (detay bilgilerin elde edilmesi) sağlanırken eğitim süresi uzamaktadır. Nöron sayısının azalması ile ise ağ daha genel

bilgiler sağlamakta ve eğitim süresi azalmaktadır (Gabrielsson, Gabrielsson, 2006). Nöron sayısı için ilgili yazınında standart bir kural bulunmamasına rağmen kabul görmüş bazı uygulamalar bulunmaktadır. Örneğin, gözlem sayısından küçük (Beccali, vd., 2004), gözlem sayısının kareköküne yakın (Hagan, vd., 1996), gözlem sayısının karekökünün 5 katına yakın (Tian, Azarian, Pecht, 2014; Chicco, vd., 2004) ve gözlem sayısının karekökünün 5 katı ile 20 katı arasında (Panapakidis, vd., 2014) olabileceği belirtilmektedir. Kohonen (2013)'e göre ÖÖH'ler, veri setindeki gözlemleri çıktı katmanındaki nöronlar ile eşleştirdiği ve dolayısı ile nöronları kullanarak girdileri bölümlediği için bir nevi histogram görevi görmektedir. Bu yüzden özellikle istatistiksel analizler ile birlikte kullanılacak olan ÖÖH'lerin yapılandırılmasında her bir nörona 50 civarında gözlem düşecek şekilde nöron sayısı tespit edilmelidir. Ciampi, Lechevallier (2000)'e göre ise  $a$  iki boyutlu haritanın enini oluşturan nöron adedi,  $b$  boyunu oluşturan nöron adedi ve  $r$  haritadaki toplam nöron adedi olmak üzere olmak üzere  $a$  ve  $b$  veri seti  $X$ 'in birinci  $\lambda_1$  ve ikinci  $\lambda_2$  en büyük eigen değerlerinin karekökleri ile orantılı olmalıdır. Bu yaklaşım, veri setindeki gözlemlerin nümerik öznitelikler olması durumunda uygulanabilir.

$$r = ab \quad (71)$$

$$\frac{a}{b} = \sqrt{\frac{\lambda_1}{\lambda_2}} \quad (72)$$

Sonuç olarak, alan yazınında çeşitli harita yapılarının küçükten büyüğe olacak şekilde denenmesi önerilmektedir. Bu yaklaşım verinin genel yapısının daha rahat gözlemlenmesini sağlamakta olup, tam tersi yaklaşıma nazaran daha etkili (cost-effective) olarak değerlendirilmektedir (Wehrens, Kruisselbrink, 2018a).

İterasyon sayısının çevrim içi öğrenen ağlarda çıktı katmanındaki nöron sayısının en az 500 katı kadar olması gerektiği (Su, Liu, Chang, 2002; Kohonen, 2001), yığın öğrenme algoritması ile eğitilen ağlarda ise birkaç iterasyonun yeterli olduğu ifade edilmektedir (Kohonen, 2001; Cottrell, vd., 2016; Hagan, vd., 1996).

ÖÖH'lerin çıktı katmanındaki nöron komşuluk alanlarının geometrik yapısı (haritanın topolojisi) bu bölümde kısım 3.2'de de bahsedildiği üzere genellikle kare ve altıgen olarak tasarlanmaktadır. Kohonen, bu yapılardan görsel olarak daha kesin

sonuçlar sağlayan altıgen dizilimin kullanılmasını önermektedir. Tüm haritanın genel şeklinin ise kareden ziyade dikdörtgensel (oblong/boyu eninden uzun) bir şekilde tasarlanmasını önermiştir. Çünkü dikdörtgen şekle sahip olan haritalar daha hızlı ve güvenilir bir şekilde yakınsamaktadır/öğrenmektedir (Kohonen, 2013).

Öğrenme oranının başlangıç değerinin  $\eta_0 = 0.1$ 'e olabildiğince yakın olması ve her iterasyonda azalması fakat 0.01'den daha küçük bir değere sahip olmaması ve bu doğrultuda Eş. (65)'te yer alan  $\tau_2 = 1000$  olması önerilmektedir (Haykin, 2009). Öğrenme oranının (algoritmanın başlangıcında) 1 gibi çok yüksek bir değer olması durumunda ise ağırlık eğitimi hızlanmakta fakat ağırlık vektörü bir kümeyi temsil ettikten sonraki iterasyonda (öğrenme oranı çok büyük olduğu için) değişerek, başka bir kümenin merkezi olabilir. Bu durum haritanın kararlı/stabil bir yapıya kavuşmasına engeldir. Diğer taraftan öğrenme oranının (algoritmanın başlangıcında) 0'a çok yakın bir değer olması durumunda ağırlık vektörü, bir kümeyi temsil ettikten sonraki iterasyonlarda (öğrenme oranı çok küçük olduğu için) değişmeyebilir, başka bir kümenin merkezi olmayabilir ve harita kararlı bir yapıya kavuşur. Fakat ağırlık eğitimi yavaşlar.

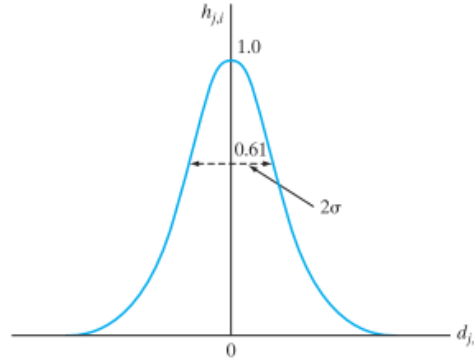
Komşuluk mesafesinin ise, algoritmanın başlangıcında, çıktı katmanındaki neredeyse tüm nöronları kapsayacak genişlikte büyük olması ve her iterasyonda zamanla azalması beklenmektedir. Örneğin 1000 iterasyonluk bir eğitimde komşuluk fonksiyonunun değerinin neredeyse kazanan nöron etrafındaki birkaç nöronu ya da sadece kazanan nöronu kapsayacak kadar küçük bir değere kadar azalması gibi.

Komşuluk mesafesinin kullanıldığı komşuluk fonksiyonunda ise fonksiyonun büyük bir değerden çok küçük bir değere kadar azalabilmesi için Eş. (57)'de yer alan birinci zaman sabitinin Eş. (73) ile hesaplanması önerilmektedir. (Haykin, 2009).

$$\tau_1 = \frac{1000}{\log \sigma_0} \quad (73)$$

Komşuluk fonksiyonundaki bir diğer önemli nokta ise fonksiyonun konveks olması gerektiğidir. Çünkü fonksiyonun konkav olması, ÖÖH'nin meta-stabil/yarı-kararlı durumda kalmasına neden olmaktadır. Bu durum, haritanın yerel minimuma takılması anlamına gelir. Bunu önlemek için algoritmanın ikinci aşamasında komşuluk fonksiyonu için kullanılan Gauss fonksiyonunun  $2\sigma$  genişliğindeki orta kısmının kullanılmasının yeterli olduğu ifade edilmektedir (Kohonen, 1989). Çünkü fonksiyonun tepe noktası

çevresindeki  $\sigma$  yarıçaplık mesafede kalan çan yapısındaki kısmı konveks yapıdadır, Şekil 27. Ayrıca kazanan nöronun bu mesafedeki komşuları hem kendisi ile yüksek derecede benzeşmekte hem de varyansın büyük bir kısmını temsil etmektedir (%68). Fonksiyonun kuyruk kısmına doğru ilerledikçe ise komşu nöronların varyansı açıklama gücü daha da artarken kazanan nöron ile benzerlikleri azalmakta ve fonksiyon konkavlaşmaktadır.



Şekil 27. Gauss komşuluk fonksiyonu (Haykin, 2009)

Başlangıç ağırlık vektörlerinin tespitinde ise çeşitli yaklaşımlar bulunmaktadır. Bu yaklaşımlardan en sık kullanılanlarından bir tanesi rassal ağırlıklandırma “random initialization”. Bu yaklaşımda tüm sinaptik ağırlıklara 0-1 aralığında rassal değerler atanmaktadır. Bir diğer yaklaşım ise sinaptik ağırlıkların değerlerinin girdi vektörü değerlerinden seçilmesidir (Su, Liu, Chang, 2002). Bu yaklaşımın uygulanabilmesi için girdi vektörü değerlerinin 0-1 aralığına çekilerek normalleştirilmesi tavsiye edilmektedir. Sıklıkla kullanılan bir diğer yaklaşım ise doğrusal ağırlıklandırma. Bu yaklaşımda öncelikle veri setinin ortalaması alınır. Daha sonra en büyük iki eigen değeri tespit edilir. Ağırlıklar, ortalama değere birinci ve ikinci eigen değerlerinin lineer bileşkesi eklenerek hesaplanır. Böylelikle ağırlık vektörü girdi vektörünün birinci ve ikinci bileşenlerinin uzayına ölçeklenmiş olur (Kohonen, 2001; Hagan, vd., 1996). Bu yaklaşımın öğrenme algoritmasını hızlandırdığı fakat algoritmanın yerel minimuma takılma ihtimalini arttırdığı da ifade edilmektedir. Ayrıca Akinduko, Mirkes, Gorban (2015) çalışmasında rassal belirlenen başlangıç ağırlık vektörlerinin doğrusal olmayan veri setlerinde doğrusal ağırlıklandırma yöntemine kıyas ile çok daha iyi sonuçlar sağladığı tespit edilmiştir.

### 3.4.Öz-Örgütlemeli Haritaların Doğrulaması

Öz-Örgütlemeli Haritaların doğrulanmasında ise çeşitli ölçütler incelenmektedir. Doğrulamada kullanılan ölçütlerden ilki nicemleme hatasıdır (NH) “quantization error”.

NH, her bir girdi vektörü ile bu vektörlerin ait oldukları/eşleştikleri kazanan nöronlara ait sinaptik ağırlıklar arasındaki ortalama mesafedir. Her bir  $i$ 'inci girdi ile eşleştiği  $j$ 'inci kazanan nöron için,  $x_j$  girdi vektörü,  $w_i^*$  kazanan nöronun ağırlık vektörü,  $m$  gözlem (girdi vektörü) sayısı olmak üzere NH, Eş. (74) ile hesaplanmaktadır.

$$NH = \frac{1}{m} \sum_{j=1}^m \|x_j - w_i^*\| \quad (74)$$

Çıktı katmanı ile girdi katmanı arasındaki uyumu “fitting” değerlendirmeye yarayan bir ölçüttür. NH ne kadar küçük olursa, girdi katmanı ile çıktı katmanı arasındaki mesafenin ortalaması da o kadar küçük olur (Chabane, Islam, Zerr, 2017; Uriarte, Martin, 2008). Dolayısı ile farklı topolojilere veya başlangıç parametrelerine sahip iki haritadan NH'si düşük olanın, diğerine kıyas ile daha kesin sonuç sağladığı ifade edilir (Gabrielsson, Gabrielsson, 2006).

Topografik hata (TH), bir  $x_j$  girdi vektörü için birinci ve ikinci kazanan nöronların komşu nöronlar olmama oranını temsil etmektedir. Haritadaki nöronların diziliminin sürekliliğini temsil etmeye yarayan ve haritanın yapısının veri setine uygun olup olmadığının incelenmesini sağlayan (Gabrielsson, Gabrielsson, 2006; Pözlzbauer, 2004) bir ölçüttür. Sürekliliğin düşük olduğu haritalarda, haritanın yapısının korunamadığı ifade edilir. Kazanan nöronlar birbirine uzak ve harita yoğun/sıkı bir yapıda değildir. Kazanan nöron dizilimleri seyrekleştiği için hiç atama yapılmamış (hiçbir girdi vektörü ile benzeşmeyen) nöron sayısı fazla olabilmektedir. Bu durum haritanın gereksiz büyük tasarlandığı anlamına gelir ve eğitim süresinin uzamasına neden olur. Bu sebeplerden ötürü, TH ne kadar düşük olursa, haritanın topolojisi de o kadar korunur.

$$TH = \frac{1}{m} \sum_{j=1}^m U(x_j) \quad (75)$$

$$U(x_j) = \begin{cases} 1 & \text{Birinci ve ikinci kazanan nöronlar komşu değil ise} \\ 0 & \text{Komşu ise} \end{cases} \quad (76)$$

Eş. (75) ile tanımlanmış olan TH, kare komşuluk alanı ile tasarlanmış olan haritalar için yanıltıcı sonuçlar verebilir. Kare komşuluk alanında kazanan nöronun dört komşusu bulunmaktadır. Fakat örneğin, bir girdi vektörüne ait ikinci kazanan nöron birinci kazanan nöronun köşegeninde kalıyor ise söz konusu girdi vektörüne ait  $U(x_j) = 1$ 'dir. Dolayısı ile bu şekildeki bir değerlendirme sanki söz konusu bu iki nöronun birbirinden

çok uzak olduğu ve haritanın kopuk bir yapıya sahip olduğu anlamına gelmektedir.<sup>127</sup> Fakat aslında bu değerin sıfır veya sıfıra yakın bir değer olması daha uygundur. Başka bir ifade ile kare komşuluk alanları için köşegen nöronun da değerlendirilmesini sağlayan bir eşitlik ile TH'nin hesaplanması daha gerçekçidir. Bu amaç ile alfa topografik hata ölçütü Uriarte, Martin (2008) tarafından geliştirilmiştir (Chabane, Islam, Zerr, 2017). Yazarlar kendi çalışmaları için  $Z = \sqrt{2}$  değerini kullanmışlardır.<sup>128</sup>

$$ATH = \frac{1}{m} \sum_{j=1}^m \alpha(x_j) \quad (77)$$

$$\alpha(x_j) = \begin{cases} 1 & \text{Birinci ve ikinci kazanan nöronlar komşu değil ise} \\ 1/Z & \text{Birinci ve ikinci kazanan nöronlar köşegen ise} \\ 0 & \text{Diğer tüm durumlar} \end{cases} \quad (78)$$

NH ve TH ölçütlerinin birlikte değerlendirilmesi önerilmektedir (Gabrielsson, Gabrielsson, 2006). Açıklamak gerekir ise NH değeri, çıktı katmanındaki nöron sayısının artırılması ile sıfıra yaklaşmaktadır. Çünkü girdi katmanındaki her bir girdi vektörü daha fazla nöron ile temsil edilerek, nöronlar ile girdiler arasındaki benzerlik artmakta ve dolayısı ile NH azalmaktadır. Fakat bu durumda ise TH değeri artacaktır. Çünkü girdi vektörlerinin haritadaki dağılımı seyrekleşecektir. (Pözlbauer, 2004)

ÖÖH sonuçlarının doğrulanmasında, yukarıda sayılan NH ve TH parametrelerinin incelenmesine ek olarak, veri seti bölümlenmeli veya hiyerarşik kümeleme algoritmaları gibi farklı algoritmalarından bir veya birkaç tanesi ile kümelenecek kümeleme sonuçları karşılaştırılmaktadır (Costa, De Carvalho, De Souto, 2004). Bu yaklaşım, çapraz doğrulama olarak adlandırılmaktadır. Aynı zamanda bootstrap algoritması,  $t$  veya  $F$  testleri de ÖÖH sonuçlarının değerlendirilmesinde kullanılabilir. Bu bölümde Kısım 1.2 ve 1.3'te uygun küme adedinin tespiti ve küme kalitesinin incelenmesi konuları altında söz konusu teknikler/algoritmalar hakkında bilgi edinilebilir.

---

<sup>127</sup> ...ve belki de bu yüzden Kohonen, kare komşuluk alanından ziyade altıgen komşuluk alanının kullanılmasını önermektedir.

<sup>128</sup> Bir birim kare nöronun merkezinden, köşegeninde kalan nöronun merkezine olan uzaklık, oluşan 1 birimlik dik üçgenin hipotenüsüdür  $=\sqrt{2}$

### 3.5.Karma Veri Kümelemede Kullanılan Öz-Örgütlemeli Harita Çeşitleri

ÖÖH'lerin en önemli ön şartı girdi uzayını oluşturan özneliklerin sürekli nitelikte olması zorunluluğudur (Haykin, 2009). Çünkü kazanan nöron (küme merkezleri), uzaklık kriterine göre tespit edilmektedir ve söz konusu uzaklık (genellikle) Öklid mesafesi ile hesaplanmaktadır. Fakat bu şart günümüzde bazı özel ÖÖH türleri için geçerli değildir.

Bu amaç ile geliştirilen ÖHH'ler incelendiğinde çoğunluğunun üç ÖÖH algoritması esas alınarak geliştirildiği dikkati çekmektedir. Söz konu algoritmaların en bilinenleri ve onların bazı türevleri aşağıda kısaca açıklanmaya çalışılmıştır.

Söz konusu temel algoritmalarından ilki Büyüyen ÖÖH'dir (BÖÖH) "Growing SOM" (Alahakoon, Halgamuge, Srinivasan, 2000). Yazarlara göre ağırlık eğitimi süresince sabit tutulan nöron sayısı ÖÖH'lerin bir dezavantajıdır. Eğitim süresince sabit kalan nöron sayısı yerine, çok küçük bir harita ile eğitime başlanmalı ve eğitim süresince nöron sayısı arttırılmalıdır. Yazarlar bu amaç ile değeri  $[0, 1]$  aralığında değişen bir yayılma faktörünün "spread factor" kullanılmasını önermiştir. Algoritmanın ilk aşamasında daha genel kümeler, ikinci aşamasında ise nöron sayısının artışı ile daha hassas kümeler elde edilmektedir. Haritanın eğitimi süresince hangi nöronun yeni bir komşu nöron yaratacağı, başlangıç haritasındaki bu nöronların hata değerine (girdi vektörü ile ağırlık vektörleri arasındaki Öklid mesafesine) ve büyüme eşik "growth threshold" değeri adı verilen bir parametreye bağlıdır. Büyüme eşik değeri veri setindeki öznelik sayısı ve yayılma faktörü dikkate alınarak hesaplanan bir değerdir. Hata değeri en büyük olan ve aynı zamanda büyüme eşik değerinden de büyük olan nöron yeni bir komşu oluşturan nöron olarak belirlenir. Yeni nöronun sinaptik ağırlığı ise haritadaki kendi pozisyonuna göre komşu olduğu nöronların ağırlıkları dikkate alınarak atanır. Yayılma faktörünün 1'e yakın seçilmesi büyüme eşik değerini küçültmekte ve 0'a yakın seçilmesi ise arttırmaktadır. İlk aşamadaki nöron sayısı 4 ve civarı olmalıdır ve bu yüzden komşuluk yarıçapı da klasik ÖÖH'lere nazaran tüm haritayı kapsar büyüklükte olmamalıdır. Dört nöron ile eğitime başlanan bir harita için yayılma faktörünün başlangıç değeri 0.3 gibi küçük bir değer olarak seçilebilir. Yayılma faktörünün zamanla artışı, eşik değerini küçülttüğü için eğitim sürecinde haritadaki nöron sayısını arttırmaktadır. Böylelikle eğitimin başında daha genel ve kapsayıcı nitelikte olan kümeler, eğitimin sonunda daha

hassas bölümlere ayrılmış alt kümelere dönüşecektir.<sup>129</sup> Bu algoritma, sadece nümerik öznitelikler içeren veri setleri için uygundur, karma veri setleri için uygun değildir.

İkinci temel algoritma Görselleştirme Kaynaklı ÖÖH'dir (GKÖÖH) "Visualization-Induced SOM (ViSOM)" (Yin, 2002). ÖÖH'ler büyük boyutlu verilerin görüntülenmesi için kullanıldığında, nöronlar arasındaki mesafeleri görselleştirmek için U matrisinden faydalanmaktadır. Fakat bazen veri setinin sahip olduğu gizli örüntü(ler) tespit edilebilir değildir ya da verilerin harita üzerindeki iki boyutlu izdüşümlerinin geometrik yapıları biçimsiz ya da bozuk olabilmektedir. Klasik ÖÖH'ler bu türdeki yapıları zor tespit etmekte ya da tespit edememektedir. GKÖÖH, bu problemi çözmek için tasarlanmış bir algoritmadır. Algoritma, bir parametre<sup>130</sup> ile orijinal veri setindeki gözlemler arası mesafeyi ve topolojiyi dikkate alarak, çıktı katmanındaki nöronlar arası mesafeyi sınırlandırıp harita topolojisini düzenlemeye çalışır. Yazarlar, önerdikleri algoritma ile eğitilen haritaların topolojisinin, orijinal veri setinin topolojisi ile neredeyse tamamen eşleştiğini ifade etmektedir.<sup>131</sup> Bu algoritma, sadece nümerik öznitelikler içeren veri setleri için uygundur, karma veri setleri için uygun değildir.

Üçüncü temel algoritma ise Genelleştirilmiş ÖÖH (GÖÖH)'dir "Generalized SOM (GSOM)" (C.-C. Hsu, 2006). Kavram ya da mesafe hiyerarşisi esas alınarak geliştirilmiş bir algoritmadır. Bu algoritmada, kategorik gözlemler arası benzerlik sezgisel olarak ölçülerek analize dahil edilmektedir. Örneğin çalışmada gaz ve kahve içeren içecekler için bir kavram hiyerarşisi oluşturulmuş ve Pepsi, Coca-Cola, Nescafe ve Mocca içecekleri arasındaki benzerlik maksimum 1 olacak şekilde sezgisel olarak belirlenmiştir. Dolayısı ile örneğin hiyerarşi ağacındaki kök "Pepsi" ile bir üst hiyerarşide bağlı olduğu yaprak "gazlı içecek" arasındaki mesafe 1'dir. Diğer taraftan, kategorik verilerde

---

<sup>129</sup> Elde edilen kümeler, birbirleri ile hiyerarşik ilişkiye sahiptir. Hiyerarşik küme analizinde kullanılan dendrogramın farklı kesim seviyelerinde elde edilen kümeler ve alt kümeler gibi farklı yayılma faktörü değerleri için küme-alt küme grupları elde edilmektedir.

<sup>130</sup> Söz konusu parametre çözünürlük "resolution" parametresi olup, Eş. ()'te yer alan sönümlenme terimi iki bileşene ayrılarak oluşturulmuştur. Çözünürlük parametresi, komşu nöronlar arası mesafeyi yakın tutup, komşu olmayan nöronlar arasındaki mesafeyi büyütmemektedir. Parametre değerinin sıfıra yaklaşması çözünürlüğü artırarak çok boyutlu verilerde farklı kümeleme algoritmalarının tespit edemediği kümeleri yakalayabilmektedir.

<sup>131</sup> Algoritmanın Sammon'nun Haritası ile çok benzeştiği ve bu haritadan farkının, algoritmanın yeni gözlemleri de işleyebilir yeteneğe olması ifade edilmektedir.



gerçekleştirilen bu dönüşüm, nümerik verilerde de bir dönüşümü gerekli kılmaktadır. Örneğin; satın alınan gazlı içeceğin miktarı nümerik bir özneliktir. Bu öznelik için tek seviyeli bir mesafe hiyerarşisi oluşturulur. Hiyerarşinin kökü, söz konusu “miktar” özneliğinin minimum gözlem değeri ve yaprağı ise maksimum gözlem değeri olarak tespit edilip, tüm gözlemler kök-yaprak bağlantısı arasında göreceli olarak konumlandırılır. Böylelikle hem kategorik özneliklerin nitel değerleri, örneğin Manhattan veya Öklid mesafelerinin kullanılabilmesi için, nümerik niteliğe kavuşur hem de bu dönüşümdeki yaklaşım, nümerik özneliklere de uygulanarak tüm öznelikler bir nevi tek tipleştirilmiş olur. Öznelik dönüşümleri gerçekleştirildikten sonra ise tüm veri seti normalleştirilir ve ÖÖH çevrim içi öğrenme algoritması ile ağın eğitimi gerçekleştirilir. Aslında bu yüzden GÖÖH algoritması tam olarak klasik çevrim içi algoritmayı kullanmakta fakat mesafelerin Öklit veya Manhattan uzaklıkları ile hesaplanabilmesi için bir öznelik dönüşüm yaklaşımı sunmaktadır.

C. C. Hsu, Wang, Wang (2006) çalışmasında önerilen algoritma GKÖÖH ve GÖÖH algoritmalarının entegre edilmiş versiyonudur, Genelleştirilmiş Görselleştirme Kaynaklı ÖÖH (GGKÖÖH) “Generalized Visualization-Induced SOM”. Bu algoritmada öncelikle GÖÖH’de önerilen öznelik dönüşümü yaklaşımı ile gerekli dönüşümler yapılır, veriler min-maks yaklaşımı ile standardize edilir ve GKÖÖH ile ağ eğitilir.

C. C. Hsu, Wang (2006) çalışmasında GÖÖH algoritması ile Genişletilmiş Öznelik Odaklı Tümevarım (GÖOT) “Extended Attribute-Oriented Induction” algoritması entegre edilmiştir. GÖOT, öznelik çıkarımı yöntemlerinden bir tanesidir. Yazarların önerdiği kümeleme yaklaşımında veri seti önce GÖÖH algoritması ile kümelere ayrılmış ve daha sonra GÖOT algoritması ile her bir kümenin ayırt edici öznelikleri tespit edilmiştir.<sup>132</sup> GÖOT algoritmasında bir  $\beta, \epsilon [0, 1]$  çoğunluk eşik değeri “majority threshold” araştırmacı tarafından belirlenir ve her bir kümenin ayırt edici öznelikleri  $\beta$  değeri referans alınarak tespit edilir. Bu algoritmaya göre bir  $C_i$  kümesindeki  $m$  tane gözlemden  $t$  tanesinin  $A_j$  özneliği değerinin aynı olması durumunda  $t/m$  oranı  $\beta$  değerine eşit veya ondan büyük ise söz konusu öznelik, incelenen küme için ayırt edici özelliktedir. Örneğin  $\beta \geq .70$  için bir işletmenin onaylı

---

<sup>132</sup> Yazarların ifadesi ile her bir kümenin karakteristik özellikleri tespit edilmiştir.

tedarikçi havuzundaki işletmeler çeşitli kriterler dikkate alınarak kümelensin. Bir kümedeki gözlemlerin 0.80'inin Avrupa'dan olması (kategorik öznelik) ve ppm değerlerinin 0.75'inin 100 ve 100 den küçük olması (nümerik öznelik), söz konusu küme için coğrafi bölge ve ppm öznelikleri, "Avrupa" ve "...  $\leq 100$ " değerleri için, ayırt edici özelliktedir.

Büyüyen Karma ÖÖH (BKÖÖH) "Growing Mixed Type SOM (GMixSOM)" ise BÖÖH ve GÖÖH algoritmalarını temel alan bir algoritmadır (Tai, Hsu, Chen, 2010; Tai, Hsu, 2012). Yazarlara göre BKÖÖH, kapsadığı GÖÖH algoritması ile haritanın kategorik özneliklerini dikkate almakta ve BÖÖH ile ise ÖÖH'lerin sabit nöron kısıtını bertaraf etmektedir. Ayrıca algoritmanın GÖÖH algoritmasını içeren kısmı aynen korunmuş olup, BÖÖH algoritmasında bir değişiklik önerilmiştir. BÖÖH algoritmasında başlangıç nöron sayısı dört iken önerilen algoritmada beştir ve BÖÖH algoritmasında yeni komşu nöron, kazanan nöronun yatay veya dikey doğrultusunda konumlanabilirken önerilen algoritmada çapraz olarak da konumlanabilmektedir. Çapraz konumlama, başlangıç nöron sayısının beş olması ile sağlanmaktadır. Dört nöronlu bir yapıda nöronlar bir karenin köşelerinde konumlanırken, beş nöronlu yapıda bir artı işaretinin uçlarında ve merkezinde konumlanmaktadır. Yazarlara göre dört nöronlu başlangıç, haritanın gereksiz büyümesine neden olmakta ve önerdikleri beş nöronlu yapı, gereksiz büyümeyi engellemektedir.

Prasad, Punithavalli (2012) çalışmasında önerilen Büyüyen Hiyerarşik ÖÖH (BHÖÖH) "Growing Hierarchical SOM" algoritması ismen BÖÖH algoritmasını andırsa da süreç olarak BÖÖH'den tamamen farklıdır. BÖÖH algoritmasında, her bir eğitim sürecinde (bir önceki haritanın genişlemiş versiyonu olan) tek bir harita ve dolayısıyla tek bir katman eğitilmektedir. BHÖÖH ise çok katmanlı bir yapıya sahiptir ve her katmanda birden fazla ve birbirinden farklı haritalar eşanlı ve birbirinden bağımsız bir şekilde eğitilmektedir. Eğitime 2x2'lik bir harita ile başlanması önerilmektedir.  $i$ 'inci nöron ile eşleşen  $x$  girdi vektörleri ve  $i$ 'inci nörona ait  $w_i$  ağırlık vektörü dikkate alınarak her bir nöron için ortalama NH, Öklit mesafesi kullanılarak hesaplanır.  $t$ 'inci iterasyonda, ortalama NH, başlangıç NH'nin  $r$  oranı kadarlık kısmından küçük olan nöronlar ve bu nöronlara en benzemeyen komşu "the most dissimilar neighbor" nöronlar tespit edilir. Örneğin,  $t$ 'inci iterasyonda NH şartını sağlayan bir  $i$  nöronu, haritanın ikinci satırı ve

üçüncü sütununda ve  $i$  nöronunun kendisine en benzemeyen komşusu  $b$  ise birinci satır ve üçüncü sütununda bulunuyor olsun. Söz konusu nöronlar birbirlerine komşu olmalarına rağmen benzeşimleri düşük olduğu için bir sonraki iterasyonda söz konusu iki nöron arasına yeni bir nöron satırı veya nöron sütunu eklenmelidir. Dolayısı ile  $i$  ve  $b$  aynı sütunda ve farklı satırlarda bulunduğu için  $i$ 'nin ve  $b$ 'nin bulunduğu satırların arasına yeni bir komşu nöron satırı oluşturulur ve  $t + 1$ 'inci iterasyonda  $i$ & $b$  ikilisi ve yeni oluşturulan komşu satır  $t$ 'inci iterasyondaki haritadan ayrılarak yeni bir harita oluşturur. Yeni nöronların ağırlıkları ise en yakın komşularına ait ağırlık vektörlerinin ortalaması alınarak belirlenir. Algoritma, NH şartını sağlayan nöron kalmayana kadar genişlemeye ve eğitime devam eder. BHÖÖH ile elde edilen kümeler, C. C. Hsu, Wang (2006) çalışmasındaki gibi, GÖOT algoritması ile incelenmiş ve her bir kümenin ayırt edici öznelikleri tespit edilmiştir. Dolayısı ile BHÖÖH algoritması ile aslında sadece nümerik öznelikler dikkate alınarak küme analizi yapılmıştır. Nümerik olmayan öznelikler ise her bir kümeye uygulanan GÖOT algoritması ile incelenmiştir.

Yukarıda açıklanan algoritmalara ek ve onlardan bağımsız olarak ayrıca Nümerik-Kategorik ÖÖH (NKÖÖH) “Numerical-Categoric SOM”, Olasılıksal Karma Topolojik Haritalar (OKTH) “Probabilistic Mixed Topological Maps”, Kategorik Olasılıksal ÖÖH (KOÖÖH) “Categorical Probabilistic SOM”, Frekans Nöronlu Karma ÖÖH (FNKÖÖH) “Frequency Neuron Mixed SOM”, ve Karma Öz Örgütlemeli Artımlı Sinir Ağı (KÖÖASA) “Mixed Self-Organizing Incremental Neural Network” algoritmaları da yazınında yer almaktadır.

NKÖÖH algoritması Chen, Marques (2005) tarafından önerilmiştir. Algoritma dört aşamadan oluşmaktadır. İlk aşamada her bir nörona ait sinaptik ağırlıklar rassal olarak belirlenir. İkinci aşamada, veri setindeki tüm gözlemler sırası ile ağa verilir ve nümerik öznelikler için Öklit mesafesi ile ve nümerik olmayan öznelikler için ise 2. Bölüm 5.3 numaralı kısımda açıklanan basit eşleşme katsayısı benzeri bir hesaplama<sup>133</sup> ile ağırlık ve girdi vektörleri arasındaki mesafeler hesaplanır. Üçüncü aşamada yığın algoritması esas alınarak sinaptik ağırlıklar güncellenir. Nümerik öznelikler için sinaptik ağırlıkların

---

<sup>133</sup> Basit eşleşme katsayısında, gözlemler arası toplam benzeşme  $(a + b)$ ,  $a: (1 - 1)$  ve  $b: (0 - 0)$  temsil etmek üzere, nümerik olmayan toplam öznelik sayısına bölünerek hesaplanmaktadır. Söz konusu çalışmada ise nümerik olmayan öznelikler için mesafe sadece  $(a + b)$  toplamı ile hesaplanmıştır.

güncellenmesi klasik yığın algoritmasındaki gibi gerçekleştirilirken, nümerik olmayan öznitelikler için ağırlık güncelleme işlemi iki türlü yapılabilmektedir. Öncelikle, her bir özniteliğin tanım kümesinde alabileceği her bir kategorik değer için frekansı hesaplanır. Akabinde ise

1. Frekansı, diğer kategorik değerlerin toplam frekansından daha büyük olan kategorik değer tespit edilir ve frekans, komşuluk fonksiyonuna bölünür. Bölüm, söz konusu sinaptik ağırlığın  $t + 1$ 'inci iterasyondaki güncel değeri olur ya da

2. Bir  $\theta$  oranı ve frekansı  $\theta$  değerinden büyük olan kategorik değerler tespit edilir. Söz konusu kategorik değerlerden frekansı en büyük olan seçilir. Maksimum frekans değeri komşuluk fonksiyonuna bölünerek elde edilen değer, söz konusu sinaptik ağırlığın  $t + 1$ 'inci iterasyondaki güncel değeri olur. Aksi halde söz konusu nöronun  $t + 1$ 'inci iterasyondaki sinaptik ağırlığı değişmeden kalır. Açıklamak gerekirse, örneğin ısı işlem görmüş parçalar çeşitli kriterler dikkate alınarak kümelenebilir ve bu kriterlerden bir tanesi parça yüzeyinin renk tonu olsun. Renk tonu kategorik değişkeninin tanım kümesi ise çok koyu, koyu, normal, açık ve çok açık olacak şekilde beş farklı değere sahip olsun. Bu durumda ÖÖH çıktı katmanındaki bir nöron ile eşleşen  $m$  tane girdi vektörünün renk tonu kategorik değişkeninin ölçüm değerlerinden bir veya bir kaçının frekansı  $\theta = 0.5$  için ( $m * 0.5$ ) değerinden büyük ise söz konusu nöronun kategorik değişkeni ile olan sinaptik bağlantısının ağırlığı yukarıda açıklanan kural dikkate alınarak güncellenmelidir. Eğer renk tonu kategorik değişkeninin ölçüm değerlerinin frekanslarının hiçbirisi  $m * 0.5$ 'ten büyük değil ise söz konusu nöronun kategorik değişkeni ile olan sinaptik ağırlığı güncellenmeden bir sonraki iterasyonda aynı sinaptik ağırlık ile eğitim devam etmelidir.

Yazarlar, sıralayıcı (ordinal) öznitelikler için de ayrı bir ağırlık güncelleme yaklaşımı önermişlerdir. Bu yaklaşıma göre ilgili nöronun sıralayıcı öznitelik ile olan sinaptik bağlantısının  $t + 1$ 'inci iterasyondaki güncel değeri  $t$ 'inci iterasyondaki ağırlıklandırılmış frekans toplamına eşittir.

Algoritmanın üçüncü aşamasında ağırlık güncelleme işlemleri tamamlandıktan sonra, dördüncü ve son aşamada harita kararlı bir yapıya ulaşana kadar, ikinci ve üçüncü aşamaların tekrarlanması gerektiği ifade edilmektedir. Önerilen bu algoritmanın alan yazınında, kategorik özniteliklere daha fazla ağırlık verildiği ve bu yüzden nümerik özniteliklerin kategorik özniteliklere kıyas ile önemsiz kaldığı belirtilmiştir.

OKTH modeli Rogovschi, Lebbah, Bennani (2008) çalışmasında önerilmiştir. OKTH'nin alan yazınındaki diğer tüm ÖÖH'lerden en önemli farkı haritadaki her bir nöronun bir olasılık dağılımını temsil etmesidir. Yazarlar, nümerik türdeki özniteliklerin koşullu olasılık dağılımı fonksiyonu için Gauss dağılımını, kategorik öznitelikler için ise Bernoulli dağılımını kullanmışlardır. Her bir nöronun temsil ettiği dağılım modelini en iyilemek için ise Beklenti Maksimizasyonu algoritması uygulanmıştır. Önerilen OKTH modelinin hesaplama yoğun bir model olduğu belirtilmektedir. Ayrıca kategorik özniteliklerin (Bernoulli dağılımı kullanıldığı için) (0-1) tipi ikili türde özniteliklere dönüştürülmesi gerekmektedir. Dönüşümün, alan yazınında bilgi kaybına neden olduğu ifade edilmektedir.

Del Coso, vd. (2015) çalışmasında yer alan FNKÖÖH algoritması ise NKÖÖH algoritması ile Lebbah, Benabdeslem (2010) çalışmasında önerilen ve sadece kategorik verilerin kümelenmesinde kullanılan bir ÖÖH algoritması esas alınarak geliştirilmiştir. FNKÖÖH algoritması, bu bölümdeki Kısım 3.2'de açıklanan klasik çevrim içi öğrenme algoritması ile benzer üç aşamaya sahiptir. Rekabetçi aşamada girdi vektörü ile ağırlık vektörleri arasındaki mesafe nümerik öznitelikler için Öklit uzaklığı ile ve kategorik öznitelikler için ağırlık vektörünün girdi vektörüne benzememe olasılığı ile ölçülmektedir. Olasılığın hesaplanmasında her bir kategorik değişkenin aldığı değerlerin frekansı kullanılmakta ve hesaplanan mesafeler dikkate alınarak kazanan nöronlar tespit edilmektedir. Akabindeki işbirlikçi aşamada komşuluk mesafesi, klasik yığın algoritması kullanılarak, tüm veriler ağa tanıtıldıktan sonra güncellenmektedir. Son olarak yakınsama aşamasında ise ağırlık vektörleri güncellenmektedir. İkinci aşamada komşuluk mesafesi hesaplanırken kullanılan eşitliğin pay kısmında birinci aşamada hesaplanan Öklit mesafesi ve benzememe olasılıkları yer aldığı için bu aşamada sinaptik ağırlıklar güncellenirken hem nümerik hem de kategorik öznitelikler ayrı ayrı dikkate alınmış olmaktadır. Ağırlık güncelleme işleminde nümerik öznitelikler için klasik yığın algoritmasındaki Eş. () dikkate alınarak gerçekleştirilmektedir. Kategorik öznitelikler için ise nümerik öznitelikler için kullanılan eşitliğin bir benzeri kullanılmaktadır. Eşitlik,  $j$ 'inci kategorik değişkenin  $r$  adet kategorik değerleri için hesaplanan komşuluk fonksiyonlarının toplamının,  $i$ 'inci nöron için hesaplanan komşuluk fonksiyonlarının toplamına bölümü şeklindedir. Algoritmanın NKÖÖH algoritmasından daha iyi bir

performansa sahip olduğu ifade edilmektedir. Söz konusu algoritma, kategorik özniteliklerde (0-1) dönüşümünü gerektirmediği için hem veri setinin boyutu artmamakta hem de dönüşümün neden olduğu bilgi kaybını da önlemektedir. Algoritmanın tartışmaya açık tek yönü yığın öğrenme yaklaşımı ile kullanılabilir olması olabilir.

Noorbebhani, vd. (2015) çalışmasında önerilen KÖÖASA algoritması, Shen, Hasegawa (2008) çalışmasında önerilen Düzeltilmiş Öz Örgütlemeli Artımlı Sinir Ağı (DÖÖASA) algoritmasını temel almıştır.<sup>134</sup> Tek farkı DÖÖASA algoritması sadece nümerik veri setlerine uygulanabilirken, yazarlar yeni bir uzaklık ölçütü geliştirerek bu ölçütü DÖÖASA algoritmasında kullanmış ve bu entegrasyonu KÖÖASA algoritması olarak adlandırmıştır. Önerilen uzaklık ölçütü iki parçadan oluşmaktadır. İlk parça nümerik öznitelikler için ve ikinci parça kategorik öznitelikler için gözlemler arası mesafeyi temsil etmektedir. Bir  $i$  ve  $j$  gözlemleri arasındaki mesafe bu gözlemlerin

---

<sup>134</sup> Söz konusu algoritma, Öz Örgütlemeli Artımlı Sinir Ağı (ÖÖASA) algoritmasını temel almaktadır. Bu algoritma, iki katmanlı ÖÖH yapısı ile kümeleme yapılmasına olanak sağlamaktadır. İlk katmanda, veri setinin yoğunluk dağılımı incelenmekte ve ikinci katmanda ise ilk katmanda tespit edilen yoğunluklar kullanılarak az yoğunluklu bölgelerin tespiti ve haritadan çıkartılması ile veri seti kümelerine ayrılmaktadır. Öncelikler, her bir girdi vektörü için birinci ve ikinci kazanan nöronlar Öklid uzaklığı ile tespit edilir. Eğer birinci ve ikinci kazanan nöronlar, karar verici tarafından belirlenmiş olan, birinci ve ikinci kazanan nöronlar için benzerlik eşik değerlerinden küçük ise söz konusu girdi vektörü ilgili kümeye ait olur ve ilgili ağırlık ve komşuluk fonksiyonları güncellenir. Ayrıca birinci ve ikinci kazanan nöronlar birbirine komşu değil ise (aralarında bağlantı yok ise) bu iki nöron birbirine (bir “edge” ile) bağlanır. Eğer söz konusu girdi vektörü eşik değerlerin herhangi birisinden büyük ise bir sonraki iterasyonda yeni bir kümeyi temsil etmek üzere ağa yeni bir nöron eklenmesini sağlar. Bu işlem kümeler arası nöron ekleme olarak adlandırılır “between-class insertion”. Ayrıca eğitim süreci karar verici tarafından belirlenen bir iterasyon sayısına ulaştığında ağırlık belirli bir bölgedeki (kümesindeki) toplam hata çok büyük ise söz konusu bölgeye yeni nöronlar eklenebilmektedir. Fakat yeni eklenen nöronlar ilgili bölgedeki toplam hatayı düşürmez ise bu nöronlar daha sonra silinmektedir. Bu işlem küme içi nöron ekleme olarak adlandırılır “within-class insertion”. Algoritma, ilerleyen iterasyonlarda, komşusu bir ve birden küçük olan nöronları tespit ederek bu nöronları silmektedir. Çünkü bu tip nöronlar yoğunluğun düşük olduğu alanları temsil etmektedir. Bu yüzden bu algoritmanın gürültülü veriyi bertaraf ettiği belirtilir. Birinci ÖÖH katmanında yoğunlukla kümeler arası nöron ekleme işlemi gerçekleştirirken, ikinci katmanda hem kümeler arası hem de küme içi nöron ekleme işlemi dengeli bir şekilde gerçekleştirilmektedir çünkü son katmanda amaç, haritanın kararlı bir yapıya kavuşturulmasıdır.

DÖÖASA algoritmasında ise küme içi nöron ekleme işlemi bulunmamaktadır. Sadece kümeler arası nöron ekleme işlemi ÖÖASA algoritması temel alınarak gerçekleştirilmektedir. Kümeler arası nöron ekleme işlemi ile bir nevi ön kümeleme gerçekleştirilmiş ve sinaptik ağırlıklar kabaca belirlenmiş olur. Daha sonra ön kümeleme işlemi ile elde edilen bilgiler dikkate alınarak k-ortalama algoritması ile tekrar bir kümeleme işlemi gerçekleştirilir. Bir sonraki aşamada gürültü temizlenir ve en son aşamada merkez temizleme olarak adlandırılan bir süreç işletilir. Bu süreçteki amaç, bazı kümelerin merkezlerindeki ağırlık vektörlerini tespit edip silerek, algoritmanın hızlanmasını sağlamaktır. Söz konusu ağırlık vektörleri, kendi kümeleri hariç hiçbir kümenin komşusu olmayan/olmayan vektörlerdir. Yazarlara göre, DÖÖASA ile eğitilmiş bir ağırlık, eğitim seti ile aynı dağılıma sahip bir test seti ile çalıştırılır ise merkez temizleme işlemi eğitilmiş ağırlık başarımını etkilemeyecektir.

nümerik özniteliklerinin aldığı değerlerin farklarının karelerinin toplamı ile kategorik öznitelikler için hesaplanan Hamming mesafesi<sup>135</sup> toplamının karekökü ile hesaplanmaktadır. Başka bir ifade ile önerilen yeni uzaklık ölçütü sürekli ve kategorik iki farklı mesafe ölçütünün toplamının kareköküdür. Bu yüzden önerilen ölçütün performansının alan yazınında bu amaçla geliştirilmiş olan Bölüm 2 Kısım 5.6'da açıklanan Gower ölçütü ile karşılaştırılmasında fayda bulunmaktadır. Ayrıca önerilen algoritmanın eğitim süreci DÖÖASA algoritmasını kullanmaktadır. Bu algoritmanın dipnot 134'te açıklandığı üzere en son aşaması, küme merkezlerinin temizlenerek hantal olan algoritmanın hızlandırılmasıdır. Fakat eğitilmiş ağırlık başarımlarını temizleme işleminden ötürü tartışmaya açıktır.

Söz konusu tez çalışmasında ise Wehrens, Kruisselbrink (2018a) çalışmasında önerilen SÖÖH yaklaşımı uygulanmıştır. SÖÖH'nin en önemli özelliği, çok katmanlı tasarlanabilmesi ve her bir katman için farklı bir mesafe kriteri kullanılmasına olanak sağlamasıdır (Wehrens, Kruisselbrink, 2018a, 2018b). Bu özellik, veri setinde yer alan farklı türdeki özniteliklerin kendi yapılarına uygun ölçütler ile değerlendirilmesini olanaklı kılmaktadır (Sala, vd., 2019). Başka bir ifade ile yukarıda açıklanan ÖÖH algoritmalarının çoğunluğunda gerçekleştirilen öznitelik dönüştürme işlemi bertaraf ederek bilgi kaybını engellemektedir. Ayrıca her katmana ağırlık atanabilmektedir. Yaklaşım hem çevrim içi öğrenme hem de yığın öğrenme algoritması ile çalışabilme esnekliğini sunmaktadır.

---

<sup>135</sup> Hamming mesafesi iki dizilimin birbirinden farkını veren bir uzaklık ölçütüdür. Örneğin “kapat” ve “karat” kelimeleri arasındaki mesafe 1 ve “büküm” ve “dügüm” kelimeleri arasındaki mesafe ise 2'dir. Hamming mesafesinin, yazarların önerdiği uzaklık ölçütündeki kullanımı ise şu şekildedir:  $i$  ve  $j$  gözlemlerinin ( $k$ 'ncü kategorik öznitelik için) arasındaki mesafe,  $i$  ve  $j$ 'nin  $k$  değeri aynı ise 0, farklı ise 1'dir.  $i$  ve  $j$  arasındaki toplam mesafe ise tüm  $k$  kategorik değişkenleri için  $i$  ve  $j$  arasındaki mesafeler toplamına eşittir.

## DÖRDÜNCÜ BÖLÜM

### (UYGULAMA)

Bu tez çalışmasında, Bursa halkının sürdürülebilir tüketim davranışları, değerlere bağlı olarak çevresel endişeleri (tutumları) ve doğa ile ilişkileri dikkate alınarak ekolojik kimlik profilleri aşağıdan-yukarı araştırma yaklaşımı ile keşfedilmeye çalışılacaktır. Bu kapsamda, araştırma tasarımı ve süreci aşağıda açıklanan durumlar ve ayrıca veri madenciliği bilgi keşfisüreci dikkate alınarak şekillendirilmiştir:

Kent markalaşması çalışmaları kentlerin fiziki, sosyal ve ekonomik çevrelerinin kalkınmasını sağlayan faaliyetlerdir. Ekonomik ve sosyal çevrede sınırların global çapta ortadan kalkması nedeni ile kentler, sadece kendi ülkelerindeki kentler ile değil aynı zamanda diğer ülkelerdeki kentler ile de yarış içine girmiştir. Bu yarıştaki amaç, daha çok yatırımcı, turist ve yetenekli işgücünü oluşturacak yeni vatandaşları kente çekmektir. Bu neden ile hem bir sanayi ve eğitim kenti<sup>136</sup> olan hem de bir zamanlar yeşil doğası ile anılan bir kent olan Bursa için uluslararası bilinirliği olan başarılı bir kent markasının oluşturulması önem arz etmektedir. Bu sebep ile çalışmanın uygulanacağı kent olarak Bursa seçilmiştir.

Başarılı bir markalaşma çalışması üç yapının birlikteliği ile gerçekleştirilebilir; vatandaşlar, iş çevresi ve yerel yönetimler. Alan yazınındaki birçok çalışma ve pratikteki birçok uygulama incelendiğinde markalaşma çalışmalarının genellikle yerel yönetimleri ve iş çevresini dikkate aldığı görülmektedir. Hatta bu çalışmalar, pazarlama veya halkla ilişkiler firmalarına taşere edilerek yürütülmekte ve bu yüzden kent markası, vatandaş bileşeninden uzaklaşmaktadır. Dolayısı ile kaçınılmaz olarak başarısız veya kısa ömürlü olmaktadır. Hâlbuki kentin varlığının devamlılığı vatandaşına bağlıdır. Bu neden ile tez çalışmasının araştırma yaklaşımı aşağıdan-yukarı olup, araştırma Bursa kent sakinleri üzerinde yapılmıştır.

Bir kent markalaşması çalışmasının, kenti yaşamak için iyi ve yatırım yapmak için karlı bir kent olması üzerine iki amacı bulunmaktadır. Her iki amacı tek bir potada

---

<sup>136</sup> Kentte iki devlet üniversitesi vardır ve yakın gelecekte üç tane özel üniversite açılacaktır.



harmanlayabilen markalaşma yaklaşımı ise sürdürülebilir kent markalaşmasıdır. Bu bağlamda ulaşılması hedeflenen marka, “Yeşil Bursa” olarak belirlenmiştir.

Sürdürülebilir kentler kapsamında, vatandaşların ekolojik kimlik profilleri keşfedilmeye ve anlamlandırılmaya çalışılmıştır. Bunun için, bireyin çevresel endişe bağlamında tutumu, doğa ile ilişkisi ve sürdürülebilir geri dönüşüm, ulaşım, enerji, su ve gıda tüketimi davranışları incelenmiştir. Ayrıca bireylerin yaşadıkları hanelerin elektrik, su ve doğalgaz tüketim bilgileri ve alan yazınında tüketimi etkilediği belirtilen gelir, hane halkı sayısı, evde bakıma muhtaç bireylerin olması, hane halkı türü, mesken kullanım süresi, meskendeki oda sayısı ve meskenin türü gibi özellikler de analizlere dahil edilmiştir.<sup>137</sup> Sayılan özniteliklerin bazılarının kullanım amaçlarının açıklanmasında fayda bulunmaktadır.

- “Hane halkını hangisi en iyi şekilde tanımlamaktadır” sorusu, katılımcının aile yaşam döngüsü hakkında bilgi almayı amaçlamaktadır. Aile yaşam döngüsü değişkeninin elektrik / enerji tüketimini etkilediği bilirse de sayılı çalışmada dikkate alındığı tespit edilmiştir. Fritzsche, (1981)’in de belirttiği üzere hane halkının büyüklüğü kesinlikle enerji tüketim davranışını ve dolayısı ile tüketimi etkilemektedir. Fakat hanenin ya da ailenin yaşamsal olarak hangi döngüde olduğu da önemlidir (Van Raaij, Verhallen, 1983; Frederiks, Stenner, Hobman, 2015b). Nitekim Fritzsche, (1981) çalışmasında, ailenin elektrik tüketiminin genç çocuklu orta yaşlı evli çift döngüsüne kadar sürekli olarak arttığını ve devamındaki evrelerde yavaşça azalmaya başladığını tespit etmiştir. Dolayısı ile söz konusu ifade bu çalışmada, başlı başına tek bir değişken olarak yer almayacaktır. Bu kavram altında, ailenin yaşam döngüsündeki konumu (stage of family life cycle), yaş grubuna bağlı olarak hane halkı nüfusu ve bakıma muhtaç (hasta, engelli ve yaşlı) kişi değişkenleri yer alacaktır. Ek2’den de inceleneceği üzere soru formundaki ilgili madde 13 seçenekten oluşmakta olup detayda ebeveynlerin yaş grubunu içerip, çocukların yaş gruplarını içermemektedir. Bu nedenle aile yaşam döngüsü değişkenine ek olarak her yaş grubundan kaç kişinin bulunduğu “family composition” değişkeni de incelenecektir.

---

<sup>137</sup> Eğitim değişkeni dikkate alınmamıştır. Çünkü alan yazınında bu değişkenin örneğin sürdürülebilir enerji tüketimi, ya da davranışında olumlu etkisi olduğu, olumsuz etkisi olduğu ve aynı zamanda etkisiz olduğu şeklinde çok farklı bulgular vardır (Guerin, Yust, Coopet, 2000). Bu yüzden bu çalışma kapsamında eğitim değişkeni dikkate alınmamıştır.

- Ailede bakıma muhtaç kişilerin bulunmasının da çeşitli araştırmalarda enerji tüketimine etkisi olduğu tespit edilmiştir. Çünkü bakıma muhtaç kişinin varlığı, meskenin gün içi kullanım süresini ve dolayısıyla hanenin elektrik / enerji ihtiyacını arttırmaktadır (Frederiks, Stenner, Hobman, 2015b; Guerin, Yust, Coopet, 2000; Longhi, 2015).

- İlgili yazınında meskenin gün içi kullanım süresi “time spent at home / presence at home / occupancy patterns” değişkeninin de enerji talebini etkilediği şeklinde bulgular yer almaktadır (Hiller, 2014; Kipping, Trømborg, 2016; Tao Zhang, Siebers, Aickelin, 2012). Çünkü evin kullanım süresi veya hane halkının evde bulunma süresi arttıkça enerji talebi de artmaktadır.

- Gelir değişkeni ilgili yazınındaki birçok çalışmada elektrik /enerji talebini etkilediği veya tüketimdeki varyansı kuvvetli bir şekilde açıkladığı belirtilmiştir (Frederiks, Stenner, Hobman, 2015b; Güloğlu, Akın, 2014). Örneğin Şangay’da yapılan bir araştırmada, hane halkı gelirinde yaşanacak %10’luk bir artışın enerji tüketiminde %2’lik bir artışa neden olduğu tespit edilmiştir (Xie, Ouyang, Gao, 2016). Çünkü gelirin artması hane halkının refahının ve alım gücünün artması anlamına gelmekte ve dolayısıyla evde daha fazla sayıda elektrikli alet bulundurma ve daha fazla sıklıkta bu aletleri kullanma sonucunu doğurmaktadır (Wiedenhofer, 2011).

Alan yazınında, çevresel endişe (tutum), doğayla ilişki ve STD arasındaki ilişkinin daha önce analiz edilmediği ve ayrıca kullanılan ölçme araçlarının Türkiye bağlamında hane halkı gibi genel bir kitle üzerinde daha önce hiç test edilmediği dikkati çekmektedir.<sup>138</sup> Bu yüzden, doğa ile ilişkinin tutum ve davranış arasındaki aracı etkisine ait YEM, KEKK yaklaşımı ile analiz edilmiştir. Analizler için PLS-SEM yazılımı kullanılmıştır:

Vatandaşların ekolojik kimlik profillerinin keşfedilmesinde kullanılacak olan veri seti nominal, ordinal ve nümerik türde değişkenler içerdiği için karma veri setidir. Bu yüzden, keşif sürecinde karma veri kümeleme yöntemlerinden SÖÖH algoritması kullanılmıştır. Analizler ise R-Studio ve PRIMER-e yazılımları ile gerçekleştirilmiştir:

---

<sup>138</sup> Bu kanıya varmak için ilgili yazın taraması yapılmış ve bütünsel olarak modelin testini yapan bir çalışmaya rastlanılmamıştır. Ek olarak ilgili ölçeği geliştirmiş olan ekipten Elisabeth Lisa Nisbet ile irtibata geçilmiştir.

Böylelikle sürdürülebilir bir kent olma ve AYB ödülünü alarak “Yeşil Kent” markasını yaratmanın ilk adımı niteliğindeki vatandaşların ekolojik kimliklerinin tanımlanması, açıklanması ve anlaşılması aşaması tamamlanmış olacaktır.

Her araştırmada olduğu gibi bu tez çalışmasında da gerek coğrafi açıdan, gerek ise kavramsal ve uygulama açısından kısıtlar bulunmaktadır.

- Zaman, mekân ve ekonomik sınırlılıktan ötürü ve ayrıca kent ve kırsal hane halklarının enerji ihtiyaçlarının ve tüketim kalıplarının farklılaşmasından ötürü araştırma, Bursa merkez ilçelerindeki merkez mahallelerde yaşayan haneleri kapsamaktadır.

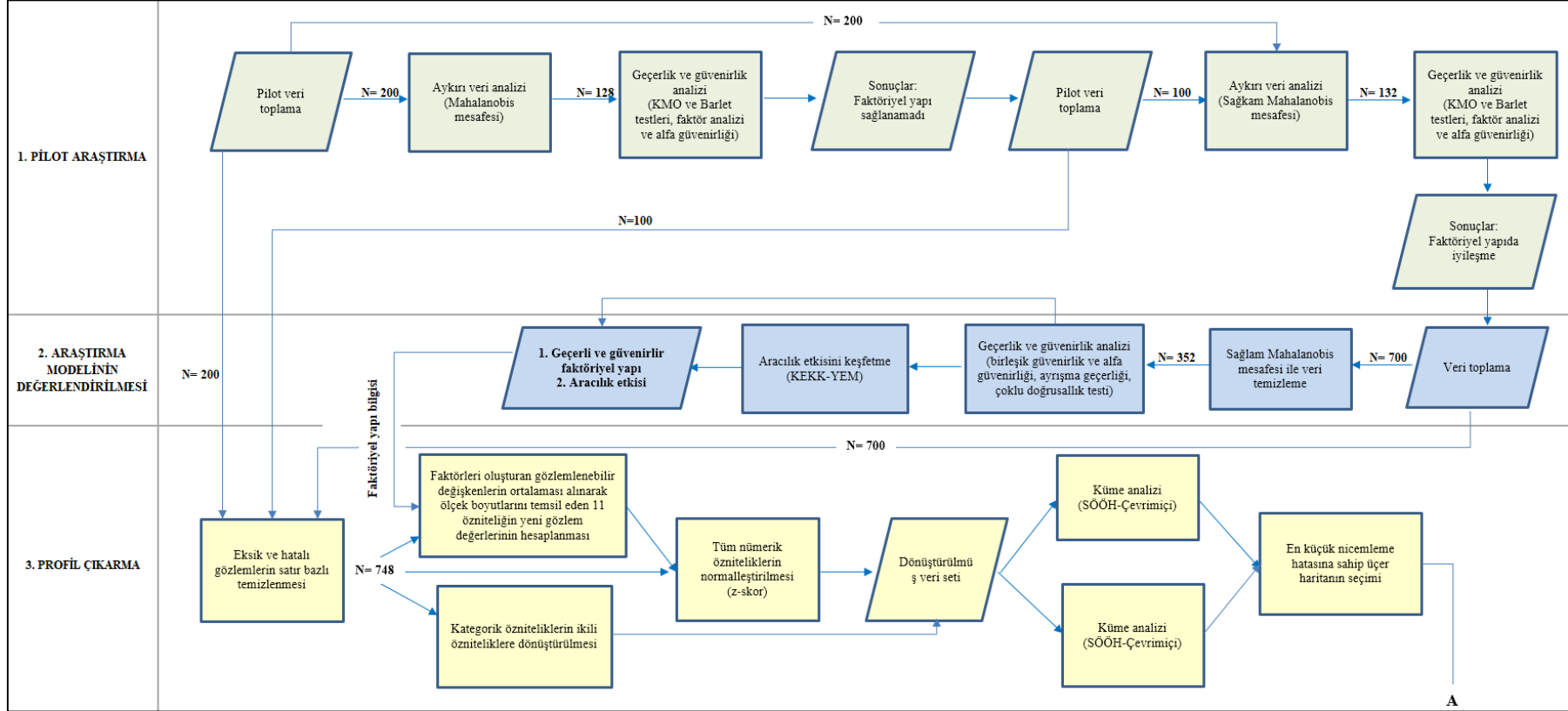
- Enerji ve su tüketim verileri sayaç verileri değildir. Tüketim özniteliklerinin değerleri, deneklerin beyan ettikleri fatura meblağlarıdır.

- Vatandaşların sürece dâhil edilmesi tek yönlü olarak ele alınmıştır. Vatandaşların perspektifinden kentin “yeşil” algısı, vatandaşların ihtiyaçları, önerileri ve onların davranışlarını olumlu yönde manipüle edecek uygulamaların (politika ve yatırımların) araştırılması söz konusu tez kapsamına dâhil edilmemiştir. Sayılan bu çalışmaların, gelecek yıllarda birer devam çalışması/araştırması şeklinde gerçekleştirilmesi planlanmaktadır.

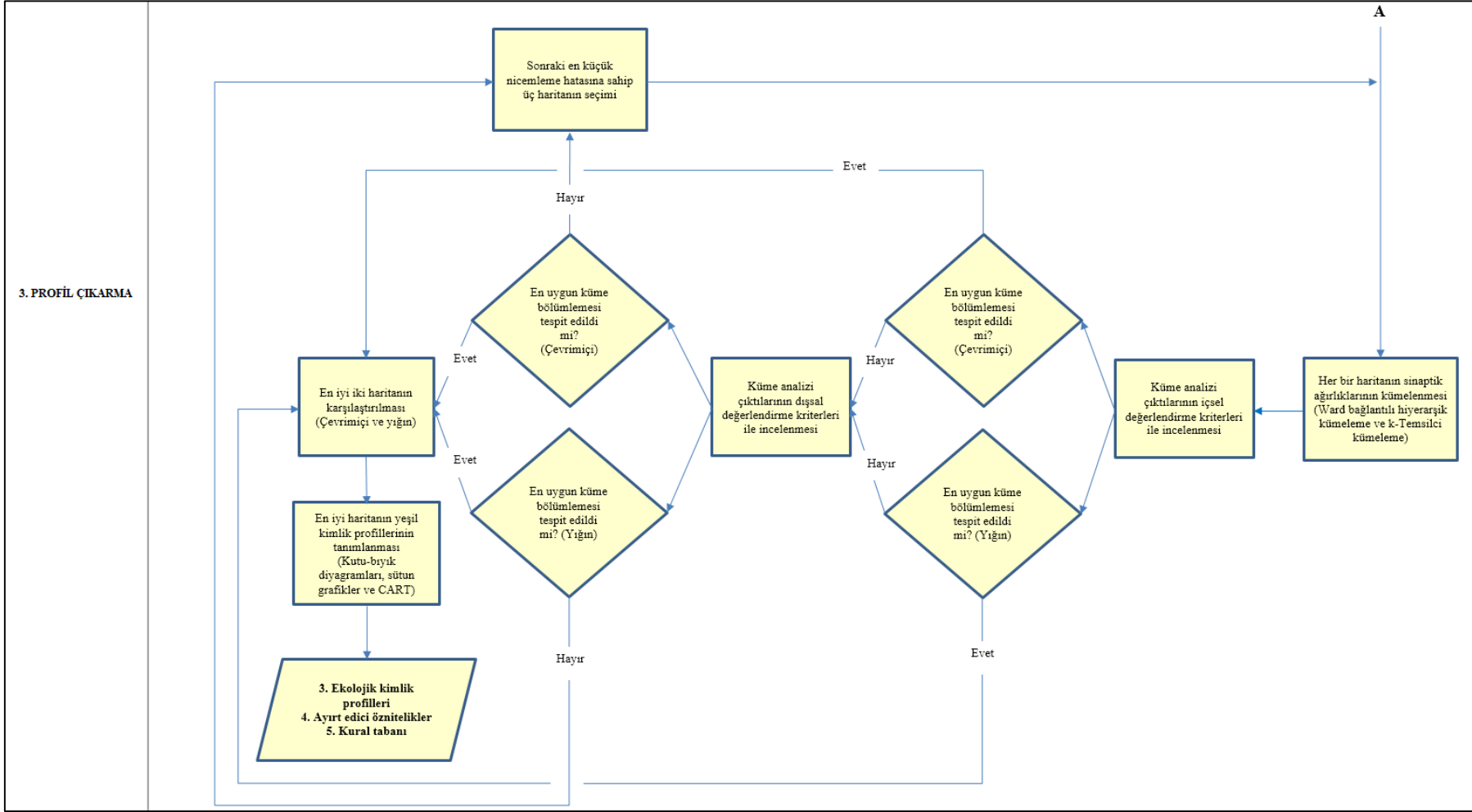
- ÇDD alan yazınında değerler, duygular, hisler, normlar, davranışsal kontrol, bilgi, önceki deneyimler, farkındalık-biliş vb. gibi çeşitli faktörlerin de davranışı etkilediği belirtilmektedir. Bu çalışmada sadece değerlere bağlı çevresel tutumlar ve bireyin bilişsel, duyuşsal (duygu) ve deneyimsel olarak doğayla ilişkisi dikkate alınmıştır.

## **1. ARAŞTIRMA TASARIMI VE SÜRECİ**

Çalışmanın araştırma ve analiz süreci toplam üç ana aşama olmak üzere dokuz alt aşamadan oluşmaktadır, Şekil 28. Aşamalar oluşturulmadan önce, bir dizi alt yapı faaliyeti gerçekleştirilmiştir. Öncelikle kavramsal çerçeve ile kullanılması planlanan ölçeklere karar verilip, ölçeği geliştiren ve gerekli ise uyarlayan bilim insanları ile irtibata geçilerek ölçeklerin kullanım izinleri alınmıştır. Akabinde Bursa Uludağ Üniversitesi Rektörlüğüne bağlı Sosyal ve Beşeri Bilimler Araştırma ve Yayın Etik Kurulu’ndan deneklere uygulanacak ölçeklerin uygunluğu için onay ve izin alınmıştır. Tez kapsamındaki araştırma Bursa’da gerçekleştirildiği, konusu “Yeşil Kent” markalaşması olduğu ve Bursa Büyükşehir Belediyesi (BBB) daha önce AYB ödülüne başvurmuş



A



Şekil 28. Araştırma süreci

olduğu için BBB Çevre Koruma ve Kontrol Dairesi Başkanlığı ile irtibata geçilmiş ve BBB'nin AYB için hazırlamış olduğu en son dosya temin edilmiştir.

**1. Ana Aşama (Pilot Araştırma):** Tez çalışmasında kullanılan ölçeklerden ÇEÖ ve DİÖ, Türkçe alan yazınına kazandırılmış ölçeklerdendir. Fakat bu ölçeklerin geçerlik ve güvenilirlikleri ilköğretim ve üniversite öğrencilerinden oluşan örneklemeler kullanılarak test edilmiştir. Bu çalışmanın ana kütlesi ise Bursa halkı kent merkezi sakinleridir ve belirli bir grup öğrenci kitlesinden tutum, çevresel farkındalık ve davranışsal açıdan farklılıklara sahip olacağı aşikârdır.

Diğer taraftan, STD ölçeği ise Türkçe oluşturulmuş, uygulama için İngilizceye çevrilerek redaksiyon yapılmış ve ABD'de hane hakları örnekleminde test edilmiştir. ABD ve Türkiye sosyoekonomik ve kültürel açıdan farklı yapıda iki ülkedir. Bu iki ülkenin aynı zamanda sürdürülebilirlik ve çevresel farkındalık açısından da farklı olacağı aşikârdır.

Yukarıda açıklanan farklılıklardan ötürü bu aşama sadece, sahadan veri toplamak için kullanılan ölçeklerin ön testi için gerekli olan çalışmaları kapsamakta olup üç alt aşamaya sahiptir. Bu aşamanın çıktısı, geçerlik ve güvenilirliği sınanmış olan ölçme modelidir. Analizler, SPSS ve R Studio yazılımları ile gerçekleştirilmiştir.

**Pilot Veri Toplama:** Veri toplamada kullanılan ölçeklerden STD ölçeğinin Çevreye Duyarlı Gıda Tercihi boyutunda yer alan bazı ifadelerin<sup>139</sup> sürdürülebilir davranış bilinci ile ilişkilendirilememeye ihtimalinden dolayı bu alt aşamada, ölçeklerin kapsam geçerliği kontrol edilmiştir. Bu amaç ile ölçekler, Bursa Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi İşletme bölümü dördüncü sınıf öğrencilerine ve yaşları 50 ve üstü olan çalışan veya emekli kadın ve erkeklere pilot veri toplanmadan önce uygulanmıştır. Elde edilen geri dönütler doğrultusunda, beklenildiği üzere, söz konusu boyuttaki bazı ifadelerin Türkiye bağlamında problemlili bir yapıya sahip olduğu tespit edilmiştir.

Bu yüzden ifadeler, genişletilerek güncellenmiştir. Örneğin kırmızı et tüketimi ile ilgili ifade “Büyükbaş hayvanların yaydıkları sera gazları hava kirliliğinde yüksek derecede etkili olduğu için çok fazla kırmızı et tüketmemeye özen gösteririm” şeklinde

---

<sup>139</sup> Örneğin “Çok fazla kırmızı et tüketmemeye özen gösteririm” gibi...

geniřletilmiřtir. Gncellenmiř lekler, ana ktleyi temsil ettięi dřnlen rneklem hacmindeki deneęe yklenici firma tarafından uygulanmıřtır.

Yazınındaki pilot arařtırma uygulamaları incelendięinde kapsamlı bir analiz iin 100 – 200 deneklik bir rneklem uygun grldęi belirtilmektedir (Johanson, Brooks, 2010). Bu yzden pilot uygulama iin 200 deneklik rneklem ile alıřılmasına karar verilmiřtir.

**Veri niřleme:** Bu alt ařamada, aykırı veriler tespit edilerek, veri setinin kalitesi, kolay ve etkin bir řekilde iřlenebilirlięi arttırılmaya alıřılmıřtır (Kirchner, Zec, Delibařić, 2016).

**Analiz:** Bu alt ařamada, leklerin i tutarlılıęını lmek iin gvenirlik analizi ve boyutlarını incelemek iin aımlayıcı FA gerekleřtirilmiřtir.

**2.Ana Ařama (Arařtırma Modelinin Deęerlendirilmesi):** Bu ana ařama, arařtırma modelinin analizi iin gerekli verilerin sahadan toplanmasını, verilerin temizlenmesini ve modelin analiz edilmesini kapsamakta olup,  alt ařamadan oluřmaktadır. Bu ařamadaki analizler, R Studio ve SmartPLS yazılımını kullanarak gerekleřtirilmiřtir.

**Arařtırma Verisi Toplama:** Veri toplamak iin kullanılan soru formunda, Ek 2, enerji tketimi ve enerji tketim davranıřları gibi deęiřkenler yer almaktadır. Enerji tketimi alan yazınında, kent ve kırsal alanda bulunan hanelerin tketim kalıplarının farklı olduęu belirtilmektedir (Khanna, Guo, Zheng, 2016; Wiedenhofer, 2011). Aynı zamanda, kent ve kırsal alanda yařayan bireylerin evre hakkındaki endiřeleri de farklılařabilmektedir. Bu iki neden dolayısı ile kentin kırsal alanlarında yařayan vatandaşlar ana ktleye dhil edilmemiřtir. Ana ktledeki bir dięer kısıtlama ise kıř aylarında ısınma ihtiyacını doęalgazdan karřılamayan hanelere getirilmiřtir. Bařka bir ifade ile rneklem homojen olabilmesi adına arařtırmanın erevesi kıř aylarındaki ısınma ihtiyacını doęalgazdan karřılayan ve Bursa merkezinde yařayan haneler olarak belirlenmiřtir. Arařtırmanın rnekleme birimi konut olup, gzlem birimi her bir konutta yařayan 18 yař ve zeri bir yetiřkindir. Dolayısıyla yklenici firma Osmangazi, Yıldırım ve Nilfer ilelerinin merkez mahallelerinde yařayan yetiřkinlerden yz yze grřme yntemi ile veri toplamıřtır. Ayrıca toplanan verilerin miktar aısından ileler arası denk olmasına dikkat edilmeye alıřılmıřtır.

Örneklem hacmi  $n = \frac{(Z_{\alpha/2})^2 p(1-p)}{E^2}$  formülasyonu esas alınarak hesaplanmıştır.

Popülasyondan çekilen örneklemin normal dağıldığı varsayıldığında ve  $p = 0.5$  iken, 0.95 olasılıkla ( $Z_{\alpha/2} = 1.96$ ), örnekleme hatası ( $E$ ) 0.05 ve altında istenirse örneklem hacminin en az (her haneden bir birey olmak üzere) 384 hane olması gerekmektedir. Fakat ilgili yazınında örneklem büyüklüğüyle ilgili farklı yaklaşımlar da bulunmaktadır. Örneğin 20 maddelik bir ölçek için 100 denek az, 90 maddelik bir ölçek için ise 400 denek yeterli olabilmektedir (DeVellis, 2014). Genel olarak ise bir araştırma için 300 denek iyi ve 500 denek ise çok iyi olarak nitelendirilmektedir. İlgili yazınında aşağıdan-yukarı yaklaşımla gerçekleştirilen çalışmalar da incelendiğinde 859 kişi (Çakır, vd., 2015), 924 hane (Huebner, vd., 2015), 1538 kişi (Dean, vd., 2018), 1550 hane (Kipping, Trømborg, 2016) ve 8000 hane (Räsänen, Ruuskanen, Kolehmainen, 2008) şeklindeki örneklem hacimlerine de rastlanmıştır. Çalışmanın örneklem hacmi, veri toplama maliyetini de dikkate alarak, pilot hariç 800<sup>140</sup> kişi olarak belirlenmiştir.

**Veri Önışleme:** Veri setindeki eksik veri içeren gözlemler ve SMM ölçütü kullanılarak aykırı veriler tespit edilip, temizlenmiştir.

**Yapısal Analiz:** Bu alt aşamada, gözlenen (manifest) değişkenler kullanarak, ölçeklerin gizil (latent) değişkenleri ölçme gücünü değerlendirmek amacıyla Doğrulayıcı FA gerçekleştirilmiştir. Amaç, çevresel endişe (tutumlar) ile STD arasındaki ilişkide İnsan-Doğa İlişkisinin aracı etkisini keşfetmektir. Bu yüzden, Kısmi En Küçük Kareler esaslı Yapısal Eşitlik Modellemesi (KEKK-YEM) yöntemi kullanılmıştır (Henseler, Hubona, Ray, 2016). Yapısal analizde pilot araştırma verisi analiz dışında tutulmuştur.

**3.Ana Aşama (Profil Çıkarma):** Bu ana aşamada, araştırmaya katılan kent sakinlerinin, çevresel tutumları, doğayla ilişkileri, STD'leri, enerji tüketimleri ve bazı sosyo-demografik özellikleri dikkate alınarak küme analizi yapılmıştır. Küme analizindeki amaç, bir hipotezi test etmekten ziyade, mevcut veri setini kullanarak gözlemler ya da öznitelikler arasındaki bağlantıları ve gizli bilgiyi keşfetmektir. Bu nedenle, bu aşamada 1000 gözlemlik veri setinin tamamı kullanılarak, sayılan öznitelikler bağlamında vatandaş tipolojileri keşfedilmeye çalışılmıştır. Süreç, üç alt

---

<sup>140</sup> Araştırmada kullanılan verilere ait demografik bilgilere Ek 3'ten ulaşılabilir.



aşamadan oluşmaktadır. Gerekli tüm analizler, R Studio ve PRIMER-e yazılımları ile gerçekleştirilmiştir.

**Veri Önleme:** Bu alt aşamada, veri indirgeme ve veri dönüştürme işlemleri ile veri seti küme analizine hazır hale getirilmiştir.

**Küme Analizi:** Kümeleme ile ilgili yazınında çeşitli yöntemler bulunmaktadır. Çalışmanın veri seti kategorik, ikili ve nümerik türdeki verilerden oluştuğu için karma niteliktedir. Bu yüzden, küme analizi, Kohonen Ağlarının özel bir türü olan ve karma nitelikteki veri setlerinin kümelenmesine olanak tanıyan Wehrens ve Kruisselbrink tarafından önerilen SÖÖH kullanılarak gerçekleştirilmiştir. Veri önleme alt aşamasında oluşturulan ve ölçek boyutlarını temsil eden yeni öznitelikler, su ve enerji tüketimleri ve sosyo-demografik özellikler çeşitli ebatlarda oluşturulan üç katmanlı SÖÖH’lerde birer öznitelik olarak kullanılmıştır. Vatandaşlar ise kümelenecek nesnelere olmuştur. Veri setinin SÖÖH’ler ile kümeleme sonuçları ısı haritaları ile görselleştirilmiş ve küme sayıları sezgisel olarak tespit edilmiştir. Akabinde küme analizi sonuçlarından nöron ağırlıkları da ayrı bir küme analizine tabi tutularak, SÖÖH’ler ile sezgisel olarak tahmin edilen küme sayısı netleştirilmiştir. Bu kapsamda nöron ağırlıkları Ward mesafeli hiyerarşik küme analizi ve ayrıca k-temsilli küme analizi yöntemleri ile kümeleneştir.

Uygun küme sayısının tespitinde ( $k - KT$ ) grafiği, Silüet, Davies-Bouldin, Dunn ve Calinski-Harabasz endeksleri içsel değerlendirme kriterleri olarak kullanılmıştır. Akabinde, elde edilen küme bölümlenmeleri SÖÖH haritalarına yansıtılarak küme bütünlükleri kontrol edilmiştir. Bu aşamada küme bütünlüğü sağlamayanlar elenmiş, sağlayanlara ise TBA uygulanmıştır. Bu işlem iki amaç ile yapılmıştır. İlk amaç, söz konusu harita için birinci ve ikinci temel bileşenlerin varyansın ne kadarlık kısmını açıklayabildiği bilgisine ulaşarak bu bilgi ışığında haritaları karşılaştırmaktır. İkinci amaç ise gözlemlerin ve küme bölümlenmelerinin saçılımını inceleyerek küme bütünlüklerini tekrar kontrol etmektir.

Son olarak aday küme bölümlenmeleri dışsal değerlendirme amacı ile parametrik olmayan çok değişkenli varyans analizine tabi tutulmuştur. Küme analizi alt aşamasının bitişi ile kimlik tanımlama aşamasında kullanılmak üzere, en uygun küme bölümlenmesini sağlayan harita seçilmiştir.

**Kimlik Tanımlama:** Bu alt aşamada küme analizi ile tespit edilen kümelerin her biri bir ekolojik kimlik profilini temsil etmek üzere öncelikle kutu-bıyık diyagramları ve sütun grafikler ile incelenmiştir. Ardından, bir karar ağacı algoritması olan CART algoritması sınıflandırma amaçlı olarak kullanılmış ve hem kümeleri birbirinden ayırt eden en önemli öznitelikler elde edilmiş hem de her kümeye ait bir kural tabanı oluşturulmuştur.

## 2. TEZ ÇALIŞMASININ UYGULAMA AŞAMASINDA KULLANILAN DİĞER YÖNTEMLER

### 2.1. Güvenirlik Analizi

Güvenirlik analizi, ölçülmek istenen bir boyutun o boyut altında yer alan maddeler tarafından ne derecede ölçüldüğü ile ilgilidir. Bu tez çalışmasında “aynı zamanda verilen cevaplar arasında tutarlılık” kapsamında gerçekleştirilen Cronbach Alfa ( $\alpha$ ) güvenilirliği incelenmiştir (Büyüköztürk, 2005).

Cronbach Alfa bir ölçekte yer alan maddeler arası korelasyona bağlı uyum değeri olarak tanımlanmaktadır (Durmuş, Yurtkoru, Çinko, 2016). Ölçekte yer alan maddelerin ölçtüğü özelliklerin (örneklediği davranışların) benzeşmesi ölçeğin güvenilirliğini yükseltmektedir (Büyüköztürk, 2005). İlgili yazınında Alfa katsayısı için 0.70 değeri alt eşik (asgari) değer olarak tanımlanmaktadır. Keza 0.70-0.80 aralığındaki alfa değerleri kayda değer, 0.80-0.90 aralığındaki alfa değerleri ise çok iyi olarak kabul edilmektedir. Diğer yandan alfa değerinin 0.90 üzerinde olması pek arzu edilen bir durum değildir. Cronbach alfa ölçekte yer alan maddeler arasındaki kovaryansın büyüklüğü ve madde sayısı ile ilgilidir (DeVellis, 2014). Eğer ölçekte birbirine benzer madde sayısı çok ise bu maddelerin birbirleri ile korelasyonları bire yakın olacak ve alfa değerini arttıracaktır. Bu durum deneklerin maddelere verdikleri cevaplarda varyansın gözlemlenememiş olması olarak açıklanmaktadır.

Cronbach alfa değeri ile ilgili bir diğer önemli nokta, bu değer ölçüğün uzunluğunun düzenlenmesinde kullanılabilmesidir. Ölçekte yer alan her hangi bir maddenin ölçekten çıkarılması durumunda alfa değerinin değişimi incelenerek ölçek uzunluğuna karar verilebilmektedir. Örneğin, ölçekte yer alan bir maddenin ortalama korelasyonu “average item correlation”, ölçekte yer alan diğer maddelerle olan ortalama

korelasyondan “inter-item correlation” daha düşük ise o maddenin ölçekten çıkartılması ölçeğin alfasını yükseltmektedir (DeVellis, 2014). Çünkü çıkarılan madde, ölçeğin güvenilirliğini (Cronbach alfa değerini) düşürmüştür. Aynı yaklaşım, faktörler için de geçerlidir. Eğer bir faktöre ait alfa değeri asgari değerin altında kaldı ise bu faktör tamamı ile ölçekten çıkartılabilmektedir (Durmuş, Yurtkoru, Çinko, 2016). Bu sebep ile Cronbach Alfa güvenilirliği ölçeğin yapı geçerliliği sağlandıktan sonra test edilmelidir.

## **2.2. KEKK-YEM Yöntemi**

KEKK-YEM yöntemi, gizil değişkenler tarafından temsil edilen teorik yapıların ölçümünde kullanılan gelişmiş bir doğrusal regresyon analizi olup, varyans temelli bir yöntemdir. Klasik regresyon analizinde, korelasyon matrisi dikkate alınarak, bağımlı değişkendeki değişimin ne kadarlık kısmının bağımsız değişken(ler) tarafından açıklandığı bilgisi elde edilmektedir. Sadece gözlenen değişkenlere uygulanabilmekte olup, doğrudan ilişkiler tespit edilebilmektedir (Civelek, 2018). YEM’de ise regresyondan farklı olarak kovaryans matrisi esas alınmakta ve birden fazla sayıda gözlenen ve gizil değişkenler arasındaki hem doğrudan hem de dolaylı ilişkiler incelenebilmektedir. YEM’in amacı, var olan teorik bir yapının veri ile uyumunu araştırmaktır. Bu uyumu araştırırken gözlemlenen ve tahmin edilen kovaryans matrisi arasındaki farkı minimize etmeye çalışır. Bu yüzden doğrulayıcı bir analiz yöntemi olarak kabul edilmektedir. YEM’de hem ölçüm modeline hem de yapısal modele ait regresyon denklemleri eş anlamlı olarak bir arada çalışmaktadır (Astrachan, Patel, Wanzenried, 2014; Joe F. Hair Jr, vd., 2014). Bu yüzden YEM’de regresyon analizinin tüm varsayımları geçerlidir (Civelek, 2018). Bunlar, bağımlı ve bağımsız değişkenler arasındaki ilişkinin doğrusal olması, hata terimlerinin normal dağılması, bağımsız değişkenler arasında çoklu doğrusal ilişki olmaması, bağımsız değişkenler ile hata terimleri arasında ilişki olmaması ve hata terimleri arasında otokorelasyon olmamasıdır. YEM’in kendi varsayımları ve koşulları ise gözlenen ve gizil değişkenlerin çok değişkenli normal dağılıma sahip olması, gizil değişkenlerin kendi arasında ve gizil-gözlenen değişkenler arasında doğrusal ilişki olması, veri setinin aykırı değer içermemesi, gözlem sayısının gözlemlenen değişken sayısının en az 5 katı aralığında olması ve bir gizil değişkenin en az üç gözlenen değişken ile ölçülmesidir (Astrachan, Patel, Wanzenried, 2014; Civelek, 2018).

KEKK-YEM ise keşfedici bir analiz yöntemidir. Varyans temellidir. Klasik YEM’de yapı-veri arasındaki teorik uyum, kovaryans matrisleri arasındaki fark ile ölçülürken, KEKK-YEM’de açıklanan varyans ile ölçülmektedir. Bu yüzden amaç, açıklanamayan varyansı minimize etmek veya bağımlı değişkenlerin açıklanan varyansını maksimize etmektir (Amaro, Saeber, Abrantes, 2015; Joe F. Hair Jr, vd., 2014). Örneklem küçük olduğu, verinin normal dağılmadığı, bir gizil değişkenin iki veya bir gözlenen değişken ile ölçüldüğü, gözlem sayısının değişken sayısında az olduğu durumlarda, eksik veri içeren gözlemlere sahip olan veri setlerinde, formatif yapılar içeren modellerde ve ayrıca yapılar arası ilişkilerin varlığının, yönünün ve büyüklüğünün keşfedilmesi amaçlandığı zaman kovaryans temelli YEM’den ziyade KEKK-YEM’in kullanılması tavsiye edilmektedir (Astrachan, Patel, Wanzenried, 2014; Joe F. Hair Jr, vd., 2014; Rigdon, Sarstedt, Ringle, 2017).

### 2.3. Hiyerarşik Küme Analizi

Hiyerarşik küme analizinde, iki temel yaklaşım bulunmaktadır. Birleştirici yaklaşımda (Altunkaynak, 2017; Akpınar, 2018), veri setindeki  $n$  tane gözlem, her birisi birer kümeyi temsil etmek üzere kümelenir. Daha sonra, kümeler arası uzaklıklar hesaplanır ve en yakın iki küme birleştirilir. Bu süreç, tüm gözlemler bir kümede toplanana kadar devam eder. Bölümleyici yaklaşımda ise (Altunkaynak, 2017; Akpınar, 2018) veri setindeki  $n$  tane gözlemin hepsi bir kümeyi oluşturur. En geniş küme seçilir.<sup>141</sup> Fakat ilk iterasyonda tek küme kendisi olduğu için kendisi en geniş küme olarak seçilir. Gözlemlerin her birinin, diğer gözlemlere olan uzaklıklarının ortalaması hesaplanır. En büyük ortalamaya sahip olan gözlem kümeden ayrılır ve yeni bir küme oluşturur. Eski kümede kalan gözlemlerin her birisi için hem kendi kümesindeki gözlemlere olan uzaklıkların ortalaması hem de diğer küme(ler)deki gözlem(ler)e olan uzaklıkların ortalaması hesaplanır. Eski kümedeki her bir gözlemin kendi kümesindekilere olan uzaklık ortalamasından, diğer küme(ler)deki gözlemlere olan uzaklık ortalaması çıkartılır. Eski kümedeki gözlemler için hesaplanan fark değerlerinden en büyük pozitif

---

<sup>141</sup> Genişlik kavramı Macnaughton, Williams, Dale ve Mockett (1965) çalışmasındaki algoritmada gözlem sayısını ve Kaufman, Rousseeuw (2005) çalışmasında kümenin çapını temsil etmektedir. Friedman, Hastie, Tibshirani (2008) ise genişlik olarak, küme içi benzeşmezliklerin ortalama değerinin kullanılabileceğini önermiştir.

değere sahip olan gözlem eski kümeden ayrılarak, yeni kümeye atanır. Bu süreç, en büyük kümede pozitif fark değeri kalmayana kadar tekrarlanır. Bir gözlem için hesaplanan farkın pozitif olması, o gözlemin kendi kümesindeki gözlemlere uzak, diğer kümedeki gözlemlere daha yakın olduğu anlamına gelmektedir. Alan yazınında, bölümleyici yaklaşım, birleştirici yaklaşıma kıyas ile çok daha seyrek kullanılmaktadır.<sup>142</sup> Diğer taraftan küme analizindeki amaç, az küme sayısı elde etmek ise bölümleyici yaklaşımın birleştirici yaklaşımdan daha başarılı olduğu ifade edilmektedir (Friedman, Hastie, Tibshirani, 2008).

Hiyerarşik kümelemede her iki yaklaşımda da küme sayısı önceden belli değildir ve küme yapıları bir ağaç şeklinde dendrogramlar ile görselleştirilerek, küme sayısı, tespit edilir.

Birleştirici hiyerarşik küme analizinde bölünmeye tabi tutulan kümedeki gözlemlerin diğer kümelere olan uzaklığının hesaplanması, küme analizi sonuçlarının kalitesini etkilemektedir. Örneğin, bölünen  $A$  kümesindeki bir gözlemin, küçük olan bir  $B$  kümesindeki tüm gözlemlere olan uzaklığının ortalaması mı dikkate alınmalıdır; yoksa  $B$  kümesinden kümeyi temsil eden bir ya da daha fazla gözlem tespit edilip, temsilci gözlem(ler)e olan uzaklığı mı dikkate alınmalıdır? Yazında, bu soruya cevap olarak bazı yöntemler geliştirilmiştir.

Bu yöntemlerden ilki, en yakın komşu/tekli bağlantı yöntemidir “nearest neighbour/single linkage method”. En yakın komşu yöntemi en eski ve en kolay bağlantı yöntemidir. Bu yöntemde,  $A$ 'da yer alan bir  $x_A$  gözleminin  $B$  kümesine uzaklığı,  $B$  kümesindeki gözlemlerden  $x_A$ 'ya en yakın olan kadardır. Bu yöntem, aykırı ve gürültülü veriye duyarlıdır (Akpınar, 2018; Han, Kamber, Pei, 2012; Tan, Steinbach, Kumar, 2014). Birbirinden çok farklı ve dairesel dağılan iki kümeyi bile tek bir küme olarak birleştirebilmektedir. Bu sebeple, lineer bir çizgi veya uzun eliptik geometrilere az

---

<sup>142</sup> Bunun nedeni, veri setindeki herhangi iki gözlemin birleştirilebilmesi için denenmesi gerekli kombinasyon sayısı karesel bir fonksiyon iken bölümleyici yaklaşımda, verilerin tüm boş olmayan iki alt kümeye bölünmesinin bir üssel fonksiyon olmasıdır. Fakat bu kısımda açıklanan bölümleyici algoritma tüm bölümlenmeleri dikkate almamaktadır. Bu sebeple, birleştirici ve bölümleyici küme analizi yaklaşımları ile elde edilen dendrogramlar birbirinden farklıdır.

sayıda küme oluşturma ve benzerlikleri düşük olan gözlemleri bir kümeye atama eğilimi bulunmaktadır. Kullanımı pek önerilmemektedir (Kaufman, Rousseeuw, 2005).

En uzak komşu/tam bağlantı yönteminde “furthest neighbour/complete linkage method”  $x_A$  gözleminin  $B$  kümesine uzaklığı,  $B$  kümesindeki gözlemlerden  $x_A$ 'ya en uzak olan kadardır. Aykırı ve gürültülü veriye duyarlıdır (Akpınar, 2018; Han, Kamber, Pei, 2012; Tan, Steinbach, Kumar, 2014). İterasyon sayısı arttıkça küme içi benzeşmezlikleri az olan çok sayıda küme oluşturma eğilimindedir. Bu sebep ile benzerlikleri yüksek olan gözlemler farklı kümelere atanabilmektedir (Kaufman, Rousseeuw, 2005). Dairesel dağılan veri setleri için uygundur (Han, Kamber, Pei, 2012; Tan, Steinbach, Kumar, 2014). Eğer, veri setindeki kümelerin küme için varyanslarının küçük ve kümelerdeki gözlem sayısının birbirine yakın olduğu biliniyor ise bu yöntemin yüksek kalitede sonuçlar sağladığı belirtilmektedir (Han, Kamber, Pei, 2012).

En yakın komşu yönteminin benzerliği düşük olan gözlemleri birleştirerek az sayıda küme oluşturma ve en uzak komşu yönteminin benzer gözlemleri farklı kümelere atayarak küme sayısını gereksiz arttırma eğilimi problemine bir çözüm olarak gruplar arası bağlantı /ortalama bağlantı “between-groups linkage/average linkage” yöntemi geliştirilmiştir (Kaufman, Rousseeuw, 2005). Bu yöntemde,  $x_A$  gözleminin  $B$  kümesine uzaklığı,  $x_A$ 'nın  $B$  kümesindeki tüm gözlemlere olan uzaklığının ortalaması hesaplanarak tespit edilir. Tekli ve tam bağlantı yöntemlerine kıyas ile daha sağlıklı sonuçlar sağlamaktadır ve aykırı veri duyarlılığı yoktur. Hem kategorik hem de nümerik verilerin işlenmesine imkan tanımaktadır (Han, Kamber, Pei, 2012).

Merkezi bağlantı “centroid linkage” yönteminde öncelikle hem  $A$ 'nı hem de  $B$ 'nin ağırlık noktası hesaplanır. Akabinde  $x_A$  gözleminin  $B$ 'ye uzaklığı,  $B$ 'nin merkezi referans alınarak tespit edilir. Tekli ve tam bağlantı yöntemlerine kıyas ile daha sağlıklı sonuçlar sağlamaktadır ve aykırı veri duyarlılığı yoktur (Han, Kamber, Pei, 2012). Bu yöntem, küçük bir küme, kendisinden daha büyük bir küme ile birleştirildiği zaman elde edilen yeni kümenin ağırlık noktasının, büyük olan kümenin ağırlık noktasına daha yakın olmasına neden olmaktadır (Akpınar, 2018). Ayrıca, diğer tüm bağlantı yöntemlerinde kümeler arası mesafe her iterasyonda artarken, sadece merkezi bağlantı yönteminde tam tersi durum söz konusu olabilmektedir (Tan, Steinbach, Kumar, 2014).

Medyan bağlantı “median linkage” yönteminde küme orta noktaları, kümenin merkezi olarak ele alınır ve uzaklık hesaplaması söz konusu merkezler dikkate alınarak gerçekleştirilir.

Sayılan bu yöntemlere ek olarak ve onlardan çok farklı olarak Ward bağlantı “Ward’s linkage” yöntemi de mevcuttur. Bu yöntemde, küme varyansları dikkate alınır. İki küme birleştirildiği zaman hata kareler toplamındaki artış hesaplanır. Amaç bu artışı minimize edecek bağlantıları seçerek kümeleme işlemini gerçekleştirmektir. Keza bu yöntemin de ortalama bağlantı yöntemi gibi diğer yöntemlere kıyas ile daha sağlıklı sonuçlar sağladığı belirtilmektedir. Gözlemler arası mesafe öklit uzaklığı ile hesaplandığı zaman ortalama bağlantı yöntemi ile benzer sonuçlar sağlamaktadır (Tan, Steinbach, Kumar, 2014). Bu tez çalışmasında Ward bağlantılı birleştirici hiyerarşik küme analizi kullanılmıştır.

#### **2.4. k-Temsilci Küme Analizi**

k-Temsilci küme analizi yönteminde, her küme medoid adı verilen bir gözlem ile temsil edilir. Algoritma, küme sayısı  $k$ 'nın önsel bilgi olarak atanması ile başlar. Daha sonra  $n$  tane gözlem içeren veri setinden  $k$  tane gözlem rassal bir şekilde temsilci olarak seçilir ve geriye kalan  $n - k$  tane gözlem ile temsilci olarak seçilen  $k$  tane gözlem arasındaki uzaklıklar hesaplanır. Akabinde,  $n - k$  tane gözlem, kendisine en yakın olan  $k$ 'ncü gözlemin temsil ettiği kümeye atanır. Atama işlemi sonucunda küme içi benzeşmezlik (KT) hesaplanır. Sonrasında,  $k$  tane kümenin her birinde yeni bir temsilci rassal olarak seçilir ve ikinci atama gerçekleştirilir. İkinci atama işlemi için de veri setindeki tüm gözlemlerin yeni temsilcilere olan uzaklıkları kullanılarak küme içi benzeşmezlikler hesaplanır. Her iki atamadaki toplam küme içi benzeşmezlikler dikkate alınarak, birinci ve ikinci atamalardaki temsilciler için yer değiştirme maliyeti “total swapping cost” incelenir. İkinci atamanın toplam küme içi benzeşmezliği, birinci atamanın toplam küme içi benzeşmezliğinden küçük ise temsilciler yer değiştirir ve yer değiştirme durana kadar bu süreç tekrar eder. Bu yöntemin amacı, temsilci ile temsil ettiği kümedeki diğer gözlemler arasındaki mesafeyi asgari düzeyde tutmaktır.

k-Temsilci yöntemi, özellikle küçük veri setlerinde aykırı verilerden etkilenmemektedir. Fakat veri setinin büyüklüğü arttıkça, aykırı gözlemler etkilerini

yitireceği için bu yöntemin sağlıklı sonuç veremeyebileceği ifade edilmektedir (Altunkaynak, 2017; Akpınar, 2018). Ayrıca, küme sayısı arttıkça toplam küme içi benzeşmezlik değeri azalmaktadır. Bu yüzden en uygun  $k$  sayısının 3. Bölüm Kısım 1.3.1’de açıklanan içsel değerlendirme kriterleri kullanılarak (Kaufman, Rousseeuw, 2005), başka yöntemler ile çapraz doğrulanarak ya da 3.Bölüm Kısım 1.2’de açıklanan  $(k - KT)$  grafiği kullanılarak tespit edilmesi önerilmektedir. Örneğin veri setindeki gerçek küme adedi  $k^*$  için  $k < k^*$  olmak üzere tüm  $k$ ’lar için hesaplanan küme içi benzeşmezlik değerlerinin toplamı,  $k = k^*$  olana kadar belirgin bir şekilde azalacaktır. Bunun nedeni,  $k < k^*$  için elde edilen kümelerin  $k^*$ ’ı oluşturan kümelerin birkaçının birleşiminden meydana gelmesi ve  $k^*$ ’a kıyas ile daha heterojen nitelikte olmasıdır. Diğer taraftan  $k > k^*$  olmak üzere bütün  $k$ ’lar için hesaplanan küme içi benzeşmezlik değerlerinin toplamındaki azalış ise belirgin miktarda olmayacaktır. Çünkü elde edilecek olan kümelerin her birisi aslında orijinal kümelerin birer alt kümesidir (Friedman, Hastie, Tibshirani, 2008).

Çapraz doğrulama, içsel değerlendirme kriterleri ve  $(k - KT)$  grafiği dışında  $k$  sayısının veya başka bir ifade ile kaç temsilcinin ve hangi gözlemlerin temsilci olacağıın tespiti için, örnekleme (gözlem sayısının birkaç bin ve üzeri gibi büyük veri setlerinde),<sup>143</sup> dal-sınır (gözlem sayısının en fazla 50-60 olması durumunda) ve tam sayılı programlama yöntemleri ve ayrıca veri setini temsil eden her bir öznelik için ortalama değer hesabının kullanılması gibi çeşitli yaklaşımlar bulunmaktadır. Kaufman, Rousseeuw (2005) çalışmasından söz konusu yaklaşımlar hakkında bilgi edinilebilir.

## 2.5. Permütasyonel Çok Değişkenli Varyans Analizi

Ekoloji araştırmalarındaki verinin çok değişkenli normal dağılım göstermemesi, değişkenlerin kesikli değerler alması ve örneğin doğadaki bazı türlerin ender olmasından ötürü veri setinin sıfır değerli çok sayıda gözlem içermesi ve uzaklık hesabında yarı-metrik ölçütlerin kullanılma gerekliliği, bilinen varyans analizlerinin sağlıklı bir şekilde kullanılamamasına neden olmaktadır. PERMANOVA “Permutational Multivariate

---

<sup>143</sup> CLARA “Clustering Large Applications” algoritması olarak adlandırılmaktadır. k-temsilci algoritmasını geliştiren ekip tarafından geliştirilmiştir (Kassambara, 2015).



Analysis of Variance”, bu ve benzer nitelikli verilerin analizi için geliştirilmiş bir yöntemdir (M. J. Anderson, 2001).

PERMANOVA, parametrik varyans analizinin normal dağılım<sup>144</sup> ve varyansların homojenliği varsayımlarını ve gözlem sayısının öznitelik sayısından fazla olma ve gruplar arası gözlem sayısının dengeli olması koşullarını gerektirmeyen çok yeni bir yöntemdir.<sup>145</sup> Yapay  $F$  “pseudo- $F$ ” istatistiğini kullanır.  $p$  değerini ise permütasyon testleri ile hesaplar. Bu yüzden, her hangi bir dağılım varsayımı gerektirmemektedir. Bağımlı değişkenler arasındaki korelasyona duyarlıdır. Ayrıca, Bray-Curtis ya da Ki-kare gibi yarı metrik benzerlik/benzeşmezlik ölçütleri PERMANOVA ile varyans analizinde kullanılabilir (Pasin, vd., 2016; McArdle, Anderson, 2001; M. J. Anderson, vd., 2017).

PERMANOVA'nın benzerlik, değişebilirlik ve bağımsızlık olmak üzere üç varsayımı bulunmaktadır. Benzerlik “homogeneity of variances”, grup dağılımlarının benzer olmasıdır. Benzerlik varsayımının PERMANOVA için katı bir varsayım olmadığı ve fakat heterojen varyansa ait veri setlerinin söz konusu varyans farklılıklarından etkilenebileceği de ifade edilmektedir. Değişebilirlik “exchangable objects”, gözlemlerin yokluk hipotezi altında gruplar arası değiştirilebilir olması durumudur. Bağımsızlık “independence” ise veri setindeki gözlemlerin birbirinden bağımsız olmasıdır (M. J. Anderson, 2017).

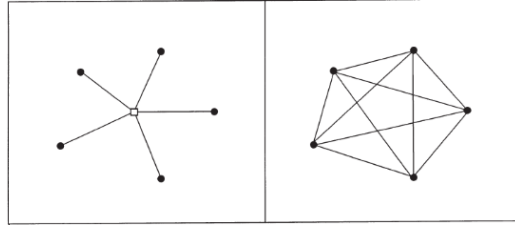
Varyans analizi,  $F$ - istatistiği oranını kullanarak, grup içi değişkenliğin gruplar arası değişkenlik ile karşılaştırılması esasına dayanmaktadır.  $F$  değeri büyüdükçe, grup ortalamaları arasında farklılık olmadığını var sayan yokluk hipotezinin reddi, olası muhtemeldir.  $F$ - istatistiğinde grup içi ve gruplar arası varyansın hesaplanabilmesi için hem grup merkezlerinin hem de veri setinin genel merkezinin hesaplanabilir olması

---

<sup>144</sup> Mills (1997)'den aktaran Önder (2005)'e göre normal dağılıma sahip olma koşulu sağlanmadığı durumlarda ham veri ya da kalıntılar, logaritma, normalleştirme ya da standardizasyon gibi veri dönüştürme yöntemleri ile dönüştürülmektedir. Fakat normallik, örneğin logaritmik veri dönüştürme işleminden sonra elde edildiyse aritmetik ortalamaların eşitliği varsayımı ile oluşturulan yokluk hipotezi, geometrik ortalamaların eşitliği şeklinde ele alınmalıdır. Böyle bir durumda ise analiz sonuçlarının yorumlanması karmaşık hale gelebilmektedir.

<sup>145</sup> Genellikle ekoloji ve biyoloji araştırmalarında kullanılan PERMANOVA yeni geliştirilmiş bir yöntemdir. Nadir de olsa sağlık alanında (Pasin, vd., 2016) ve mühendislikte önleyici bakım araştırmalarında kullanılmaya başlanmıştır (Amruthnath, Gupta, 2019).

gerekmektedir. Merkez noktaların hesaplanabilmesi ise uzaklık hesabında öklit gibi metrik ölçütler kullanıldığı zaman mümkündür. Bray-Curtis mesafesi gibi yarı metrik ölçütler için veri setinin merkez noktalarının hesaplanması mümkün değildir. Bu problem,  $F$  istatistiği hesabında gözlemler arası benzerlik matrisinin kullanımı ile küme merkezleri hesaplanmasına gerek kalmadan giderilmiştir. Çünkü Şekil 29’da gösterildiği üzere, gözlemlerin grup merkezine uzaklıklarının KT, gözlemlerin kendi arasındaki uzaklıklarının KT'nin gözlem sayısına bölümüne eşittir. Bu mantık, yapay  $F$ -istatistiğinin temelini oluşturmaktadır. Bu sebeple grup merkezinin ve veri setinin genel merkezinin hesaplanmasına gerek yoktur ve yarı metrik ölçütler için bile bir  $F$  değeri hesaplanabilmektedir.



Şekil 29. Grup içi varyansın merkez nokta dikkate alınarak ve dikkate alınmayarak hesaplanabilirliğinin geometrik gösterimi (M. J. Anderson, 2001)

PERMANOVA’da kullanılan yapay  $F$  istatistiği ve  $p$  değeri ise  $SS_T$  veri setine ait genel KT,  $SS_W$  grup içi KT (kalıntılar toplamı),  $N$  veri setindeki toplam gözlem sayısı,  $n$  bir gruptaki gözlem sayısı,  $i = 1, \dots, N$  ve  $j = 1, \dots, N$  olmak üzere  $d_{ij}$   $i$ 'inci ve  $j$ 'inci gözlemler arası mesafe,  $F$  orijinal veri setinin  $F$ -oranı,  $F^\pi$  permütasyon sonucu elde edilen  $F$ -oranı,  $a$  grup sayısı ve  $p$ -değeri olmak üzere yapay  $F$  istatistiği ve karşılık geldiği  $p$  değeri Eş. (81, 82) ile temsil edilmektedir. Eş. (80)'de yer alan  $\epsilon_{ij}$  bir ikili değişken olup  $ij$  gözlem çiftinin aynı grupta olma durumunda 1, aksi durumda ise 0 değerini almaktadır.  $SS_A$  ise gruplar arası KT olup,  $SS_A = SS_T - SS_W$  eşitliği ile hesaplanmaktadır. Bu değerlerin matrisel gösterimi Şekil 30.'daki gibidir.

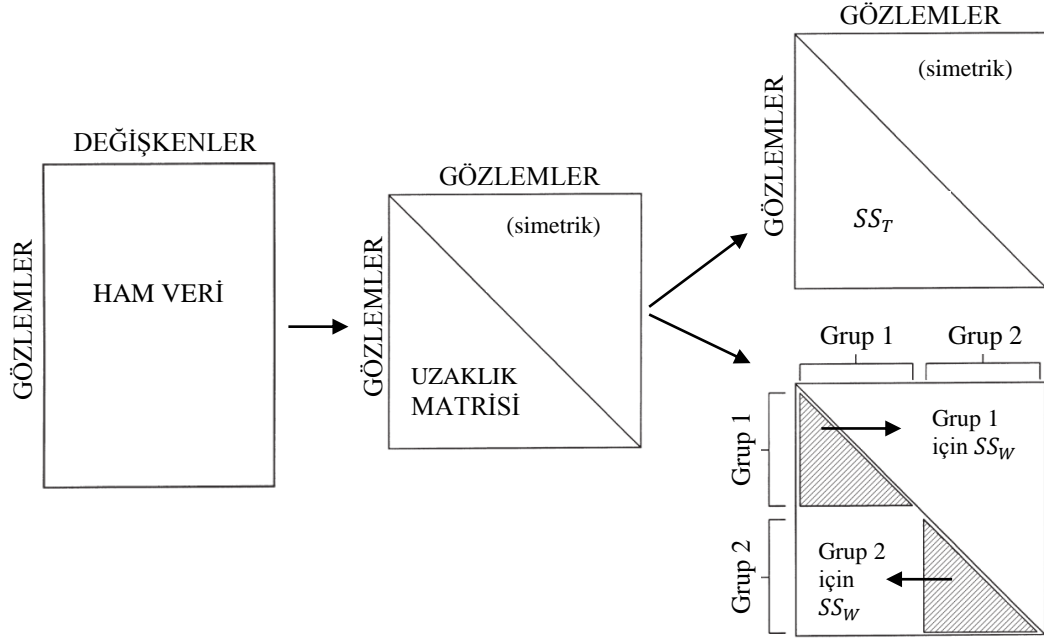
$$SS_T = \frac{1}{N} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2 \quad (79)$$

$$SS_W = \frac{1}{n} \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^2 \epsilon_{ij} \quad (80)$$

$$F = \frac{SS_A/(a-1)}{SS_W/(N-a)} \quad (81)$$

$$p = \frac{F^{\pi} \geq F \text{ koşulunu sađlayan permütasyon sayısı}}{\text{Toplam } F^{\pi} \text{ sayısı (ya da toplam permütasyon sayısı)}} \quad (82)$$

Eş. (82)'nin paydasına orijinal veri setini temsilen +1 eklenmemiştir. Çünkü orijinal veri seti de olası permütasyonlardan bir tanesi olarak değerlendirilmektedir.



Şekil 30. Grup içi, gruplar arası ve genel kareler toplamının şematik gösterimi

PERMANOVA analizinde ham verideki gözlemlerin permütasyonu ve indirgenmiş modelde veya tam modelde kalıntıların permütasyonu ile yapay- $F$  ve  $p$  değerleri hesaplanabilmektedir. Bu çalışmada, bağımsız değişkene ait gruplar az sayıda gözleme sahip olduğu için ve model tek yönlü olduğu için PERMANOVA, ham verinin permütasyonu kullanılarak gerçekleştirilmiştir. Ham verinin permütasyonunda öncelikle Eş. (79-81) kullanılarak veri setine ait bir  $F$  değeri hesaplanır. Akabinde, gözlemlerin grup etiketleri değiştirilir (permütasyon) ve bir  $F^{\pi}$  değeri hesaplanır. Bu işlem  $\alpha = 0.05$  için 1000 ve  $\alpha = 0.01$  için 5000 permütasyon için tekrarlanır (M. J. Anderson, 2001). Elde edilen tüm  $F^{\pi}$  değerleri,  $F$  değeri ile karşılaştırılarak, sonuçları Eş. (82)'de kullanılır ve  $p$  istatistiği hesaplanır.

İndirgenmiş ve tam modelde kalıntıların permütasyonu, kalıntı değerleri esas alınarak gerçekleştirildiği için bir ortalama hesabı yapılmasını gerektirmektedir. Bu neden ile, bağımsız değişkendeki gruplar, ortalama hesabının sağlıklı bir şekilde yapılmasına izin verecek sayıda gözleme sahip olduğu zaman, bu iki yaklaşımın kullanılması önerilmektedir. Kalıntıların permütasyonu yöntemlerinin kullanımı, aynı

zamanda çok faktörlü modellerde ve aykırı gözlem içeren veri setlerinde de tavsiye edilmektedir (M. Anderson, Gorley, Clarke, 2008). Kalıntıların permütasyonu yaklaşımları hakkında M. Anderson, Gorley, Clarke (2008) ve Good (2000) çalışmalarından bilgi edinilebilir.

Yazın ile paralel olarak Önder (2005) ve Önder, Cebeci (2009) çalışmalarında çoklu doğrusal regresyon analizi için bağımsız değişkenler arası yüksek korelasyon olduğu, veri setinde aykırı gözlem olduğu ve verilerin normal dağılıma sahip olmadığı durumlarda tam modelde kalıntıların permütasyonu, veri setinde aykırı gözlem olmadığı durumlarda ham verinin permütasyonu ve gruplardaki gözlem sayısının en az 10 olduğu durumda ise indirgenmiş modelde kalıntıların permütasyonu yaklaşımlarının kullanılması önerilmiştir.

## **2.6. Karar Ağaçları**

Karar ağaçları, çoğunlukla sınıflandırma ve aynı zamanda tahmin problemlerinde kullanılan klasik yöntemlerden bir tanesi olup, kökü yukarıda ve yaprakları aşağıda duran bir ağaç yapısına sahiptir. Tüm yaprakların ve kökten yaprağa giden yolda oluşturulan ara düğümlerin birer etiketi/sınıfı bulunmaktadır. Ayrıca, her bir ara düğüm bir girdiyi ifade etmekte ve gizli bilginin kökten yapraklara erişimini sağlamaktadır. Düğümler arası bağlantıyı sağlayan dallar ise bir işlemi temsil etmektedir.

Ağacın eğitilmesi sürecinde, veri seti kendisini tanımlayan özneliklerin çeşitli değerleri dikkate alınarak alt kümelere bölünür. Bu işlem özyinelemeli olarak tekrarlanır “recursive”. Tekrarlama, tahminde/sınıflandırmada bir iyileşme olmayana kadar veya iyileşme bir eşik değerinin altında kalana kadar devam eder. Bu süreç, özyinelemeli parçalama “recursive partitioning” olarak adlandırılmaktadır.

Sınıflandırma amaçlı ağaçlarda “Classification Trees” veri seti olabildiğince homojen olarak alt kümelere bölünerek sınıflandırılmaktadır. Tahmin amaçlı ağaçlarda ise “Regression Trees” tahmin modelleri oluşturulmaktadır. Her iki karar ağacı türünde de günümüze kadar çeşitli algoritmalar geliştirilmiştir. ID3, C4.5, CHAID, CART, SPRINT, QUEST, MARS, sıklıkla kullanılan algoritmalarındandır. Ayrıca Rassal Orman “Random Forest”, Hızlandırılmış Ağaçlar “Boosted Trees” ve Döndürme Ağaçları “Rotation Forest” gibi karar ağaçlarının performansını artırma amacı ile

Topluluk Öğrenme “Ensemble Learning” algoritmaları da bulunmaktadır. Bu tür algoritmalar hakkında Seni, Elder (2010), Zhou (2012) ve Cha Zhang, Ma (2012) çalışmalarından bilgi edinilebilir.

VM’de kullanılan diğer bir çok yöntem ile karşılaştırıldığında, karar ağaçlarının çeşitli avantajları bulunmaktadır (Şeker, 2013; Akpınar, 2018; Özkan, 2008; Köse, 2018; Boehmke, Greenwell, 2020). Öncelikle, karar ağaçlarının anlaşılması ve sonuçlarının yorumlanması diğer VM yöntemlerine kıyas ile çok daha kolaydır. Karar ağaçlarında normal dağılım koşulu bulunmamaktadır; eksik ya da aykırı gözlemlerden çok fazla etkilenmezler ve hem sürekli hem de kategorik verilerin işlenmesinde kullanılabilirler. Sayılan bu özelliklerinden ötürü, karar ağacı analizinde veri setinin ön işlenmesi veya dönüştürülmesi gerekmemektedir. Karar ağaçları, beyaz kutu modelini kullandıkları için eğitim sürecindeki her adım görüntülenip, yorumlanabilmektedir. Düşük hesaplama karmaşıklığına sahip oldukları için ise eğitim süreci basit ve hızlı olup yüksek miktardaki veriyi bile kısa sürede işleyebilmektedir.

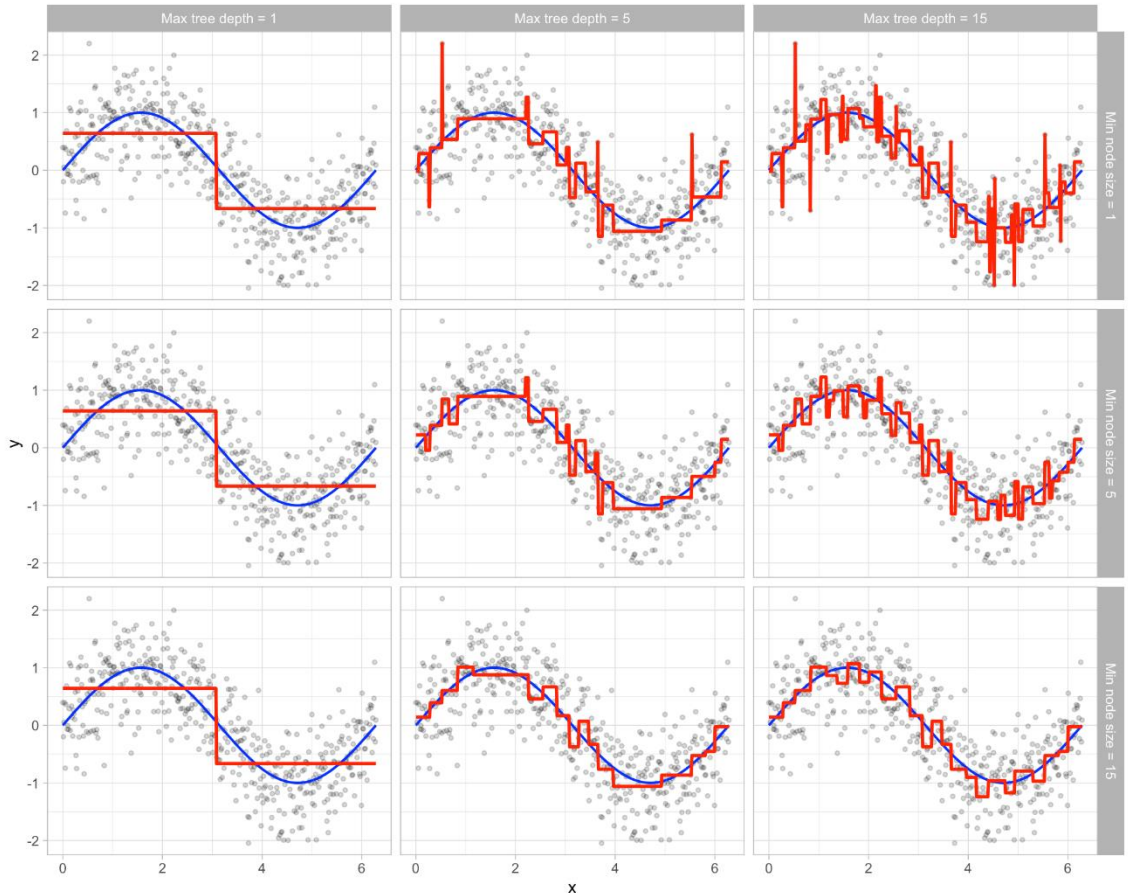
Karar ağaçlarının sayılan avantajlarına karşılık bazı dezavantajları da bulunmaktadır. Örneğin, karar ağaçları ile global optimum çözümün bulunması çoğu zaman imkansız olup, genellikle lokal optimum çözüm elde edilmektedir. Çünkü ağacın performansı, veri setini en iyi bölerek dallanmayı sağlayacak olan öznelik seçim kriterine “splitting criteria” bağlıdır ve bu alanda kullanılan seçim kriterleri sezgisel özelliğe sahiptir (Akpınar, 2018). Dolayısı ile karar ağaçlarının genelleştirme yeteneğinin düşük olduğu ifade edilebilir. Ayrıca, ağacın derinliği (kök ve yapraklar arasındaki maksimum mesafe) arttıkça veriyi öğrenmek yerine ezberleyebilmektedir. Fakat bu durumu engellemek için çeşitli kriterler dikkate alınarak budama işlemi yapılması önerilmektedir.

Bir karar ağacının büyümesi veya ara düğümlerin alt düğümler oluşturarak hiyerarşik bir şekilde bölünmesi, eğitimin durdurulması için kullanılan kriterlerden en az bir tanesi sağlanana kadar devam etmektedir. Durdurma kriteri sağlandığı zaman, söz konusu düğüm için dallanma sona ererek ağaç budanmaktadır. Durdurma kriterlerinden ilki, bir ara düğümdeki nesnelere hepsinin aynı kategoriden olması durumudur. Bu kriter sağlandığı zaman, söz konusu düğüm budanarak yaprak düğüm niteliğine kavuşmaktadır. Bu kriterin kullanımı tüm karar ağaçlarında geçerli olan standart bir uygulamadır. Diğer

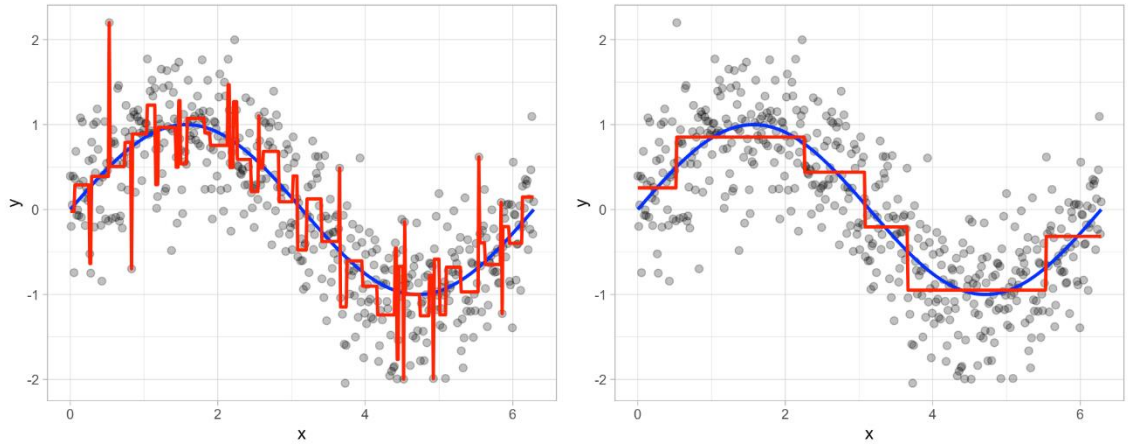
kriterlerin hepsi için ise birer eşik değer sezgisel olarak belirlenmektedir. Eşik değerler, veri analistinin öngörüsüne veya ele alınan problemin doğasına bağlı olarak tespit edilebilmektedir. Ağaç;

- Önsel olarak belirlenen maksimum derinliğe ulaştığı zaman,
- Bir ara düğümdeki gözlem sayısı çocuk düğümler oluşturabilmek için önsel olarak belirlenen minimum gözlem sayısından (*minsplit*) daha az olduğu zaman,
- Bir ara düğüm dallanarak yeni çocuk düğümler oluşturduğunda, çocuk düğümlerdeki gözlem sayısı önsel olarak belirlenen minimum gözlem sayısından (*minbucket*) daha az olduğu zaman veya
- Bir dallanma işleminde, öznelik seçim kriterinin değerindeki iyileşme, önsel olarak belirlenen bir minimum değerden (*cp*) daha küçük olduğu zaman eğitim sonlandırılarak dallanma durdurulabilir.

Yukarıda kısaca sayılan durdurma kriterlerinden *cp*, “complexity parameter”, karmaşıklık parametresidir. Bu parametre, iterasyonlar arası iyileşmenin belirli bir oranda olup olmadığını kontrol etmektedir. Örneğin,  $cp = 0.3$  olması durumunda regresyon ağaçlarında kullanılan  $R^2$  değeri ve sınıflandırma ağaçlarında kullanılan Gini, Entropi, Sınıflandırma Hatası veya Twoing endeks değerleri iterasyonlar arası en az 0.3 oranında iyileştiği sürece ağaç büyümeye devam edecektir. Dolayısı ile bu parametrenin yüksek tutulması, ağacın derinliğini azaltarak yaprak düğümlerin safsızlığını arttırmakta, düşük tutulması ise derinliği arttırarak safsızlığı azaltmaktadır. Ağacın artan derinliğine bağlı olarak ise aşırı uyum riski meydana geldiği için karar ağacı analizinde safsızlık ile ağaç derinliği arasında bir ödünleşim yapılması gerekmektedir. Örneğin, aşırı uyumu önlemek için düşük safsızlık sağlayan çok derin bir ağaç yerine, Şekil 31(a)’da yer alan örnekteki gibi safsızlığın daha fazla ve derinliğin daha az olduğu bir ağaç çözüm olarak kabul edilebilmektedir. Ayrıca, Şekil 31(b)’de yer alan sol taraftaki görselin *cp* değeri, sağ tarafta yer alan görselin *cp* değerinden çok daha küçüktür. Bu yüzden sol taraftaki görsele ait CART ağacı veriyi nerede ise ezberlemiş ve yeni gelen bir gözlemi doğru tahmin etme olasılığı, sağ tarafta yer alan ağaca kıyas ile düşüktür.



(a)



(b)

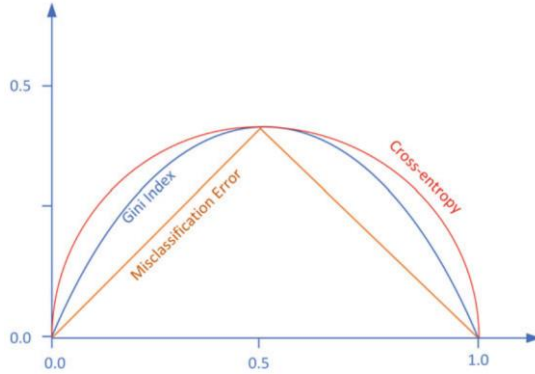
**Şekil 31.** Gauss gürültüsü eklenmiş sinüs fonksiyonunun CART algoritması ile tahmininde (a) ağaç derinliği (x-ekseni) ile minbucket (y-ekseni) miktarının ve (b) cp değerinin öğrenmeye etkisi (Boehmke, Greenwell, 2020)

Karmaşıklık parametresi, Ture, Kurt Omurlu (2018) simülasyon çalışmasında da ifade edildiği üzere, büyük veri setleri için ( $n = 1000$ ) genellikle  $cp = 0.01$  olarak kabul edilmekte ve budama işlemi gerçekleştirilmektedir. Budama için en uygun cp'nin

tespitindeki bir diğer yaklaşım ise tam ağaç için hesaplanmış olan minimum çapraz doğrulama hatasının, kendisine ait bir standart hata genişliği içerisinde kalan en küçük ağaç için budamanın gerçekleştirilmesidir (Breiman, vd., 1984). Bu yaklaşım, alan yazınında bir standart hata kuralı olarak adlandırılmaktadır, “1-SE rule”. Söz konusu yaklaşımlar, her zaman olmasa da genellikle birbirine yakın sonuçlar sağlamaktadır (Boehmke, Greenwell, 2020).

Bu tez çalışmasında, Breiman, vd. (1984) tarafından geliştirilen CART “Classification and Regression Trees” algoritması sınıflandırma amaçlı kullanılmıştır. CART ağacı, ikil dallanan ağaçlardandır ve kök düğüm de dahil olmak üzere, her bir karar düğümü iki dala ayrılarak büyümektedir. Hem kategorik, hem de sürekli değişkenler ile çalışabilmekte ve hem sınıflama hem de tahminleme amaçlı kullanılabilir.

CART algoritmasında, veri setini en iyi bölen özneliliğin seçiminde ise Gini endeksi kullanılmıştır. Gini endeksi rassal bilgi miktarını ölçen bir belirsizlik ölçüsüdür. Şekil 32.’den inceleneceği üzere entropi ile benzer sonuçlar sağlamaktadır. Aynı zamanda, logaritma işlemi gerektirmediği için hesap yükü açısından entropiden daha hızlıdır.



Şekil 32. x eksenini bir sınıftan olma olasılığı ve y eksenini endeks değeri olmak üzere entropi değeri 1/2 ile ölçeklendirildiğinde iki sınıflı bir karar problemi için endeks değerlerinin benzerliği (Joshi, 2020)

Öznitelik  $t$ ,  $i$ 'inci özneliliğin sahip olduğu kategori sayısı  $c$ ,  $i$ 'inci kategorinin olasılığı  $p_i$ ,  $i$ 'inci kategorideki gözlem sayısı  $N_i$  ve bir düğümdeki gözlem sayısı  $T$  olmak üzere gini endeksi Eş.(83) ile ifade edilmektedir.

$$g(t) = 1 - \sum_{i=1}^c p_i^2 = 1 - \sum_{i=1}^c \left(\frac{N_i}{T}\right)^2 \quad (83)$$



Eş. (83) kullanılarak,  $t$  özniteliğindeki kategorilerin her biri için bir gini endeksi hesaplanır. Daha sonra endeks değerlerinin, veri seti büyüklüğü dikkate alınarak ağırlıklı ortalaması hesaplanır. Elde edilen ağırlıklı ortalama,  $t$  özniteliğinin gini değerini temsil eder. Bu işlem, tüm öznitelikler için tekrarlanır. Elde edilen gini endekslerinden en küçük değere sahip olan  $t$  özniteliği, ilgili düğümdeki en iyi bölen öznitelik olarak tespit edilmiş olur. Çünkü gini değeri,  $t$  özniteliği için düğümün safsızlık bilgisini sağlamaktadır. Safsızlık değeri sıfıra yaklaştıkça, incelenen düğüm öznitelik  $t$ ' açısından (aslında) saflaşmaktadır. Ayrıca, bir  $T$  düğümünde en iyi bölen öznitelik olarak tespit edilmiş olan bir  $i$  özniteliği, hiyerarşik olarak  $T$ 'den sonra gelen diğer düğümlerde de en iyi bölen öznitelik olabilmektedir.

### 3. PİLOT ARAŞTIRMANIN ANALİZİ

Pilot veri seti toplanmadan önce, orijinal ölçekler farklı ülkelerde geliştirilmiş olduğu için kapsam geçerliliği amacı ile kolayda örnekleme yöntemi kullanılarak üniversite öğrencileri ve 50 yaş üstü emekli ya da çalışan kadın ve erkekler ile görüşmeler yapılmıştır. Denekler STD ölçeğinin Çevreye Duyarlı Gıda Tercihi boyutunda yer alan dört ifadesi hakkında örneğin, ekonomik sebeplerden ötürü kırmızı eti “zaten” tüketemediklerini, süt ve süt ürünlerini “zaten” merkez köylerdeki tanıdıklardan ya da mandıralardan satın aldıklarını, kendi hazırladıkları konserveleeri yıl boyunca tükettiklerini belirtmişler ve hatta “egzotik meyve” ifadesini ti'ye almışlardır. Bu ve benzeri geri bildirimler dolayısı ile hedef kitlenin söz konusu ifadeleri sürdürülebilirlik kavramı ile bağdaştıramadığı kanaatine varılmış ve söz konusu dört ifade Tablo 9 'da yer aldığı üzere genişletilerek güncellenmiştir.

STD ölçeğinde yer alan ifadelerin genişletilmesinden sonra yüklenici firma tarafından Ek 2'de yer alan soru formu kullanılarak 200 adetlik veri, sahadan yedi hafta süresinde toplanmıştır. Elde edilen pilot veri setinin sadece ölçek verisi, her biri bir ölçek (ÇEÖ, DİÖ, STD ölçeği) olmak üzere, üç bölüme ayrılmış ve her bir bölüm veri önışlemeye tabi tutulmuştur. Bu kapsamda sadece aykırı veri analizi gerçekleştirilmiştir. Aykırı veriler hem Mahalanobis mesafesi hem de SMM ölçütleri kullanılarak R-Studio yazılımı ile tespit edilmiştir.

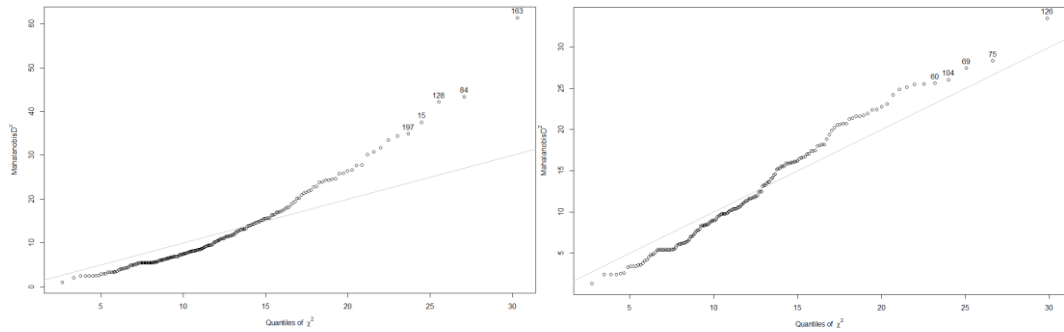
**Tablo 9.** STD ölçeğinin Çevreye Duyarlı Gıda Tercihi boyutunda yer alan genişletilerek güncellenen ifadeler

Orijinal ifade	Genişletilmiş ifade
Çok fazla kırmızı et tüketmemeye özen gösteririm	Büyükbaş hayvanların yaydıkları sera gazları hava kirliliğinde yüksek derecede etkili olduğu için çok fazla kırmızı et tüketmemeye özen gösteririm.
Süt ve süt ürünlerini (şarküteri) yerel işletmelerden almayı tercih ederim	Uzak yerlerden taşınan süt ve süt ürünlerinin nakliyesinin yarattığı kirlilik nedeni ile bu ürünleri yerel işletmelerden almayı tercih ederim
Çeşitli egzotik meyveler gibi ithal edilen gıdaları tüketmekten kaçınırım	Çeşitli egzotik meyveler gibi ithal edilen gıdaların nakliyesinin yarattığı kirlilik nedeni ile bu tür gıdaları tüketmekten kaçınırım.
Konserve hazır gıdaları tüketmekten kaçınırım	Konserve hazır gıdaları, nakliye ve ambalaj atıklarının yarattığı kirlilik nedeni ile tüketmekten kaçınırım

Mahalanobis mesafesi ile temizlendiği var sayılan veri seti SPSS yazılımı kullanılarak güvenilirlik ve geçerlik analizine tabi tutulmuştur. Geçerlik analizi kapsamında KMO ve Barlett testleri ile verilerin faktör analizine uygunluğu değerlendirilmiştir. Güvenirlik analizi kapsamında ise ölçeklerde yer alan ifadelerin toplanacağı boyutların ve boyut sayılarının belirlenebilmesi amacı ile Varimax dik döndürme<sup>146</sup> yöntemi ile AFA gerçekleştirilmiştir.

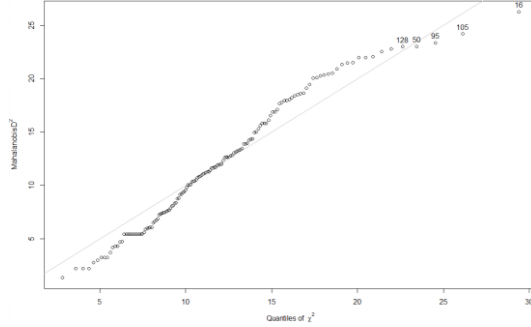
### 3.1. Birinci Pilot Araştırmanın Aykırı Veri Analizi (Veri Önleme)

ÇEÖ'ye ait (200x12) boyutlu endişe veri seti, Mahalanobis mesafesi kullanılarak temizlenmiş ve veri setinde 28 gözlem aykırı değer olarak tespit edilmiştir, Şekil 33. Veri temizleme işlemi SMM ölçütü kullanılarak gerçekleştirildiğinde ise 55 gözlem aykırı değer olarak tespit edilmiştir, Şekil 34.



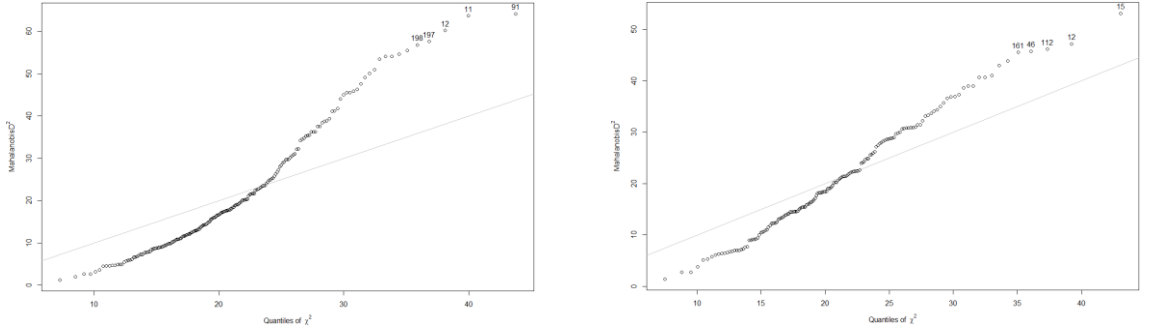
**Şekil 33.** Gözlem sayısı 200 olan endişe veri setinin aykırı veri analizi öncesi ve sonrası  $\alpha = 0.05$ , serbestlik derecesi= 12,  $\chi^2 = 21.026$  için hesaplanan Mahalanobis mesafeleri

<sup>146</sup> İkinci bölüm, kısım 2.3.4.1'de açıklandığı üzere dik döndürme yöntemi, eğik döndürme yöntemi ile benzer sonuçlar sağladığı için, faktörlerin daha kolay anlamlandırılabilmesi adına, dik döndürme yöntemi ile elde edilen sonuçlar raporlanmıştır. Sadece birinci pilot veri seti için eğik döndürme yöntemi kullanılarak gerçekleştirilen FA sonuçları örnek teşkil etmesi adına Ek 4'te ayrıca raporlanmıştır.

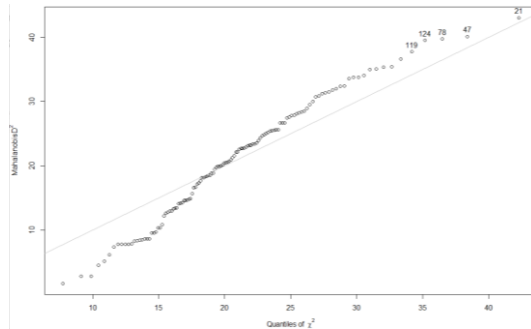


**Şekil 34.** Gözlem sayısı 200 olan endişe veri setinin aykırı veri analizi sonrası  $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinanı  $-7.87$  için SMM

Dİ ölçeğine ait (200x21) boyutlu ilişki veri seti Mahalanobis mesafesi kullanılarak temizlendiğinde veri setinde 38 gözlem, Şekil 35 ve SMM kullanılarak temizlendiğinde ise 71 gözlem aykırı değer olarak tespit edilmiştir, Şekil 36.

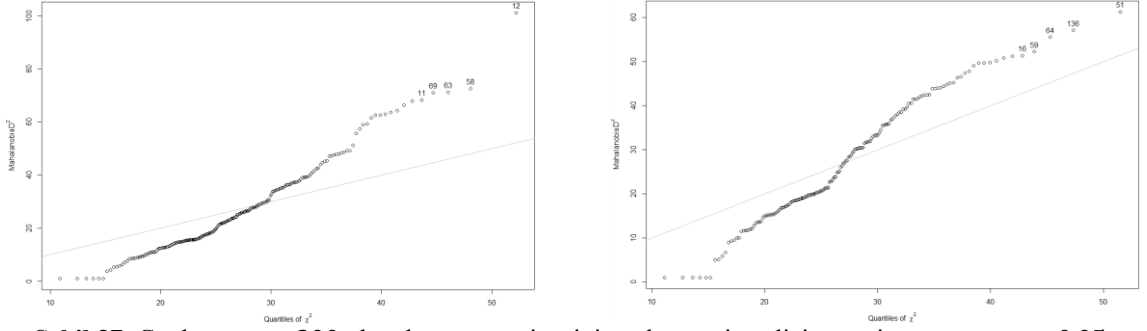


**Şekil 35.** Gözlem sayısı 200 olan ilişki veri setinin aykırı veri analizi öncesi ve sonrası  $\alpha = 0.05$ , serbestlik derecesi = 21,  $\chi^2 = 32.670$  için hesaplanan Mahalanobis mesafeleri



**Şekil 36.** Gözlem sayısı 200 olan ilişki veri setinin aykırı veri analizi sonrası  $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinanı  $-29.1$  için SMM

STD ölçeğine ait (200x27) boyutlu davranış veri seti Mahalanobis mesafesi kullanılarak temizlendiğinde veri setinde 36 gözlem aykırı değer olarak tespit edilmiştir, Şekil 37. Fakat SMM hesaplanamamıştır. Çünkü veri setindeki bir veya birden fazla özneliğin IQR'si sıfır olarak hesaplanmıştır.

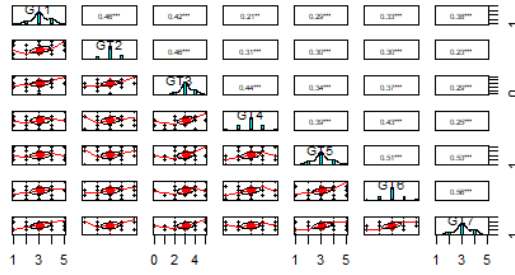
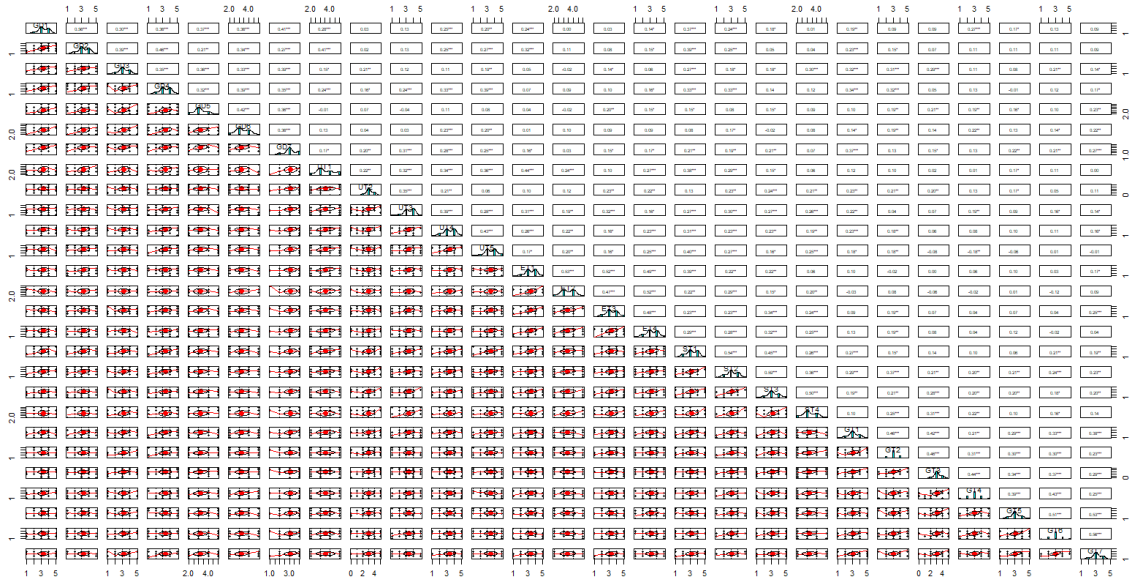


**Şekil 37.** Gözlem sayısı 200 olan davranış veri setinin aykırı veri analizi öncesi ve sonrası  $\alpha = 0.05$ , serbestlik derecesi= 27,  $\chi^2 = 40.113$  için hesaplanan Mahalanobis mesafeleri

IQR'nin sıfır olması, söz konusu özniteliğin gözlem değerlerinin en az %50'sinin aynı olması anlamına gelmektedir. Bu durumu tespit edebilmek için her bir özniteliğe ait (iki değişkenli) saçılım grafikleri ve histogramları Şekil 38.'den incelenebilir. Örneğin; bir kesit alınarak, GT2, GT4 ve GT6 özniteliklerine ait grafikler incelenirse birinci ve üçüncü kartil değerlerinin her ikisinin de üç olduğu ve dolayısı ile IQR'nin sıfır hesaplandığı tespit edilir.

IQR'nin sıfır olması, söz konusu özniteliklerin varyansının sıfıra yakın olması ve dolayısı ile ölçülmeye çalışılan özniteliğe ait gözlemlerin bir çoğunun yanlı olduğu, bu problemin giderilmesi için örneklem hacmi küçük ise eğer artırılması gerektiği, örneklem hacmi büyük ise eğer söz konusu özniteliğin bir ayrt edici özellik olmadığı ve veri setinden ya da ölçekten çıkartılması gerektiği anlamına gelir. Aynı zamanda  $N$  gözlem sayısı olmak üzere Eş. (84)'teki kovaryans formülü dikkate alındığında, eşitliğin payının alacağı değerlerin örneğin, çarpanlardan bir tanesinin değerinin sıfır olmasından dolayı azalacağı ve bu azalmanın kovaryans değerini de azaltacağı anlamına gelmektedir. Kovaryansın değerindeki düşüş ise SMM ile hesaplanan kovaryans determinantını olumsuz yönde etkileyerek Mahalanobis mesafelerinin eliptik geometrisini bozacaktır çünkü elipsin merkezi, verinin ortalama değeri ile belirlenirken elipsin genişliği, kovaryans ile belirlenmektedir.

$$cov_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N-1} \quad (84)$$



GT2		GT4		GT6	
Min.	:1.000	Min.	:1.000	Min.	:1.000
1st Qu.	:3.000	1st Qu.	:3.000	1st Qu.	:3.000
Median	:3.000	Median	:3.000	Median	:3.000
Mean	:3.065	Mean	:3.020	Mean	:3.035
3rd Qu.	:3.000	3rd Qu.	:3.000	3rd Qu.	:3.000
Max.	:5.000	Max.	:5.000	Max.	:5.000

Şekil 38. Özniteliklere ait iki değişkenli saçılım grafikleri ve histogramları

Özet ile, endişe ve ilişki veri setleri için hem Mahalanobis mesafesi hem de SMM'ler hesaplanmıştır. Fakat davranış veri seti için SMM'ler hesaplanamamıştır. Bu yüzden 200'er gözlemden oluşan pilot endişe, ilişki ve davranış veri setleri sadece Mahalanobis mesafesi kullanılarak temizlenmiştir. Temizleme işleminde, her bir veri setinin aykırı olarak tespit edilen gözlemleri birbirinden farklı olabildiği için, tüm veri setlerinde aykırı olarak tespit edilen gözlemlerin birleşim kümesi, 200 gözlemlik pilot veri setinden çıkartılmıştır. Örneğin; 12 numaralı gözlem endişe veri setinde bir aykırı değer ise hem endişe hem de ilişki ve davranış veri setlerinden çıkartılmıştır. Bu durumda 200 gözlemlik pilot veri seti 128'e inmiştir.

## 3.2. Birinci Pilot Araştırmanın Geçerlik ve Güvenirlik Analizi

### 3.2.1. ÇE Ölçeğine Ait Analiz Sonuçları

Bu kısımda gerçekleştirilen KMO ve Barlett testlerinin sonuçları Tablo 10.'da özetlenmiştir. KMO testindeki 0,934 değeri örnek büyüklüğünün yeterli olduğunu gösterirken, Barlett testinin sonucu 1820.322 olup,  $p = 0,000$  güven seviyesinde anlamlı bulunmuştur. Dolayısı ile verilerin faktör analizine uygun olduğu ifade edilebilir.

**Tablo 10.** Birinci pilot veri setinin ÇE ölçeği için KMO ve Barlett testi sonuçları

Test		Değer
KMO örneklem uyum testi		0.934
Barlett küresellik testi	Yaklaşık Ki-Kare	1820.322
	Df	66
	Sig.	0.0

AFA sonucunda ÇE ölçeğine ait ifadelerin tümü, varyansı %100 açıklayacak şekilde tek boyut altında toplanmıştır. Bu yüzden faktör sayısının tespitinde Kaiser kriteri kullanılmadan ve faktör sayısı üç olacak şekilde AFA tekrar edilmiştir. Tablo 11.'den inceleneceği üzere üç faktörlü model tarafından toplam varyansın %82.180'i açıklanmıştır. Birinci faktör toplam varyansın % 28.574'ünü, ikinci faktör % 26.855'ini ve üçüncü faktör % 26.751'ini açıklamaktadır.

**Tablo 11.** Birinci pilot veri setinin ÇE ölçeği için varyans açıklama yüzdeleri

Faktörler	Varimax döndürme sonrası		
	Özdeğerler	Açıklanan varyans	Birikimli yüzde
1	3.429	28.574	28.574
2	3.223	26.855	55.429
3	3.210	26.751	82.180
4			
:	:	:	:
12			

AFA sonucu ortaya çıkan faktör grupları, faktör gruplarına karşılık gelen ifadeler, faktör yükleri ve Cronbach Alfa katsayıları Tablo 12.'de özetlenmiştir. AFA sonucunda, her bir ifadenin ilgili faktördeki yükünün 0.50 ve üstü olduğu görülmektedir. Ayrıca her bir ifadenin ilgili faktördeki yükü ile diğer faktörlerdeki yükleri arasındaki fark, 0.10'dan büyüktür. Elde edilen bu değerler ile üç faktörlü ölçeğin yapı ve ayırsama geçerliklerinin

sağlandığı ifade edilebilir. Güvenirlilik analizi sonuçlarına göre Cronbach alfa katsayıları incelendiğinde bütün faktörlerin alfa değerlerinin 0.90'dan büyük olduğu görülmektedir.

**Tablo 12.** Birinci pilot veri setinin ÇE ölçeği için faktör yükleri

İfadeler	Faktör yükleri			Cronbach Alfa
	1	2	3	
Bio1	0.6	.488	.421	0.947
Bio2	0.68	.423	.385	
Bio3	0.87		.320	
Bio4	0.82	.318	.316	
Ego1	.319	.402	0.79	0.958
Ego2	.483	.412	0.65	
Ego3	.436	.397	0.72	
Ego4	.347	.410	0.76	
Alt1	.472	0.57	.303	0.924
Alt2	.344	0.73	.405	
Alt3	.346	0.75	.368	
Alt4		0.75	.400	

Alfa değerinin bu kadar yüksek olması aslında pek arzu edilen bir durum değildir. Çünkü Cronbach alfa bir ölçekte yer alan maddeler arasındaki kovaryansın büyüklüğü ve madde sayısı ile ilgilidir (DeVellis, 2014). Eğer ölçekte birbirine benzer madde sayısı çok ise bu maddelerin birbirleri ile korelasyonları da 1'e yakın olacak ve alfa değerini arttıracaktır. Bu durumda örneklem sayısının arttırılması düşünülebilir. Fakat ölçeğin geliştirildiği çalışmada ve ölçeğin geliştiricisi araştırmacı tarafından gerçekleştirilen bazı diğer çalışmalar Tablo 13.'ten incelendiğinde, (\*işareti, örneklemin üniversite öğrencisi olduğu anlamına gelmek üzere) 0.90 ve üzeri alfa değerleri tespit edilmiştir.

**Tablo 13.** Schultz tarafından farklı çalışmalarda kullanılan ÇE ölçeği boyutlarının ait alfa değerleri

Yayın	Kitle	Örneklem büyüklüğü	Cronbach alfa		
			Biyosferizm	Altruizm	Egoizm
P. W. Schultz (1999)	ABD*	180	<b>0.94</b>	<b>0.92</b>	<b>0.91</b>
W. P. Schultz (2001)	ABD*	148	<b>0.92</b>	0.78	0.83
	ABD-Kaliforniya (Yetişkin)	1005	<b>0.90</b>	0.90	0.88
Schmuck, vd. (2005)	ABD*	1010	0.86	0.64	0.71
	Brezilya*	208	<b>0.92</b>	0.79	0.79
	Çek Cumhuriyeti*	113	<b>0.95</b>	0.83	0.87
	Almanya*	120	<b>0.93</b>	0.73	0.84
	Hindistan*	210	0.73	0.61	0.75
	Yeni Zelanda*	217	<b>0.90</b>	0.82	<b>0.91</b>
Bruni, Chance, Schultz (2012)	Rusya*	120	0.87	0.70	<b>0.90</b>
	ABD-Kaliforniya (9-18 yaş öğrenci)	305	0.81	0.80	0.86

### 3.2.2. Dİ Ölçeğine Ait Analiz Sonuçları

Bu kısımda gerçekleştirilen KMO ve Barlett testlerinin sonuçları Tablo 14.'te özetlenmiştir. KMO testindeki 0,876 değeri örnek büyüklüğünün yeterli olduğunu gösterirken, Barlett testinin sonucu, 1820.547 ve  $p = 0,000$  güven seviyesinde anlamlı bulunmuştur. Elde edilen bu değerler ile verilerin faktör analizine uygun olduğu ifade edilebilir.

**Tablo 14.** Birinci pilot veri setinin Dİ ölçeği için KMO ve Barlett testi sonuçları

Test		Değer
KMO örneklem uyum testi		0,876
Barlett küresellik testi	Yaklaşık Ki-Kare	1820.547
	df	210
	Sig.	0.0

AFA sonucunda ifadeler, varyansı %64.72 açıklayacak şekilde Tablo 15.'ten de inceleneceği üzere beş boyut altında toplanmıştır.

**Tablo 15.** Birinci pilot veri setinin ÇEÖ için varyans açıklama yüzdeleri

Faktörler	Özdeğerler	Varimax döndürme sonrası	
		Açıklanan varyans	Birikimli yüzde
1	4.870	23.191	23.191
2	3.575	17.022	40.213
3	2.555	12.165	52.378
4	1.586	7.555	59.933
5	1.014	4.829	64.762
⋮	⋮	⋮	⋮
21			

P6 ve P7 ifadeleri beklenildiği üzere perspektif boyutunda yüklenmemiştir, Tablo 16. Çünkü bu ifadeler hem orijinal (Nisbet, Zelenski, Murphy, 2009) hem de Türkçeye uyarlanmış (Çakır, vd., 2015) ölçeklerde özbenlik boyutu altında yüklenmiş ve fakat her iki ifadenin de kavramsal olarak perspektif boyutu altında değerlendirilmesi önerilmiştir. Deneyim boyutu altında yer alan D4 ve D5 ifadeleri nerede ise eşit yükler ile iki faktöre yüklenmiş ve D6 ifadesi ise başka bir faktöre güçlü bir şekilde yüklenmiştir. Boyutların alfa güvenilirlikleri incelendiğinde perspektif boyutunun alfa değeri 0.616 gibi düşük bir değer hesaplanmıştır. Fakat diğer taraftan deneyim boyutunun alfa değeri 0.851 gibi yüksek bir değer hesaplanmış ve D5 ve D6 ifadeleri çıkartıldığında ise bu değer 0.833'e düşmüştür. Orijinal ve Türkçeye uyarlanmış ölçeğe ait alfa değerleri ise sırası ile,



özbenlik boyutu için 0.84 ve 0.87, perspektif için 0.66 ve 0.74 ve deneyim için 0.80 ve 0.73 olarak belirtilmiştir. Dolayısı ile söz konusu ifadeleri ölçekten doğrudan çıkarmak yerine, faktör sayısının tespitinde Kaiser kriteri kullanılmadan ve faktör sayısı üç olacak şekilde AFA tekrar edilmiştir.

**Tablo 16.** Birinci pilot veri setinin Dİ ölçeği için birinci faktör analizi faktör yükleri

İfadeler	Faktör yükleri					Cronbach alfa
	1	2	3	4	5	
Ö1	.728					0.916
Ö2	.762					
Ö3	.785					
Ö4	.740					
Ö5	.764					
Ö6	.790					
Ö7	.639					
Ö8	.609					
P1		.749				0.616
P2		.714	-.307			
P3		.819				
P4		.734				
P5		.697		-.340		
<b>P6</b>	<b>.354</b>	<b>-.335</b>				
<b>P7</b>					<b>.742</b>	
D1		-.337	.719			0.851
D2			.615			
D3			.706			
<b>D4</b>			<b>.577</b>	<b>.544</b>		
<b>D5</b>		<b>-.344</b>	<b>.453</b>	<b>.431</b>		
<b>D6</b>				<b>.873</b>		

Tekrarlanan AFA ile üç faktörlü model tarafından toplam varyansın %56.481'inin açıklandığı tespit edilmiştir, Tablo 17. Varimax döndürme ile birinci faktör toplam varyansın % 23.948'ini, ikinci faktör % 16.663'ünü ve üçüncü faktör % 15.870'ini açıklamıştır.

Faktör yükleri tekrar incelendiğinde, P6 ve P7 ifadeleri hariç olmak üzere her bir ifadenin ilgili faktördeki yükünün 0.50 ve üstü olduğu görülmektedir, Tablo 18. Ayrıca her bir ifadenin ilgili faktördeki yükü ile diğer faktörlerdeki yükleri arasındaki fark, 0.10'dan büyüktür. Elde edilen bu değerler ile üç faktörlü ölçeğin (P6 ve P7 ifadeleri hariç olmak üzere) yapı ve ayırsama geçerliklerini sağladığı ifade edilebilir.

**Tablo 17.** Birinci pilot veri setinin DİÖ için ikinci AFA varyans açıklama yüzdeleri

Faktörler	Özdeğerler	Varimax döndürme sonrası	
		Açıklanan varyans	Birikimli yüzde
1	5.029	23.948	23.948
2	3.499	16.663	40.610
3	3.333	15.870	56.481
4			
:	:	:	:
21			

**Tablo 18.** Birinci pilot veri setinin DİÖ için ikinci AFA faktör yükleri

İfadeler	Faktör yükleri		
	1	2	3
Ö1	0.724		
Ö2	0.713		
Ö3	0.821		
Ö4	0.750		
Ö5	0.752		
Ö6	0.743		
Ö7	0.680		
Ö8	0.653		
P1		0.763	
P2		0.707	-0.350
P3		0.818	-0.321
P4		0.738	
P5		0.628	-0.317
<b>P6</b>	<b>0.396</b>	<b>-0.321</b>	<b>0.358</b>
<b>P7</b>	<b>0.372</b>		
D1		-0.376	0.553
D2			0.575
D3		-0.302	0.674
D4			0.850
D5		-0.337	0.580
D6			0.609

### 3.2.3. STD Ölçeğine Ait Analiz Sonuçları

Bu kısımda gerçekleştirilen KMO ve Barlett testlerinin sonuçları Tablo 19.'da özetlenmiştir. KMO testindeki 0,811 değeri örnek büyüklüğünün yeterli olduğunu gösterirken, Barlett testinin sonucu da 1898.205 ve  $p = 0.000$  güven seviyesinde anlamlı bulunmuştur. Dolayısı ile verilerin faktör analizine uygun olduğu ifade edilebilir.

Akabinde AFA gerçekleştirilmeye çalışılmış fakat başarısız olmuştur. Çünkü yedi faktörlü bir yapı keşfedilmeye çalışılmış ve ölçek ifadelerinden bir tanesinin ortak varyansı<sup>147</sup> birden büyük olarak hesaplanmıştır. Kısım 3.1’de STD ölçeği için yapılan aykırı veri analizi SMM kullanılarak gerçekleştirilememiştir. Çünkü ölçekte yer alan bazı ifadelerin IQR’si sıfır idi. Dolayısı ile deneklerin STD ölçeğine verdikleri cevapların problemlili olduğu düşünülebilir. Fakat yine de genel bir fikir edinmek adına mevcut veri seti kullanılarak Kaiser kriteri dikkate alınmadan, faktör sayısı beş olacak şekilde AFA tekrarlanmıştır.

**Tablo 19.** Birinci pilot veri setinin STD ölçeği için KMO ve Barlett testi sonuçları

Test		Değer
KMO örneklem uyum testi		0,811
Barlett küresellik testi	Yaklaşık Ki-Kare	1898.205
	df	351
	Sig.	0.0

Tekrarlanan AFA ile beş faktörlü model tarafından toplam varyansın %53.650’sinin açıklandığı tespit edilmiştir. Birinci faktör toplam varyansın % 14.017’ini, ikinci faktör % 13.709’ünü, üçüncü faktör % 9.392’ini, dördüncü faktör % 9.244’ünü ve beşinci faktör % 7.288’ini açıklamıştır, Tablo 20.

**Tablo 20.** Birinci pilot veri setinin STD ölçeği için ikinci AFA varyans açıklama yüzdeleri

Faktörler	Özdeğerler	Varimax döndürme sonrası	
		Açıklanan varyans	Birikimli yüzde
1	3.785	14.017	14.017
2	3.702	13.709	27.727
3	2.536	9.392	37.119
4	2.496	9.244	46.363
5	1.968	7.288	53.650
⋮	⋮	⋮	⋮
27			

<sup>147</sup> Ortak varyans, bir değişkenin (ölçekte yer alan bir ifadenin) varyansının faktörler tarafından ne kadarlık kısmının açıklandığıdır. Bu değer birden büyük olması durumu Heywood vakalarındandır. Bu durumun olası sebepleri, aykırı veriler, örneklem büyüklüğünün 100 – 150’den küçük olması, bir faktörde sadece iki değişkenin yer alması ve mevcut veri setinin çok yüksek veya çok düşük korelasyonlu olmasıdır (Kline, 2011).

Faktörler ve faktör yükleri incelendiğinde geri dönüşüm ve su tüketimi ifadelerinin birden fazla faktöre yüklendiği, ulaştırma tercihinde UT2'nin hiçbir boyuta yüklenmediği ve ayrıca GD6, UT3, ST2'nin aynı anda iki farklı boyuta zayıf bir şekilde yüklendiği görülmektedir, Tablo 21. Sonuç olarak, mevcut veri seti ile STD ölçeğinin tamamen problemli bir yapıya sahip olduğu gözlemlenmiştir.

**Tablo 21.** Birinci pilot veri setinin STD ölçeği için ikinci AFA faktör yükleri

İfadeler	Faktör yükleri					Cronbach alfa
	1	2	3	4	5	
GD1		.614		.478		0.863
GD2		.701				
GD3				.718	.302	
GD4		.595				
GD5	.322			.670		
<b>GD6</b>		<b>.437</b>		<b>.437</b>		
GD7				.611		
UT1		.674				
<b>UT2</b>						
<b>UT3</b>		<b>.307</b>		<b>.300</b>		
UT4		.546				
UT5		.623				
ET1		.321	.691			
ET2			.727			0.844
ET3			.688			
ET4			.726			
ST1		.553			.456	
<b>ST2</b>		<b>.447</b>			<b>.496</b>	0.810
ST3					.736	
ST4					.585	
GT1	.593					
GT2	.618					
GT3	.605					
GT4	.452					
GT5	.810					
GT6	.708					
GT7	.862					

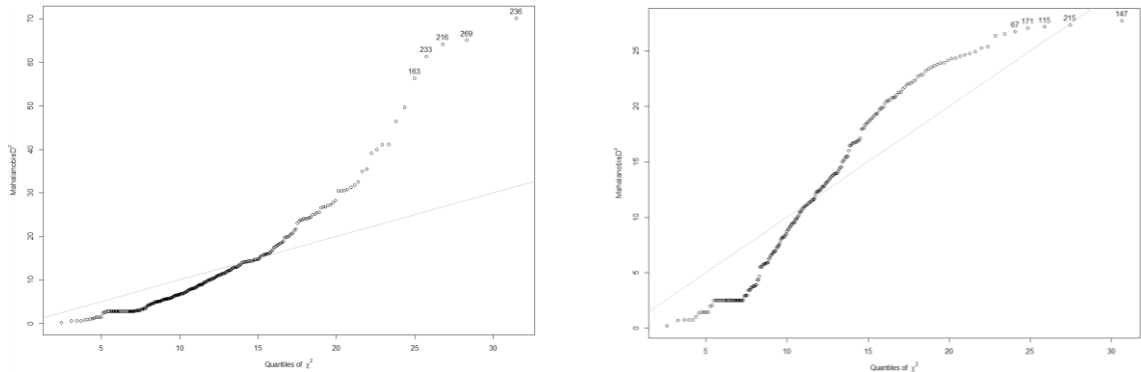
Özetle, gerçekleştirilen AFA sonuçlarına göre Kaiser kriteri dikkate alındığında ÇEÖ ve DİÖ ilgili teorik yapılarından farklı bir faktör yapısına sahip olmuştur. STD ölçeği ise AFA'ya tabi tutulamamıştır çünkü ölçek ifadelerinden bir tanesinin ortak varyansı birden büyük olarak hesaplanmıştır. Akabinde tüm ölçekler, Kaiser kriteri kullanılmadan ve beklenen faktör yapısı ön bilgi olarak kabul edilerek AFA'ya tabi

tutulmuştur. Tekrarlanan AFA sonuçlarına göre üç faktörlü ÇEÖ ve DİÖ'nin yapı ve ayırsama geçerliklerini sağladığı fakat STD ölçeğinin problemlili olduğu tespit edilmiştir. STD ölçeğindeki problemin, ölçeğin yapısından kaynakladığı düşünülmektedir. Aksine, örneklemin küçük olduğu veya örnekleme oluşturan deneklerin tutarsız ya da yanlış bilgi verdiği düşünülmektedir. Bu yüzden kullanılan ölçeklerin yapılarında hiçbir değişiklik yapılmadan pilot veri setinin büyüklüğü 300'e çıkartılmış ve sahadan 100 deneklik gözlem daha toplanmıştır. Ek veri toplama süreci üç hafta sürmüştür. Akabinde veri temizleme, verinin AFA'ya uygunluğunun testi ve AFA tekrarlanmıştır.

### 3.3.İkinci Pilot Araştırmanın Aykırı Veri Analizi (Veri Önışleme)

Bu aşamada, aykırı veri analizinde hem Mahalanobis mesafesi hem de SMM ölçütleri kullanılmış fakat sadece SMM ölçütü kullanılarak gerçekleştirilen aykırı veri analizi sonuçları detaylı raporlanmıştır.<sup>148</sup>

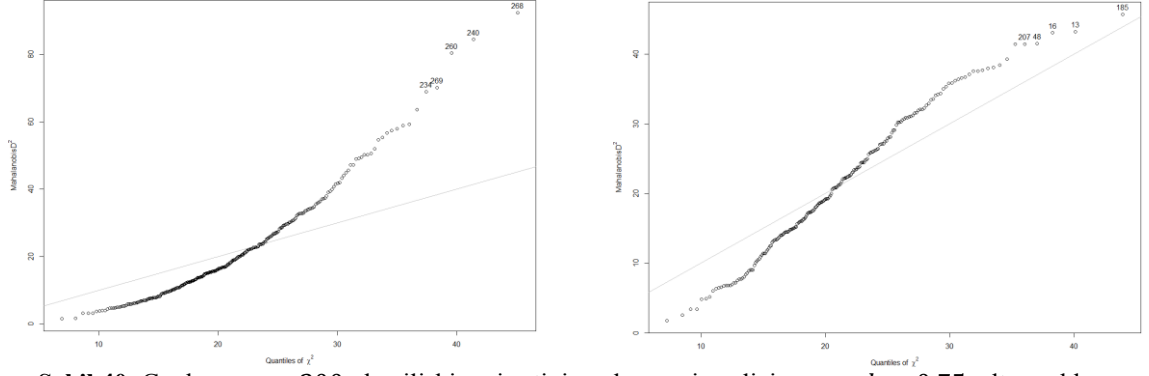
ÇE ölçeğine ait (300x12) boyutlu endişe veri seti SMM kullanılarak temizlendiğinde 75 gözlem aykırı değer olarak tespit edilmiştir, Şekil 39.



Şekil 39. Gözlem sayısı 300 olan endişe veri setinin aykırı veri analizi sonrası  $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinanı  $-6.74$  için SMM

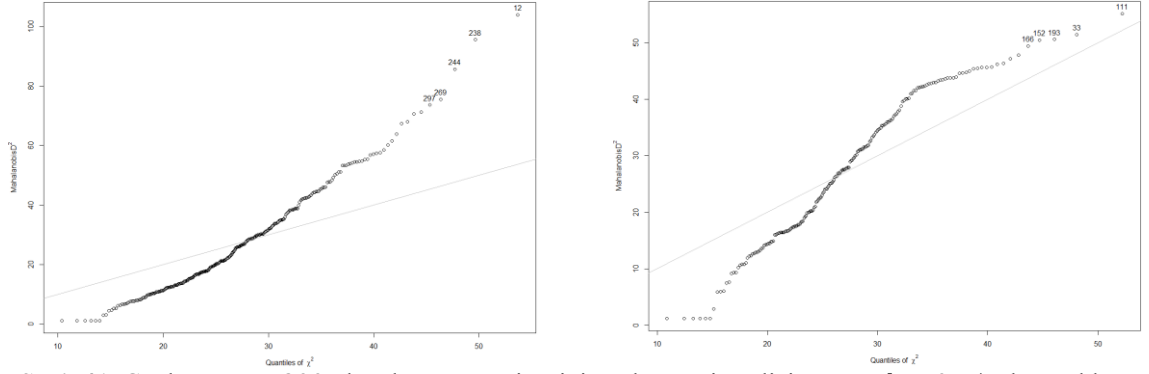
Dİ ölçeğine ait (300x21) boyutlu ilişki veri seti SMM kullanılarak temizlendiğinde ise 92 gözlem aykırı değer olarak tespit edilmiştir, Şekil 40.

<sup>148</sup> Mahalanobis mesafesi kullanılarak gerçekleştirilen aykırı veri analizi sonucu ve temizlenmiş veri setine ait güvenilirlik geçerlik analizi sonuçlarına ait özet bilgi Ek 5'ten incelenebilir.



**Şekil 40.** Gözlem sayısı 300 olan ilişki veri setinin aykırı veri analizi sonrası  $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinantı  $-20.2$  için SMM

Son olarak, STD ölçeğine ait (300x27) boyutlu davranış veri seti SMM kullanılarak temizlendiğinde 95 gözlem aykırı değer olarak tespit edilmiştir, Şekil 41.



**Şekil 41.** Gözlem sayısı 300 olan davranış veri setinin aykırı veri analizi sonrası  $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinantı  $-26.6$  için SMM

Örneklem büyüklüğü 300 gözleme çıkartıldığında tüm ölçeklere ait veri setleri için hem Mahalanobis mesafeleri hem de SMM ölçütleri hesaplanabilmektedir. Ölçeklere ait her bir veri setinin aykırı olarak tespit edilen gözlemleri birbirinden farklı olduğu için, tüm veri setlerinde aykırı olarak tespit edilen gözlemlerin birleşim kümesi, 300 gözlemlik pilot veri setinden çıkartılmıştır. Böylelikle yeni pilot veri seti 133 gözlem olarak oluşturulmuş ve akabinde yeni pilot veri seti SPSS yazılımı kullanılarak güvenilirlik ve geçerlik analizine tabi tutulmuştur.

### 3.4.İkinci Pilot Araştırmanın Geçerlik ve Güvenirlik Analizi

#### 3.4.1. ÇE Ölçeğine Ait Analiz Sonuçları

Bu kısımda gerçekleştirilen KMO ve Barlett testlerinin sonuçları Tablo 22'. de özetlenmiştir. KMO testindeki 0.918 değeri örnek büyüklüğünün yeterli olduğunu

gösterirken, Barlett testinin sonucu 2199.851 olup  $p = 0,000$  güven seviyesinde anlamlı bulunmuştur. Dolayısı ile mevcut veri seti AFA'ya uygun niteliktedir.

**Tablo 22.** İkinci pilot veri setinin ÇE ölçeği için KMO ve Barlett testi sonuçları

Test		Değer
KMO örneklem uyum testi		0.918
Barlet küresellik testi	Yaklaşık Ki-Kare	2199.851
	df	66
	Sig.	0.0

AFA sonucunda ÇEÖ'ye ait ifadeler, varyansı %81.560 açıklayacak şekilde, iki faktör altında toplanmıştır, Tablo 23. Birinci faktör toplam varyansın % 43.766'sını ve ikinci faktör % 37.794'ünü açıklamaktadır. Birinci faktör biyosferik endişe ifadelerinden oluşurken, ikinci faktör altruistik ve egoistik endişe ifadelerinden oluşmaktadır. Söz konusu yapı Schultz'un çevresel endişe yapısı ile benzeşmesine rağmen aslında, alan yazınındaki çevre etiği teorisi ile uyuşmaktadır. Çevre etiği teorisine göre çevreye duyarlılık bağlamında egoizm ya da egoistik tutum, bireyin tüketim yaparken sadece kendi konforunu dikkate alarak davranmasıdır. Altruizmde ise toplumsal fayda, bireysel faydadan önemlidir. Fakat bu görüş aynı zamanda çevrenin ve doğal kaynakların toplumsal fayda için sonuna kadar kullanılabilmesini de savunmaktadır. Biyosferizm ya da biyosferik tutum ise tüm evrene önem atfetmektedir. Çevre, tüm bileşenleri ile birlikte, bir değerler bütünüdür ve insani faaliyetler bu değerler arasındaki dengeyi bozmayacak ve onu koruyacak şekilde planlanmalıdır. Bu bağlamda Merchant (2005) biyosferizmi, altruizm ve egoizmden ayrı tutarak, altruizm ve egoizmi eş değerlendirmektedir. Dolayısı ile söz konusu iki faktörlü yapı, ifadelerin faktör yükleri ve çapraz yükler istatistiksel açıdan uygun olduğu sürece çevre etiği teorisi bağlamında değerlendirilebilir.

**Tablo 23.** İkinci pilot veri setinin ÇEÖ için varyans açıklama yüzdeleri

Faktörler	Özdeğerler	Varimax döndürme sonrası	
		Açıklanan varyans	Birikimli yüzde
1	5.252	43.766	43.766
2	4.535	37.794	81.560
⋮	⋮	⋮	⋮
12			

AFA sonucu ortaya çıkan faktör grupları, faktör gruplarına karşılık gelen ifadeler, faktör yükleri, ve Cronbach Alfa katsayıları Tablo 24'te özetlenmiştir. İnceleneceği

üzere, her bir ifadenin ilgili faktördeki yükünün 0.50 ve üstündedir. Ayrıca Alt1 ve Alt4 ifadeleri hariç olmak üzere her bir ifadenin ilgili faktördeki yükü ile diğer faktörlerdeki yükleri arasındaki fark, 0.10'dan büyüktür. Alt1 ve Alt4 ifadeleri ise iki farklı faktörde nerede ise eşit şiddette çapraz yüklendiği için söz konusu iki faktörlü ÇEÖ yapısının çevre etiği teorisini de desteklemediği ve ölçeğin geçerliğinin sağlanamadığı sonucuna varılmaktadır.

**Tablo 24.** İkinci pilot veri setinin ÇEÖ için faktör yükleri

İfadeler	Faktör yükleri		Cronbach Alfa
	1	2	
Bio1	.412	0.81	0.962
Bio2	.407	0.84	
Bio3	.367	0.83	
Bio4	.357	0.86	
Ego1	0.87	.339	0.973
Ego2	0.85	.376	
Ego3	0.86	.382	
Ego4	0.85	.369	
<b>Alt1</b>	<b>0.6</b>	<b>.559</b>	0.945
Alt2	0.7	.513	
Alt3	0.66	.561	
<b>Alt4</b>	<b>0.66</b>	<b>.548</b>	

Fakat diğer taraftan, örneklem büyüklüğünün artırılması, 200 örneklemlik pilot veri setine kıyas ile faktör yapısında ciddi bir iyileşmeye neden olmuştur. Bu sebep ile, 200 gözlemlik pilot veri setinde uygulandığı gibi bu veri setinde de Kaiser kriteri dikkate alınmadan, faktör sayısı üç olacak şekilde AFA tekrar edilmiştir. Tekrarlanan AFA sonucunda üç faktörlü ÇE modeli tarafından varyansın %86'sı açıklanmıştır. Birinci faktör toplam varyansın % 31.323'ünü, ikinci faktör % 30.728'ini ve üçüncü faktör % 24.241'ini açıklamaktadır, Tablo 25.

**Tablo 25.** İkinci pilot veri setinin ÇE ölçeği için ikinci AFA varyans açıklama yüzdeleri

Faktörler	Özdeğerler	Varimax döndürme sonrası	
		Açıklanan varyans	Birikimli yüzde
1	3.759	31.323	31.323
2	3.687	30.728	62.051
3	2.909	24.241	86.292
⋮	⋮	⋮	⋮
12			



AFA sonucu ortaya çıkan faktör grupları, faktör gruplarına karşılık gelen ifadeler ve faktör yükleri, Tablo 26.'da özetlenmiştir. Her bir ifadenin ilgili faktördeki yükü 0.50 ve üstü olup, her bir ifadenin ilgili faktördeki yükü ile diğer faktörlerdeki yükleri arasındaki fark, 0.10'dan büyüktür.

**Tablo 26.** İkinci pilot veri setinin ÇE ölçeği için faktör yükleri

İfadeler	Faktör yükleri		
	1	2	3
Bio1	.326	0.75	.384
Bio2	.325	0.79	.381
Bio3	.329	0.83	
Bio4	.309	0.85	
Ego1	0.84	.313	.341
Ego2	0.81	.350	.348
Ego3	0.8	.350	.377
Ego4	0.8	.340	.360
Alt1	.408	.439	0.59
Alt2	.465	.359	0.7
Alt3	.398	.396	0.74
Alt4	.402	.381	0.74

Güvenirlilik analizi sonuçlarına göre ise Cronbach alfa katsayıları biyosferizm, altruizm ve egoizm boyutları için sırası ile 0.962, 0.973 ve 0.945 tir. Alan yazınında 0.90 üstü alfa değerleri problemlili olarak kabul edilmesine rağmen, orijinal ölçeğin uygulandığı ve aynı zamanda uyarlandığı bazı çalışmalarda 0.90 üstü alfa değerleri Tablo 13.'teki gibi gözlemlenmiştir. Bu sebeple elde edilen bu değerler ile üç faktörlü ölçeğin yapı ve ayırsama geçerliklerinin sağlandığı ifade edilebilir.

### 3.4.2. DİÖ'ye Ait Analiz Sonuçları

Bu kısımda gerçekleştirilen KMO ve Barlett testlerinin sonuçları Tablo 27.'de özetlenmiştir. KMO testindeki 0.858 değeri örnek büyüklüğünün yeterli olduğunu gösterirken, Barlett testinin sonucu da 2187.169 ve  $p = 0,000$  güven seviyesinde anlamlı bulunmuştur. Dolayısı ile verilerin faktör analizine uygun olduğu ifade edilebilir.

AFA sonuçları Tablo 28.'den inceleneceği üzere ifadeler, varyansın %64.805'ini açıklayacak şekilde dört boyut altında toplanmıştır. Birinci faktör toplam varyansın % 25.943'ini, ikinci faktör % 20.217'ünü ve üçüncü faktör % 14.392'ini ve dördüncü faktör % 4.454'ünü açıklamıştır.

**Tablo 27.** İkinci pilot veri setinin Dİ ölçeği için KMO ve Barlett testi sonuçları

Test		Değer
KMO örneklem uyum testi		0.858
Barlet küresellik testi	Yaklaşık Ki-Kare	2187.169
	df	210
	Sig.	0.0

**Tablo 28.** İkinci pilot veri setinin Dİ ölçeği için ikinci AFA varyans açıklama yüzdeleri

Faktörler	Özdeğerler	Varimax döndürme sonrası	
		Açıklanan varyans	Birikimli yüzde
1	5.448	25.943	25.943
2	4.246	20.217	46.159
3	3.022	14.392	60.551
4	0.893	4.454	64.805
⋮	⋮	⋮	⋮
12			

P6 ve P7 ifadeleri birinci pilot veri setinde olduğu gibi perspektif boyutunda yüklenmemiştir, Tablo 29. Çünkü bu ifadeler hem orijinal (Nisbet, Zelenski, Murphy, 2009) hem de uyarlanmış (Çakır, vd., 2015) ölçeklerde özbenlik boyutu altında yüklenmiş ve fakat her iki ifadenin de kavramsal olarak perspektif boyutu altında değerlendirilmesi önerilmiştir. Deneyim boyutu altında yer alan D5 ve D6 ifadeleri ayrı bir faktöre yüklenmiştir. Birinci pilot veri setinin aksine D4 ifadesi beklendiği şekilde teoriye uygun olarak deneyim boyutunda yüklenmiştir. Boyutların alfa güvenilirlikleri incelendiğinde ise birinci pilot veri setine kıyas ile iyileşme gözlemlenmiştir. Dolayısı ile söz konusu iyileşmeler ve dördüncü boyutun açıkladığı varyansın %10'dan düşük olması dolayısı ile ölçekten ifade çıkarmak yerine, faktör sayısının tespitinde Kaiser kriteri kullanılmadan ve faktör sayısı üç olacak şekilde AFA tekrar edilmiştir, Tablo 30.

Faktör yükleri tekrar incelendiğinde, P6 ve P7 ifadeleri hariç olmak üzere her bir ifadenin ilgili faktördeki yükünün 0.50 ve üstü olduğu görülmektedir, Tablo 31. Ayrıca her bir ifadenin ilgili faktördeki yükü ile diğer faktörlerdeki yükleri arasındaki fark, 0.10'dan büyüktür. D6 ifadesi hiçbir faktöre yüklenmemiş ve D5 ise iki faktöre birden yüklenmiştir. İkinci pilot veri seti için elde edilen ikinci AFA sonuçları, birinci pilot veri setinde gerçekleştirilen ikinci AFA sonuçları ile karşılaştırıldığında D5 ve D6 ifadeleri problemli olarak yorumlanabilir.

**Tablo 29.** İkinci pilot veri setinin DİÖ için birinci AFA faktör yükleri

İfadeler	Faktör yükleri				Cronbach alfa
	1	2	3	4	
Ö1	.776				0.928
Ö2	.779				
Ö3	.850				
Ö4	.772				
Ö5	.620				
Ö6	.828				
Ö7	.783				
Ö8	.704				
P1		.866			0.826
P2		.868			
P3		.911			
P4		.895			
P5		.888			
<b>P6</b>	<b>.484</b>		<b>.478</b>		0.818
<b>P7</b>	<b>.485</b>				
D1			.651		
D2			.774		
D3	.353		.806		
D4			.792		
<b>D5</b>			.408	<b>.656</b>	
<b>D6</b>				<b>.380</b>	

**Tablo 30.** İkinci pilot veri setinin DİÖ için ikinci AFA varyans açıklama yüzdeleri

Faktörler	Varimax döndürme sonrası		
	Özdeğerler	Açıklanan varyans	Birikimli yüzde
1	5.484	26.113	26.113
2	4.355	20.739	46.852
3	3.124	14.877	61.729
4			
⋮	⋮	⋮	⋮
21			

**Tablo 31.** İkinci pilot veri setinin Dİ ölçeği için ikinci AFA faktör yükleri

İfadeler	Faktör yükleri		
	1	2	3
Ö1	.781		
Ö2	.789		
Ö3	.855		
Ö4	.778		
Ö5	.615		
Ö6	.841		
Ö7	.783		
Ö8	.698		

İfadeler	Faktör yükleri		
	1	2	3
P1		.872	
P2		.856	
P3		.930	
P4		.894	
P5		.887	
<b>P6</b>	<b>.460</b>		<b>.443</b>
<b>P7</b>	<b>.474</b>		
D1			.674
D2			.756
D3	.354		.821
D4			.805
D5		<b>.368</b>	<b>.446</b>
<b>D6</b>			

### 3.4.3. STD Ölçeğine Ait Analiz Sonuçları

Bu kısımda gerçekleştirilen KMO ve Barlett testlerinin sonuçları Tablo 32.'de özetlenmiştir. KMO testindeki 0.877 değeri örnek büyüklüğünün yeterli olduğunu gösterirken, Barlett testinin sonucu da 2681.870 olup,  $p = 0,000$  güven seviyesinde anlamlı bulunmuştur. Dolayısı ile verilerin faktör analizine uygun olduğu ifade edilebilir.

**Tablo 32.** İkinci pilot veri setinin STD ölçeği için KMO ve Barlett testi sonuçları

Test		Değer
KMO örneklem uyum testi		0.877
Barlet küresellik testi	Yaklaşık Ki-Kare	2681.870
	df	351
	Sig.	0.0

AFA sonucunda ifadeler, varyansı %63.523'ünü açıklayacak şekilde beş boyut altında toplanmıştır. Tablo 33.'ten de inceleneceği üzere birinci faktör toplam varyansın % 21.400'ünü, ikinci faktör % 17.239'unu ve üçüncü faktör % 11.989'unu, dördüncü faktör % 6.516'sını ve beşinci faktör % 6.379'unu açıklamıştır.

Tablo 34.'ten faktörler ve faktör yükleri incelendiğinde birinci pilot veri setine kıyas ile ciddi bir iyileşme görülmüştür. Öncelikle, ifadeler Kaiser kriteri altında beş boyut altında faktörleşebilmiştir. Geri dönüşüm ifadeleri, sadece GD2 ve GD4 hariç olmak üzere, toplanarak nerede ise tek bir boyut altında faktörleşmiş ve tüm ifadelerin faktör yükleri artmıştır. Elektrik ve su tüketimi ifadeleri, orijinal ölçeğin aksine, tek bir

boyut altında faktörleşmiştir ve ifadelerin faktör yükleri artmıştır. Çalışmanın devamında da aynı faktörleşme gözlemlenir ise tek yapı altındaki faktörleşmenin, enerji tüketimi boyutu olarak değerlendirilmesi düşünülebilir. Gıda tüketimi boyutunda ise teori ile paralel bir faktörleşme gözlemlenmiş ve her bir ifadenin faktör yükleri birinci pilot veri setine kıyas ile artmıştır. Son olarak ulaştırma tercihi boyutunda UT1 eş anlamlı olarak üç boyut altında faktörleşmiş ve UT2 ve UT5 yeni oluşmuş olan enerji boyutuna yüklenmiştir. Orijinal ölçeğe ait her bir faktörün toplam varyansı açıklama gücü geri dönüşüm, ulaşım tercihi, elektrik tüketimi, su tüketimi ve gıda tercihi boyutları için ise sırası ile % 15.883, % 12.535, %10.463, %10.051 ve % 14.818'dir.

**Tablo 33.** İkinci pilot veri setinin STD ölçeği için varyans açıklama yüzdeleri

Faktörler	Özdeğerler	Varimax döndürme sonrası	
		Açıklanan varyans	Birikimli yüzde
1	5.778	21.400	21.400
2	4.654	17.239	38.639
3	3.237	11.989	50.628
4	1.759	6.516	57.144
5	1.722	6.379	63.523
⋮	⋮	⋮	⋮
27			

Güvenirlilik analizi sonuçlarına göre Cronbach alfa katsayıları incelendiğinde her bir boyutun tatmin edici veya yüksek derecede alfa değerlerine sahip olduğu gözlemlenmektedir. Güvenirlilik analizi GD2 ve GD4'ün çıkartılarak tekrarlandığında alfa değeri 0.898'e, UT1 çıkartıldığında 0.771'e ve UT5 çıkartıldığında 0.765'e düşmüştür. UT2 çıkartıldığında ise alfa değeri 0.824'e yükselmiştir. Ek olarak elektrik ve su tüketimi boyutlarına ayrı ayrı iki boyut olarak güvenirlilik analizi tekrarlandığında alfa değeri sırası ile 0.913 ve 0.885 olarak hesaplanmıştır. Orijinal ölçeğe ait alfa değerleri ise geri dönüşüm, ulaşım tercihi, elektrik tüketimi, su tüketimi ve gıda tercihi boyutları için sırası ile 0.813, 0.843, 0.820, 0.862 ve 0.802 şeklindedir.

**Tablo 34.** İkinci pilot veri setinin STD ölçeği için faktör yükleri

İfadeler	Faktör yükleri					Cronbach alfa
	1	2	3	4	5	
GD1			.623		.490	
<b>GD2</b>			<b>.358</b>		<b>.702</b>	
GD3		.301	.801			
<b>GD4</b>			<b>.454</b>		<b>.404</b>	0.905
GD5		.308	.775			
GD6		.353	.528		.370	
GD7			.619			
<b>UT1</b>	<b>.374</b>			<b>.429</b>	<b>.419</b>	
<b>UT2</b>	<b>.483</b>					
UT3	.392			.664		0.801
UT4	.368			.695		
<b>UT5</b>	<b>.471</b>			.348	.334	
ET1	.667			.329		
ET2	.744					
ET3	.773					
ET4	.759					
ST1	.710					0.923
ST2	.713					
ST3	.795					
ST4	.691					
GT1		.726				
GT2	.410	.631				
GT3		.795				
GT4		.677				0.913
GT5		.862				
GT6		.837				
GT7		.774				

Özetle, ikinci pilot veri seti için gerçekleştirilen AFA sonuçlarına göre:

- Kaiser kriteri dikkate alındığında tüm ölçekler, ilgili teorik yapılarından farklı bir faktör yapısına sahip olmuştur. Fakat birinci pilot veri setine ait birinci AFA sonuçlarına kıyas ile faktör yapısında ciddi bir iyileşme tespit edilmiştir. ÇEÖ birinci pilot veri setinin birinci AFA’da tek boyuta sahip iken ikinci pilot veri setinin birinci AFA’da iki boyuta sahip olmuştur. DİÖ ve STD ölçekleri ise sırası ile beş boyuttan dört boyuta ve yedi boyuttan beş boyuta düşmüştür.

- Kaiser kriteri kullanılmadan ve teorik faktör yapısı ön bilgi olarak kabul edilerek tekrarlanan AFA sonuçlarına göre üç faktörlü ÇEÖ’nin yapı ve ayırsama geçerliklerini sağladığı tespit edilmiştir. Birinci pilot veri setinin ikinci AFA sonuçlarına

kıyas ile açıklanan varyans %82'den %86'ya çıkmıştır. Diğer taraftan, DİÖ ve STD ölçeklerinin problemlili olduğu fakat birinci pilot veri setine ait ikinci AFA sonuçlarına kıyas ile ölçeklerin faktör yapılarında ciddi bir iyileşme olduğu tespit edilmiştir. Sonuçlara göre, DİÖ ve STD ölçekleri birinci pilot veri setinde varyansın %56'sı ve %53'ünü açıklayabilmiş iken, ikinci pilot veri setinde %61'i ve %56'sını açıklayabilmiştir. Ayrıca STD ölçeğinde faktörleşmede problem yaşanan ifade sayısı sekizden beşe düşmüştür.

### **3.5.Pilot Araştırma Süreci Hakkında Genel Değerlendirme**

Pilot araştırma sürecindeki veri setleri kullanılarak gerçekleştirilen güvenilirlik ve geçerlik analizleri sonuçları dikkate alındığında, ölçeklerin faktör yapılarında problemler tespit edilmiştir. Yaşanan problemlerin, ölçeğin yapısından kaynaklandığı düşünülmektedir. Açıklamak gerekirse birinci ve ikinci veri setlerinde başlangıçta 200 ve 300 gözlem olmasına rağmen, veri temizleme işlemi sonrasında gözlem sayısı 128 ve 133'e inmiştir. Daha sonra, her iki pilot veri seti için indirgenmiş veri setleri kullanılarak analizler gerçekleştirilmiştir. Başka bir ifade ile birinci ve ikinci veri setleri gözlem sayısı açısından nerede ise birbirine eşittir. Fakat temizleme işlemi ile ikinci pilot veri setindeki gözlemlerin değişmesi, ölçeklerin faktör yapılarındaki problemleri kısmen ve ciddi bir şekilde gidermiştir. Bu sebep ile analiz sonuçlarındaki ciddi iyileşmenin, veri setindeki gözlemlerin değişmesinden kaynaklandığı ve ölçeklerdeki faktörleşme probleminin, örnekleme oluşturan deneklerin tutarsız ya da yanlış bilgi vermesinden kaynaklandığı düşünülmektedir.

Pilot veri setlerinin sırası ile %36 ve %55'i aykırı veri olarak tespit edilmiştir. Bu oranlar, özellikle %55 oranı, aykırı veri analizi için çok yüksek bir değerdir. Fakat bu bölümde Kısım 2.2 ve 2.4'te raporlanan sonuçlar ve Ek 5'te yer alan Tablo 61. incelendiğinde, Mahalanobis ölçütünün veri setinden daha az gözlemi aykırı veri olarak tespit ettiği ama bu ölçüt ile temizlenen veri setlerinin, SMM ölçütü ile temizlenen veri setlerine kıyas ile daha kötü faktör yapısına sahip olduğu görülmektedir. Zira birinci pilot veri setine veri temizleme işlemi gerçekleştirilmeden güvenilirlik ve geçerlik analizi uygulandığında, ÇEÖ iki faktöre sahip olup varyansı açıklama gücü %70.6'ya, DİÖ beş faktöre sahip olup varyansı açıklama gücü %55.2'ye ve STD ölçeği sekiz faktöre sahip olup varyansı açıklama gücü %53.86'ya düşmektedir. Ölçeklerin teorik faktör yapıları

ön bilgi olarak kullanıldığında ise açıklanan varyanslar sırası ile %75.1, %46.89 ve %44.18'dir. Söz konusu değerler, bu bölümdeki Kısım 2.2'de raporlanan değerlerin çok altındadır (%82, %56 ve %53). Cronbach alfa değerlerindeki düşüş, açıklanan varyans oranlarındaki gibi ciddi olmasa da mutlak değerce maksimum 0.062'ye ulaşmaktadır. Ayrıca son olarak, ölçeklerdeki her bir ifadenin ilgili faktördeki yükleri de azalmıştır.

Yukarıda sayılan sebepler nedeni ile kullanılan ölçeklerin yapısında hiçbir değişiklik yapılmadan pilot dahil toplam 1000 gözlemden geriye kalan 700 gözlemlik tüm verinin toplanması ve SMM ölçütü kullanılarak, aykırı veri analizinin gerçekleştirilmesinin gerekli olduğu düşünülmektedir.

#### **4. ARAŞTIRMA MODELİNİN DEĞERLENDİRİLMESİ**

Bu aşamada ana araştırma için 700 gözlemlik veri seti sahadan toplanmıştır. Veri toplama 15 hafta sürmüştür. Veri seti temin edildikten sonra hem eksik veri içeren hem de aykırı nitelikli gözlemler veri setinden çıkartılmıştır (veri temizleme aşaması) ve araştırma modeli KEKK-YEM ile PLS-SEM yazılımı kullanılarak keşfedilmiştir (yapısal analiz aşaması). Ayrıca, bütün veri setine ait bazı demografik bilgiler Ek 3'te özetlenmiştir.

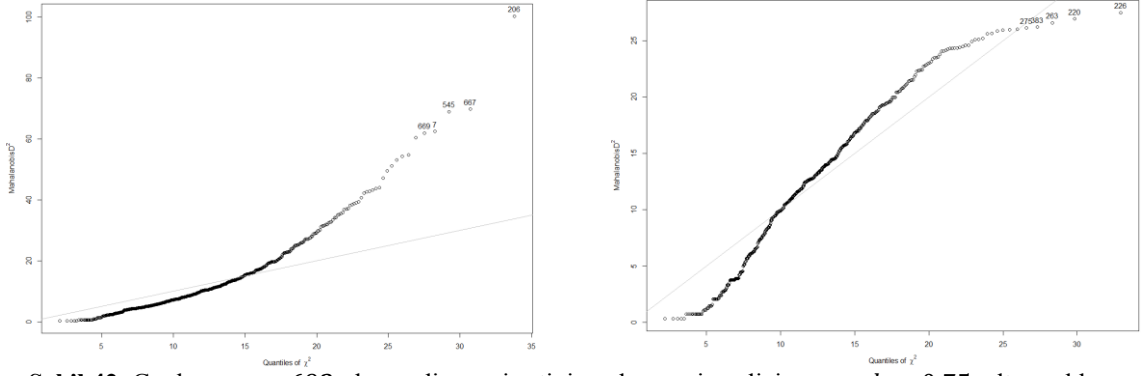
##### **4.1. Veri Ön işleme**

Yüklenici firma tarafından temin edilen 700 gözlemlik veri setinden 8 adet gözlem, eksik veri içerdiği için çıkartılmıştır ve sonrasında R-Studio yazılımı kullanılarak SMM ölçütü ile aykırı veri analizi her bir ölçek için ayrı ayrı gerçekleştirilmiştir.

Pilot araştırma aşamasında SMM ölçütü kullanılarak temizlenen veri setleri, Mahalanobis mesafesi kullanılarak temizlenen veri setlerine kıyas ile daha iyi sonuçlar sağladığı için aykırı veri analizinde sadece SMM ölçütlü kullanılmıştır.

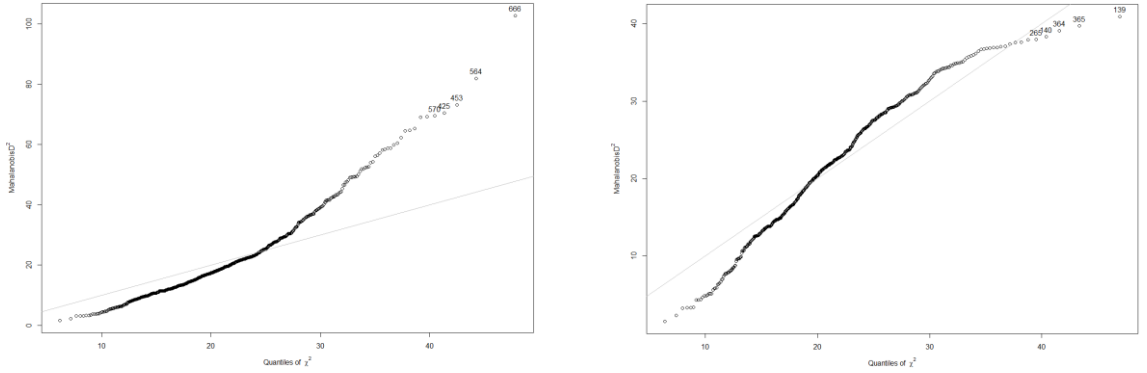
ÇEO'ye ait (692x12) boyutlu endişe veri seti SMM kullanılarak temizlendiğinde 187 gözlem aykırı değer olarak tespit edilmiştir, Şekil 42.





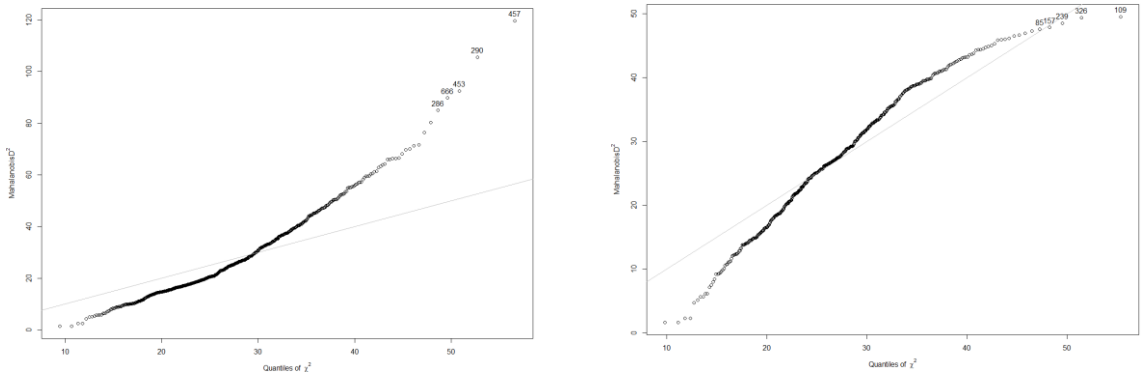
**Şekil 42.** Gözlem sayısı 692 olan endişe veri setinin aykırı veri analizi sonrası  $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinanı  $-7.57$  için SMM

DİÖ'ye ait  $(692 \times 21)$  boyutlu ilişki veri seti SMM kullanılarak temizlendiğinde ise 167 gözlem aykırı değer olarak tespit edilmiştir, Şekil 43.



**Şekil 43.** Gözlem sayısı 692 olan ilişki veri setinin aykırı veri analizi sonrası  $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinanı  $-16.9$  için SMM

Son olarak, STD ölçeğine ait  $(692 \times 27)$  boyutlu davranış veri seti SMM kullanılarak temizlendiğinde 204 gözlem aykırı değer olarak tespit edilmiştir, Şekil 44.

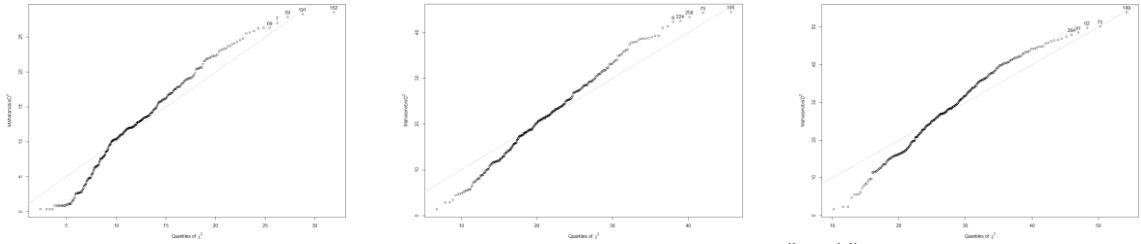


**Şekil 44.** Gözlem sayısı 692 olan davranış veri setinin aykırı veri analizi sonrası  $h = 0.75$ , alt örneklem sayısı 3000 ve hesaplanan minimum kovaryans determinanı  $-23.5$  için SMM

Ölçeklere ait her bir veri setinin aykırı olarak tespit edilen gözlemleri birbirinden farklı olabildiği için, tüm veri setlerinde aykırı olarak tespit edilen gözlemlerin birleşim kümesi, 692 gözlemlik veri setinden çıkartılmıştır. Böylelikle yeni veri seti 352 gözlem olarak oluşturulmuş ve akabinde yapısal modelin analizi gerçekleştirilmiştir.

#### 4.2.Yapısal Analiz

Bu aşamada öncelikle veri setinin çok değişkenli normal dağılıma sahip olup olmadığı test edilmiştir. Çünkü veri setinin normal dağılmadığı durumlarda, yapısal analiz klasik YEM (kovaryans temelli YEM) ile gerçekleştirilememektedir. Yerine KEKK-YEM yönteminin uygulanması önerilmektedir (Astrachan, Patel, Wanzenried, 2014; Rigdon, Sarstedt, Ringle, 2017; Joe F. Hair Jr, vd., 2014). Şekil 45.'ten inceleneceği üzere, 352 gözlemlik veri seti, her bir ölçek için normallik testine tabi tutulduğunda veriye ait Mahalanobis mesafesi teorik ki-kare dağılımına uymamaktadır ve bu yüzden çok değişkenli normal dağılıma sahip değildir (Burdenski, 2000).<sup>149</sup>



Şekil 45. Gözlem sayısı 352 olan ana araştırmaya ait veri setinin ÇEÖ, DİÖ ve STD ölçekleri için Mahalanobis mesafelerinin normallik testi

Veri seti normal dağılıma sahip olmadığı için, çevresel endişe ile STD arasındaki ilişkide İnsan-Doğa İlişkisinin aracı etkisi KEKK-YEM yöntemi kullanılarak araştırılmıştır.

Araştırma modelinin analizinden önce, modelde yer alan yapıların iç tutarlılık güvenilirliği, birleşme ve ayrışma geçerliği değerlendirilmiştir.<sup>150</sup> Araştırma modelinin analiz edilmesinde ise yapılar arası çoklu doğrusallık, yol katsayıları, modelin bağımlı

<sup>149</sup> Veri setinin çokdeğişkenli normal dağılıma uyup uymadığının testi ayrıca Mardia katsayılarının kontrolü ile de gerçekleştirilmiştir (Korkmaz, Gökşülük, Dinçer, Zararsız, 2014; Lee, Yom, 2013). Veri seti, Mardia katsayıları dikkate alındığı zaman da çok değişkenli normal dağılıma sahip değildir. Söz konusu test, R-Studio yazılımı kullanılarak gerçekleştirilmiş ve sonuçları Ek 6'da paylaşılmıştır.

<sup>150</sup> SPSS programı kullanılarak ölçme modelleri için gerçekleştirilen geçerlik ve güvenilirlik analizi sonuçları Ek 7'de özetlenmiştir.

değişkeni açıklama gücü ve etki büyüklüğü KEKK algoritması ile incelenmiştir. Yol katsayılarının anlamlılıkları ise yeniden örnekleme ile ana örneklemden alt örneklemeler çekilerek (bootstrap) değerlendirilmiştir.

#### ***4.2.1. Güvenirlik ve Geçerlik Analizleri***

İç tutarlılık güvenirliliği için Cronbach alfa ve birleşik güvenirlik (CR) “Composite Reliability” katsayıları incelenmiştir. Birleşme geçerliğinin tespitinde, faktör yükleri ile açıklanan ortalama varyans (AVE) “Average Variance Extracted” değerleri kullanılmıştır. Alan yazınında faktör yüklerinin, alfanın ve CR değerlerinin 0.70’in ve AVE değerinin ise 0.50’nin üzerinde olması gerektiği ifade edilmektedir (J. F. Hair, vd., 2006; Joe F. Hair Jr, vd., 2017; Fornell, Larcker, 1981). Tablo 35.’te araştırmada yer alan yapıların iç tutarlılık güvenirliliği ve birleşme geçerliği ile ilgili sonuçlar özetlenmiştir.

Sonuçlara göre Ö5, P6 ve P7 ifadelerinin faktör yüklerinin 0.70’nin altında, GD1 ve ST2 ifadelerinin faktör yüklerinin ise 0.70 sınırında olduğu görülmektedir. Alan yazınında KEKK-YEM temelli analiz yöntemlerinde faktör yüklerinin en az 0.70 olması beklendiği ifade edilmektedir. Fakat güncel çalışmalarda, ortalama açıklanan varyansın 0.50’nin ve alfa değerinin 0.70’in üstünde olması koşulu ile, 0.60 – 0.70 arası faktör yüklerinin kabul edilebilir olduğu belirtilmektedir (Joe F. Hair Jr, vd., 2017). Bu sebeple sadece Ö5, P6 ve P7 ifadeleri DİÖ’den çıkartılarak analiz tekrar edilmiş ve Tablo 36.’da yer alan sonuçlar elde edilmiştir.

Tablo 36.’daki değerler incelendiğinde ise GD1 ve ST2 ifadelerinin faktör yüklerinin 0.70 sınırında kaldığı tespit edilmiştir. Fakat bu ifadelerin yüklendiği boyutların AVE ve alfa değerlerinin sırası ile 0.70 ve 0.50’nin üstünde olmasından ötürü, söz konusu ifadelerin modelde korunmasına karar verilmiştir. Ayrıca, revize edilen ölçekte Ö5, P6 ve P7 ifadelerinin yer almaması, özbenlik ve perspektif boyutlarının alfa, CR ve AVE değerlerinde artışa neden olduğu tespit edilmiştir.

**Tablo 35.** Orijinal ölçekler için ölçme modelinin sonuçları

Değişkenler	İfade	Faktör Yüğü	Cronbach Alfa	CR	AVE	
Çevresel Endişe	Biosferizm	Bio1	0.910	0.954	0.967	0.876
		Bio2	0.946			
		Bio3	0.964			
		Bio4	0.930			
	Egoizm	Ego1	0.951	0.963	0.973	0.900
		Ego2	0.939			
		Ego3	0.953			
		Ego4	0.951			
	Altruizm	Alt1	0.923	0.956	0.968	0.883
		Alt2	0.943			
		Alt3	0.950			
		Alt4	0.942			
Doğa ile İlişki	Özbenlik	Ö1	0.721	0.891	0.917	0.590
		Ö2	0.795			
		Ö3	0.837			
		Ö4	0.862			
		Ö5	0.329			
		Ö6	0.808			
		Ö7	0.821			
		Ö8	0.833			
	Perspektif	P1	0.853	0.830	0.821	0.531
		P2	0.884			
		P3	0.845			
		P4	0.834			
		P5	0.811			
		P6	0.022			
		P7	-0.375			
	Deneyim	D1	0.832	0.881	0.910	0.627
		D2	0.824			
		D3	0.774			
D4		0.809				
D5		0.776				
D6		0.732				
Sürdürülebilir Tüketim Davranışı	Geri dönüşüm	GD1	0.695	0.902	0.922	0.630
		GD2	0.811			
		GD3	0.759			
		GD4	0.844			
		GD5	0.794			
		GD6	0.848			
		GD7	0.796			
	Ulaştırma tercihi	UT1	0.818	0.875	0.907	0.660
		UT2	0.788			
		UT3	0.803			
		UT4	0.832			
		UT5	0.821			
	Elektrik tüketimi	ET1	0.827	0.886	0.921	0.745
		ET2	0.886			
ET3		0.884				
ET4		0.855				
Su tüketimi	ST1	0.883	0.866	0.909	0.716	
	ST2	0.681				
	ST3	0.915				
	ST4	0.886				

Değişkenler	İfade	Faktör Yüğü	Cronbach Alfa	CR	AVE
Gıda tüketimi	GT1	0.778	0.907	0.926	0.643
	GT2	0.791			
	GT3	0.762			
	GT4	0.851			
	GT5	0.822			
	GT6	0.840			
	GT7	0.764			

**Tablo 36.** Revize edilmiş ölçekler için ölçme modelinin sonuçları

Değişkenler	İfade	Faktör Yüğü	Cronbach Alfa	CR	AVE	
Çevresel Endişe	Biosferizm	Bio1	0.911	0.954	0.967	0.879
		Bio2	0.946			
		Bio3	0.964			
		Bio4	0.930			
	Egoizm	Ego1	0.951	0.963	0.973	0.900
		Ego2	0.939			
		Ego3	0.953			
		Ego4	0.951			
	Altruizm	Alt1	0.924	0.956	0.968	0.883
		Alt2	0.943			
		Alt3	0.950			
		Alt4	0.942			
Doğa ile İlişki	Özbenlik	Ö1	0.724	0.915↑	0.932↑	0.663↑
		Ö2	0.796			
		Ö3	0.834			
		Ö4	0.863			
		Ö6	0.812			
		Ö7	0.824			
		Ö8	0.838			
		Perspektif	P1			
	P2		0.925			
	P3		0.917			
	P4		0.871			
	P5		0.850			
	Deneyim	D1	0.832	0.881	0.910	0.627
		D2	0.824			
		D3	0.774			
		D4	0.809			
		D5	0.776			
		D6	0.732			
Sürdürülebilir Tüketim Davranışı	Geri dönüşüm	<b>GD1</b>	<b>0.693</b>	0.902	0.922	0.631
		GD2	0.809			
		GD3	0.763			
		GD4	0.841			
		GD5	0.798			
		GD6	0.847			
		GD7	0.797			
	Ulaştırma tercihi	UT1	0.817	0.875	0.907	0.661
		UT2	0.790			
		UT3	0.808			
UT4		0.835				
UT5		0.816				

Değişkenler	İfade	Faktör Yüğü	Cronbach Alfa	CR	AVE
Elektrik tüketimi	ET1	0.827	0.886	0.921	0.745
	ET2	0.885			
	ET3	0.885			
	ET4	0.855			
Su tüketimi	ST1	0.882	0.866	0.909	0.716
	ST2	0.679			
	ST3	0.915			
	ST4	0.887			
Gıda tüketimi	GT1	0.772	0.907	0.927	0.644
	GT2	0.784			
	GT3	0.759			
	GT4	0.856			
	GT5	0.829			
	GT6	0.845			
	GT7	0.765			

Özetle, yapıların Cronbach alfa katsayılarının 0.866 üzeri olması; CR katsayılarının da 0.907 üzeri olması nedeni ile iç tutarlılık güvenilirliğinin, faktör yüklerinin 0.679 üzeri ve AVE değerlerinin de 0.627 üzeri olması nedeni ile birleşme geçerliğinin sağlandığı belirtilebilir.

Ayrışma geçerliğinin tespitinde, Fornell-Larcker ve HTMT "Heterotrait-Monotrait Ratio" kriterleri kullanılmıştır (Fornell, Larcker, 1981; Henseler, Ringle, Sarstedt, 2015). Fornell-Larcker kriterine göre, modelde yer alan her bir boyuta ait AVE'nin karekökü, boyutlar arasındaki korelasyon değerlerinden yüksek olmalıdır. Tablo 37.'de yer alan matrisin köşegen değerleri ilgili boyutların AVE'lerinin karekökünü ve köşegen dışı değerleri boyutlararası korelasyonu temsil etmek üzere, köşegen değerlerinin, ilgili köşegen dışı değerlerden daha büyük olduğu görülmektedir.

Henseler, Ringle, Sarstedt (2015) çalışmasında önerilen HTMT kriteri ise aslında bir orandır. Oranın pay kısmında, farklı boyutlara ait olan ifadeler arası korelasyonların (the heterotrait-heteromethod correlations) ortalaması, payda kısmında ise aynı boyutta yer alan ifadelerin kendi aralarındaki korelasyonlarının (the monotrait-heteromethod correlations) ortalamalarının geometrik ortalaması yer almaktadır. Şekil 46.'da yer alan matris, HTMT ayrışma geçerliği için kullanılmaktadır. Matris incelendiğinde, en iyi ayrışmanın farklı boyutlara ait ifadeler arası korelasyonların küçük ve aynı boyuta ait ifadelerin kendi aralarındaki korelasyonların büyük olduğu zaman elde edileceği aşikârdır. Bu neden ile, HTMT değerinin birden küçük bir değer olması beklenmektedir.

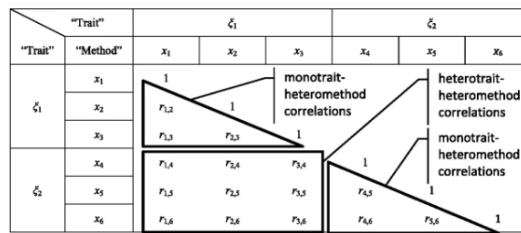
Yazarlar bu değerin 0.85'ten küçük olması gerektiğini, fakat bir modelde yer alan boyutlar birbirine kavramsal olarak yakın ise eşik değerin 0.90 olarak da belirlenebileceği ifade edilmektedir. Tablo 38.'de yer alan HTMT değerleri incelendiğinde hepsinin 0.85'ten küçük olduğu ve ayırışma geçerliğinin sağlandığı görülmektedir.

**Tablo 37.** Ayırışma geçerliği (Fornel-Larcker kriteri)

	Alt	Bio	Deneyim	ET	Ego	GD	GT	Ozbenlik	Perspektif	ST	UT
Alt	<b>0.940</b>										
Bio	0.605	<b>0.938</b>									
Deneyim	0.146	0.059	<b>0.792</b>								
ET	0.322	0.382	0.217	<b>0.863</b>							
Ego	0.650	0.545	0.192	0.302	<b>0.949</b>						
GD	0.391	0.285	0.442	0.465	0.387	<b>0.794</b>					
GT	0.282	-0.005	0.376	0.195	0.138	0.343	<b>0.802</b>				
Özbenlik	0.742	0.555	0.215	0.446	0.507	0.502	0.322	<b>0.814</b>			
Pers.	-0.140	-0.308	0.430	-0.168	-0.032	0.046	0.458	-0.123	<b>0.895</b>		
ST	0.389	0.408	0.177	0.642	0.296	0.418	0.321	0.424	-0.144	<b>0.846</b>	
UT	0.322	0.357	0.237	0.615	0.255	0.502	0.240	0.437	-0.080	0.583	<b>0.813</b>

**Tablo 38.** Ayırışma geçerliği (HTMT kriteri)

	Alt	Bio	Deneyim	ET	Ego	GD	GT	Ozbenlik	Perspektif	ST	UT
Alt											
Bio	0.632										
Deneyim	0.158	0.078									
ET	0.344	0.407	0.245								
Ego	0.675	0.566	0.210	0.324							
GD	0.400	0.288	0.504	0.515	0.408						
GT	0.298	0.094	0.412	0.216	0.155	0.367					
Özbenlik	0.791	0.588	0.237	0.490	0.536	0.533	0.347				
Pers.	0.155	0.325	0.479	0.173	0.047	0.139	0.479	0.157			
ST	0.419	0.438	0.193	0.725	0.299	0.442	0.386	0.466	0.178		
UT	0.326	0.362	0.278	0.686	0.265	0.553	0.255	0.461	0.130	0.632	



**Şekil 46.** Multitrait-multimethod (MTMT) matrisi (Henseler, Ringle, Sarstedt, 2015)

Modelde yer alan boyutlar arası çoklu doğrusallığı test etmek için ise Varyans Arttırıcı Faktör “Variance Inflation Rate-VIF” kriteri incelenmiştir. Varyans arttırıcı faktör, modelde yer alan bir değişkenin varyansının, modeldeki diğer değişkenler tarafından ne kadar etkilendiğini temsil eden bir orandır  $\left(\frac{1}{1-R_i^2}\right)$ . Paydada yer alan  $r$ -kare,  $i$ 'inci değişkenin determinasyon katsayısı olup,  $i$ 'inci değişkenin bağımlı değişken

olduğu bir regresyon denkleminde, diğer değişkenlerin  $i$ 'inci değişkenin varyansının yüzde kaçını açıklayabildiği bilgisini vermektedir. Bu oranın değerinin beşten küçük olması beklenmektedir (Joe F. Hair Jr, vd., 2017). Fakat Garson (2016)'a göre bu değer dörtten ve Diamantopoulos, Siguaw (2006)'a göre ise 3.3'ten küçük olmalıdır. Bu oranın büyüklüğü değişkenler arası çoklu doğrusal ilişkinin büyüklüğü ile ilgilidir. Açıklamak gerekirse, modelde yer alan boyutlar arası VIF değerleri Tablo 39.'dan inceleneceği üzere beşten küçüktür ve en büyük varyans arttırıcı faktör değeri 3.006'dır. Varyans arttırıcı faktör oranı dikkate alındığında, altruizm bağımsız değişken iken ET, GD, GT, ST ve UT boyutlarının  $3.006 = \left(\frac{1}{1-R^2}\right)$  eşitliğinden 0.66'sını açıkladığı bilgisine ulaşılmaktadır. Dolayısı ile varyans arttırıcı faktörün olabildiğince küçük olması gerekmektedir. Aşağıda yer alan tablodaki değerler Diamantopoulos, Siguaw (2006)'ın eşik değeri dikkate alındığında, modeldeki boyutlar arasında var olan çoklu doğrusallığın model analizini etkilemeyecek veya kabul edilebilir seviyede (orta seviyede) olduğu ifade edilebilir.

**Tablo 39.** Çoklu doğrusallık (VIF kriteri)

	Alt	Bio	Deneyim	ET	Ego	GD	GT	Ozbenlik	Perspektif	ST	UT
Alt			2.047	3.006		3.006	3.006	2.047	2.047	3.006	3.006
Bio			1.684	1.944		1.944	1.944	1.684	1.684	1.944	1.944
Deneyim				1.363		1.363	1.363			1.363	1.363
ET											
Ego			1.847	1.908		1.908	1.908	1.847	1.847	1.908	1.908
GD											
GT											
Ozbenlik				2.393		2.393	2.393			2.393	2.393
Perspektif				1.442		1.442	1.442			1.442	1.442
ST											
UT											

#### 4.2.2. Model Analizi

Şekil 47.'de yer alan araştırma modelinin analizinde KEKK-YEM kullanılmıştır. Çalışmada, ÇEÖ ile ölçülen tutumun STD üzerinde oluşturduğu etkide doğa ile ilişkinin aracı etkisi olduğu savunulmaktadır. Baron, Kenny (1986)'ye göre aracılık etkisinin var olması için aşağıdaki koşulların sağlanması gerekmektedir:

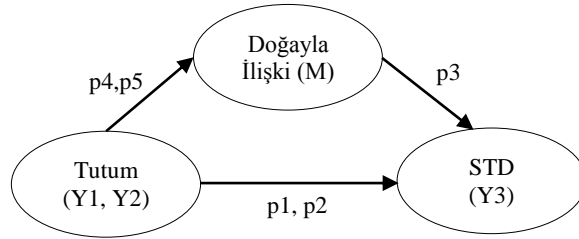
- Bağımlı değişken ile bağımsız değişken arasındaki ilişki anlamlı olmalıdır ( $p1, p2 \neq 0$ ). Bu durum, bağımsız değişkenin, bağımlı değişkenin tahminçisi olma



durumdur. Bağımlı ve bağımsız değişken arasında herhangi bir ilişki yok ise aracı değişkenin aracılık edeceği herhangi bir ilişki de yoktur.

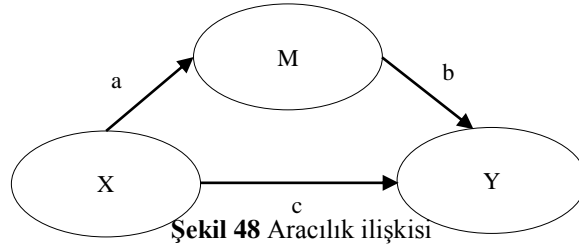
- Aracı değişken ile bağımsız değişken arasındaki ilişki anlamlı olmalıdır ( $p4, p5 \neq 0$ ). Bu durum, bağımsız değişkenin, aracı değişkenin tahmincisi olma durumudur. Eğer aracı değişken ile bağımsız değişken arasında bir ilişki yok ise aracı değişkenin mevcut modelde aracılık etkisinin kesinlikle var olmadığı kabul edilmektedir.

- Aracı değişken ile bağımlı değişken arasındaki ilişki anlamlı olmalıdır ( $p3 \neq 0$ ). Bu durum, bağımsız değişken tarafından tahmin edilen aracı değişkenin, bağımlı değişkenin tahmincisi olma durumudur. Dolayısı ile hem bağımsız değişken hem de bağımsız değişken tarafından tahminlenen aracı değişken, bağımlı değişkenin birer tahmincisidir.



Şekil 47. Araştırma modeli

Aracı değişkenin bir modelde kısmi ya da tam aracılık şeklinde iki rolü olabilir. Aracı değişken modele dahil edildiği zaman, bağımlı ve bağımsız değişkenler arasındaki ilişki anlamlılığını kaybeder ise söz konusu aracı değişken, tam aracı olarak nitelendirilir. Kısmi aracılık durumu ise bağımlı ve bağımsız değişkenler arasındaki ilişki zayıfladığı zaman gerçekleşmektedir. Aracılık etkisinin veya başka bir ifade ile bağımlı değişkenin, aracı değişken üzerinden bağımsız değişkeni dolaylı olarak etkilemesinin anlamlılığını Eş.(85) ve Şekil 48. ile tanımlanan açıklanan varyans VAF “Variance Accounted For” kriteri ile kontrol edilir. Bu oranın pay kısmı, X’in M üzerinden Y’ye olan dolaylı etkisini ve payda kısmı ise dolaylı etki ile X’in Y üzerindeki direkt etkisinin toplamını temsil etmektedir (başka bir ifade ile toplam etki). Bu yüzden, oranın pay kısmı büyüdükçe, aracılık etkisinin artarak M’nin tam aracı olmasına, küçüldükçe ise aracılık etkisinin azalmasına/yok olmasına neden olmaktadır. Alan yazınında,  $VAF < \%20$  ise aracı etkinin olmadığı,  $\%20 \leq VAF \leq \%80$  ise kısmi aracılık,  $VAF > \%80$  ise tam aracılık etkisi olduğu kabul edilmektedir.



$$VAF = \frac{ab}{(ab)+c} \quad (85)$$

Modele ilişkin; yol katsayıları ve R<sup>2</sup>' yi hesaplamak için PLS algoritması çalıştırılmıştır. PLS yol katsayılarının anlamlılıklarını değerlendirmek için yeniden örnekleme (bootstrapping) ile örneklemden 5000 alt örneklem alarak t-değerleri hesaplanmıştır. Araştırma modeline ilişkin sonuçlara Tablo (41.-43.)'te yer verilmiştir.

Aracılık etkisinin araştırılması için ilk aşamada aracı değişken doğa ile ilişki, araştırma modelinden çıkarılarak aşağıda yer alan H1 hipotez(ler)i için yol katsayılarının anlamlılığı test edilmiştir. Test sonuçları Tablo 40.'ta özetlenmiştir.

*H1: Çevresel endişeye ait biyosferik, egoistik ve altruistik tutum boyutlarının STD boyutları üzerinde anlamlı bir etkisi vardır.*

STD beş boyuta sahip olduğu için, biyosferik, egoistik ve altruistik tutumlar bağımsız değişken ve her bir STD boyutu bağımlı değişkenler olmak üzere H1 hipotezi 15 alt hipoteze sahiptir. Tablo 40.'ta yer alan sonuçlara göre H1 altındaki tüm alt hipotezlerin alternatif hipotezlerinden biyosferik tutumun GD davranışı üzerindeki etkisi, egoistik tutumun UT, ET, ST ve GT davranışları üzerindeki etkisi ve altruistik tutumun ET davranışı üzerindeki etkisi hariç olmak üzere istatistiksel olarak anlamlı bulunmuştur, yokluk hipotezleri reddedilmiştir. Dolayısı ile H1 hipotezi kısmen desteklenmiştir. Ayrıca egoistik tutum ile GD davranışı arasında pozitif yönlü bir ilişki bulunduğu ve biyosferik tutum ile GT davranışı arasında negatif yönlü bir ilişki bulunduğu dikkati çekmektedir.

**Tablo 40.** Araştırma modeli katsayıları I (Çevresel endişenin STD üzerindeki direkt etkisi)

Boyutlar		□	St.Hata	t-değeri	p
Bio	GD	0.032	0.060	0.513	<b>0.608</b>
Ego		<b>0.217</b>	0.072	3.014	0.003
Alt		0.244	0.073	3.296	0.001
Bio	UT	0.263	0.068	3.870	0.000
Ego		0.006	0.070	0.085	<b>0.932</b>
Alt		0.168	0.076	2.179	0.029
Bio	ET	0.276	0.065	4.269	0.000
Ego		0.091	0.062	1.455	<b>0.146</b>

Boyutlar		□	St.Hata	t-değeri	p
Alt		0.099	0.067	1.426	<b>0.154</b>
Bio		0.276	0.067	4.114	0.000
Ego	ST	-0.005	0.069	0.069	<b>0.945</b>
Alt		0.225	0.072	3.120	0.002
Bio		<b>-0.277</b>	0.061	4.464	0.000
Ego	GT	-0.010	0.073	0.138	<b>0.890</b>
Alt		0.485	0.063	7.631	0.000

**Not:** Kırmızı renkli p değerleri, reddedilen alternatif hipotezleri temsil etmektedir.

İkinci aşamada, aracı değişken modele dahil edilerek yol katsayılarının anlamlılığı test edilmiştir.

*H2: Çevresel endişeye ait biyosferik, egoistik ve altruistik tutum boyutlarının doğa ile ilişkinin boyutları üzerinde anlamlı bir etkisi vardır.*

DİÖ üç boyuta sahip olduğu için, biyosferik, egoistik ve altruistik tutumlar bağımsız değişken ve her bir DİÖ boyutu bağımlı değişkenler olmak üzere H2 hipotezi 9 alt hipoteze sahiptir. Tablo 41.'de yer alan sonuçlara göre H2 altındaki tüm alt hipotezlerin alternatif hipotezlerinden biyosferik tutumun deneyim üzerindeki etkisi, egoistik tutumun özbenlik üzerindeki etkisi ve altruistik tutumun perspektif ve deneyim üzerindeki etkisi hariç olmak üzere istatistiksel olarak anlamlı bulunmuştur, yokluk hipotezleri reddedilmiştir. Dolayısı ile H2 hipotezi kısmen desteklenmiştir. Ayrıca biyosferik tutum ile perspektif arasında negatif yönlü, egoistik tutum ile perspektif ve deneyim arasında ise pozitif yönlü bir ilişki bulunduğu dikkati çekmektedir.

**Tablo 41.** Araştırma modeli katsayıları II (Çevresel endişenin Doğayla ilişki üzerindeki etkisi)

Boyutlar		□	St.Hata	t-değeri	p
Bio		0.166	0.047	3.522	0.000
Ego	Özbenlik	-0.001	0.051	0.012	<b>0.991</b>
Alt		0.642	0.048	13.415	0.000
Bio		<b>-0.406</b>	0.055	7.392	0.000
Ego	Perspektif	<b>0.208</b>	0.069	2.984	0.003
Alt		-0.029	0.065	0.429	<b>0.668</b>
Bio		-0.095	0.062	1.510	<b>0.131</b>
Ego	Deneyim	<b>0.193</b>	0.077	2.505	0.012
Alt		0.079	0.076	1.018	<b>0.309</b>

**Not:** Kırmızı renkli p değerleri, reddedilen alternatif hipotezleri temsil etmektedir.

*H3: DİÖ'nün özbenlik, perspektif ve deneyim boyutlarının STD boyutları üzerinde anlamlı bir etkisi vardır.*

STD ölçeği beş boyuta sahip olduğu için özbenlik, perspektif ve deneyim, bağımsız değişken ve her bir STD ölçeği boyutu bağımlı değişkenler olmak üzere H3 hipotezi 15

alt hipoteze sahiptir. Tablo 42.'de yer alan sonuçlara göre H3 altında yer alan tüm alt hipotezlerin alternatif hipotezleri (perspektifin GD, UT ve ST davranışları üzerindeki etkisi hariç olmak üzere) istatistiksel olarak anlamlı bulunmuştur, yokluk hipotezleri reddedilmiştir. Dolayısı ile H3 hipotezi kısmen desteklenmiştir. Ayrıca perspektif ile ET davranışı arasında negatif yönlü bir ilişki bulunduğu dikkati çekmektedir.

**Tablo 42.** Araştırma modeli katsayıları III (Doğayla ilişkinin'nin STD üzerindeki etkisi)

Boyutlar		□	St.Hata	t-değeri	p
Özbenlik		0.377	0.063	6.021	0.000
Perspektif	GD	-0.077	0.053	1.416	<b>0.157</b>
Deneyim		0.380	0.051	7.372	0.000
Özbenlik		0.346	0.075	4.620	0.000
Perspektif	UT	-0.078	0.062	1.206	<b>0.228</b>
Deneyim		0.206	0.063	3.179	0.001
Özbenlik		0.363	0.074	4.858	0.000
Perspektif	ET	<b>-0.189</b>	0.062	3.010	0.003
Deneyim		0.220	0.060	3.615	0.000
Özbenlik		0.212	0.076	2.752	0.006
Perspektif	ST	-0.117	0.062	1.807	<b>0.071</b>
Deneyim		0.162	0.057	2.797	0.005
Özbenlik		0.275	0.072	3.844	0.000
Perspektif	GT	0.446	0.054	8.262	0.000
Deneyim		0.116	0.057	2.002	0.045

**Not:** Kırmızı renkli p değerleri, reddedilen alternatif hipotezleri temsil etmektedir.

Son olarak aşağıda yer alan hipotez ile dolaylı etkiler, doğayla ilişkinin çevresel endişe-STD ilişkisindeki aracı etkisi, incelenmiştir.

*H4: ÇEÖ boyutlarının, STD ölçeği boyutları üzerindeki etkisinde doğayla ilişki boyutlarının aracı rolü vardır.*

Biyosferik, egoistik ve altruistik tutum bağımsız değişkenler, her bir STD boyutu bağımlı değişkenler (beş boyut) ve Dİ üç boyuta sahip olmak üzere her bir Dİ boyutu hem bağımlı hem de bağımsız değişken (aracı) olmak üzere H4 hipotezi 45 alt hipoteze sahiptir. Tablo 43.'te yer alan sonuçlara göre: biyosferik ve altruistik tutumun özbenlik üzerinden GD davranışını, egoistik tutumun deneyim üzerinden GD davranışını, biyosferik ve altruistik tutumun özbenlik üzerinden UT davranışını, egoistik tutumun deneyim üzerinden UT davranışını, biyosferik ve altruistik tutumun özbenlik üzerinden ET davranışını, biyosferik ve egoistik tutumun perspektif üzerinden ET davranışını, egoistik tutumun deneyim üzerinden ET davranışını, biyosferik ve altruistik tutumun özbenlik üzerinden ST davranışını, biyosferik ve altruistik tutumun özbenlik üzerinden GT davranışını ve son olarak biyosferik ve egoistik tutumun perspektif üzerinden GT

davranışını dolaylı olarak etkilediği tespit edilmiştir. Dolayısı ile H4 hipotezi kısmen desteklenmiştir.

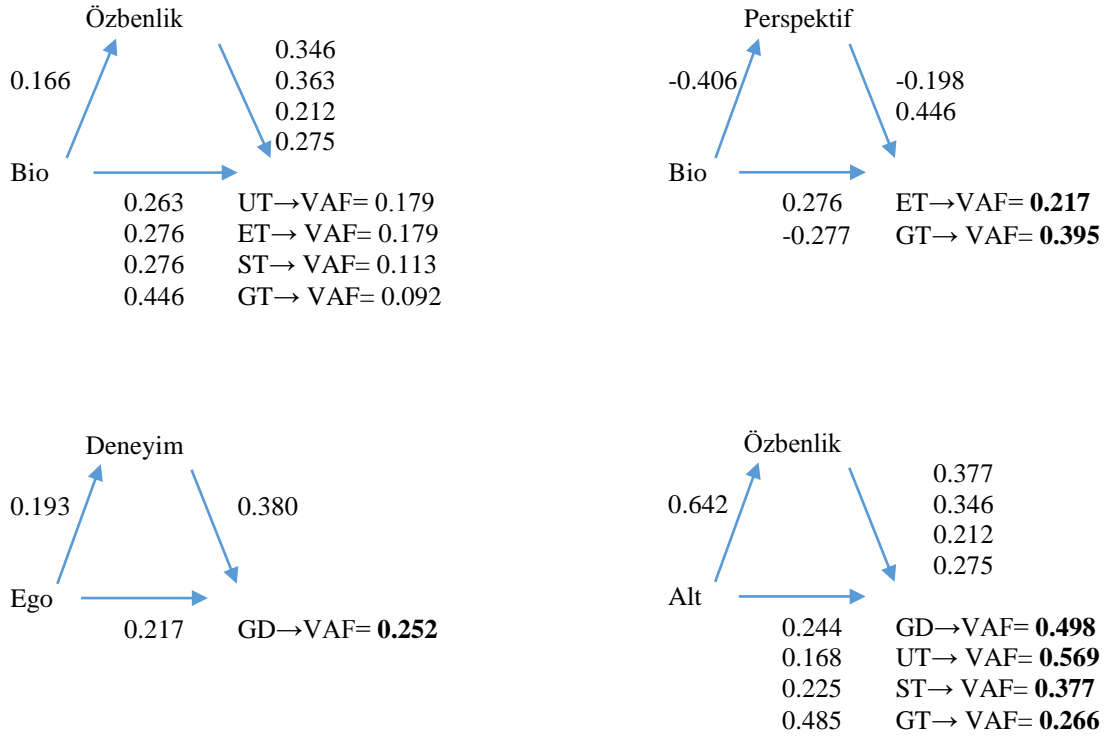
Tablo 43. incelendiğinde CEO boyutlarının, STD boyutları üzerinde doğa ile ilişki aracılığı ile toplam 17 farklı yoldan dolaylı etkisi var olduğu görülmektedir. Fakat Baron, Kenny (1986)'nin ifade ettiği üzere, aracı değişken modelden çıkartıldığı zaman bağımsız değişkenin, bağımlı değişken üzerinde bir etkisi yok ise aracı değişken modele girdiğinde aracı değişkenin aracılık edeceği herhangi bir ilişki de yoktur. Bu sebep ile aracılık etkisi sadece 11 yolda bulunmaktadır ve bu etkinin anlamlılığı ve türü (DİÖ boyutlarının kısmi aracı ve tam aracı olma durumu) Eş. (85) ile tanımlı olan VAF değeri ile incelenmiştir.

**Tablo 43.** Araştırma modeli katsayıları III (Çevresel endişenin STD üzerindeki dolaylı etkisi)

Boyutlar		$\beta$	St.Hata	t-değeri	p
Bio		0.063	0.021	3.062	0.002
Ego	Özbenlik	-0.000	0.020	0.011	<b>0.991</b>
Alt		0.242	0.044	5.470	0.000
Bio		0.031	0.022	1.364	<b>0.173</b>
Ego	Perspektif	-0.016	0.013	1.186	<b>0.236</b>
Alt		0.002	0.006	0.328	<b>0.743</b>
Bio		-0.036	0.024	1.487	<b>0.137</b>
Ego	Deneyim	0.073	0.031	2.337	0.019
Alt		0.030	0.029	1.004	<b>0.316</b>
Bio		0.057	0.021	2.804	0.005
Ego	Özbenlik	-0.001	0.018	0.011	<b>0.991</b>
Alt		0.223	0.053	4.196	0.000
Bio		0.032	0.026	1.160	<b>0.246</b>
Ego	Perspektif	-0.017	0.015	1.028	<b>0.304</b>
Alt		0.002	0.007	0.309	<b>0.757</b>
Bio		-0.020	0.015	1.259	<b>0.208</b>
Ego	Deneyim	0.039	0.020	1.959	0.050
Alt		0.017	0.018	0.880	<b>0.379</b>
Bio		0.060	0.020	3.034	0.002
Ego	Özbenlik	-0.000	0.019	0.011	<b>0.991</b>
Alt		0.234	0.053	4.321	0.000
Bio		0.077	0.027	2.786	0.005
Ego	Perspektif	-0.039	0.019	2.074	0.038
Alt		0.006	0.013	0.397	<b>0.691</b>
Bio		-0.021	0.015	1.367	<b>0.172</b>
Ego	Deneyim	0.042	0.021	2.053	0.040
Alt		0.018	0.018	0.925	<b>0.355</b>
Bio		0.035	0.017	2.100	0.036
Ego	Özbenlik	0.000	0.011	0.011	<b>0.991</b>
Alt		0.136	0.051	2.644	0.008
Bio		0.048	0.027	1.698	<b>0.090</b>
Ego	Perspektif	-0.025	0.016	1.428	<b>0.153</b>
Alt		0.004	0.009	0.355	<b>0.723</b>
Bio		-0.015	0.012	1.245	<b>0.213</b>
Ego	Deneyim	0.031	0.017	1.810	<b>0.070</b>
Alt		0.013	0.014	0.869	<b>0.385</b>
Bio	Özbenlik	0.046	0.018	2.610	0.009

Boyutlar		$\beta$	St.Hata	t-değeri	p
Ego		0.000	0.014	0.011	<b>0.991</b>
Alt		0.177	0.048	3.675	0.000
Bio		-0.181	0.031	5.767	0.000
Ego	Perspektif	0.092	0.032	2.884	0.004
Alt		-0.013	0.029	0.426	<b>0.670</b>
Bio		-0.011	0.010	1.114	<b>0.265</b>
Ego	Deneyim	0.022	0.015	1.502	<b>0.133</b>
Alt		0.009	0.011	0.808	<b>0.419</b>

**Not:** Kırmızı renkli p değerleri, reddedilen alternatif hipotezleri temsil etmektedir.



**Şekil 49.** ÇE boyutlarını, STD boyutları üzerindeki etkisinde, Dİ boyutlarının anlamlı bulunan aracılık etkileri

Şekil 49.'da yer alan VAF değerleri doğrultusunda altruistik tutumun GD, UT, ST ve GT davranışları üzerindeki etkisinde özbenliğin; biyosferik tutumun ET ve GT davranışları üzerindeki etkisinde perspektifin ve son olarak egoistik tutumun GD davranışı üzerindeki etkisinde deneyimin kısmi aracı etkisi görülmektedir.

Modele ait elde edilen düzeltilmiş  $R^2$  değerleri incelendiğinde özbenliğin %56, perspektifin % 12, deneyimin %4, GD davranışının %39, UT davranışının %24, ET davranışının %27, ST davranışının %24 ve GT davranışının %40 oranında açıklandığı; aracı değişken doğayla ilişkinin modelde olmadığı durumda ise GD davranışının %19, UT davranışının %15, ET davranışının %16, ST davranışının %20 ve GT davranışının

%14 oranında açıklandığı tespit edilmiştir. Dolayısı ile, doğayla ilişkinin modele aracı değişken olarak eklenmesi, STD boyutlarının açıklanan varyansını arttırmıştır.

## **5. PROFİL ÇIKARMA**

Bu aşamanın amacı, Bursa il merkezindeki vatandaşların yeşil kimlik profillerini keşfetmektir. Örneğin vatandaşların sürdürülebilir tüketim kalıplarını, bu kalıpların çevresel endişe, ekolojik perspektif, ekolojik özbenlik ya da doğadaki deneyim ile uyumu ya da uyumsuzluğunu, gelir, hane halkı türü ya da belirli bir ilçede yaşayan vatandaşların diğer vatandaşlardan ayrışmasını sağlayan bir tipoloji olup olmadığını incelemek gibi. Bu sebeple bu aşamada bin gözlemlik veri setinin tamamı kullanılmıştır.

Bu aşamadaki analizler gerçekleştirilmeden önce, bazı kontroller yapılarak veri seti temizlenmiştir. Örneğin, hane halkının türü olarak çocuklu orta yaşlı çift ya da çocuklu genç çift seçildi ise hanede 0-18 yaş aralığında bir çocuk olması, yok ise 18 yaş üstü nüfusun en az üç olması, tek başına yaşayan genç-orta yaşlı-yaşlı hane halklarında nüfus değişkeninin değerinin bir olması, çocuksuz hanelerde çocuk olmaması ve benzeri gibi durumlar kontrol edilmiş ve tutarsız gözlemler satır bazında işlem yapılarak tamamen silinmiştir. Ayrıca, 111 TL, 222 TL 333 TL ve benzeri gibi elektrik, su ve doğal gaz tüketim değerlerine sahip gözlemler de kontrol edilerek satır bazlı silinmiştir. Ek olarak aylık geliri 200,000 TL olup su ve enerji tüketim meblağı 100-200 TL arasında olan tek bir gözlem de veri setinden çıkartılmıştır. Söz konusu kontrol ve temizleme işlemleri MS Excel'de gerçekleştirildikten sonra veri setinde 748 gözlem kalmıştır.

### **5.1. Veri İndirgeme (Boyut Azaltma)**

Joseph F. Hair Jr., vd. (2014) çalışmasında belirtildiği üzere AFA, gözlemlenebilir değişkenler arasındaki ilişkiyi ve var ise bu değişkenler topluluğunun oluşturduğu gizil yapıyı tanımlamada kullanılmaktadır. Fakat AFA'nın bir diğer kullanım amacı verinin boyutunu indirgemektir. Eğer amaç, veri indirgemek ve sonuçlarını diğer istatistiksel yöntemlerde girdi olarak kullanmak ise üç yaklaşım bulunmaktadır. Bu yaklaşımlar, her bir faktörde en büyük yüke sahip olan bir ifadenin faktör temsilcisi olarak kullanılması "surrogate variable", faktörü oluşturan ifadelerin aritmetik ortalamasının "summed scale" kullanılması ve faktör skorlarının "factor scores" kullanılması şeklindedir. Söz konusu yaklaşımlar ile indirgenen veriler örneğin regresyon ya da diskriminant analizinde

bağımsız değişken, çok değişkenli varyans analizinde bağımlı değişken ya da küme analizinde kümeleme değişkenleri olarak kullanılabilir.

Bu tez çalışmasının, birinci ve ikinci ana aşamalarında geçerlik ve güvenilirlikleri test edilen ve nihai yapıya kavuşturulan ölçeklerin boyutları (faktörler), küme analizinde birer öznitelik olarak kullanılmıştır. Faktörleri temsil eden özniteliklerin girdi değerleri ise yukarıda açıklanan “summed scale” toplanmış değerler ölçeği yaklaşımı ile faktörü oluşturan ifadelerin aritmetik ortalaması alınarak hesaplanmıştır. Bu uygulama ile ölçeklerde yer alan 60 gözlemlenebilir değişken, 11 (gizil) değişkene indirgenmiş ve bu değişkenler, küme analizinde birer öznitelik olarak kullanılmıştır.

Joseph F. Hair Jr., vd. (2014), bu yaklaşımın kullanılması durumunda güvenilirlik ve geçerlik testi yapılması gerektiğini ifade etmektedir. Toplanmış değerler ölçeğinin güvenilirliği için öncelikle ilgili ölçeğe açılımlı faktör analizi uygulanarak faktörlerin tek boyutluluğu ve Cronbach alfa ile iç tutarlılığı test edilmelidir. Sonrasında birleşme ve ayrışma geçerlikleri incelenmelidir. Birleşme geçerliği, söz konusu ölçeğin benzer ölçekler ile korelasyonu, ayrışma geçerliği ise ilgili alan yazınındaki diğer ölçeklerden farklılığı test edilerek incelenmelidir.

Dolayısı ile Kısım 4’te yer alan güvenilirlik ve geçerlik analizleri kabul edilebilir sonuçlar sağladığı için, toplanmış değerler yaklaşımının bu tez çalışmasında kullanılabilmesi adına gerekli olan güvenilirlik ve geçerlik testleri tekrardan gerçekleştirilmemiştir.<sup>151</sup>

## **5.2. Veri Dönüştürme**

Bu alt aşamada, nümerik nitelikte olan nüfus, meskenin gün içi kullanım süresi (uyku dahil), meskendeki salon dahil oda sayısı, gelir ve bir yıl içinde ödenmiş en yüksek elektrik, su ve doğalgaz fatura tutarı öznitelikleri ve faktörleri temsil eden öznitelikler 2. bölüm kısım 2.3.2.3 ‘te açıklanan ve Eş. (8) ile tanımlanan, z-score normalleştirme yöntemi ile standardize edilmiştir. Standardizasyondaki amaç, özniteliklerin değerlerini benzer aralığa çekerek, uzaklık hesabında örneğin (1000) bin’ler ile ölçülen “gelir” özniteliği, (1-5) aralığında ölçülen faktör öznitelikleri veya 20-800 aralığında değişen

---

<sup>151</sup> Zaman ve bütçe kısıtları sebebi ile ölçeklerin, benzer ölçekler ile korelasyonuna bakılmamıştır. Kısım 4’te yer alan birleşme geçerliği test sonuçları dikkate alınmıştır.



doğalgaz tüketim öznitelikleri arasında bir denge sağlamaktır. Veri dönüştürme işlemi için R-Studio yazılımı kullanılmıştır.

### 5.3.Küme Analizi

Bu alt aşamada, önişleme tabii tutulmuş olan ana araştırma veri seti Wehrens, Kruisselbrink (2018a, 2018b) tarafından önerilen SÖÖH ile küme analizine tabii tutulmuştur. Gerekli tüm analizler R-Studio ve PRIMER-e yazılımları kullanılarak gerçekleştirilmiştir.

Kohonen ağlarının özel bir türü olan SÖÖH'ler ile küme analizinde üç katmanlı bir ağ yapısı oluşturulmuştur. İlçe, mesken türü ve hane halkı türü kategorik öznitelikleri, "one hot encoding" yöntemi ile ikili özniteliklere dönüştürülerek elde edilen yeni öznitelikler birinci katmana yerleştirilmiştir.<sup>152</sup> İkinci katmana, orijinalinde ikili türde olan bakım, bebek, çocuk, ergen ve yaşlı öznitelikleri; üçüncü katmana ise nümerik türde olan gelir, elektrik, su ve doğalgaz tüketim bedelleri, meskendeki oda sayısı, hane halkı büyüklüğü, meskenin gün içi kullanım süresi ve AFA'nın çıktısı olan faktörleri temsil eden öznitelikler yerleştirilmiştir. Dolayısı ile birinci ve ikinci katmanda tanimoto mesafesi<sup>153</sup> ve üçüncü katmanda öklit mesafesi kullanılmıştır.

Oluşturulan katmanlar, sahip olduğu öznitelik sayısı ile doğru orantılı olarak ağırlıklandırılmıştır.<sup>154</sup> Birinci, ikinci ve üçüncü katmanların ağırlıkları sırası ile 0.425, 0.125 ve 0.45 şeklindedir.

Veri setinin kontrolü ve veri indirgeme ve dönüştürme işlemlerinden sonra elde edilen (748x40) boyutlu yeni veri setinin gözlem sayısına bağlı olarak çeşitli büyüklükte haritalar oluşturulmuştur. Veri setinde 748 gözlem olduğu için haritalardaki nöron sayıları aşağıda yer alan kurallar çerçevesinde tespit edilmiştir:

---

<sup>152</sup> Böylelikle, profil çıkarmanın kimlik tanımlama aşamasında kullanılacak olan ikili dallanan CART ağaçları için de veri seti hazır hale getirilmiştir.

<sup>153</sup> Tanimoto mesafesinin alan yazınında çeşitli versiyonları bulunmaktadır. Bu tez çalışmasında Wehrens, Kruisselbrink (2018a)'in önerdiği, iki ikili vektörün birbirine benzeşme oranı olan versiyonu kullanılmıştır. Bu oran ikinci bölüm kısım 5.3'te simetrik ikili benzeşmezlik ölçütü olarak yer almaktadır.

<sup>154</sup> Rough training süresinde gerçekleştirilen denemelerde katman ağırlıklandırılması yapılmadığı zaman küme bütünlüğü olmayan küme bölümlenmeleri elde edilmiştir. Örneğin bir kümenin haritanın farklı lokasyonlarında olacak şekilde parçalı konumlanması gibi.

- Gözlem sayısından küçük olmalı (Beccali, vd., 2004).
- Nöron sayısı, nöron başına yaklaşık 50 adet gözlem düşecek şekilde olmalı (Kohonen, 2013).
- Ayrıca  $\sqrt{748} = 27.349$  olduğu için, haritaların  $x$  ve  $y$  boyutları (2:27) nöron aralığında olmalı (Hagan, vd., 1996).

Yukarıda yer alan bir numaralı kural, büyük haritaların oluşmasına ve dolayısı ile eğitilmiş haritalarda boş nöron olmasına neden olmaktadır. Bu sebeple, bir haritadaki nöron sayısı, nöron başına 5 gözlem düşecek şekilde bir üst limit ile sınırlandırılmıştır. Üst limit,  $\frac{748}{5} = 149.6$  eşitliğinden, 149 adet olarak hesaplanmıştır. Diğer taraftan, iki numaralı kural ise çok küçük haritaların oluşmasına neden olarak, kümelerin sağlıklı ve detaylı bir şekilde ayrışmasına engel olabilmektedir. Bu sebeple bir haritadaki minimum nöron sayısı,  $\frac{748}{50} = 14.96 \cong 15$  eşitliğinden, 16 adet ile sınırlandırılmıştır. Sonuç olarak sayılan kurallar çerçevesinde birbirinden farklı ebatlarda (Wehrens, Kruisselbrink, 2018a) 293 adet harita oluşturulmuştur. Bu haritalardan en küçüğü 16 ve en büyüğü ise 147 nörona sahiptir.

Komşuluk yarıçapı, haritadaki nöron sayısının  $3/4$ 'ünden başlayıp,  $1/4$ 'üne kadar incek şekilde ayarlanmıştır. Yarıçapın alt sınırı özellikle sıfırlanmamıştır. Çünkü sıfır alt limitli yarıçaplarda algoritma önceki iterasyonlarda öğrendiği bilgiyi unutarak sadece hedef nöronun kendisini güncellemektedir. Bu durum, eğitimin performansını düşürmektedir.

Haritaların nöron komşuluk alanlarının yapısı altıgen dizilim olarak tasarlanmış ve Gauss komşuluk fonksiyonu kullanılmıştır (Kohonen, 1989).

Çevrimiçi öğrenme algoritması için öğrenme oranı 0.05'ten başlayıp, 0.01'e incek şekilde ayarlanmıştır (Haykin, 2009).

İterasyon sayısı, çevrimiçi öğrenme algoritması ile eğitilen haritalarda nöron sayısının 500 katı olarak (Su, Liu, Chang, 2002; Kohonen, 2001) ayarlanmıştır. Dolayısı ile haritaların en küçüğü 16 nörona ve en büyüğü 147 nörona sahip olduğu için çevrimiçi öğrenme algoritması ile eğitilen haritaların iterasyon sayıları  $16 * 500 = 8,000$  ile  $147 * 500 = 73,500$  arasında değişkenlik göstermiştir. Diğer taraftan yığın öğrenme

algoritması ile eğitilen bütün ağların iterasyon sayısı, nöron sayısından bağımsız bir şekilde sabit (Kohonen, 2001; Cottrell, vd., 2016; Hagan, vd., 1996) 40 iterasyon olarak ayarlanmıştır.

Eğitim süreci çevrimiçi eğitilen haritalarda 36.8 saat, yığın eğitilen haritalarda ise 2.5 dakika sürmüştür. Eğitim sonrasında,<sup>155</sup> haritalardan kare olanlar ve ayrıca boş nörona sahip olanlar değerlendirmeye alınmamıştır (Kohonen, 2013). Kare veya boş nörona sahip olan eğitilmiş haritalar elendikten sonra çevrimiçi öğrenme algoritması ile eğitilen 293 haritadan geriye 15 harita kalmıştır. Bu haritaların en küçüğü 16 adet, en büyüğü ise 36 adet nörona sahiptir. Yığın öğrenme algoritması ile eğitilen 293 haritadan ise geriye 58 harita kalmıştır. Bu haritaların en küçüğü 16 adet, en büyüğü ise 42 adet nörona sahiptir.

Küme analizinde değerlendirilecek olan haritalar tespit edildikten sonra, sırası ile çevrimiçi ve yığın öğrenme algoritmaları ile eğitilen haritalardan seçilmiş olan üçer haritaya ve çıktılarına aşağıdaki prosedür uygulanmıştır:

1. Hücrelere düşen gözlem adedi “counts plot”, nöron komşuluk mesafelerini içeren U matrisi “U-matrix plot” ve kazanan nöron ile kendisine atanmış olan gözlemler arasındaki ortalama mesafe “quality plot” kullanılarak ısı haritaları oluşturulmuştur (Hongyan Liu, vd., 2011). İlk ısı haritası ile gözlemlerin harita üzerindeki saçılımları, ikinci ısı haritası ile haritadaki olası küme sayısı ve üçüncü ısı haritası ile gözlemlerin haritadaki nöronlar tarafından temsil edilip edilemediği incelenmiştir.

2. ÖÖH’lerde birbirine benzeşen gözlemlerin atandığı nöronların ağırlıklarının da birbiri ile benzeştiği belirtilmektedir. En ideal şartlar altında, ağırlıkları birbiri ile benzeşen nöronlar, topolojik olarak, haritanın belirli bir bölgesine toplanarak kümelenmelidir. Bu bağlamda nöron ağırlıkları da ayrı bir küme analizine tabi tutulabilir. Böylelikle, birinci aşamada SÖÖH’ler ile sezgisel olarak tahmin edilen küme sayısı,

---

<sup>155</sup> Bu aşamada aslında hem rough training hem de fine tuning işlemleri gerçekleştirilmiştir. Fakat rough training sonucunda elde edilen nöron ağırlıkları fine tuning aşamasında girdi olarak kullanıldığında (Nova, Estevez, 2012) fine tuning işlemi gerçekleştirilememiştir. Çünkü söz konusu nöron ağırlıkları ile girdi vektörleri arasındaki ortalama mesafe sıfır olarak tespit edilmiştir. Başka bir ifade ile eğitilmiş haritalarda fine tuning prosedürünü gerektirecek bir iyileştirme tespit edilememiştir. Bu sonuç, komşuluk yarıçapının sıfırlanmamasından kaynaklanmaktadır. Zira rough training için gerçekleştirilen eğitimlerde sadece komşuluk yarıçapı alt sınırının sıfırlanması durumunda, fine tuning işlemine gerek kalmamıştır.

netleştirilebilir. Dolayısı ile bu aşamada nöronlar gözlem olmak üzere, nöron ağırlıkları Ward bağlantılı hiyerarşik kümeleme ve öklit mesafesi kullanılarak  $k$ -temsili kümeleme algoritmaları ile kümelendi. Olası küme sayısının belirlenmesinde ( $k - KT$ ) grafiği, Silüet, Davies-Bouldin, Dunn ve Calinski-Harabasz endeksleri içsel değerlendirme kriterleri olarak kullanılmıştır. Kriterler dikkate alınarak en uygun küme bölümlenmesi tespit edilmeye çalışılmıştır.

3. Eğer, ikinci aşamada en uygun bölümlenmesi net bir şekilde tespit edilemedi ise aday bölümlenmeler, ait olduğu harita üzerine yansıtılarak, küme bütünlükleri kontrol edilmiştir. Bu aşamada, küme bütünlüğünü sağlayamayan küme bölümlenmeleri elenmiştir.

4. Akabinde, nöron ağırlıkları TBA'ya tabi tutulmuş ve ikinci aşamada elde edilen olası küme bölümlenmeleri temel bileşenler grafiğine yansıtılmıştır. Bu işlem iki amaç ile yapılmıştır. İlk amaç, birinci ve ikinci temel bileşenlerin varyansın ne kadarlık kısmını açıklayabildiği bilgisine ulaşarak bu bilgi ışığında haritaları karşılaştırmaktır (varyansı daha fazla açıklamaya imkan sağlayan harita tercih edilmelidir). İkinci amaç ise gözlemlerin ve küme bölümlenmelerinin saçılımını inceleyerek var ise örtüşen/çakışan/içiçe geçen küme bölümlenmelerine sahip olan seçenekleri tespit edip elemektir.

5. Sayılan dört aşama sonucunda uygun küme sayısı tespit edilmiş olmalıdır. Küme sayısının tespit edilememesi durumunda ise aday küme bölümlenmeleri çok değişkenli varyans analizine tabi tutulmuştur. Amaç, küme bölümlenmelerinin grup içi ve gruplar arası değişkenliğini ( $F$  değeri ile) karşılaştırmaktır (Hiller, 2014; Ribeiro, Zárate, 2019). Bu aşamada kullanılacak olan verinin gözlemleri, ikinci aşamada kullanılan nöronlar; kümeleme değişkenleri (bağımlı değişkenler), SÖÖH'lerde kullanılan öznitelikler; gözlem değerleri, nöronların her bir öznitelik için sahip olduğu ağırlıklar ve gruplar (bağımsız değişkenler) ise her bir gözlemin (nöronun) küme etiketidir.

Haritalardaki nöron sayısı 40'tan az olduğu için, kümelenecek gözlem sayısının da 40'tan az olması ve ayrıca elde edilen küme bölümlenmelerinin eşit sayıda gözlem (nöron) içermemesi, beşinci aşamanın uygulanması gerektiği durumlarda permütasyona dayalı parametrik olmayan çok değişkenli varyans analizinin kullanılmasını gerekli kılmıştır.

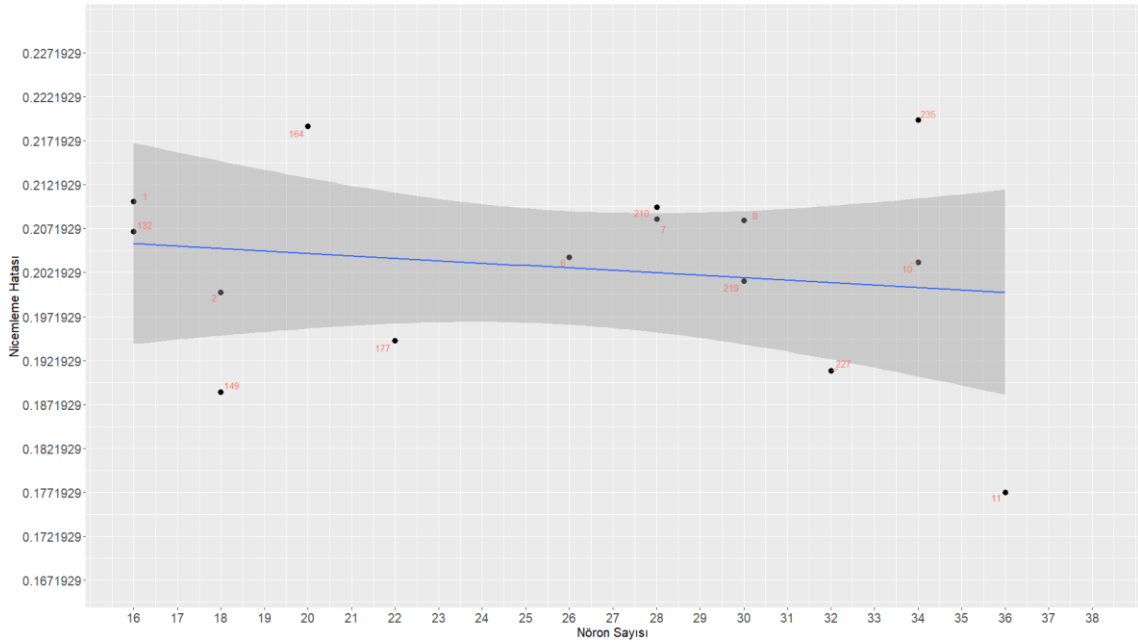
Yukarıda açıklanan beş aşamalı prosedürün uygulanması ile çevrimiçi öğrenme algoritması kullanılarak eğitilen haritalardan sadece bir tanesi kimlik tanımlama aşamasında kullanılmak üzere birinci aday olmuştur.

Bu prosedür aynı zamanda yığın öğrenme algoritması ile eğitilen haritalara da uygulanmış ve içlerinden sadece bir tanesi ikinci aday olarak tespit edilmiştir.

Son olarak iki aday harita karşılaştırılmış ve beş aşamalı prosedürden en iyi performansı sağlayan harita kimlik tanımlama aşamasında kullanılmak üzere seçilmiştir.

### 5.3.1. Çevrimiçi Öğrenme Algoritması ile Eğitilen Haritaların Analizleri

Çevrimiçi öğrenme algoritması ile eğitilen haritalardan eleme sonrası kalan 15 harita, öncelikle nicemleme hataları açısından incelenmiş<sup>156</sup> ve en küçük üç NH'ye sahip harita küme analizinde değerlendirilmiştir.



**Şekil 50.** Çevrimiçi öğrenme algoritmaları ile eğitilen haritalardan eleme sonrası kalan 15 haritanın nicemleme hatasının nöron sayısına göre değişimi

Şekil 50.'den inceleneceği üzere 11 numaralı harita en küçük NH'ye sahip olan haritadır. Haritanın boyutları  $(x, y) = (2, 18)$  olup toplam nöron sayısı 36 ve NH 0.1771'dir. İkinci en küçük NH'ye sahip olan harita 149 numaralı harita olup, boyutları

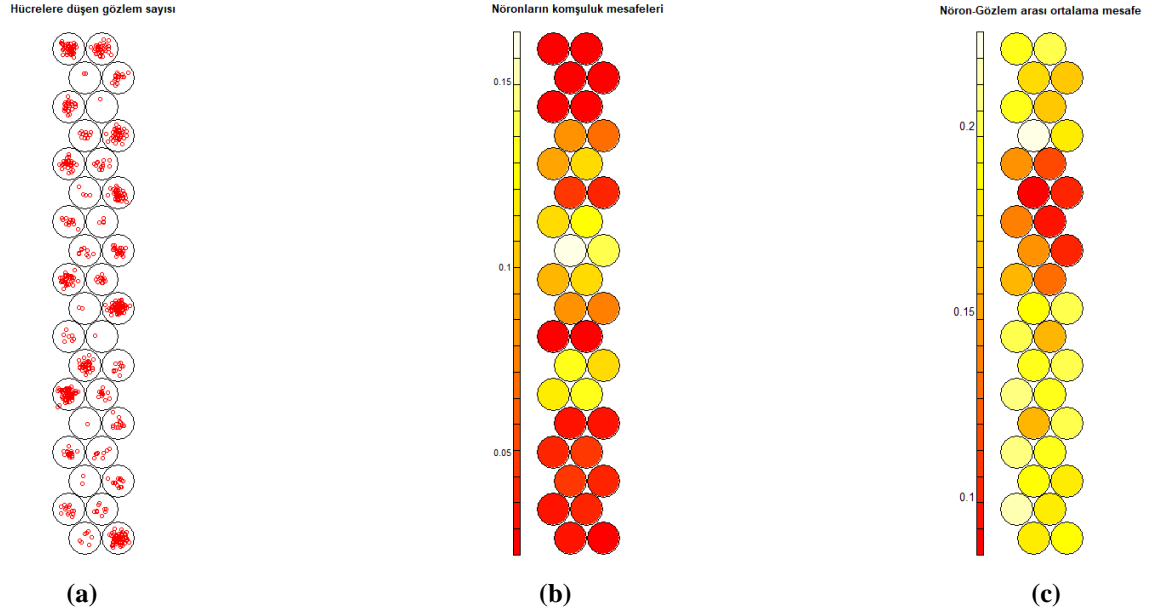
<sup>156</sup> Çevrimiçi öğrenme algoritması ile eğitilmiş olan 293 haritanın nicemleme hatasının nöron sayısına göre değişimine ait grafik Ek 8'den incelenebilir.

$(x, y) = (9,2)$  ve toplam nöron sayısı 18'dir. Bu haritanın NH ise 0.1885'tir. NH üçüncü en küçük olan harita ise 227 numaralı harita olup, boyutları  $(x, y) = (16,2)$ , toplam nöron sayısı 32 ve NH 0.1910'dur.

En küçük NH'ye sahip olarak seçilen haritaların boyutları incelendiğinde, bu haritaların çok net bir şekilde enine ya da boyuna uzayan dikdörtgen bir şekle sahip olduğu dikkati çekmektedir. Başka bir ifade ile bu haritaların şekli, kareye yakın ya da bir boyutu diğer boyutunun iki ya da üç katı gibi standart bir dikdörtgen değildir. Aksine, nerede ise tek boyutlu bir Kohonen ağı olabilecek kadar dardır.<sup>157</sup>

### 5.3.1.1. Isı haritaları ile görsel değerlendirme

**Harita 11.** için Şekil 51(a). incelendiğinde, gözlemler haritanın alt, orta ve üst kısmında yoğunlaşmıştır. Şekil 51(c).’ye göre, nöronlar ile kendilerine atanan gözlemler arasındaki ortalama mesafe (0.08 – 0.22) bandında değişmektedir. Şekil 51(b). incelendiğinde ise haritanın alt ve üst kısımlarında nöronlar arası komşuluk mesafelerinin birbirine çok yakın olduğu ve orta kısmının belirsiz bir şekilde iki ya da üç gruba ayrıldığı ifade edilebilir. Sonuç olarak, Harita 11. için toplam küme sayısı dört veya beş olabilir.

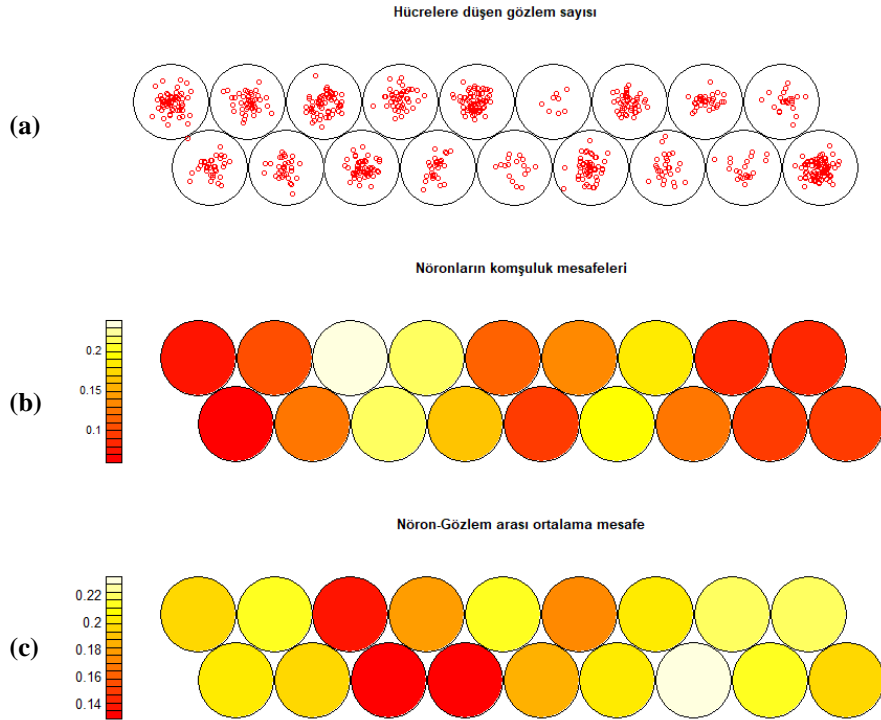


**Şekil 51.** Harita 11'in (a) hücelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe

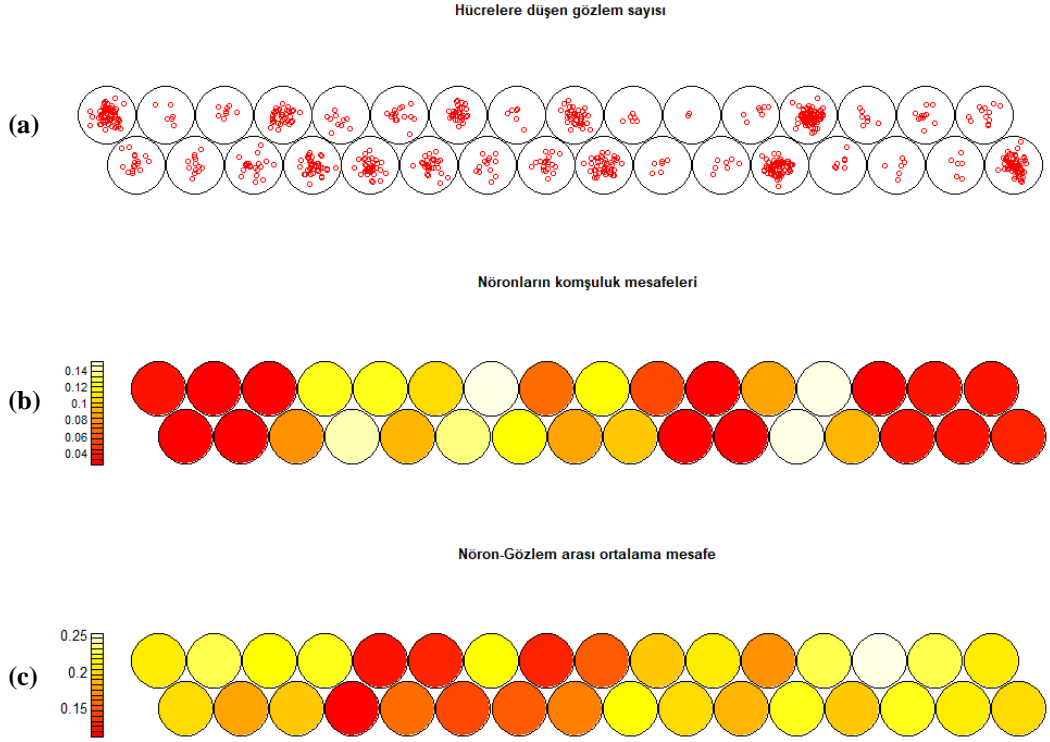
<sup>157</sup> Diğer 12 haritanın ebatlarına ait tezde bilgi paylaşılmamış olsa da bu durumun diğer 12 harita için de geçerli olduğunu belirtmekte fayda bulunmaktadır.

**Harita 149.**'da gözlemler, 5 ve 15'inci nöronlar sınır olmak üzere haritanın iki tarafına bölünmüştür, Şekil 52(a). Şekil 52(c).’ye göre, nöronlar ile kendilerine atanan gözlemler arasındaki ortalama mesafenin 11 numaralı haritadakine benzer bir şekilde, (0.13 – 0.23) bandında değişmektedir. Şekil 52(b). incelendiğinde ise haritanın sağ ve sol kısımlarında nöronlar arası komşuluk mesafelerinin birbirine çok yakın olduğu ve orta kısmının üç gruba ayrıldığı ifade edilebilir. Sonuç olarak, harita 149 için toplam küme sayısı beş olabilir.

**Harita 227.**'de ise gözlemler, haritanın iki ucunda ve öbek öbek orta kısımlarında gruplaşmıştır, Şekil 53(a). Nöronlar ile kendilerine atanan gözlemler arasındaki ortalama mesafe (0.17 – 0.25) bandında değişmektedir, Şekil 53(c). Şekil 53(b) dikkate alındığında ise nöronlar arası komşuluk mesafelerinin haritanın üç kırmızı bölgesinde birbirine çok yakın olduğu ve sarı-turuncu tonlarındaki kısımların iki bölgeye ayrılacağı ifade edilebilir. Sonuç olarak, harita 227 için toplam küme sayısı beş olabilir.



**Şekil 52.** Harita 149'un (a) hücelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe



**Şekil 53.** Harita 227'nin (a) hücelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe

Harita 11., Harita 149. ve Harita 227. eş anlı değerlendirildiğinde nöron-gözlem arası mesafeler bağlamında, veri setini en iyi temsil eden nöronlara sahip olan harita, Harita 11.'dir. Çünkü nöron-gözlem arası en küçük ortalama mesafeye sahiptir (0.08-0.22), Şekil 51(c).<sup>158</sup> Birbiri ile benzeşmezliği düşük olan gözlemlerin birbirine yakın ve yüksek olan gözlemlerin birbirine uzak nöronlarda toplanarak, yakın nöronlar arası mesafenin olabildiğince küçük ve uzak nöronlar arası mesafenin ise olabildiğince büyük değerler almasını sağlayan harita ise 227 numaralı haritadır.<sup>159</sup> Çünkü Harita 227., diğer haritalara kıyas ile yakın nöronlar arası en küçük ve uzak nöronlar arası en büyük mesafeye sahiptir (0.02 – 0.15). Bu değerler, U-matrisi değerleridir.

### 5.3.1.2. Nöron ağırlıklarının kümelenmesi

Çevrimiçi öğrenme algoritması ile eğitilen haritalardan, küme analizinde değerlendirilmek üzere seçilen üç haritanın içerdiği maksimum nöron sayısı 36'dır.

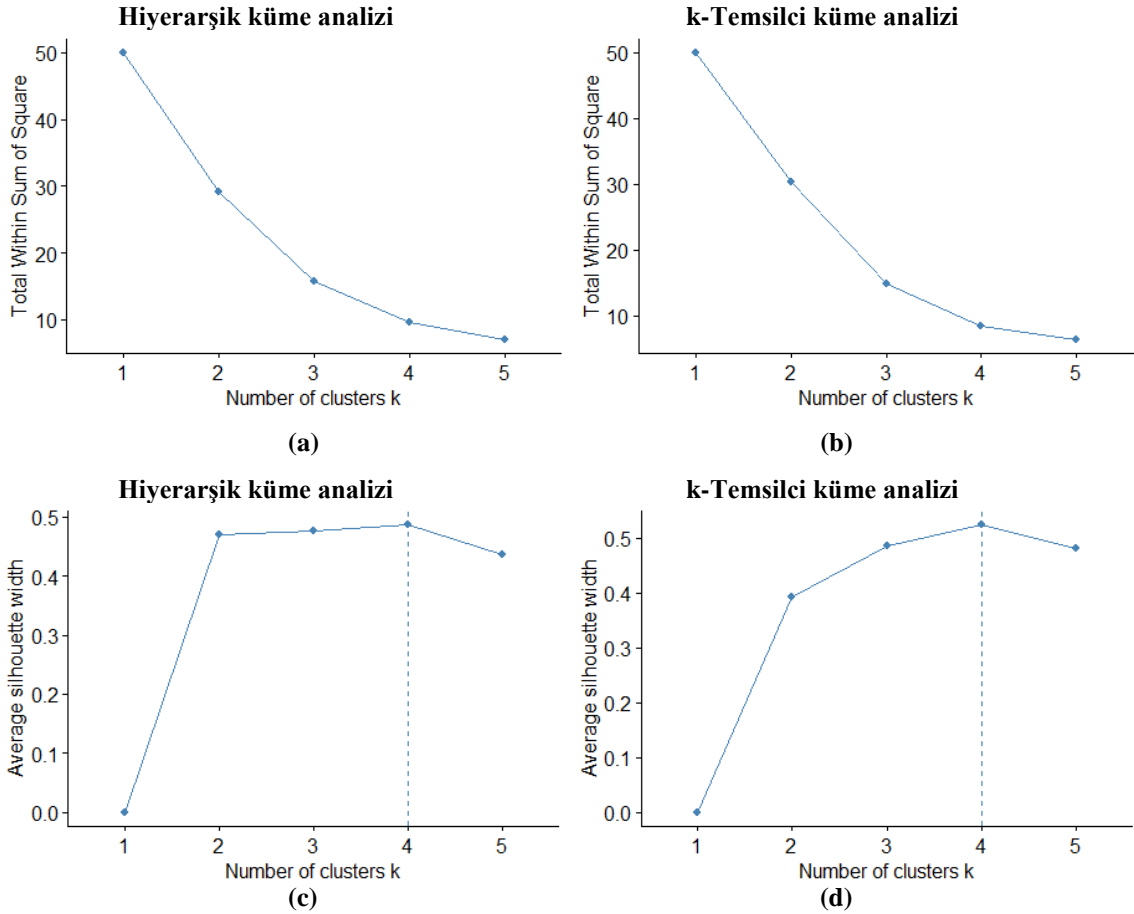
<sup>158</sup> Harita 149 için ortalama mesafe (0.13-0.23) ve harita 227 için (0.13-0.25)'tir

<sup>159</sup> U-matrisi değerleri.



Nöron ağırlıkları  $k = 2,3,4$  ve 5 olarak hem Ward bağlantılı hiyerarşik hem de k-temsilci kümeleme algoritmaları ile kümelendi.

**Harita 11.**'in hem hiyerarşik hem de k-temsilci algoritmaları ile küme analizi sonuçları Şekil 54.'te görselleştirilmiş ve Tablo 44.'te özetlenmiştir. Şekil 54(a-b)'den inceleneceği üzere, küme içi KT dikkate alındığında, haritadaki nöronlar dört kümeye ayrılabilir. Kümeleme algoritmalarının performansları karşılaştırıldığında ise k-temsilci algoritması  $k = 4$  için daha küçük küme içi KT sağlamıştır ( $KT = 8.4242$ ).<sup>160</sup> Şekil 54(c-d)'de yer alan ortalama silüet endeksinin küme sayısına göre değişim grafiği incelendiğinde haritalardaki nöronlar kesinlikle dört kümeye ayrılabilir ve k-temsilci algoritması ile elde edilen 0.5237 ortalama silüet endeksi, hiyerarşik küme analizi ile elde edilen ortalama silüet endeksinden fazladır (0.4864).



**Şekil 54.** Harita 11 için (a-b) küme içi kareler toplamının küme sayısına göre değişimi ve (c-d) ortalama silüet endeksinin küme sayısına göre değişimi

<sup>160</sup> Hiyerarşik küme analizinde  $k = 4$  için  $KT = 9.6561$

Tablo 44. incelendiğinde ise en küçük Davies-Bouldin endeksini,  $k = 4$  için 0.5913 değeri ile k-temsilci algoritması sağlamıştır. En büyük Dunn endeksini,  $k = 2$  için 0.2861 değeri ile hiyerarşik küme analizi sağlamıştır. Son olarak en büyük Calinski-Harabasz endeksini,  $k = 5$  için 52.6938 değeri ile k-temsilci küme analizi sağlamıştır. Fakat bu endeks için dikkat çeken bir önemli nokta  $k = 4$  için hesaplanan değer 52.6127,  $k = 5$  için hesaplanan değere çok yakın olmasıdır.

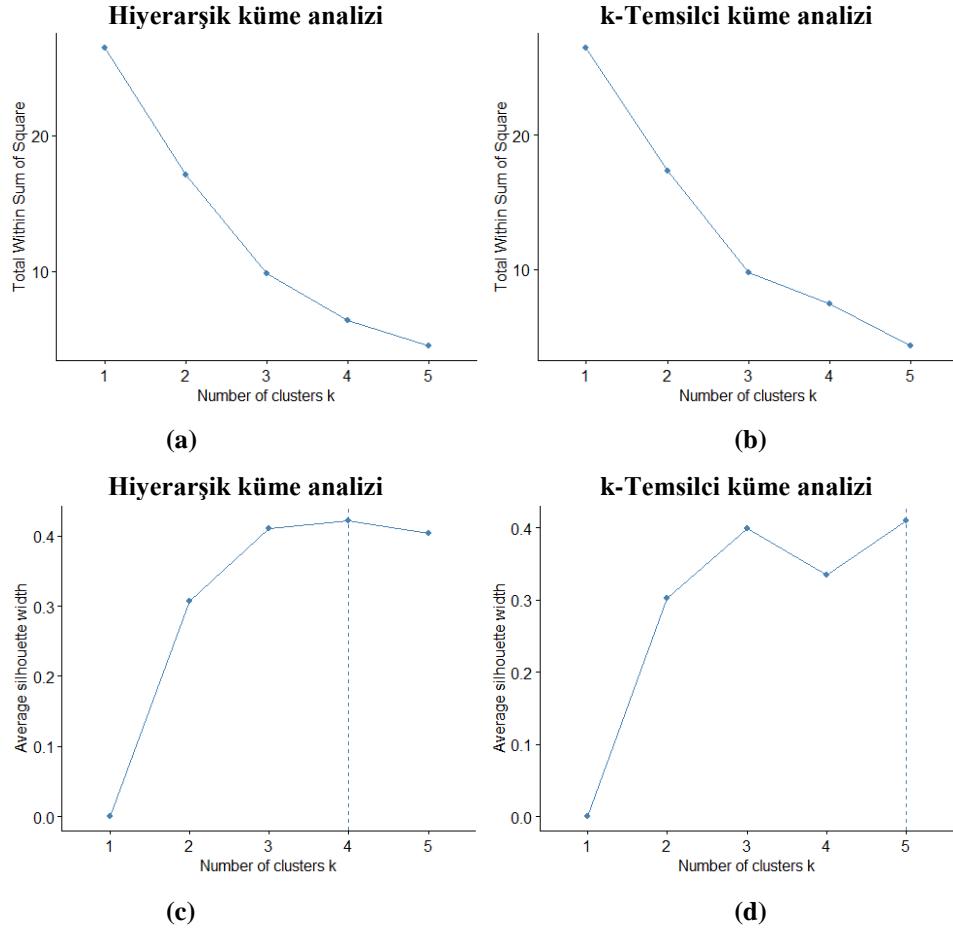
**Tablo 44.** Harita 11 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri

Endeks	Kümeleme Algoritması	k=2	k=3	k=4	k=5
Davies-Bouldin	Hiyerarşik	0.6744	0.7707	0.6183	0.6062
	k-Temsilci	1.0781	0.7572	<b>0.5913</b>	0.6244
Dunn	Hiyerarşik	<b>0.2861</b>	0.1738	0.2153	0.2352
	k-Temsilci	0.1690	0.1353	0.2360	0.1981
Calinski-Harabasz	Hiyerarşik	24.1920	38.7705	44.0995	47.2064
	k-Temsilci	21.9038	39.2048	<b>52.6127</b>	<b>52.6938</b>

Şekil 54. ve Tablo 44. dikkate alındığında, harita 11 için  $k = 2$  sadece bir (hiyerarşik-Dunn),  $k = 4$  topamda dört (k-temsilci-KT, k-temsilci silüet, k-temsilci-Davies Bouldin ve k-temsilci-Calinski Harabasz) ve  $k = 5$  bir defa (k-temsilci-Calinski Harabasz) en uygun aday küme sayısı olarak tespit edilmiştir.  $k = 4$  için k-Temsilci algoritmasının sağladığı küme bölümlenmesi çoğunlukla en iyi endeks değerlerine sahip olduğu için, harita 11'in en uygun küme sayısının dört olduğuna karar verilmiştir.

**Harita 149.**'un hem hiyerarşik hem de k-temsilci algoritmaları ile küme analizi sonuçları Şekil 55.'te görselleştirilmiş ve Tablo 45.'te özetlenmiştir. Küme içi KT dikkate alındığında, haritadaki nöronlar hiyerarşik kümeleme ile dört ( $KT = 6.3304$ )<sup>161</sup>, k-temsilci kümeleme ile beş kümeye ( $KT = 4.3195$ ) ayrılabilir. Şekil 55(c-d)'de yer alan ortalama silüet endeksinin küme sayısına göre değişim grafiği incelendiğinde ise hiyerarşik kümelemede  $k = 4$  için 0.41 silüet endeksi ve k-temsilci kümelemede  $k = 5$  için 0.42 ortalama silüet endeksi, elde edilmiştir.

<sup>161</sup> Hiyerarşik küme analizinde  $k = 5$  için  $KT = 4.4840$



**Şekil 55.** Harita 149 için (a-b) küme içi kareler toplamının küme sayısına göre değişimi ve (c-d) ortalama silüet endeksinin küme sayısına göre değişimi

Tablo 45. incelendiğinde ise  $k = 5$  için en küçük Davies-Bouldin ve En büyük Dunn endeksini, 0.6619 ve 0.4907 değeri ile hiyerarşik kümeleme algoritması sağlamıştır., En büyük Calinski-Harabasz endeksini ise  $k = 5$  için 16.6834 değeri ile k-temsilci kümeleme algoritması sağlamıştır.

Şekil 55. ve Tablo 45. birlikte değerlendirildiğinde, Harita 149. için  $k = 5$  toplamda beş defa (k-temsilci-KT, k-temsilci-silüet, hiyerarşik-Davies Bouldin, hiyerarşik-Dunn ve k-temsilci-Calinski Harabasz) en uygun aday küme sayısı olarak tespit edilmiştir ve k-temsilci algoritmasının sağladığı  $k = 5$  küme bölümlenmesi beş kriterden üçünde en iyi performansa sahiptir. Bu yüzden, Harita 149. için en uygun küme bölümlenmesinin  $k = 5$  için k-temsilci algoritması ile elde edildiğine karar verilmiştir.

**Tablo 45.** Harita 149 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri

Endeks	Kümeleme Algoritması	k=2	k=3	k=4	k=5
Davies-Bouldin	Hiyerarşik	1.1243	0.8447	0.7002	<b>0.6619</b>
	k-Temsilci	1.1526	0.8950	0.9281	0.6891
Dunn	Hiyerarşik	0.2504	0.3399	0.4369	<b>0.4907</b>
	k-Temsilci	0.1793	0.2314	0.1920	0.3579
Calinski-Harabasz	Hiyerarşik	8.7734	12.7805	14.8635	15.9522
	k-Temsilci	8.4296	12.8379	11.8897	<b>16.6834</b>

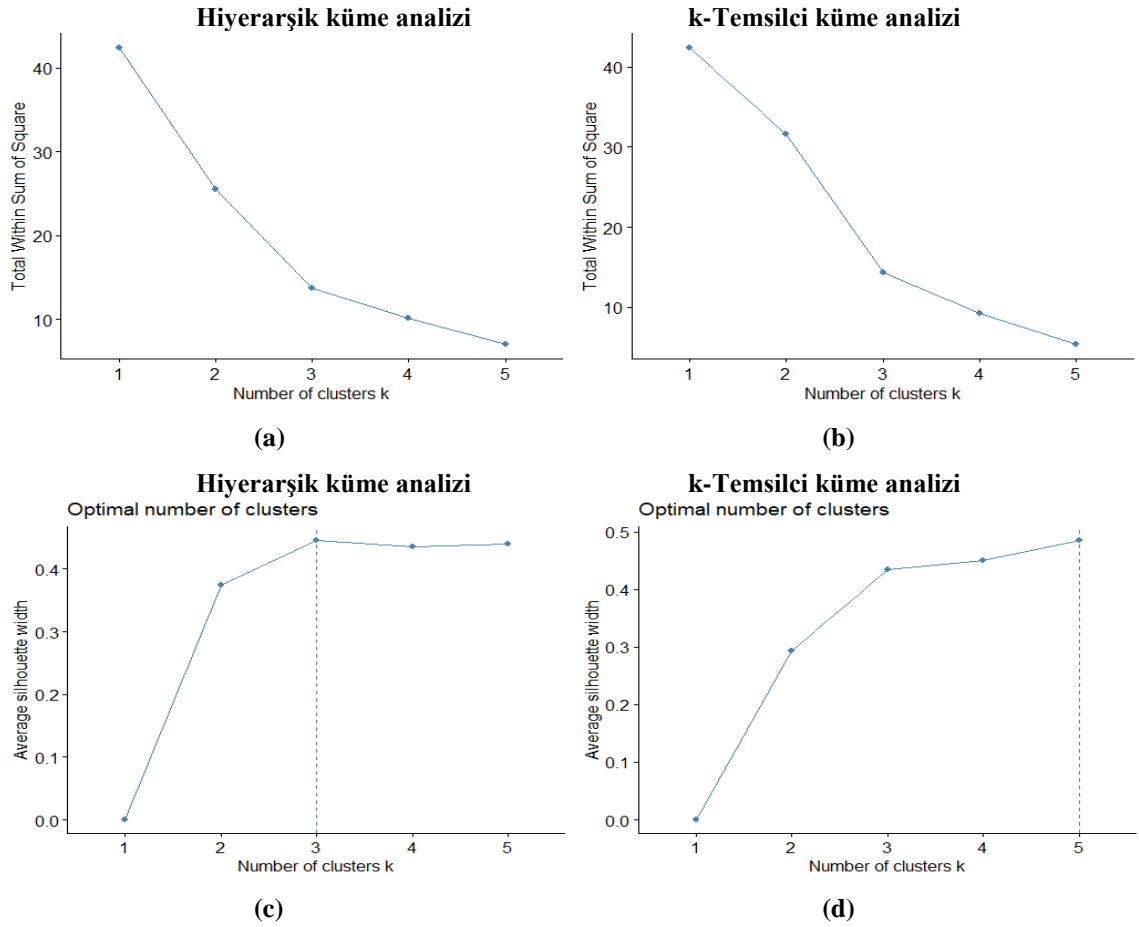
**Harita 227.**'nin hem hiyerarşik hem de k-temsilci algoritmaları ile küme analizi sonuçları Şekil 56.'da görselleştirilmiş ve Tablo 46'da özetlenmiştir. Küme içi KT Şekil 56(a-b)'den inceleneceği üzere, haritadaki nöronlar hiyerarşik kümeleme ile üç ( $KT = 13.7612$ )<sup>162</sup>, k-temsilci kümeleme ile beş kümeye ( $KT = 5.3716$ ) ayrılabilir. <sup>163</sup> Şekil 56(c-d)'de yer alan ortalama silüet endeksinin küme sayısına göre değişim grafiği incelendiğinde ise haritalardaki nöronlar hiyerarşik kümeleme ile üç, k-temsilci kümeleme ile beş kümeye ayrılabilir. Hiyerarşik kümelemede  $k = 3$  için 0.4458 silüet endeksi ve k-temsilci kümelemede  $k = 5$  için 0.4854 ortalama silüet endeksi, elde edilmiştir.

Tablo 46. incelendiğinde ise en küçük Davies-Bouldin, en büyük Dunn ve Calinski-Harabasz endeksini  $k = 5$  için k-temsilci kümeleme algoritması sağlamıştır.

Şekil 56. ve Tablo 46.'ya göre, harita 227 için k-temsilci kümeleme algoritması ile elde edilen  $k = 5$  küme bölümlenmesi içsel değerlendirmede kullanılan bütün endekslerde en iyi sonuçları sağlamıştır. Bu sebeple Harita 227. için en uygun küme bölümlenmesini  $k = 5$  için k-temsilci algoritması sağlamaktadır.

<sup>162</sup> Hiyerarşik küme analizinde  $k = 5$  için  $KT = 7.0465$

<sup>163</sup> Hiyerarşik küme analizine ait ( $k - KT$ ) grafikleri karşılaştırıldığında,  $k = 3 - 5$  arasındaki doğru parçalarının eğimi k-temsilci küme analizinde daha diktir. Bu yüzden, ( $k - KT$ ) grafiği dikkate alındığında, k-temsilci sonuçlarından  $k = 5$  en uygun küme bölümlenmesi olarak seçilmiştir.



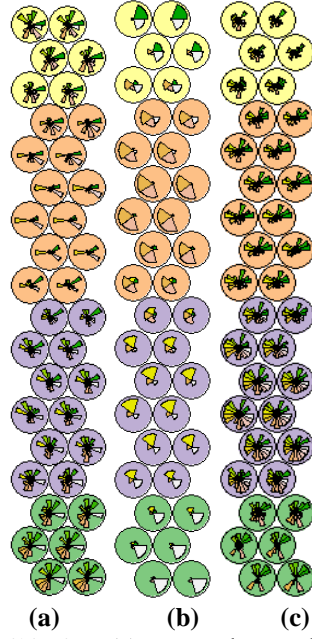
**Şekil 56.** Harita 227 için (a-b) küme içi kareler toplamının küme sayısına göre değişimi ve (c-d) ortalama silüet endeksinin küme sayısına göre değişimi

**Tablo 46.** Harita 227 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri

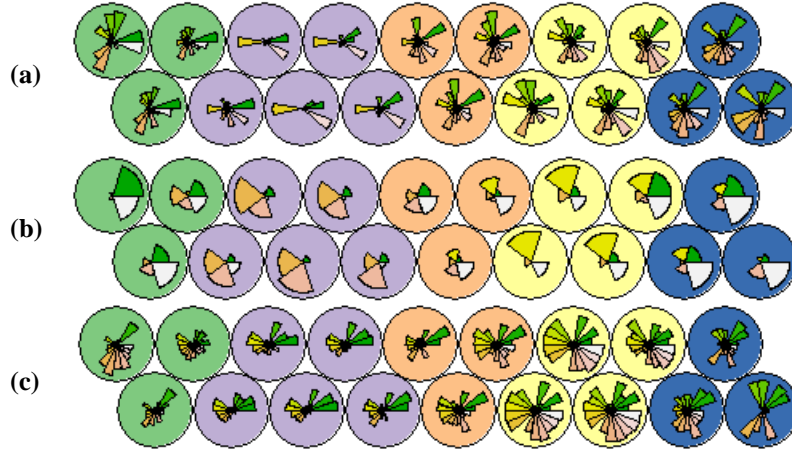
Endeks	Kümeleme Algoritması	k=2	k=3	k=4	k=5
Davies-Bouldin	Hiyerarşik	1.0638	0.7604	0.7648	0.6638
	k-Temsilci	1.4638	0.8325	0.6608	<b>0.5645</b>
Dunn	Hiyerarşik	0.0857	0.1145	0.1119	0.1347
	k-Temsilci	0.0780	0.1022	0.1139	<b>0.1819</b>
Calinski-Harabasz	Hiyerarşik	23.3313	30.1273	30.3582	35.1265
	k-Temsilci	10.1626	28.5190	33.5580	<b>46.5098</b>

### 5.3.1.3. Küme bütünlüğünün haritalara yansıtılarak kontrol edilmesi

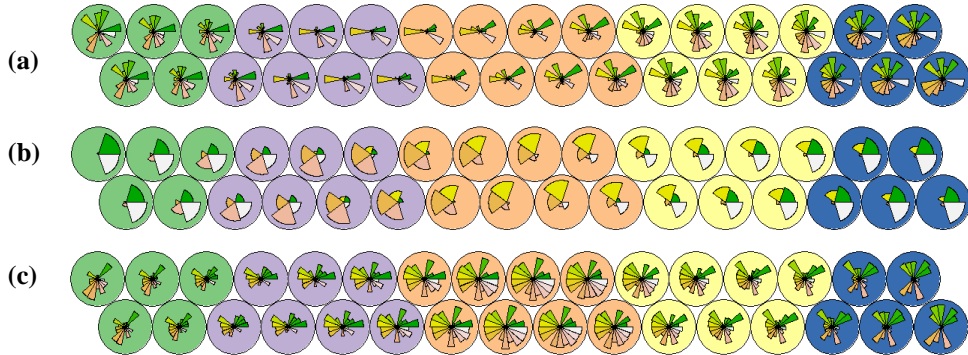
**Harita 11.**'in  $k = 4$  için, **Harita 149.**'un  $k = 5$  için ve **Harita 227.**'nin  $k = 5$  için nöron ağırlıklarına bağlı olarak k-temsilci küme analizi çıktıları Şekil (57-59)'teki gibi ilgili haritalar üzerine yansıtıldığında, tüm haritalar için küme bütünlüklerinin sağlandığı görülmektedir.



Şekil 57. Harita 11.'in (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına k-temsilli kümeleme algoritması ile elde edilen  $k=4$  için küme bölümlerinin yansıtılması



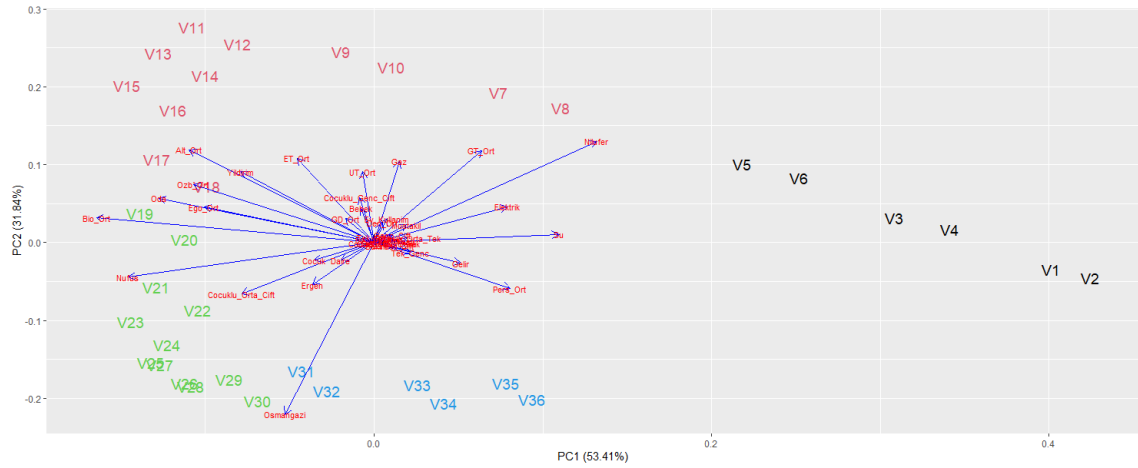
Şekil 58. Harita 149.'un (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına k-temsilli kümeleme algoritması ile elde edilen  $k=5$  için küme bölümlerinin yansıtılması



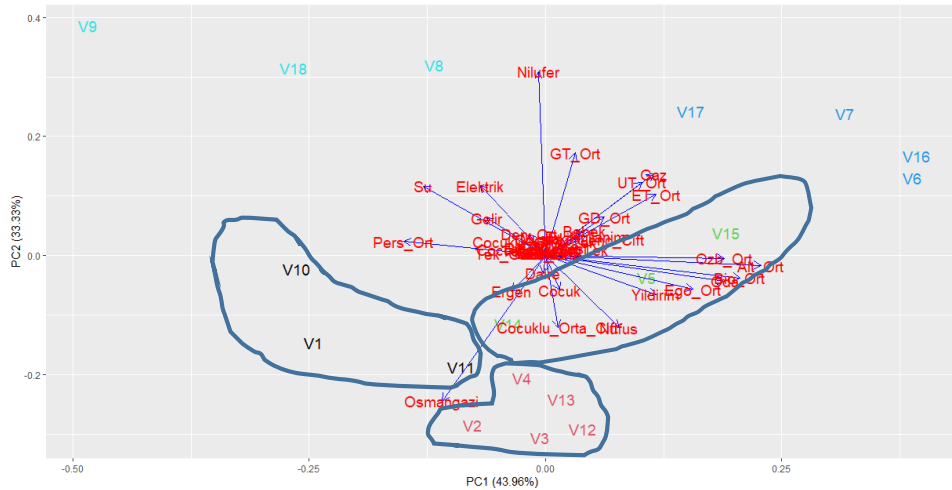
Şekil 59. Harita 227.'nin (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına k-temsilli kümeleme algoritması ile elde edilen  $k=5$  için küme bölümlerinin yansıtılması

### 5.3.1.4.TBA ile küme bütünlüğünün ve açıklanan varyansın kontrol edilmesi

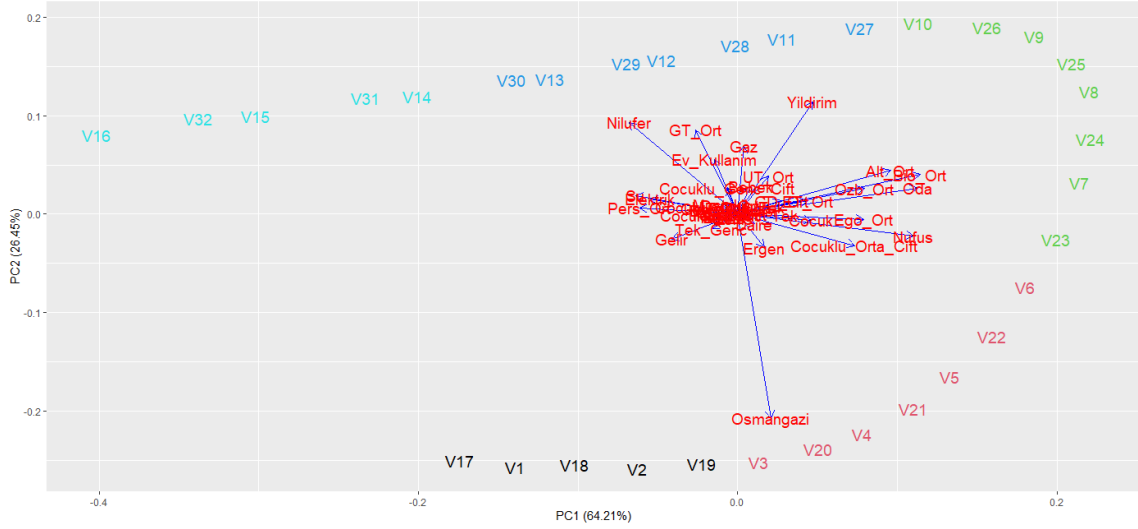
**Harita 11., Harita 149. ve Harita 227.**'in nöron ağırlıkları TBA'ya tabi tutulmuş ve analiz sonuçlarına göre birinci ve ikinci temel bileşenler varyansın sırası ile %84.89, %77.29 ve %90.86'lık kısmını açıklayabilmiştir. Harita 11. ve Harita 227.'nin TBA grafiği incelendiğinde her iki harita için elde edilen kümelerin temel bileşenler üzerinde çakışmadığı/iç içe geçmediği görülmektedir, Şekil (60-62). Fakat, Harita 149. için aynı durum geçerli değildir. Şekil 61.'den inceleneceği üzere görsel üzerinde işaretlenmiş olan birinci, ikinci ve üçüncü kümelerin temel bileşenler üzerinde iç içe geçecek kadar çok yakın olduğu dikkati çekmektedir.



Şekil 60. Harita 11.'in TBA analizi grafiği ve  $k = 4$  için k-temsili küme analizi ile elde edilen küme bölümlenmesinin TBA üzerine yansıtılması



Şekil 61. Harita 149.'un TBA analizi grafiği ve  $k = 5$  için k-temsili küme analizi ile elde edilen küme bölümlenmesinin TBA üzerine yansıtılması



**Şekil 62.** Harita 227.'nin TBA analizi grafiği ve  $k = 5$  için k-temsili küme analizi ile elde edilen küme bölümleniminin TBA üzerine yansıtılması

Harita 11, Harita 149. ve Harita 227., küme sayısı tespit etme prosedürünün iki, üç ve dört numaralı aşamaları dikkate alınarak karşılaştırmalı olarak analiz edildiğinde,<sup>164</sup> Tablo 47.'den inceleneceği üzere Harita 149. sadece tek kriter için en uygun değere sahiptir. Bu yüzden, üç harita içerisinde en kötü performansa sahip olan harita olarak değerlendirilmiş ve bu aşamada elenmesine karar verilmiştir. Harita 11. ve Harita 227. ise eşit sayıda kriterden en uygun değere sahip olduğu için, biri diğerine tercih edilmemiştir. Bu sebeple küme sayısı tespit etme prosedürünün beşinci aşaması kullanılarak söz konusu iki harita karşılaştırılmış ve içerinden sadece bir tanesi kimlik tanımlama aşamasında kullanılmak üzere aday olarak seçilmiştir.

**Tablo 47.** Harita 11, harita 149 ve harita 227'nin küme sayısı tespit etme prosedürünün iki, üç ve dört numaralı aşamaları kapsamında karşılaştırılması

Kriter	Harita 11 k=4 k-Temsilci	Harita 149 k=5 k-Temsilci	Harita 227 k=5 k-Temsilci
Silüet	<b>0.5237</b>	0.4200	0.4854
Davies-Bouldin	0.5913	0.6891	<b>0.5645</b>
Dunn	0.2360	<b>1.3579</b>	0.1819
Calinski-Harabasz	<b>52.6127</b>	16.6834	46.5098
Küme bütünlüğü	Sağlanmış	Sağlanmış	Sağlanmış
Açıklanan varyans	%84.89	%77.29	<b>%90.86</b>

<sup>164</sup>  $KT$  değerleri bu kapsamda dikkate alınmamıştır. Çünkü küme sayısı arttıkça  $KT$  değeri azalacaktır. Örneğin  $k = 4$  için elde edilen  $KT$ ,  $k = 5$  için elde edilen  $KT$ 'den daha azdır.



### 5.3.1.5. Varyans analizi ile küme ortalamaları arasındaki farklılığın kontrol edilmesi

Bu aşamada öncelikle, bağımsız değişkene ait grupların varyanslarının homojenliği test edilmiştir. Açıklamak gerekirse varyansların homojenliği PERMANOVA ile çok değişkenli varyans analizinde katı bir varsayım olmamasına rağmen, analiz sonuçlarının güvenilirliğinin varyansların homojenliği ile arttığı ifade edilmektedir. Bu yüzden, çok değişkenli varyans analizi ile grup ortalamalarının/merkezlerinin birbirinden farklılığı test edilmeden önce, varyansların homojenliği test edilmiştir.

**Harita 11** için,  $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$  ve  $\alpha = 0.05$  olmak üzere Şekil 63.'ten incelendiğinde  $F = 1.3629$  için  $p = 0.339$  olup  $\alpha = 0.05$  değerinden büyüktür ve bu durum ayrıca gruplararası ikili karşılaştırmalar için de geçerlidir. Dolayısı ile yokluk hipotezi reddedilememiştir. Başka bir ifade ile bağımsız değişkene ait grupların varyansları homojendir, Şekil 63.

DEVIATIONS FROM CENTROID		
F:	1,3629	df1: 3 df2: 32
P(perm):	0,339	
PAIRWISE COMPARISONS		
Groups	t	P(perm)
(1,2)	0,78635	0,478
(1,3)	0,27812	0,797
(1,4)	0,76451	0,493
(2,3)	1,5291	0,186
(2,4)	1,8809	0,137
(3,4)	0,84995	0,439

**Şekil 63.** Harita 11'de  $k=4$  için elde edilen kümelerin varyanslarının homojenliği testi sonuçları

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$  ve  $\alpha = 0.05$  olmak üzere Şekil 64.'ten incelendiğinde  $F = 52.613$  için  $p = 0.001$  olup  $\alpha = 0.05$  değerinden küçüktür ve bu durum ayrıca Şekil 65.'te özetlenen gruplararası ikili karşılaştırmalar için de geçerlidir.<sup>165</sup> Dolayısı ile yokluk hipotezi reddedilir ve  $Ku$  kümeleri temsil etmek üzere, bağımsız değişken  $Ku$ 'ya ait grupların ortalamaları arasında anlamlı fark vardır; grup merkezlerinin lokasyonları farklıdır.

---

<sup>165</sup> ANOVA, iki örneklem t-testine eşdeğerdir ve t-testi ile elde edilen  $t$ -değeri, ANOVA ile elde edilen  $F$ -değerinin kareköküne eşittir. PERMANOVA da benzer bir hesaplama yöntemine sahiptir. İkili karşılaştırmalarda hesaplanan  $t$ -değeri, permütasyonlara bağlı olarak hesaplanan yapay  $F$ -değerinin kareköküdür.

Factors						
Name	Abbrev.	Type	Levels			
Kume	Ku	Fixed	4			
PERMANOVA table of results						
Source	df	SS	MS	Pseudo-F	P(perm)	Unique perms
Ku	3	41,552	13,851	52,613	0,001	998
Res	32	8,4243	0,26326			
Total	35	49,976				
Details of the expected mean squares (EMS) for the model						
Source	EMS					
Ku	1*V(Res) + 8,6667*S(Ku)					
Res	1*V(Res)					
Construction of Pseudo-F ratio(s) from mean squares						
Source	Numerator	Denominator	Num. df	Den. df		
Ku	1*Ku	1*Res	3	32		
Estimates of components of variation						
Source	Estimate	Sq. root				
S(Ku)	1,5678	1,2521				
V(Res)	0,26326	0,51309				

Şekil 64. Harita 11’de  $k=4$  için elde edilen kümelerin çok değişkenli varyans analizi sonuçları

Factors						
Name	Abbrev.	Type	Levels			
Kume	Ku	Fixed	4			
PAIR-WISE TESTS						
Term 'Ku'						
Groups	t	P(perm)	Unique perms			
1, 2	7,5863	0,001	978			
1, 3	9,5754	0,001	964			
1, 4	7,0903	0,004	405			
2, 3	6,603	0,001	997			
2, 4	6,4667	0,001	964			
3, 4	5,7027	0,001	968			
Denominators						
Groups	Denominator	Den. df				
1, 2	1*Res	16				
1, 3	1*Res	16				
1, 4	1*Res	10				
2, 3	1*Res	22				
2, 4	1*Res	16				
3, 4	1*Res	16				
Average Distance between/within groups						
	1	2	3	4		
1	0,68383					
2	2,217	0,74799				
3	2,4049	1,5392	0,59325			
4	2,0594	1,8527	1,3431	0,53958		

Şekil 65. Harita 11’de  $k=4$  için elde edilen kümelerin çok değişkenli varyans analizine ait gruplar arası ikili karşılaştırma sonuçları

**Harita 227.** için  $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$  ve  $\alpha = 0.05$  olmak üzere Şekil 66.’dan incelendiğinde  $F = 0.0401$  ve  $p = 0.999$  olup,  $p$  değeri  $\alpha = 0.05$  değerinden büyüktür ve bu durum ayrıca gruplararası ikili karşılaştırmalar için de geçerlidir. Dolayısı ile yokluk hipotezi reddedilemez ve bağımsız değişkene ait grupların varyansları homojendir.

DEVIATIONS FROM CENTROID		
F:	0,040117	df1: 4 df2: 27
P(perm):	0,999	
PAIRWISE COMPARISONS		
Groups	t	P(perm)
(1,2)	0,1908	0,868
(1,3)	0,29828	0,773
(1,4)	0,30674	0,758
(1,5)	0,31912	0,713
(2,3)	0,088119	0,924
(2,4)	0,14154	0,876
(2,5)	0,18528	0,866
(3,4)	0,076324	0,942
(3,5)	0,13842	0,882
(4,5)	0,060901	0,95

Şekil 66. Harita 227’de  $k=5$  için elde edilen kümelerin varyanslarının homojenliği testi sonuçları

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$  ve  $\alpha = 0.05$  olmak üzere Şekil 67.’den incelendiğinde  $F = 46.51$  için  $p = 0.001$  olup  $\alpha = 0.05$  değerinden küçüktür ve bu durum ayrıca Şekil 68’de özetlenen gruplararası ikili karşılaştırmalar için de geçerlidir. Dolayısı ile yokluk hipotezi reddedilir. Dolayısıyla  $Ku$  kümeleri temsil etmek üzere, bağımsız değişken  $Ku$ ’ya ait grupların ortalamaları arasında anlamlı fark vardır; grup merkezlerinin lokasyonları farklıdır.

Factors						
Name	Abbrev.	Type	Levels			
Kume	Ku	Fixed	5			
PERMANOVA table of results						
Source	df	SS	MS	Pseudo-F	P(perm)	Unique perms
Ku	4	37,013	9,2532	46,51	0,001	998
Res	27	5,3717	0,19895			
Total	31	42,384				
Details of the expected mean squares (EMS) for the model						
Source	EMS					
Ku	$1^*V(\text{Res}) + 6,3437^*S(\text{Ku})$					
Res	$1^*V(\text{Res})$					
Construction of Pseudo-F ratio(s) from mean squares						
Source	Numerator	Denominator	Num. df	Den. df		
Ku	$1^*Ku$	$1^*Res$	4	27		
Estimates of components of variation						
Source	Estimate	Sq.root				
S(Ku)	1,4273	1,1947				
V(Res)	0,19895	0,44604				

Şekil 67. Harita 227’de  $k=5$  için elde edilen kümelerin çok değişkenli varyans analizi sonuçları

Factors					
Name	Abbrev.	Type	Levels		
Kume	Ku	Fixed	5		
<i>PAIR-WISE TESTS</i>					
Term 'Ku'					
Groups	t	P(perm)	Unique perms		
1, 2	5,0895	0,002	558		
1, 3	8,3798	0,002	683		
1, 4	5,711	0,005	567		
1, 5	5,8562	0,008	126		
2, 3	5,5298	0,001	930		
2, 4	6,0658	0,001	754		
2, 5	8,5684	0,002	563		
3, 4	5,6242	0,001	922		
3, 5	10,066	0,004	671		
4, 5	5,2899	0,001	568		
<i>Denominators</i>					
Groups	Denominator	Den. df			
1, 2	1*Res	10			
1, 3	1*Res	11			
1, 4	1*Res	10			
1, 5	1*Res	8			
2, 3	1*Res	13			
2, 4	1*Res	12			
2, 5	1*Res	10			
3, 4	1*Res	13			
3, 5	1*Res	11			
4, 5	1*Res	10			
<i>Average Distance between/within groups</i>					
	1	2	3	4	5
1	0,55048				
2	1,324	0,54477			
3	2,1003	1,2991	0,5379		
4	1,5934	1,5555	1,3427	0,57024	
5	1,7973	2,3508	2,6211	1,4766	0,62371

Şekil 68. Harita 227.'de  $k=5$  için elde edilen küme bölümlerinin çok değişkenli varyans analizine ait gruplar arası ikili karşılaştırma sonuçları

Harita 11. ve 227.'nin varyans analizi sonuçlarını karşılaştırmak gerekirse:

- Şekil (64, 67) inceleneceği üzere her iki harita için elde edilen  $p$  değerleri 0.01 olup eşittir. Fakat  $MS$  ile temsil edilen ortalama hata kare değerleri karşılaştırıldığında Harita 227.'nin ortalama hata karesinin (9.2532), Harita 11.'inkinden (13.851) daha küçük olduğu gözlemlenmektedir.

- Ayrıca, Şekil 68.'de "Average distance between/within groups" başlığı altında yer aldığı üzere, 227 numaralı haritada küme içi ortalama mesafeler (matrisin köşegeni) Harita 11.'inkine nispeten daha küçüktür. Bu durum, Harita 227.'deki kümelerin, Harita 11.'e kıyas ile geometrik olarak daha dar bir alana saçıldığına, kümelerin daha kompakt bir yapıya sahip olduğunun işaretidir.

- Ek olarak Harita 11.'de kümeler arası mesafe değerleri (1.3431 – 2.4049) aralığında değişim gösterirken, Harita 227.'de (1.2991-2.6211) aralığında değişmektedir. Harita 227. için hesaplanan en kısa mesafe Harita 11. için hesaplanan değerden 0.0440

birim kadar daha az iken, en uzun mesafe 0.2162 birim kadar daha fazladır. Kümeler arası mesafe değerlerinin, küme içi mesafelere kıyas ile daha fazla olması beklenmektedir. Bu bağlamda, çok bariz bir şekilde olmamak kaydı ile 227 numaralı harita, 11 numaralı haritadan biraz daha iyi performansa sahiptir.

Özetle, 227 numaralı harita, daha küçük ortalama hata kare, küme içi ortalama mesafe ve Davies-Bouldin endeksine ve daha büyük kümeler arası ortalama mesafe, Silüet, Calinski-Harabasz endekslerine ve açıklanan varyansa sahip olduğu için kimlik tanımlama aşamasında kullanılmak üzere bir numaralı aday olmuştur.

### 5.3.2. *Yığın Öğrenme Algoritması ile Eğitilen Haritaların Analizleri*

Yığın öğrenme algoritması ile eğitilen haritalardan eleme sonrası kalan 58 harita öncelikle nicemleme hataları açısından incelenmiş<sup>166</sup> ve en küçük üç NH'ye sahip harita küme analizinde değerlendirilmiştir.

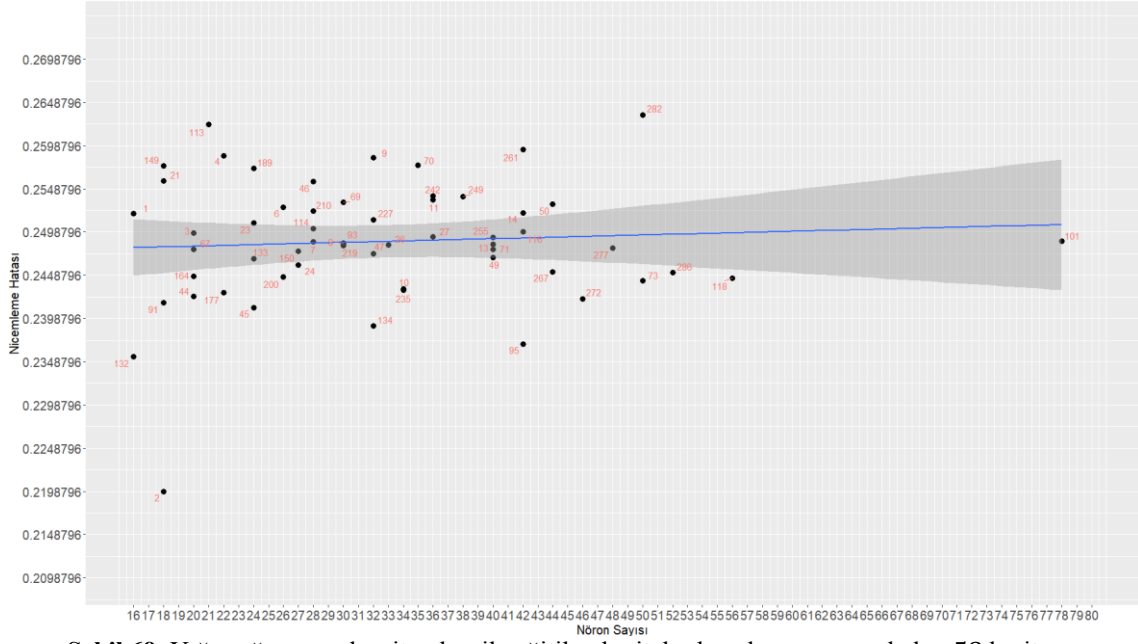
Harita 2, en küçük NH'ye sahip olan haritadır, Şekil 69. Haritanın boyutları  $(x, y) = (2, 9)$  olup toplam nöron sayısı 18 ve NH 0.2198'dir. İkinci en küçük NH'ye sahip olan Harita 132. numaralı harita olup, boyutları  $(x, y) = (8, 2)$  ve toplam nöron sayısı 16'dır. Bu haritanın NH ise 0.2354'tir. NH üçüncü en küçük olan harita ise 95 numaralı harita olup, boyutları  $(x, y) = (6, 7)$ , toplam nöron sayısı 42 ve NH 0.2369'dur.

En küçük NH'ye sahip olarak seçilen üç haritanın boyutları incelendiğinde, 95 numaralı harita hariç diğer iki haritanın, çevrimiçi öğrenme algoritması ile eğitilen haritalardaki gibi, çok net bir şekilde enine ya da boyuna uzayan dikdörtgen bir şekle sahip olduğu dikkati çekmektedir. Bir diğer husus ise Şekil (50., 69.) karşılaştırıldığında görüleceği üzere, yığın öğrenme algoritması ile eğitilen ağların NH'nin, çevrimiçi öğrenme algoritması ile eğitilen ağların NH'den daha yüksek olmasıdır. Bunun nedeni, yığın öğrenme algoritmasının çok kısa sürede (daha az iterasyonda) veya yanlış yakınsamasıdır.<sup>167</sup>

---

<sup>166</sup> Yığın öğrenme algoritması ile eğitilmiş olan 293 haritanın nicemleme hatasının nöron sayısına göre değişimine ait grafik Ek 9'dan incelenebilir.

<sup>167</sup> Çevrimiçi ve yığın öğrenme algoritmaları ile eğitilmiş olan 15 ve 58 haritaya ait eğitim grafikleri Ek 10-11'den incelenebilir.



**Şekil 69.** Yığın öğrenme algoritmaları ile eğitilen haritalardan eleme sonrası kalan 58 haritanın nicemleme hatasının nöron sayısına göre değişimi

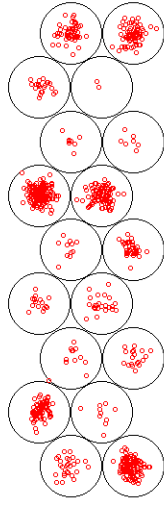
Küme analizinde değerlendirilecek olan üç harita tespit edildikten sonra, her birisine sıra ile beş aşamalı uygun küme sayısı tespit etme prosedürü uygulanmıştır.

### 5.3.2.1. Isı haritaları ile görsel değerlendirme

**Harita 2'**de gözlemler, haritanın alt, orta ve üst kısmında yoğunlaşmıştır, Şekil 70(a). Şekil 70(c). incelendiğinde nöronlar ile kendilerine atanan gözlemler arasındaki ortalama mesafenin (0.18 – 0.27) bandında değiştiği görülmektedir. Şekil 70(b). incelendiğinde ise haritanın hemen hemen tamamında nöronlar arası komşuluk mesafelerinin birbirine çok yakın olduğu ve üst orta kısmındaki bir bölgenin çok belirgin şekilde haritayı ikiye böldüğü ifade edilebilir. Sonuç olarak, Harita 2 için toplam küme sayısı üç olabilir.

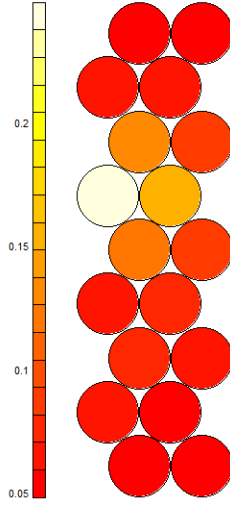
**Harita 132.'**de gözlemler, sağ ve sol uçları biraz daha yoğun olmak üzere, haritanın hemen hemen her yerine denk bir şekilde dağılmıştır, Şekil 71(a)., Şekil 71(c). incelendiğinde nöronlar ile kendilerine atanan gözlemler arasındaki ortalama mesafenin (0.18 – 0.27) bandında değiştiği görülmektedir. Şekil 71(b). incelendiğinde ise haritanın hemen hemen tamamında nöronlar arası komşuluk mesafelerinin birbirine çok yakın olduğu görülmektedir. Sağ ve sol uç birer grup ve orta kısım iki ya da üç grup olmak üzere, harita toplam dört veya beş kümeye ayrılabilir.

Hücelere düşen gözlem sayısı



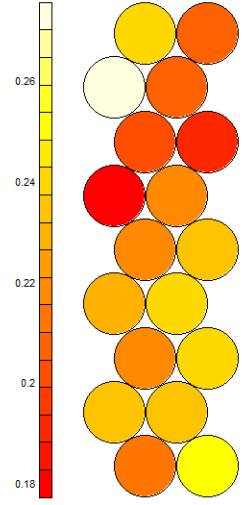
(a)

Nöronların komşuluk mesafeleri



(b)

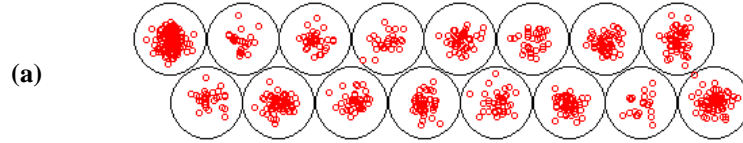
Nöron-Gözlem arası ortalama mesafe



(c)

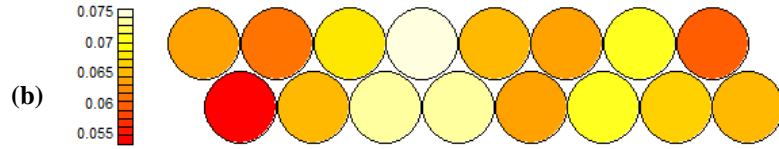
Şekil 70. Harita 2'nin (a) hücelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe

Hücelere düşen gözlem sayısı



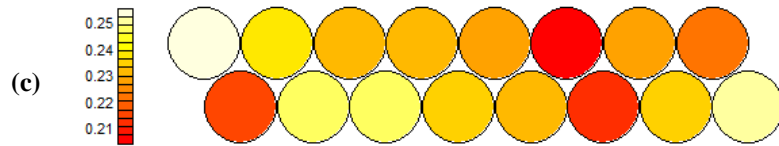
(a)

Nöronların komşuluk mesafeleri



(b)

Nöron-Gözlem arası ortalama mesafe

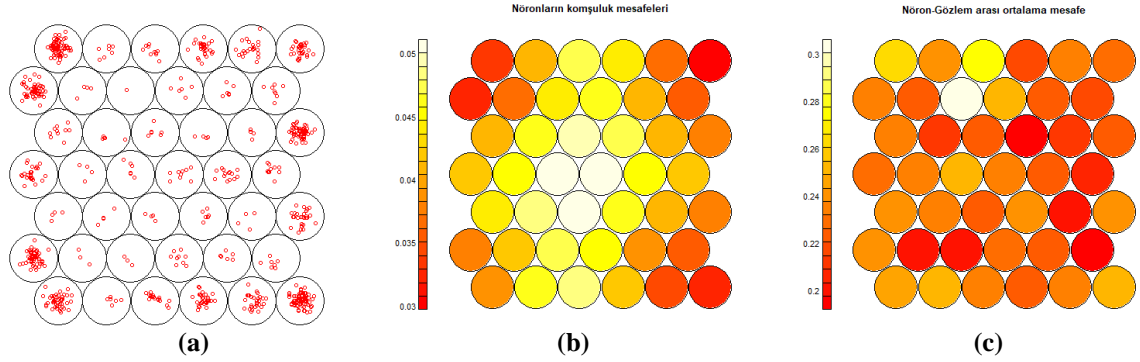


(c)

Şekil 71. Harita 132'nin (a) hücelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe

**Harita 95.** ise diğer tüm haritalardan farklı olarak kareye yakın bir yapıya sahip olup gözlemler, haritanın köşelerinde en yoğun, kenarlarında daha az yoğun ve iç kısımlarında seyrek bir şekilde dağılmıştır, Şekil 72(a), Şekil 72(c) incelendiğinde

nöronlar ile kendilerine atanan gözlemler arasındaki ortalama mesafenin (0.20 – 0.30) bandında değiştiği görülmektedir. Şekil 72(b). incelendiğinde ise haritanın köşelerinde nöronlar arası komşuluk mesafelerinin birbirine çok yakın olduğu ve tam orta kısmında ise çok uzak olduğu görülmektedir. Bu sebeple harita ya artı şeklinde dört ya da köşeler birer ve orta kısım tek olmak üzere toplamda beş kümeye ayrılabilir.



Şekil 72. Harita 95'in (a) hücelere düşen gözlem sayısı, (b) nöronların komşuluk mesafeleri ve (c) gözlemler ile atandıkları nöronlar arasındaki ortalama mesafe

Harita 11., Harita 149. ve Harita 227. eş anlı değerlendirildiğinde nöron-gözlem arası mesafeler bağlamında, veri setini en iyi temsil eden nöronlara sahip olan harita, Harita 2.'dir. Çünkü Harita 2., nöron-gözlem arası en küçük ortalama mesafeye sahiptir (0.18-0.27), Şekil 70(c).<sup>168</sup> Diğer taraftan, söz konusu haritaların U-matrisleri tam anlamı ile karşılaştırılamamıştır. Çünkü yakın nöronlar arası en küçük ve uzak nöronlar arası en büyük mesafelerin farkı her bir harita için 0.20 birim civarındadır.<sup>169</sup>

### 5.3.2.2. Nöron ağırlıklarının kümelenmesi

Yığın öğrenme algoritması ile eğitilen haritalardan, küme analizinde değerlendirilmek üzere seçilen üç haritanın içerdiği maksimum nöron sayısı 42 olduğu için maksimum küme sayısı  $k = 6$  olarak kabul edilmiştir.

**Harita 2.**'nin hem hiyerarşik hem de k-temsilli algoritmaları ile küme analizi sonuçları Şekil 73.'te görselleştirilmiş ve Tablo 48.'de özetlenmiştir. Şekil 73(a-b)'den küme içi KT değerleri incelendiğinde,  $k = 3$  için k-temsilli kümeleme algoritması daha

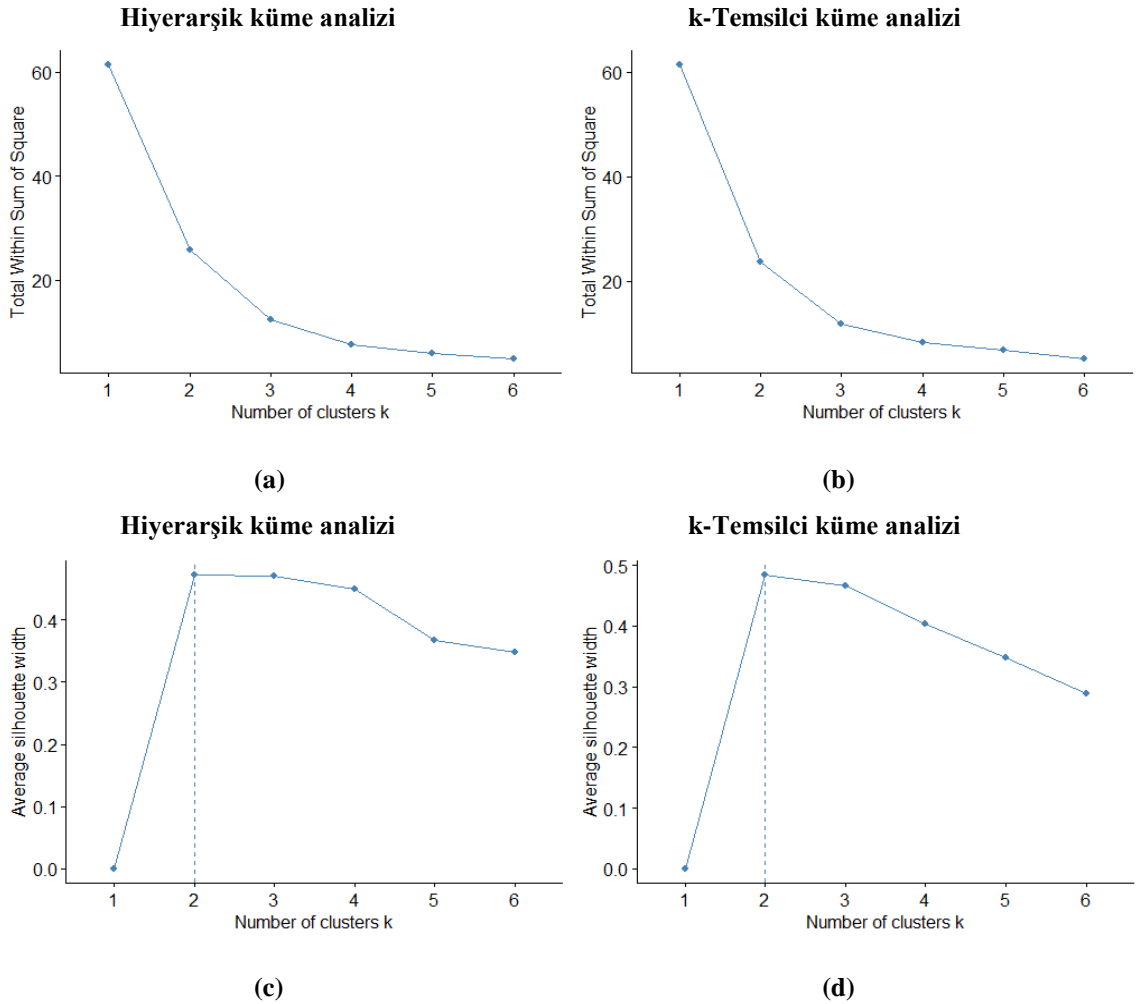
<sup>168</sup> Harita 149 için ortalama mesafe (0.13-0.23) ve harita 227 için (0.13-0.25)'tir

<sup>169</sup> Harita 2 için (0.05 – 0.25), harita 132 için (0.055 – 0.075) ve harita 95 için (0.03 – 0.05)'tir.



küçük bir değere sahiptir ( $KT = 11.8019$ ).<sup>170</sup> Şekil 73(c-d)'de yer alan silüet endeksi incelendiğinde ise  $k = 2$  için k-Temsilci algoritması ile elde edilen 0.4839 ortalama silüet endeksi,  $k = 2$  veya 3 için hiyerarşik küme analizi ile elde edilen ortalama silüet endeksinden fazladır (0.4720). Bu sebeple, silüet endeksi dikkate alındığında en uygun küme bölümlenmesini k-temsilci algoritması sağlamaktadır ve  $k = 2$ 'dir.

Tablo 48. incelendiğinde ise hiyerarşik kümeleme algoritması, en küçük Davies-Bouldin endeksini,  $k = 3$  için 0.6488 değeri ile ve en büyük Dunn endeksini,  $k = 5$  ve 6 için 0.4565 değeri ile sağlamıştır. En büyük Calinski-Harabasz endeksini ise  $k = 4$  için 33.4532 değeri ile k-temsilci kümeleme algoritması sağlamıştır.



**Şekil 73.** Harita 2 için (a-b) küme içi kareler toplamının küme sayısına göre değişimi ve (c-d) ortalama silüet endeksinin küme sayısına göre değişimi

<sup>170</sup> Hiyerarşik küme analizinde  $k = 3$  için  $KT = 12.4222$

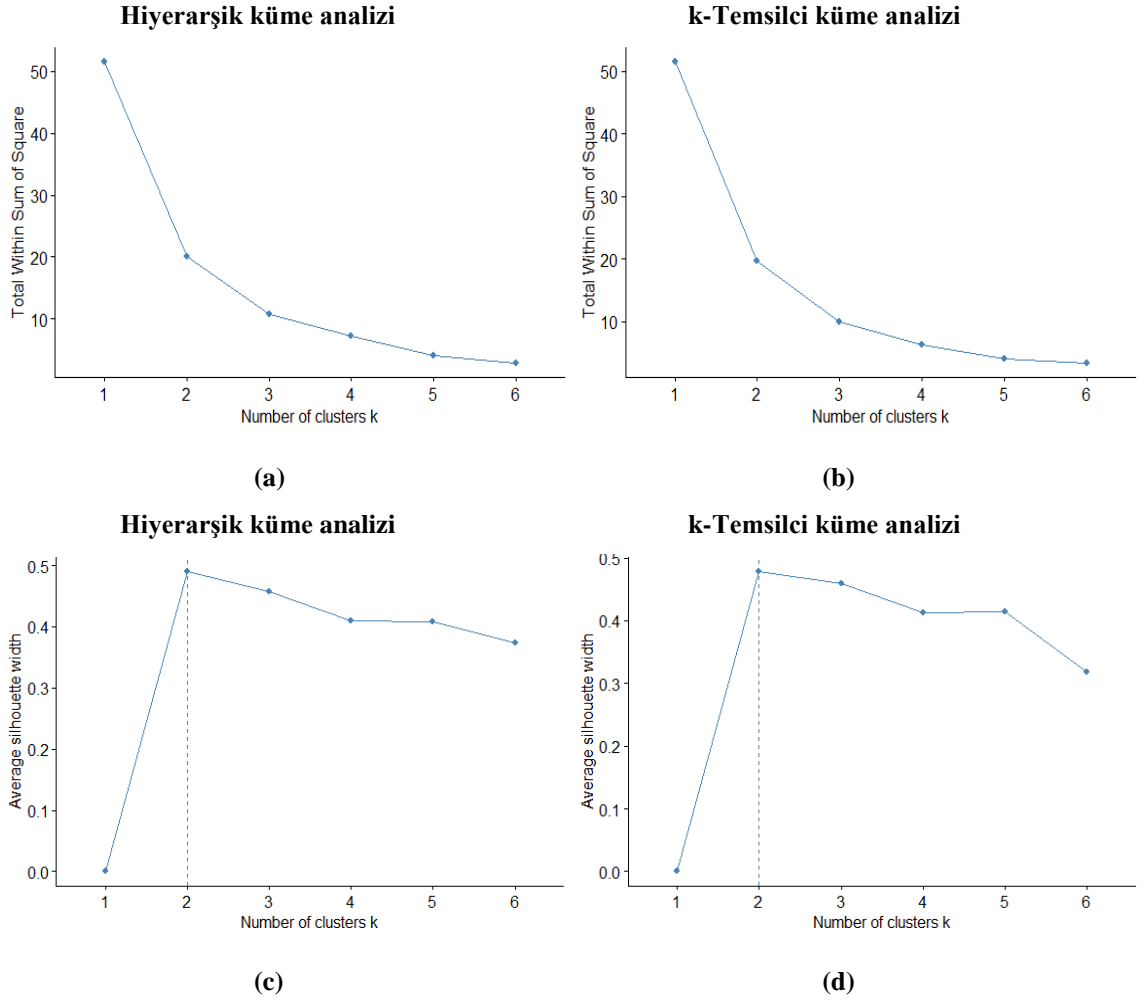
**Tablo 48.** Harita 2 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri

Endeks	Kümeleme Algoritması	k=2	k=3	k=4	k=5	k=6
Davies-Bouldin	Hiyerarşik	0.6703	<b>0.6488</b>	0.6751	0.8016	0.7685
	k-Temsilci	0.7297	0.6818	0.7294	0.7440	0.8327
Dunn	Hiyerarşik	0.3180	0.4308	0.4050	<b>0.4565</b>	<b>0.4565</b>
	k-Temsilci	0.2912	0.4159	0.3932	0.3932	0.4432
Calinski-Harabasz	Hiyerarşik	21.9978	29.5726	<b>33.4532</b>	30.3337	27.5813
	k-Temsilci	25.2160	31.5213	30.0937	26.0124	25.6978

Şekil 73. ve Tablo 48. dikkate alındığında, Harita 2. için nerede ise her bir içsel değerlendirme kriteri farklı bir sonuç sağlamıştır.  $k = 2,4,5,6$  küme bölümlenmeleri birer defa ve  $k = 3$  küme bölümlenmesi iki defa en uygun küme sayısı olarak tespit edilmiştir.  $k = 3$  diğer küme sayılarından daha fazla sayıda endeks tarafından en uygun küme sayısı olarak tespit edildiği için, Harita 2.'nin üç kümeye bölünmesi uygun görülmüştür. Ayrıca,  $k = 3$  için kümeleme algoritmalarının performansları karşılaştırıldığında k-temsilci kümeleme algoritması, hiyerarşik kümeleme algoritmasından daha iyi performansa sahiptir. Çünkü  $k = 3$  için hiyerarşik küme analizi iki kriterde (Davies-Bouldin ve Dunn endeksleri) en iyi değere sahip iken, k-temsilci küme analizi üç kriterde (KT, Calinski-Harabasz, ve silüet endeksi) en iyi değere sahiptir.

**Harita 132.**'nin hem hiyerarşik hem de k-temsilci algoritmaları ile küme analizi sonuçları Şekil 74.'te görselleştirilmiş ve Tablo 49.'da özetlenmiştir. Şekil 74(a-b)'den inceleneceği üzere, küme içi KT dikkate alındığında, haritadaki nöronlar üç veya beş kümeye ayrılabilir. Hiyerarşik küme analizi ile elde edilen kümelerin KT değeri  $k = 3$  için 10.7185'ten,  $k = 5$  için %62 oranında azalarak 3.9978'e inmiştir. KT'lerin değişimi k-temsilci küme analizinde ise 9.8562'den %60 oranında azalarak 3.9783'tür. Bu sebeple her iki grafiğin dirsek noktaları  $k = 5$  olarak alınmıştır.

Şekil 74(c.-d.).'de dikkate alındığında, haritalardaki nöronlar iki kümeye ayrılabilir.  $k = 2$  için hiyerarşik kümeleme algoritması ile elde edilen 0.4895 değerli ortalama silüet endeksi, k-temsilci kümeleme algoritması ile elde edilen ortalama silüet endeksinden büyüktür (0.4776).



Şekil 74. Harita 132 için (a-b) küme içi kareler toplamının, (c-d) ortalama silüet endeksinin küme sayısına göre değişimi

Tablo 49. incelendiğinde ise en küçük Davies-Bouldin endeksini,  $k = 6$  için 0.5736 değeri ile k-temsilci kümeleme algoritması ve en büyük Dunn ve Calinski-Harabasz endeksini,  $k = 6$  için 0.6236 ve 33.3843 değerleri ile hiyerarşik kümeleme algoritması sağlamıştır.

Şekil 74. ve Tablo 49. dikkate alındığında, Harita 132. için  $k = 2$  bir defa (KT-k-temsilci),  $k = 5$  bir defa (silüet-hiyerarşik) ve  $k = 6$  üç defa (Davies-Bouldin-k-temsilci, Dunn-hiyerarşik ve Calinski-Harabasz-hiyerarşik) en uygun küme sayısı olarak tespit edilmiştir.  $k = 6$  diğer küme sayılarından daha fazla sayıda endeks tarafından en uygun küme sayısı olarak tespit edildiği için, Harita 132.'nin en uygun küme sayısı altı olarak uygun görülmüştür.  $k = 6$  için kümeleme algoritmalarının performansları karşılaştırıldığında ise hiyerarşik kümeleme algoritması, k-temsilci kümeleme

algoritmasından daha iyi performans sağlamaktadır. Çünkü  $k = 6$  olmak üzere hiyerarşik küme analizi dört kriter (KT, silüet, Dunn ve Calinski-Harabasz endeksleri) için en iyi değere sahip iken, k-temsilci küme analizi bir kriter (Davies-Bouldin) için en iyi değere sahiptir.<sup>171-172</sup>

**Tablo 49.** Harita 132 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri

Endeks	Kümeleme Algoritması	k=2	k=3	k=4	k=5	k=6
Davies-Bouldin	Hiyerarşik	0.6820	0.6268	0.5860	0.6366	0.6260
	k-Temsilci	0.7388	0.6541	0.6386	0.6212	<b>0.5736</b>
Dunn	Hiyerarşik	0.2436	0.3197	0.3302	0.4214	<b>0.6236</b>
	k-Temsilci	0.2196	0.2822	0.3481	0.3946	0.4192
Calinski-Harabasz	Hiyerarşik	22.0646	24.7300	23.5219	32.6745	<b>33.3843</b>
	k-Temsilci	22.5312	27.4621	28.4474	32.8477	28.3781

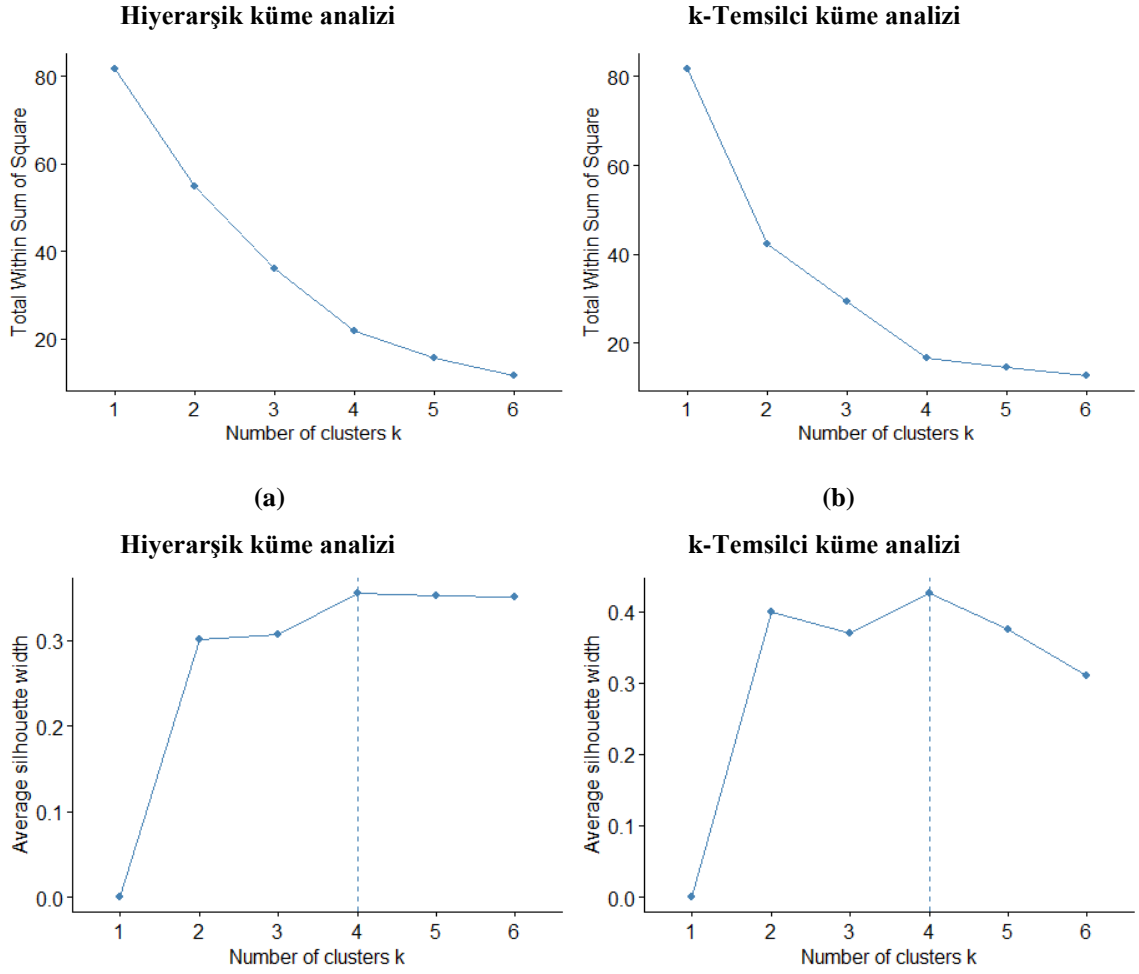
**Harita 95.**'in hem hiyerarşik hem de k-temsilci algoritmaları ile küme analizi sonuçları Şekil 75. görselleştirilmiş ve Tablo 50.'te özetlenmiştir. Haritadaki nöronlar, küme içi KT dikkate alındığında, her iki kümeleme algoritması için de en uygun küme sayısı dördüttür, Şekil 75(a-b).  $k = 4$  için en küçük *KT* değerini ise k-temsilci küme analizi sağlamıştır,  $KT = 16.7359$ .

Şekil 75(c-d)'de yer alan ortalama silüet endeksinin küme sayısına göre değişim grafiği incelendiğinde ise haritalardaki nöronlar hiyerarşik küme analizi ile dört, beş veya altı kümeye ayrılabilir. Silüet endeksi  $k = 4$  için 0.3559,  $k = 5$  için 0.3522 ve  $k = 6$  için 0.3506'dır. k-Temsilci küme analizi sonuçlarına göre ise  $k = 4$ 'tür ve silüet endeksi değeri 0.4258'dir. Bu durumda k-temsilci küme analizi ile  $k = 4$  için elde edilen silüet endeksi değeri, hiyerarşik küme analizi ile elde edilen tüm değerlerden büyük olduğu için silüet endeksine göre en uygun küme bölümlenmesini  $k = 4$  için k-temsilci algoritması sağlamıştır.

Tablo 50. incelendiğinde ise en küçük Davies-Bouldin endeksini ve en büyük Calinski-Harabasz endeksini,  $k = 4$  için k-temsilci kümeleme algoritması sağlamıştır. En büyük Dunn endeksini ise  $k = 6$  için hiyerarşik kümeleme algoritması sağlamıştır.

<sup>171</sup> Hiyerarşik kümeleme  $k = 6$  için  $KT = 2.9108$ ,  $silüet = 0.3728$

<sup>172</sup> k-Temsilci kümeleme  $k = 6$  için  $KT = 3.3904$ ,  $silüet = 0.3174$



**Şekil 75.** Harita 95 için (a-b) küme içi kareler toplamının, (c-d) ortalama silüet endeksinin küme sayısına göre değişimi

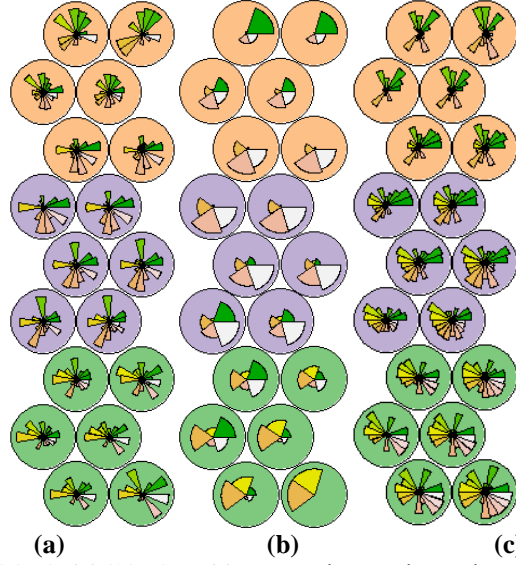
**Tablo 50.** Harita 95 için Davies-Bouldin, Dunn ve Calinski-Harabasz endekslerinin küme sayısına göre değişimleri

Endeks	Kümeleme Algoritması	k=2	k=3	k=4	k=5	k=6
Davies-Bouldin	Hiyerarşik	1.3010	1.0284	0.7884	0.8178	0.7908
	k-Temsilci	0.9593	0.8617	<b>0.7207</b>	0.8617	0.9813
Dunn	Hiyerarşik	0.1551	0.1329	0.1868	0.2026	<b>0.2983</b>
	k-Temsilci	0.1938	0.1375	0.2835	0.2818	0.2528
Calinski-Harabasz	Hiyerarşik	19.7769	24.8745	35.0046	39.1427	44.0885
	k-Temsilci	37.4549	34.9687	<b>49.2548</b>	42.8940	39.6522

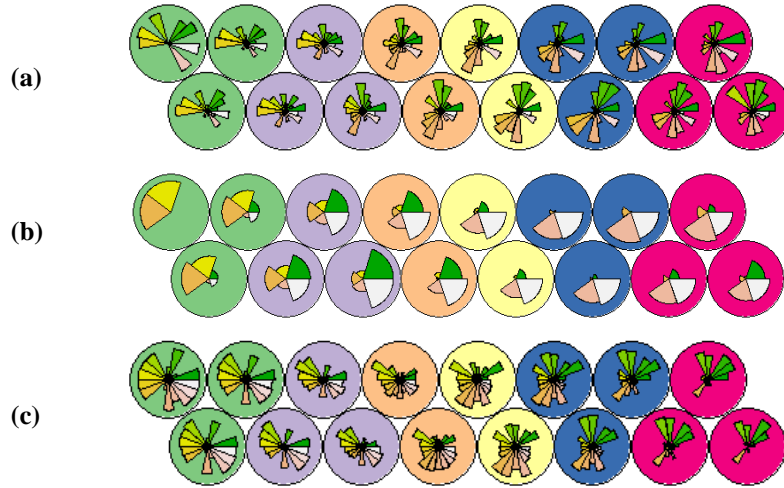
Şekil 75. ve Tablo 50. dikkate alındığında, Harita 95. için  $k = 4$  üç defa (silüet-k-temsilci, Davies-Bouldin-k-temsilci ve Calinski-Harabasz-k-temsilci) ve  $k = 6$  bir defa (Dunn-hiyerarşik) en uygun küme sayısı olarak tespit edilmiştir. Dolayısı ile Harita 95.'in,  $k = 4$  için k-temsilci küme analizi ile kümeleneceği uygun görülmüştür.

### 5.3.2.3. Küme bütünlüğünün haritalara yansıtılarak kontrol edilmesi

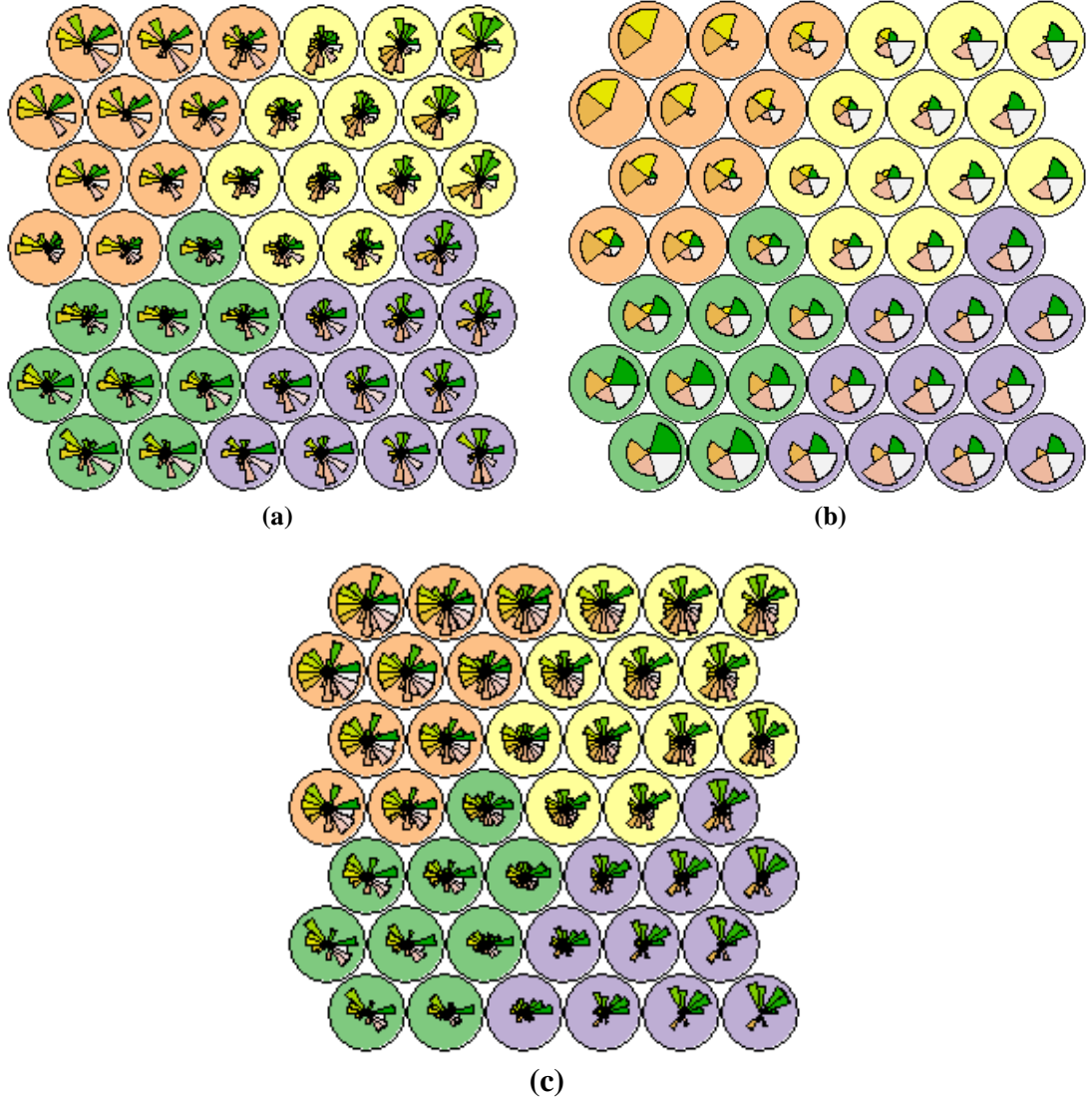
**Harita 2.**'nin  $k = 3$  için nöron ağırlıklarına bağlı olarak  $k$ -temsilci ve **Harita 132.** ve **Harita 95.**'in  $k = 6$  için hiyerarşik küme analizi çıktıları ilgili haritalar üzerine yansıtıldığında, tüm haritalar için küme bütünlüklerinin sağlandığı görülmektedir, Şekil (76-78).



**Şekil 76.** Harita 2'nin (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına  $k$ -temsilci kümeleme algoritması ile elde edilen  $k=3$  için küme bölümlerinin yansıtılması



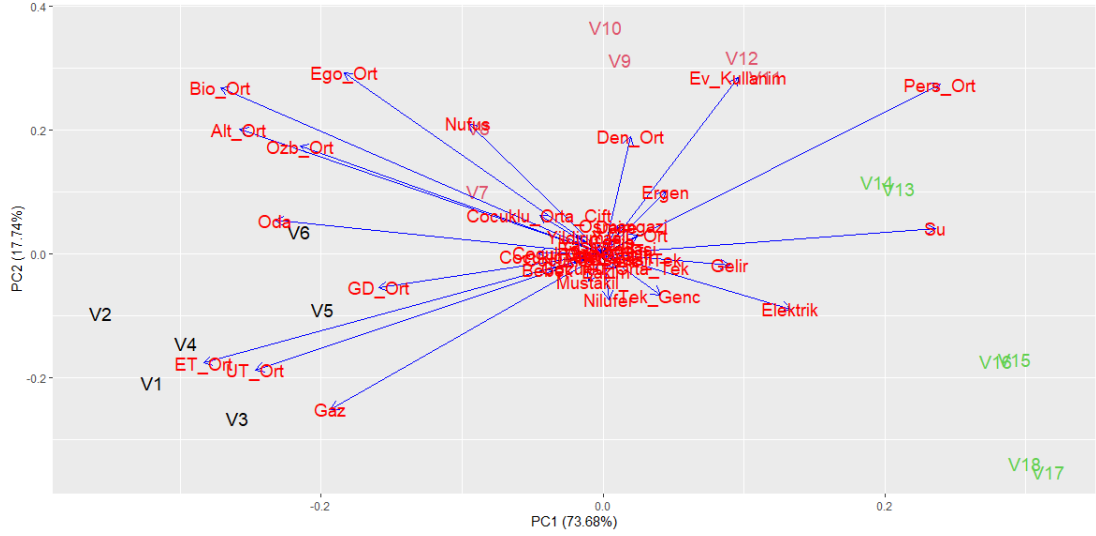
**Şekil 77.** Harita 132'nin (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına hiyerarşik kümeleme algoritması ile elde edilen  $k=6$  için küme bölümlerinin yansıtılması



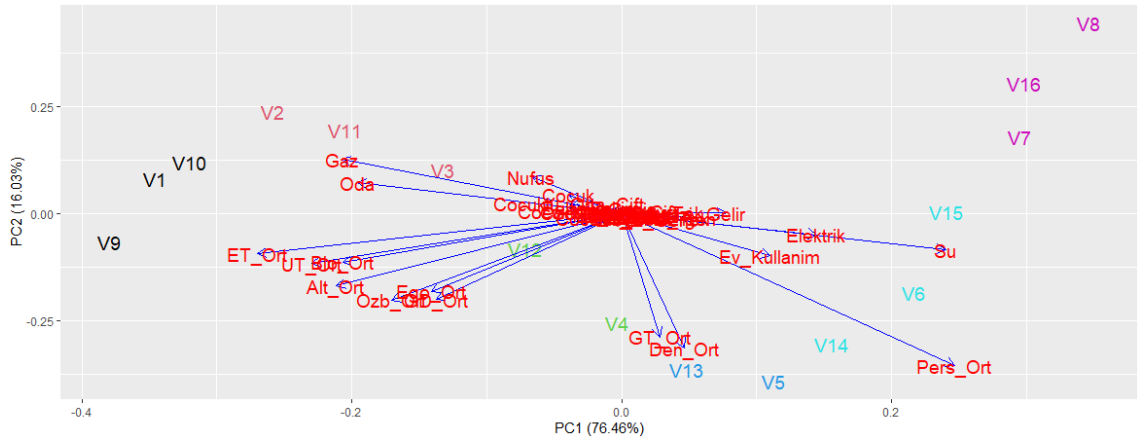
Şekil 78. Harita 95.'in (a) birinci, (b) ikinci ve (c) üçüncü katmanlarına k-temsilci kümeleme algoritması ile elde edilen  $k=4$  için küme bölümlenmelerinin yansıtılması

#### 5.3.2.4. TBA ile küme bütünlüğünün ve açıklanan varyansın kontrol edilmesi

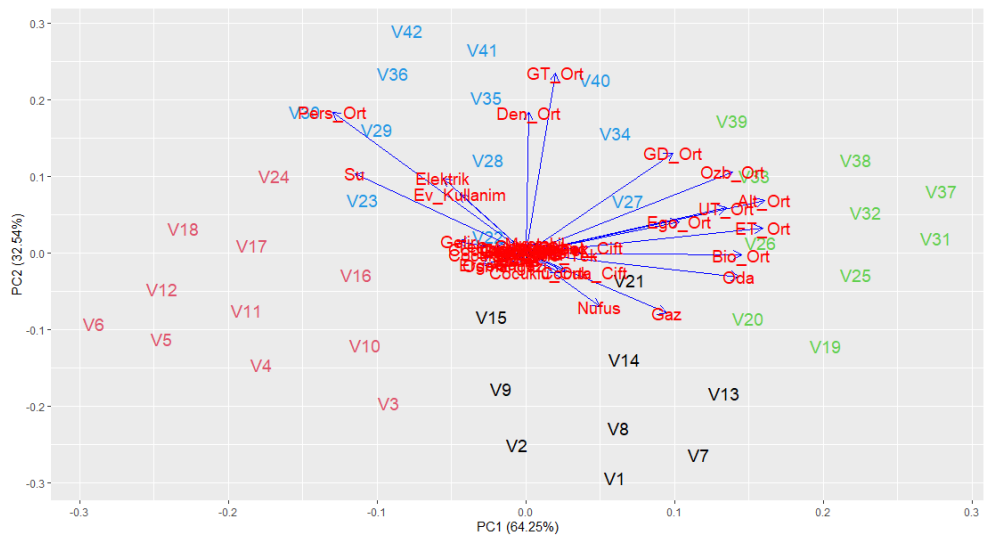
**Harita 11., Harita 132. ve Harita 95.'in** nöron ağırlıkları TBA'ya tabi tutulmuş ve analiz sonuçlarına göre birinci ve ikinci temel bileşenler varyansın sırası ile %91.42, %92.49 ve %96.79'luk kısmını açıklayabilmiştir. Tüm haritaların TBA grafiği incelendiğinde hepsi için elde edilen kümelerin temel bileşenler üzerinde çakışmadığı/iç içe geçmediği görülmektedir, Şekil (79.-81.). Fakat Harita 95. için küme sınırlarının biraz yakın olduğu dikkati çekmektedir, Şekil 81. Ayrıca Harita 2. ve Harita 132.'yi diğer tüm haritalardan farklılaştıran bir durum söz konusudur. Bu durum, her iki haritanın birinci temel bileşenlerini, varyansın %70'ten fazlasını açıklayabilmiş olmasıdır.



**Şekil 79.** Harita 2.'nin TBA analizi grafiği ve  $k = 3$  için k-temsili küme analizi ile elde edilen küme bölümlenmesinin TBA üzerine yansıtılması



**Şekil 80.** Harita 132.'nin TBA analizi grafiği ve  $k = 6$  için k-temsili küme analizi ile elde edilen küme bölümlenmesinin TBA üzerine yansıtılması



**Şekil 81.** Harita 95.'in TBA analizi grafiği ve  $k = 4$  için k-temsili küme analizi ile elde edilen küme bölümlenmesinin TBA üzerine yansıtılması



Harita 2., Harita 132. ve Harita 95., küme sayısı tespit etme prosedürünün iki, üç ve dört numaralı aşamaları kapsamında karşılaştırıldığında,<sup>173</sup> Tablo 51.'den inceleneceği üzere Harita 2. sadece tek kriter için en uygun değere sahiptir. Bu yüzden, üç harita içerisinde en kötü performansa sahip olan harita olarak değerlendirilmiş ve bu aşamada elenmiştir. Harita 132. ve Harita 95. ise eşit sayıda kriterden en uygun değere sahip olduğu için,<sup>174</sup> biri diğerine tercih edilmemiştir. Bu sebeple söz konusu iki haritanın ek olarak bir dışsal değerlendirme kriteri ile de değerlendirilmesi uygun görülmüştür.

**Tablo 51.** Harita 2., Harita 132. ve Harita 95.'in küme sayısı tespit etme prosedürünün iki, üç ve dört numaralı aşamaları kapsamında karşılaştırılması

Kriter	Harita 2 k=3 k-Temsilci	Harita 132 k=6 Hiyerarşik	Harita 95 k=4 k-Temsilci
Silüet	<b>0.4839</b>	0.3728	0.4258
Davies-Bouldin	0.6818	<b>0.6260</b>	0.7207
Dunn	0.4159	<b>0.6236</b>	0.2528
Calinski-Harabasz	31.5213	33.3843	<b>49.2548</b>
Küme bütünlüğü	Sağlanmış	Sağlanmış	Sağlanmış
Açıklanan varyans	%91.42	%92.49	<b>%96.79</b>

### 5.3.2.5. Varyans analizi ile küme ortalamaları arasındaki farklılığın kontrol edilmesi

**Harita 132.** için,  $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2$  ve  $\alpha = 0.05$  olmak üzere Şekil 82.'den incelendiğinde  $F = 0.4619$  için  $p = 0.954$  olup  $\alpha = 0.05$  değerinden büyüktür ve bu durum ayrıca gruplar arası ikili karşılaştırmalar için de geçerlidir. Dolayısı ile  $p > 0.05$  olduğu için yokluk hipotezi reddedilemez ve bağımsız değişkene ait grupların varyansları homojendir. Görselde dikkati çeken en önemli husus, küme üç ve dördün varyanslarının homejenliğinin test edilememiş olmasıdır. Söz konusu kümelere ait diğer program çıktıları incelendiğinde, her iki kümenin de standart hatalarının sıfır olduğu tespti edilmiştir. Bu durum, t-istatistiğinin paydasını sıfırlamakta ve değerinin hesaplanamamasına neden olmaktadır.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$  ve  $\alpha = 0.05$  olmak üzere Şekil 83.'ten incelendiğinde  $F = 33.384$  için  $p = 0.001$  olup  $\alpha = 0.05$  değerinden küçüktür ve bu

<sup>173</sup>  $KT$  değerleri bu kapsamda dikkate alınmamıştır. Çünkü küme sayısı arttıkça  $KT$  değeri azalacaktır. Örneğin  $k = 4$  için elde edilen  $KT$ ,  $k = 5$  için elde edilen  $KT$ 'den daha azdır.

<sup>174</sup> Harita 95'in silüet endeksi, harita 132'nin silüet endeksinden daha büyük olmasına rağmen, her iki haritanın küme bölümlenmelerinin ek olarak çok değişkenli varyans analizi ile de karşılaştırılması uygun görülmüştür.

durum ayrıca Şekil 84.'te yer alan gruplar arası ikili karşılaştırmalar için de geçerlidir. Dolayısı ile yokluk hipotezi reddedilir ve  $Ku$  kümeleri temsil etmek üzere, bağımsız değişken  $Ku$ 'ya ait grupların ortalamaları arasında anlamlı fark vardır; grup merkezlerinin lokasyonları farklıdır.

```

DEVIATIONS FROM CENTROID
F: 0,46194 df1: 5 df2: 10
P(perm): 0,954

PAIRWISE COMPARISONS
Groups      t      P(perm)
(1,2)      0,26748  0,915
(1,3)      0,019186  1
(1,4)      0,5121  1
(1,5)      0,83021  0,677
(1,6)      0,45565  0,689
(2,3)      0,30948  0,905
(2,4)      0,7943  0,706
(2,5)      0,53064  0,567
(2,6)      0,1327  1
(3,4)      No test
(3,5)      1,0339  0,704
(3,6)      0,68516  0,693
(4,5)      1,6316  0,392
(4,6)      1,456  0,501
(5,6)      0,51295  0,695

```

Şekil 82. Harita 132.'de  $k=6$  için elde edilen küme bölümlerinin varyanslarının homojenliği testi sonuçları

```

Factors
Name Abbrev. Type Levels
Kume Ku Fixed 6

PERMANOVA table of results
Source df SS MS Pseudo-F P(perm) Unique perms
Ku 5 48,588 9,7175 33,384 0,001 999
Res 10 2,9108 0,29108
Total 15 51,498

Details of the expected mean squares (EMS) for the model
Source EMS
Ku 1*v(Res) + 2,65*s(Ku)
Res 1*v(Res)

Construction of Pseudo-F ratio(s) from mean squares
Source Numerator Denominator Num.df Den.df
Ku 1*Ku 1*Res 5 10

Estimates of components of variation
Source Estimate Sq.root
s(Ku) 3,5571 1,886
v(Res) 0,29108 0,53952

```

Şekil 83. Harita 132.'de  $k=6$  için elde edilen küme bölümlerinin çok değişkenli varyans analizi sonuçları

Factors				
Name	Abbrev.	Type	Levels	
Kume	Ku	Fixed	6	

PAIR-WISE TESTS				
Term 'Ku'				
Groups	t	P(pern)	Unique	perms
1, 2	3,46	0,096	10	
1, 3	5,1157	0,126	10	
1, 4	7,0366	0,091	10	
1, 5	8,0861	0,093	10	
1, 6	10,464	0,106	10	
2, 3	2,9348	0,099	10	
2, 4	5,1316	0,106	10	
2, 5	6,1259	0,111	10	
2, 6	7,7307	0,088	10	
3, 4	2,1007	0,34	3	
3, 5	3,377	0,105	10	
3, 6	5,6014	0,081	10	
4, 5	2,2409	0,22	10	
4, 6	5,3899	0,105	10	
5, 6	3,3157	0,085	10	

Denominators		
Groups	Denominator	Den. df
1, 2	1*Res	4
1, 3	1*Res	3
1, 4	1*Res	3
1, 5	1*Res	4
1, 6	1*Res	4
2, 3	1*Res	3
2, 4	1*Res	3
2, 5	1*Res	4
2, 6	1*Res	4
3, 4	1*Res	2
3, 5	1*Res	3
3, 6	1*Res	3
4, 5	1*Res	3
4, 6	1*Res	3
5, 6	1*Res	4

Average Distance between/within groups						
	1	2	3	4	5	6
1	0,6575					
2	1,5171	0,73874				
3	2,4326	1,4905	0,76056			
4	3,1353	2,435	1,1044	0,63787		
5	3,728	2,9455	1,8947	1,2303	0,84699	
6	4,4789	3,4629	2,818	2,5672	1,6153	0,75376

Şekil 84. Harita 132.'de  $k=6$  için elde edilen küme bölümlerinin çok değişkenli varyans analizine ait gruplar arası ikili karşılaştırma sonuçları

**Harita 95.** için,  $H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$  ve  $\alpha = 0.05$  olmak üzere Şekil 85.'ten incelendiğinde  $F = 0.1756$  için  $p = 0.914$  olup  $\alpha = 0.05$  değerinden büyüktür ve bu durum ayrıca gruplar arası ikili karşılaştırmalar için de geçerlidir. Dolayısı ile  $p > 0.05$  olduğu için yokluk hipotezi reddedilemez ve bağımsız değişkene ait grupların varyansları homojendir.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$  ve  $\alpha = 0.05$  olmak üzere Şekil 86'dan incelendiğinde  $F = 49.255$  için  $p = 0.001$  olup  $\alpha = 0.05$  değerinden küçüktür ve bu durum ayrıca Şekil 87'de yer alan gruplararası ikili karşılaştırmalar için de geçerlidir. Dolayısı ile yokluk hipotezi reddedilir ve  $Ku$  kümeleri temsil etmek üzere, bağımsız değişken  $Ku$ 'nun grup ortalamaları arasında anlamlı fark vardır ve grup merkezlerinin lokasyonları farklıdır.

DEVIATIONS FROM CENTROID		
F:	0,17567	df1: 3 df2: 38
P(perm):	0,914	
PAIRWISE COMPARISONS		
Groups	t	P(perm)
(1,2)	0,071619	0,946
(1,3)	0,1457	0,887
(1,4)	0,53224	0,605
(2,3)	0,21074	0,83
(2,4)	0,44291	0,643
(3,4)	0,68563	0,49

Şekil 85. Harita 95.'te  $k=4$  için elde edilen küme bölümlerinin varyanslarının homojenliği testi sonuçları

Factors						
Name	Abbrev.	Type	Levels			
Küme	Ku	Fixed	4			
PERMANOVA table of results						
Source	df	SS	MS	Pseudo-F	P(perm)	Unique perms
Ku	3	65,079	21,693	49,255	0,001	999
Res	38	16,736	0,44042			
Total	41	81,814				
Details of the expected mean squares (EMS) for the model						
Source	EMS					
Ku	$1 * V(\text{Res}) + 10,46 * S(\text{Ku})$					
Res	$1 * V(\text{Res})$					
Construction of Pseudo-F ratio(s) from mean squares						
Source	Numerator	Denominator	Num. df	Den. df		
Ku	$1 * \text{Ku}$	$1 * \text{Res}$	3	38		
Estimates of components of variation						
Source	Estimate	Sq. root				
S(Ku)	2,0317	1,4254				
V(Res)	0,44042	0,66364				

Şekil 86. Harita 95.'te  $k=4$  için elde edilen küme bölümlerinin çok değişkenli varyans analizi sonuçları

Harita 132. ve 95.'in varyans analizi sonuçlarını karşılaştırmak gerekirse:

- Şekil (83., 86.) inceleneceği üzere her iki harita için elde edilen  $p$  değerleri 0.01 olup eşittir. Fakat  $MS$  ile temsil edilen ortalama hata kare değerleri karşılaştırıldığında harita 132'nin ortalama hata karesinin (9.7175), harita 95'inkinden (21.693) çok daha küçük olduğu gözlemlenmektedir.

- Ayrıca, Şekil 84.'te "Average distance between/within groups" başlığı altında yer aldığı üzere, 132 numaralı haritada küme içi ortalama mesafeler (matrisin köşegeni) harita 95'in küme içi ortalama mesafelere kıyas ile daha küçüktür. Bu durum, Harita 132.'deki kümelerin, Harita 95.'e kıyas ile geometrik olarak daha dar bir alana saçıldığıнын, kümelerin daha kompakt bir yapıya sahip olduğunun işaretidir.

- Ek olarak Harita 132.'de kümeler arası mesafe değerleri (1.1044 – 4.4789) aralığında değişim gösterirken, Harita 95'te (1.7671-3.0363) aralığında değişmektedir. Harita 132. için hesaplanan en kısa mesafe, Harita 95. için hesaplanan değerden 0.3427 birim kadar daha az iken, en uzun mesafe 1.4426 birim kadar daha fazladır. Kümeler arası mesafe değerlerinin, küme içi mesafelere kıyas ile daha fazla olması beklenmektedir. Bu bağlamda, 132 numaralı harita, 95 numaralı haritadan biraz daha iyi performansa sahiptir.

Özetle, 132 numaralı harita, daha küçük ortalama hata kare, küme içi ortalama mesafe, Davies-Bouldin endeksi ve NH'ye ve daha büyük kümeler arası ortalama mesafe ve Dunn değerlerine sahip olduğu için kimlik tanımlama aşamasında kullanılmak üzere iki numaralı aday harita olmuştur.

```

Factors
Name Abbrev. Type Levels
Kume Ku Fixed 4

PAIR-WISE TESTS

Term 'Ku'

Groups t P(perm) Unique
perms
1, 2 6,3803 0,001 993
1, 3 5,5512 0,001 990
1, 4 6,2694 0,001 996
2, 3 10,417 0,001 997
2, 4 5,803 0,001 996
3, 4 6,6865 0,001 997

Denominators
Groups Denominator Den. df
1, 2 1*Res 18
1, 3 1*Res 17
1, 4 1*Res 19
2, 3 1*Res 19
2, 4 1*Res 21
3, 4 1*Res 20

Average Distance between/within groups
1 2 3 4
1 0,86074
2 1,9866 0,86998
3 1,7671 3,0363 0,83484
4 1,9977 1,8048 2,0281 0,91858

```

Şekil 87. Harita 95'te  $k=4$  için elde edilen küme bölümlerinin çok değişkenli varyans analizine ait gruplar arası ikili karşılaştırma sonuçları

Aday haritalar karşılaştırıldığında, Harita 227.'nin, dokuz kriterden beşinde ve Harita 132'nin ise dokuz kriterden üçünde daha iyi sonuçlar sağladığı görülmektedir, Tablo 52.'den. Bu sebeple, hanelerin yeşil kimlik profillerinin tespit edilmesi amacı ile oluşturulan 586 adet SÖÖH'den, çevrimiçi öğrenme algoritması ile eğitilen 227 numaralı Haritanın kimlik tanımlama aşamasında kullanılmasına karar verilmiştir.

Tablo 52. Harita 227 ve harita 132'nin tüm kriterler bağlamında karşılaştırılması

Kriter	Harita 227 k=5 Çevrimiçi k-Temsilci	Harita 132 k=6 Yığın Hiyerarşik
Nicemleme hatası	<b>0.1910</b>	0.2354
Silüet	<b>0.4854</b>	0.3728
Davies-Bouldin	<b>0.5645</b>	0.6260
Dunn	0.1819	<b>0.6236</b>
Calinski-Harabasz	<b>46.5098</b>	33.3843
Küme bütünlüğü	Sağlanmış	Sağlanmış
Açıklanan varyans	%90.86	<b>%92.49</b>
MS	<b>9.2532</b>	9.7175
Küme içi mesafe	<b>0.5447-0.6237</b>	0.6378-0.8469
Kümeler arası mesafe	1.2991-2.6211	<b>1.1044-4.4789</b>

#### 5.4. Kimlik Tanımlama

Bu aşamada, 227 numaralı haritanın  $k = 5$  için k-temsili yöntemi ile kümeleneceği sonucunda elde edilen küme bölümlenmeleri, gerçek veri setine yansıtılmıştır. Küme ortalamaları ve kutu-bıyık grafikleri ile incelenmiş ve her bir kümenin ekolojik kimlik profilleri tanımlanmıştır. Akabinde karar ağaçları ile hem kümeleri birbirinden ayırt eden en önemli öznitelikler tespit edilmiş hem de her kümeye ait bir kural tabanı oluşturularak veri toplama aşamasında kullanılan ölçme araçlarının hangi boyutlarının ve hangi tüketici davranışlarının ne tür bir profil oluşturduğu tespit edilmiştir.

##### 5.4.1. Betimsel İstatistikler ile Tanımlama

Öncelikle, ikili ve kategorik özniteliklerin frekansları daha sonra ise yüzdeleri hesaplanmış ve sütun grafik haline getirilmiştir. Akabinde nümerik özniteliklerin ortalamaları dikkate alınarak sütun grafikleri ve kutu-bıyık diyagramları oluşturulmuş ve kümelerin ekolojik kimlik profilleri tanımlanmıştır.

**Birinci küme**, 111 gözleme (hane) sahip olup, sadece Osmangazi ilçesinde yaşayan ve küme analizinde hane halkı türü özneliği ile temsil edilen aile yaşam döngüsünün, başında ya da sonunda olan hanelerden oluşmaktadır, Şekil 88. Oranı %10'un üstünde olan hane halkı türlerini incelemek gerekirse kümenin %40'a yakını tek başına yaşayan genç nüfuslu hanelerden, %12'si genç çiftlerden, %15'i çocuklu genç çiftlerden ve %15'i yaşlı çiftlerden meydana gelmektedir. Dolayısı ile birinci kümede, Osmangazi ilçesinde yaşayan ve nüfusu genellikle az olan haneler yer almaktadır. Bu sebeple bu kümede yer alan hanelerin çok az bir kısmında bebek, küçük çocuk ya da ergen çocuk (%6, %8 ve %12) ve yüksek oranda yaşlı yaşamaktadır (%24). Ayrıca bu küme, bakıma muhtaç hasta, engelli ve yaşlı oranı (%21) en yüksek olan kümedir. Son olarak, bu kümede yer alan hanelerin %20'si müstakil evde yaşamaktadır. Bu oran, beşinci küme hariç, diğer kümelere nispeten yüksek bir orandır.

Şekil (89, 91) incelendiğinde ise birinci küme, Şekil 88'de yer alan bilgilere paralel olarak,

- En az ortalama nüfusa (1.8 kişi) sahip iki kümeden birisidir: Nüfus özneliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın sağa çarpık olduğu, hanelerin yaklaşık

%25'inin üç kişilik, %75'inin ise en fazla iki kişilik olduğu ve ayrıca aykırı gözlem olarak çok az sayıda dört ve beş kişilik hanelerin de bu kümede yer aldığı görülmektedir.

- Oda sayısı genellikle iki ile beş arasında değişen meskenlerde (*ortalama 3.1 oda*) yaşayan haneleri içermektedir: Oda özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın sola çarpık olduğu, hanelerin yaklaşık %75'inin üç veya dört odalı, yaklaşık %25'inin iki odalı ve ayrıca aykırı gözlem olarak az sayıda hanenin tek odalı meskende yaşadığı görülmektedir.

- İkinci en fazla gelire (*ortalama 9693 TL*) sahip olan haneleri kapsayan kümedir: Gelir özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın sivri olduğu, ikinci küme hariç olmak üzere IQR'nin diğer kümelere kıyas ile dar olduğu, hanelerin %25'inin 4,500 – 8,250 TL, %50'sinin 8,250 – 11,000 TL, %25'inin 11,000 – 15,000 TL aralığında gelire sahip olduğu ve ayrıca çok az sayıda hanenin aylık gelirin 4,500 TL'nin altında veya 15,000 TL'nin üzerinde olduğu görülmektedir.

Bu kümede yer alan hanelerde yaşayan vatandaşlar, genellikle

- Evde kalma/yaşama süresi en az olan vatandaşlardır: Ev\_kullanım özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın sola çarpık olduğu, IQR'nin diğer kümelere nispeten dar olduğu ve bu kümedeki hanelerin %50'sinin evde yaşam süresinin 13 – 17 saat arasında değiştiği ve yaklaşık %25'inin 15 – 17 saat evde yaşadığı görülmektedir. Dolayısı ile bu grupta yer alan hanelerde yaşayan ev halkının çoğunlukla çalıştığı, çocukları var ise okula ya da kreşe gittiği ve dolayısı ile evde bulunma süresinin az olduğu ifade edilebilir.

- Elektrik ve suyu diğer kümelere kıyas ile orta seviyede tüketirler: Elektrik ve Su özniteliklerine ait kutu-bıyık diyagramları incelendiğinde, her ikisinin de dağılımının sağa çarpık olduğu, diğer kümelere kıyas ile IQR'nin elektrik için ortalama, su için ise çok geniş olduğu, ayrıca her iki öz nitelik için aykırı gözlem olarak az sayıda hanenin tüketimlerinin 300 TL üzerinde olduğu görülmektedir.

- Doğal gaz tüketimleri az olan vatandaşlardır: Gaz özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın sağa çarpık olduğu, gaz tüketiminin %50'sinin 188 – 300 TL aralığında, %25'inin 90 – 188 TL arasında ve diğer %25'inin nispeten daha geniş olan 300 – 450 TL aralığında olduğu ve ayrıca çok az sayıda hanenin doğal gaz tüketiminin 450 TL'nin üzerinde olduğu görülmektedir.

Şekil (90, 91.) incelendiğinde ise bu küme, ikinci en küçük egoistik, biyosferik ve altruistik tutum puanına sahip olan kümedir. Bu kümeyi diğer kümelerden ayıran en önemli özellik, biyosferik tutum puanının, egoistik ve altruistik tutum puanından küçük olan tek küme olmasıdır (bu durum diğer kümelerde tam tersidir). Bu durum ile paralel olarak, bu kümenin özbenlik boyutundan aldığı puanların ortalaması ikinci en küçük ve doğayla ilişki bağlamında olumsuz ifadelerin bulunduğu perspektif boyutundan aldığı puanların ortalaması ise en büyüktür. Fakat diğer taraftan geri dönüşüm, gıda tüketimi, ulaştırma tercihi, elektrik ve su tüketimi davranışı özniteliklerine ait boyutlardan aldığı puanlar, diğer kümelere kıyas ile ikinci en yüksektir.

Bu tespitler denekler tarafından beyan edilen bilgilerin doğru olduğunun varsayıldığı durumlar içindir. Zira bu küme, bakıma muhtaç hasta, engelli ve yaşlıların en fazla olduğu küme olmasına rağmen, aykırı gözlemler dışında, Ev\_kullanım değişkeninin maksimum değeri 17.5 saattir. Halbu ki, bakıma muhtaç bireylere sahip olan hanelerin ev kullanım sürelerinin 24 saat olması beklenmektedir.

**İkinci küme**, 180 gözleme (hane) sahip olup, sadece Osmangazi ilçesinde yaşayan ve aile yaşam döngüsünün, ortalarında olan hanelerden oluşmaktadır, Şekil 88. Oranı %10'un üstünde olan hane halkı türlerini incelemek gerekirse kümenin %11'i çocuklu genç çiftlerden, %65'i çocuklu orta yaşlı çiftlerden ve %10'u geniş aileden meydana gelmektedir. Başka bir ifade ile ikinci kümede, Osmangazi ilçesinde yaşayan ve nüfusu en kalabalık olan haneler yer almaktadır. Bu sebeple bu kümede yer alan hanelerde yüksek oranda çocuk ya da ergen çocuk ( %48 ve %51) veya yaşlı yaşamaktadır (%16). Ek olarak, bakıma muhtaç hasta, engelli ve yaşlı oranı da %10'un üzerindedir (%13). Ayrıca bu kümede yer alan hanelerin neredeyse hepsi apartman dairesinde yaşamaktadır.

Şekil (89, 91) incelendiğinde ise ikinci küme, Şekil 88'de yer alan bilgilere paralel olarak,

- En fazla ortalama nüfusa (3.6 kişi) sahip kümedir: Nüfus özneliğine ait kutubıyık diyagramı incelendiğinde hanelerin %75'inin üç veya dört kişilik olduğu ve ayrıca aykırı gözlem olarak çok az sayıda bir ve altı kişilik hanelerin de bu kümede yer aldığı görülmektedir.



- Oda sayısı nerede ise sadece dört olan meskenlerde (*ortalama 4.1 oda*) yaşayan haneleri içermektedir: Oda özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın tamamı ile sivri olduğu “positive curtosis” ve ayrıca aykırı gözlem olarak bazı hanelerin iki, üç, beş ve altı odalı meskenlerde yaşadığı görülmektedir.

- Orta seviyede gelire (*ortalama 9,693 TL*) sahip olan haneleri kapsayan iki kümeden birisidir: Gelir özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın (aykırı gözlemler hariç) normal dağılıma yakın fakat biraz basık olduğu “negative curtosis”, IQR’nin diğer kümelere kıyas ile en dar olduğu, hanelerin %25’inin 5,000 – 8,000 TL, %50’sinin 8,000 – 10,000 TL, %25’inin 10,000 – 13,000 TL aralığında gelire sahip olduğu ve ayrıca çok az sayıda hanenin aylık gelirin 5,000 TL’nin altında veya 13,000 TL’nin üzerinde olduğu görülmektedir.

Bu kümede yer alan hanelerde yaşayan vatandaşlar, genellikle

- Evde kalma/yaşama süresi ikinci en fazla olan vatandaşlardır: Ev\_kullanım özniteliğine ait kutu-bıyık diyagramı incelendiğinde IQR’nin en fazla olduğu ve bu kümedeki hanelerin %25’inin evde yaşam süresinin 12 – 14 saat arasında değiştiği, diğer %25’inin evde yaşam süresinin 14 – 16 saat arasında değiştiği ve geriye kalan %50’sinin 16 – 24 saat evde yaşadığı görülmektedir. Dolayısı ile bu grupta yer alan hanelerde yaşayan ev halkının bir kısmının çalıştığı, çocukların okula gittiği ve evin olası muhtemel gün içerisinde boş kaldığı, diğer kısmının ise gün içerisinde evde kalan ya da evde kalması gereken aile bireylerine sahip olduğu ifade edilebilir (annenin ev hanımı olması, ergen çocukların okul sonrası ya da öncesi evde bulunması, bakıma muhtaç bireylerin veya geniş ailelerde büyük anne-babanın evde kalması gibi).

- Elektrik ve suyu diğer kümelere kıyas ile orta ve düşük seviyede tüketirler: Elektrik ve Su özniteliklerine ait kutu-bıyık diyagramları incelendiğinde, elektrik tüketiminin ortalamasının minimum ve tüketim değerlerinin %50’sinin 145-200 TL arasında olduğu, su tüketiminin sağa çarpık ve en düşük alt ve üst sınır değerlerine sahip olduğu, ayrıca her iki öz niteliğin de aykırı gözlem içermediği görülmektedir.

- Doğal gaz tüketimleri az olan vatandaşlardır: Gaz özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın sağa çarpık olduğu, gaz tüketiminin %50’sinin 100 – 200 TL aralığında, %25’inin 200 – 320 TL arasında ve diğer %25’inin nispeten

daha geniş olan 320 – 500 TL aralığında olduğu ve ayrıca çok az sayıda hanenin doğal gaz tüketiminin 500 TL'nin üzerinde olduğu görülmektedir.

Şekil (90., 91.) incelendiğinde ise bu küme, ikinci en büyük egoistik, biyosferik ve altruistik tutum puanına sahip olan kümedir. Biyosferik tutum puanı, egoistik ve altruistik tutum puanından büyüktür. Çünkü biyosferik tutum puanlarının %50'si altı veya yedidir. Diğer taraftan altruistik ve egoistik tutum puanlarının %50'si beş ve üzeridir. Hatta 5 – 6 aralığı, söz konusu %50'nin yarısını temsil etmektedir. Bu durum ile paralel olarak, bu kümenin özbenlik boyutundan aldığı puanların ortalaması ikinci en büyük ve doğayla ilişki bağlamında olumsuz ifadelerin bulunduğu perspektif boyutundan aldığı puanların ortalaması ise en ikinci en küçüktür. Fakat diğer taraftan geri dönüşüm, gıda tüketimi, ulaştırma tercihi, elektrik ve su tüketimi özniteliklerine ait boyutlardan aldığı puanlar, diğer kümelere kıyas ile ya en küçük ya da orta seviyededir.

**Üçüncü küme**, 164 gözleme (hane) sahip olup, %80'i Yıldırım ve %20'si Nilüfer ilçelerinde yaşayan ve küme analizinde hane halkı türü özneliği ile temsil edilen aile yaşam döngüsünün, başında ve ortalarında olan hanelerden oluşmaktadır, Şekil 88. Oranı %10'un üstünde olan hane halkı türlerini incelemek gerekirse kümenin %21'i çocuklu genç çiftlerden ve %65'i çocuklu orta yaşlı çiftlerden meydana gelmektedir. Başka bir ifade ile üçüncü kümede, Yıldırım (çoğunluk) ve Nilüfer (azınlık) ilçelerinde yaşayan ve nüfusu en kalabalık olan haneler yer almaktadır. Bu sebeple bu kümede yer alan hanelerde yüksek oranda bebek, çocuk ya da ergen çocuk (%27, %52 ve %28) yaşamaktadır. Ek olarak, bakıma muhtaç hasta, engelli ve yaşlı oranı da %10'un üzerindedir (%11). Ayrıca bu kümede yer alan hanelerin %16'sı müstakil evlerde yaşamaktadır.

Şekil (89, 91) incelendiğinde ise üçüncü küme, Şekil 88'de yer alan bilgilere paralel olarak,

- İkinci en fazla ortalama nüfusa (3.5 kişi) sahip kümedir: Nüfus özneliğine ait kutu-bıyık diyagramı incelendiğinde hanelerin %75'inin üç veya dört kişilik olduğu ve ayrıca aykırı gözlem olarak çok az sayıda bir, altı ve yedi kişilik hanelerin de bu kümede yer aldığı görülmektedir.

- En büyük meskenlerde (ortalama 4.6 oda) yaşayan haneleri içermektedir: Oda özneliğine ait kutu-bıyık diyagramı incelendiğinde, hanelerin %75'inin dört veya

beş odalı ve ayrıca aykırı gözlem olarak az sayıda gözlemin üç, yedi veya sekiz odalı meskenlerde yaşadığı görülmektedir.

- En az gelire (*ortalama 8,424 TL*) sahip olan haneleri kapsayan kümedir: Gelir özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın sağa yatık olduğu, hanelerin %25'inin 2,000 – 6,000 TL, %50'sinin 6,000 – 10,000 TL, %25'inin 10,000 – 16,000 TL aralığında gelire sahip olduğu ve ayrıca çok az sayıda hanenin aylık gelirinin 16,000 TL'nin üzerinde olduğu görülmektedir.

Bu kümede yer alan hanelerde yaşayan vatandaşlar, genellikle

- Evde kalma/yaşama süresi ikinci en fazla olan vatandaşlardır: Ev\_kullanım özniteliğine ait kutu-bıyık diyagramı incelendiğinde bu kümedeki hanelerin %50'sinin evde yaşam süresinin 13 – 16 saat arasında değiştiği, diğer %50'sinin evde yaşam süresinin 16 – 24 saat arasında daha geniş bir bantta değiştiği görülmektedir. Dolayısı ile bu grupta yer alan hanelerde yaşayan ev halkının ikinci kümedeki hanelere benzer şekilde bir kısmının çalıştığı, çocukların okula gittiği ve evin olası muhtemel gün içerisinde boş kaldığı, diğer kısmının ise gün içerisinde evde kalan ya da evde kalması gereken aile bireylerine sahip olduğu ifade edilebilir.

- Elektrik ve suyu diğer kümelere kıyas ile en az seviyede tüketirler: Elektrik ve Su özniteliklerine ait kutu-bıyık diyagramları incelendiğinde, IQR'nin diğer kümelere kıyas ile her iki öznitelik için de en az olduğu ve tüketim değerlerinin %50'sinin sırası ile 120 – 180 TL ve 100 – 200 TL aralığında değiştiği, ayrıca her iki öz niteliğin de az sayıda aykırı gözlem içerdiği görülmektedir.

- Doğal gaz tüketimleri en fazla olan vatandaşlardır: Gaz özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın sağa çarpık olduğu, gaz tüketiminin %50'sinin 200 – 400 TL aralığında, %25'inin 25 – 200 TL arasında ve diğer %25'inin nispeten daha geniş olan 400 – 700 TL aralığında olduğu görülmektedir.

Şekil (90., 91.) incelendiğinde ise bu küme, en büyük egoistik, biyosferik ve altruistik tutum puanına sahip olan kümedir. Tutum puanları büyükten küçüğe biyosferik, altruistik ve egoistik olmak üzere sıralanmaktadır. Bu durum ile paralel olarak, bu kümenin özbenlik boyutundan aldığı puanların ortalaması en büyük ve doğayla ilişki bağlamında olumsuz ifadelerin bulunduğu perspektif boyutundan aldığı puanların ortalaması ise en küçüktür. Ayrıca geri dönüşüm, gıda tüketimi, ulaştırma tercihi, elektrik

ve su tüketimi özniteliklerine ait boyutlardan aldığı puanlar, diğer kümelere kıyas ile en büyüktür. Başka bir ifade ile bu kümede yer alan hanelerin (hane temsilcilerinin) çevresel endişeleri doğa ile ilişkileri ve beyan ettikleri sürdürülebilir tüketim davranışları yüksektir.

**Dördüncü küme**, 187 gözleme (hane) sahip olup, %48'i Yıldırım ve %52'si Nilüfer ilçelerinde yaşayan ve küme analizinde hane halkı türü özneliği ile temsil edilen aile yaşam döngüsünün çoğunlukla başında olmak üzere her aşamasından olan hanelerden oluşmaktadır, Şekil 88. Oranı %10 civarı olan hane halkı türlerini incelemek gerekirse kümenin %9'u tek gençlerden %15'i çocuksuz genç çiftlerden, %28'i çocuklu genç çiftlerden, %10'u çocuklu orta yaşlı çiftlerden, %9'u çocuksuz yaşlı çiftlerden ve %12'si geniş ailelerden meydana gelmektedir. Başka bir ifade ile üçüncü küme, sadece Nilüfer ve Yıldırım ilçelerinde yaşayan, aile yaşam döngüsü açısından her türden aileyi içeren bir kümedir. Aile çeşitliliğinin yüksek olmasından ötürü, Bakım, Bebek, Çocuk, Ergen ve Yaşlı değişkenleri yüksek oranlara sahiptir. Bu oranlar sırası ile %17, %14, %19, %23, ve %24 şeklindedir. Ayrıca bu kümede yer alan hanelerin %16'sı müstakil evlerde yaşamaktadır.

Şekil (89., 91.) incelendiğinde ise dördüncü küme, Şekil 88.'de yer alan bilgilere paralel olarak,

- Diğer kümelere kıyas ile orta derecede (üçüncü en fazla) nüfusa (2.7 kişi) sahiptir: Nüfus özneliğine ait kutu-bıyık diyagramı incelendiğinde, aykırı gözlemler hariç, dağılımın normale yakın olduğu, hanelerin %50'sinin iki veya üç kişilik olduğu ve ayrıca aykırı gözlem olarak çok az sayıda beş, altı ve yedi kişilik hanelerin de bu kümede yer aldığı görülmektedir.

- Orta büyüklükte meskenlerde (ortalama 3.7 oda) yaşayan haneleri içermektedir: Oda özneliğine ait kutu-bıyık diyagramı incelendiğinde, hanelerin %50'sinden fazlasının üç veya dört odalı ve ayrıca aykırı gözlem olarak az sayıda gözlemin bir, altı veya yedi odalı meskenlerde yaşadığı görülmektedir.

- Orta seviyede gelire (ortalama 9,128 TL) sahip olan haneleri kapsayan kümedir: Gelir özneliğine ait kutu-bıyık diyagramı incelendiğinde, hanelerin %25'inin 1.500 – 7,000 TL, %50'sinin 7,000 – 11,000 TL, %25'inin 11,000 – 16,000 TL aralığında gelire sahip olduğu ve ayrıca çok az sayıda hanenin aylık gelirinin

16,000 TL'nin üzerinde olduğu görülmektedir. Bu küme, minimum ve maksimum gelir arasındaki farkın en fazla olduğu kümedir. Bu durumun, kümenin aile yaşam döngüsü özniteliğine ait değerlerin çeşitliliğinden kaynaklandığı düşünülmektedir.

Bu kümede yer alan hanelerde yaşayan vatandaşlar, genellikle

- Evde kalma/yaşama süresi en fazla olan vatandaşlardır: Ev\_kullanım özniteliğine ait kutu-bıyık diyagramı incelendiğinde, dağılımının ikinci küme ile benzeştiği ve ikinci kümeden farklı olarak medyan ile maksimum arasında çok sayıda gözleme sahip olmasıdır. Bu durumun, kümenin aile yaşam döngüsü özniteliğine ait değerlerin çeşitliliğinden kaynaklandığı düşünülmektedir.

- Elektrik ve suyu diğer kümelere kıyas ile orta ve az seviyelerde tüketirler: Elektrik ve Su özniteliklerine ait kutu-bıyık diyagramları incelendiğinde, her iki özniteliğin de dağılımı iki tepeye sahiptir. Bu durumun da, Ev\_kullanım ve Gelir değişkenlerinde gözlemlendiği gibi, kümenin aile yaşam döngüsü özniteliğine ait değerlerin çeşitliliğinden kaynaklandığı düşünülmektedir.

- Doğal gaz tüketimleri fazla olan vatandaşlardır: Gaz özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın sağa çarpık olduğu, gaz tüketiminin %50'sinin 200 – 400 TL aralığında, %25'inin 20 – 200 TL arasında ve diğer %25'inin nispeten daha geniş olan 400 – 680 TL aralığında olduğu görülmektedir.

Şekil (90, 91) incelendiğinde bu kümede yer alan vatandaşların çevresel endişe, doğayla ilişki ve sürdürülebilir tüketim davranışı özniteliklerine ait boyutların puanları, diğer kümelere kıyas ile orta seviyededir. Bu durumla paralel olarak, elektrik ve su tüketimleri orta-az seviyede fakat doğalgaz tüketimleri yüksek seviyededir.

**Beşinci küme**, 106 gözleme (hane) sahip olup, sadece Nilüfer ilçesinde yaşayan ve dördüncü kümeye benzer şekilde aile yaşam döngüsünün, çoğunlukla başında olmak üzere her aşamasından olan hanelerden oluşmaktadır, Şekil 88. Oranı %10 civarı olan hane halkı türlerini incelemek gerekirse kümenin %21'i tek gençlerden, %10'u çocuksuz genç çiftlerden, %13'ü çocuklu genç çiftlerden, %16'sı çocuklu orta yaşlı tek ebeveynlerden, %11'i tek başına yaşlı birey ve %9'u yaşlı çiftlerden meydana gelmektedir. Bu kümede ayrıca %7 oranında çocuklu genç tek ebeveynler de yer almaktadır. Özetle, beşinci küme, sadece Nilüfer ilçesinde yaşayan ve ailede genellikle bir veya iki yetişkinin

bulunduđu hanelerden oluřmaktadıř. Ek olarak, bu kúmede yer alan hanelerin %23'ü müstakil evlerde yařamaktadır.

řekil (89, 91) incelendiđinde ise beřinci kúme, řekil 88'de yer alan bilgilere paralel olarak,

- En az ortalama nüfusa (2 kiři) sahip iki kúmeden birisidir: Nüfus özniteliđine ait kutu-bıyık diyagramı incelendiđinde dađılımin sađa çarpık olduđu, hanelerin yaklaşık %50'sinin iki kiřilik ve diđer %50'sinin ise bir veya üç kiřilik olduđu ve ayrıca aykırı gözlem olarak çok az sayıda dört, beř ve altı kiřilik hanelerin de bu kúmede yer aldıđı görölmektedir.

- Oda sayısı genellikle üç olan meskenlerde (*ortalama 2.9 oda*) yařayan haneleri içermektedir: Oda özniteliđine ait kutu-bıyık diyagramı incelendiđinde IQR'nin sıfır olduđu dikkati çekmektedir. Hanelerin çok büyük bir kısmı üç odalı ve geri kalan haneler ise çođunlukla 2 odalı meskenlerde yařamaktadır. Az sayıda da olsa ayrıca bir, dört, beř ve altı odalı meskenlerde yařayan haneler de bu kúmede yer almaktadır.

- En fazla gelire (*ortalama 10,204 TL*) sahip olan haneleri kapsayan kúmedir: Gelir özniteliđine ait kutu-bıyık diyagramı incelendiđinde dađılımin sađa çarpık olduđu, hanelerin yaklaşık %25'inin 4,000 – 8,000 TL, %25'inin 8,000 – 9,180 TL, %25'inin 9,180 – 12,000 TL ve yaklaşık %25'inin 12,000 – 18,000 TL aralıđında gelire sahip olduđu görölmektedir.

Bu kúmede yer alan hanelerde yařayan vatandaşlar genellikle;

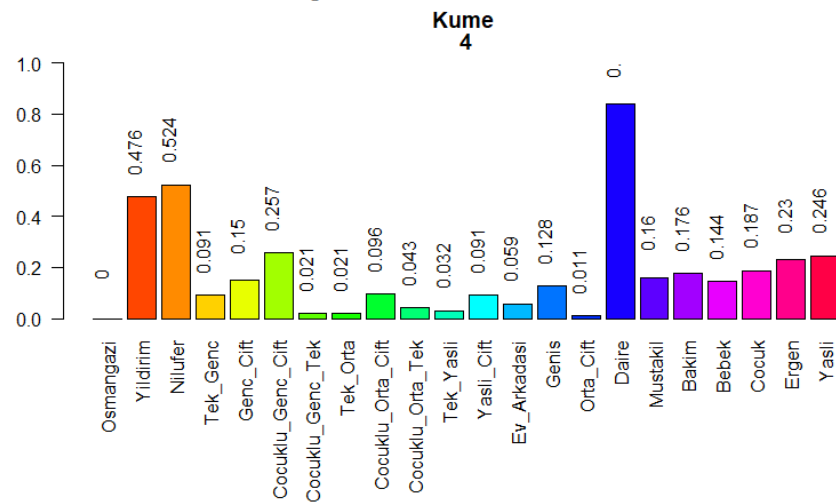
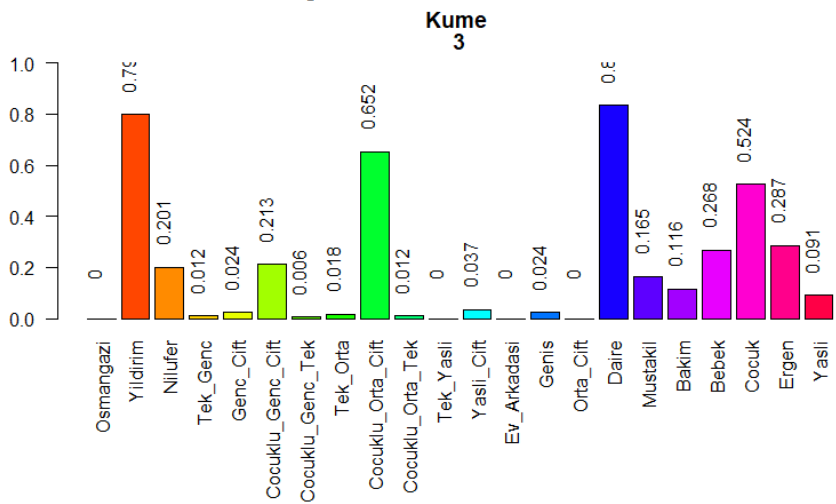
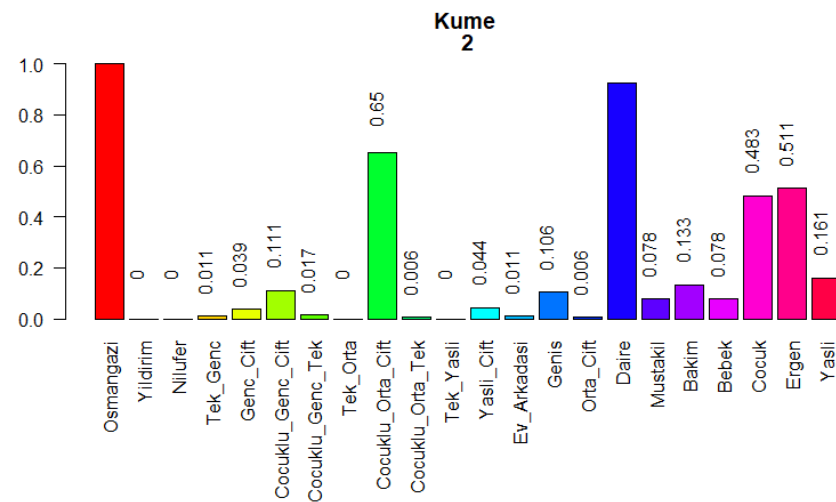
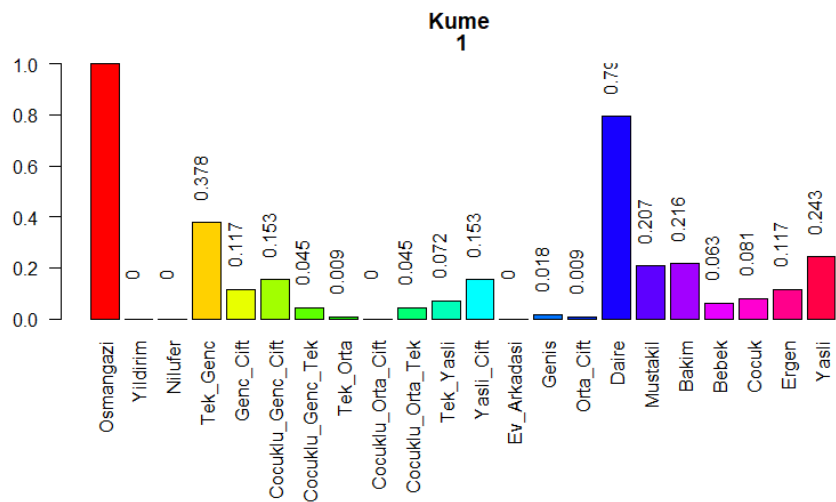
- Evde kalma/yařama süresi ikinci en fazla olan vatandaşlardır: Ev\_kullanım özniteliđine ait kutu-bıyık diyagramı incelendiđinde dađılımin sađa çarpık olduđu, IQR'nin diđer kümelere nispeten dar olduđu ve bu kúmedeki hanelerin %50'sinin evde yařam süresinin 16 – 20 saat arasında deđiřtiđi, yaklaşık %25'inin 12 – 16 saat ve diđer %25'inin 20 – 24 saat evde yařadıđı görölmektedir. Dolayısı ile bu grupta yer alan hanelerde yařayan ev halkının bir kısmının çalıřtıđı ya da okula gittiđi diđer kısmının evde kaldıđı ifade edilebilir.

- Elektrik ve suyu en fazla tüketen vatandaşlardır: Elektrik özniteliđine ait kutu-bıyık diyagramı incelendiđinde, bu kümenin en fazla IQR'ye sahip olduđu dađılımin sađa çarpık olduđu, aykırı gözlem içermediđi fakat diđer kümelere kıyas ile üçüncü kartil ile

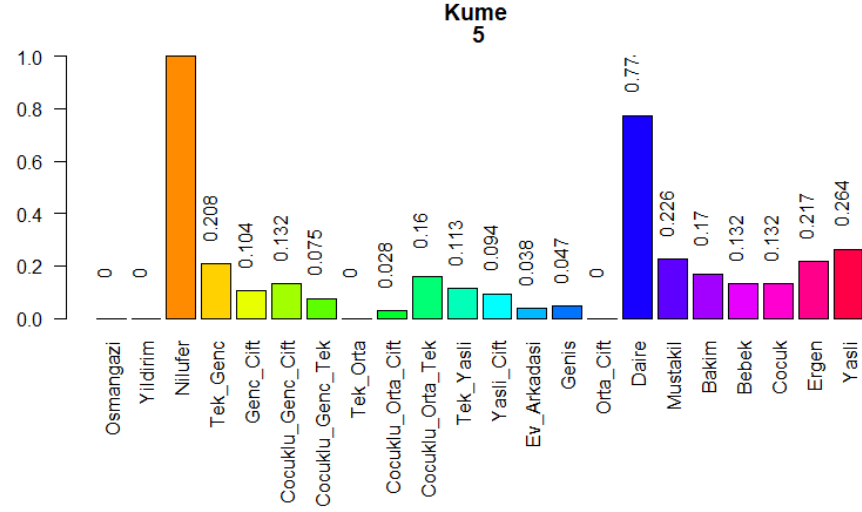
sağ (üst) bıyık arası mesafenin 300 – 600 TL gibi geniş bir aralığa sahip olduğu görülmektedir. Su özniteliğine ait kutu bıyık diyagramı incelendiğinde ise hanelerin yaklaşık %70'inin su tüketimlerinin çoğunlukla 120 – 200 TL aralığında olduğu dikkati çekmektedir.

- Doğalgaz tüketimleri orta seviyede olan vatandaşlardır: Gaz özniteliğine ait kutu-bıyık diyagramı incelendiğinde dağılımın sağa çarpık olduğu, gaz tüketiminin %50'sine yakın kısmının 200-300 TL olduğu ve %25'ine yakın kısmını 120 – 200 TL aralığında diğer %25'inin 300 – 450 TL arasında olduğu ve ayrıca çok az sayıda hanenin doğal gaz tüketiminin 450 TL'nin üzerinde olduğu görülmektedir.

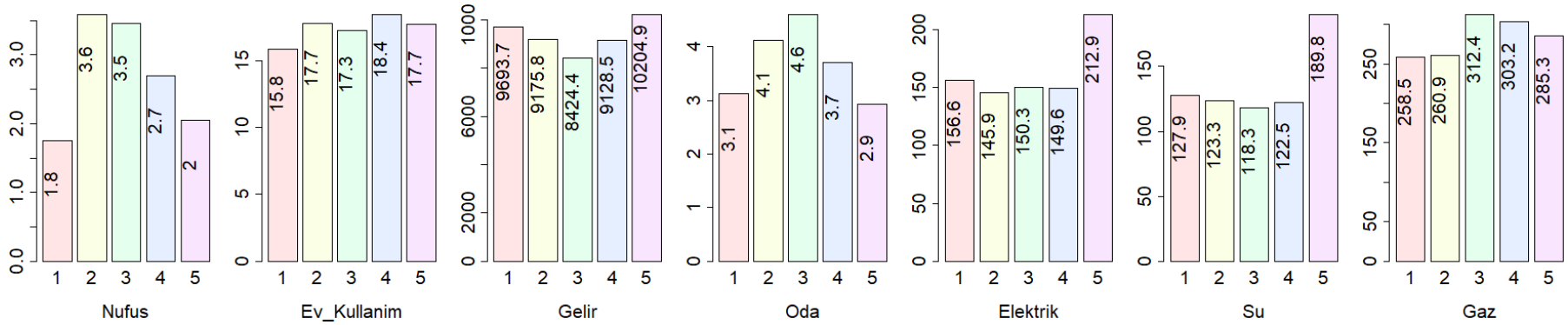
Şekil (90, 91) incelendiğinde bu küme, en küçük ve birbirine eşit egoistik, biyosferik ve altruistik tutum puanına sahip olan kümedir. Bu durum ile paralel olarak, bu kümenin özbenlik boyutundan aldığı puanların ortalaması en küçük ve doğayla ilişki bağlamında olumsuz ifadelerin bulunduğu perspektif boyutundan aldığı puanların ortalaması ise en büyüktür. Ayrıca, geri dönüşüm, ulaştırma tercihi, elektrik ve su tüketimi boyutlarına ait özniteliklerin puanlar, diğer kümelere kıyas ile en küçük ve gıda tüketimi için ise en büyüktür.



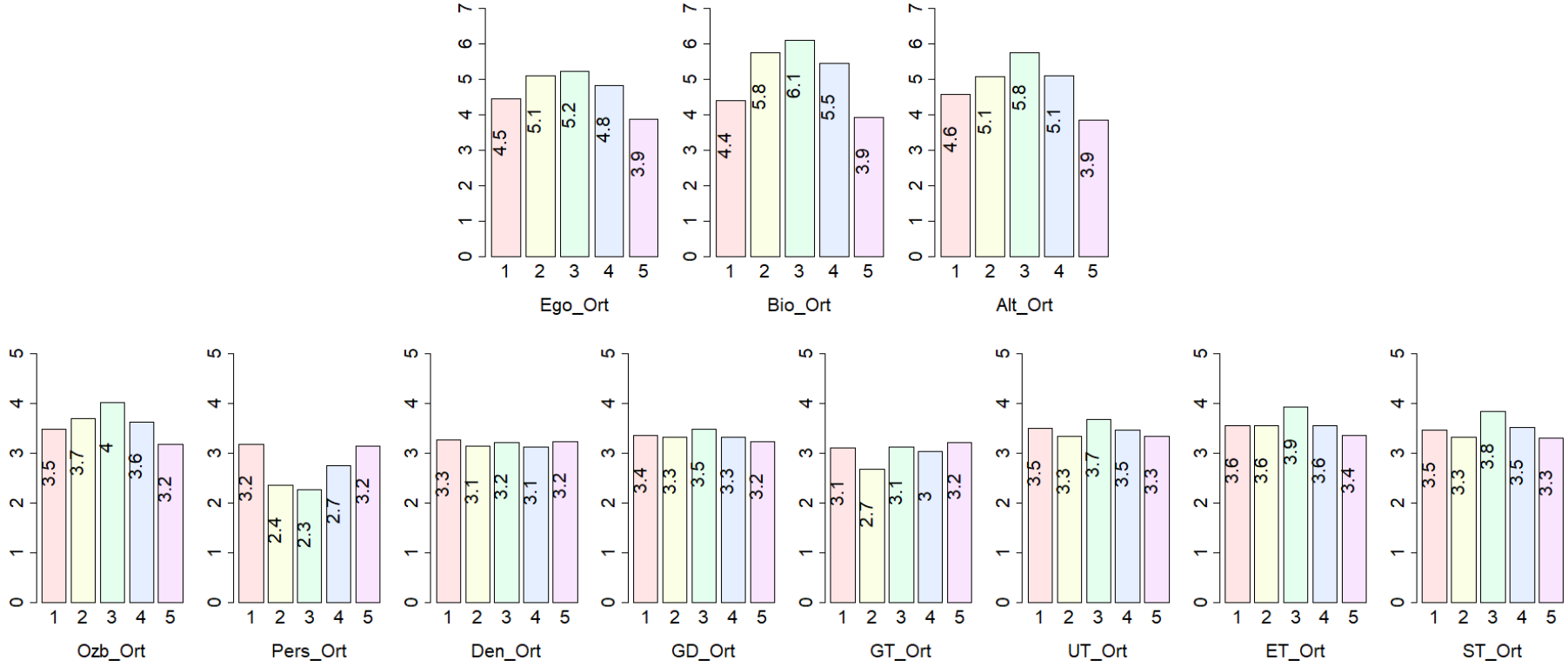




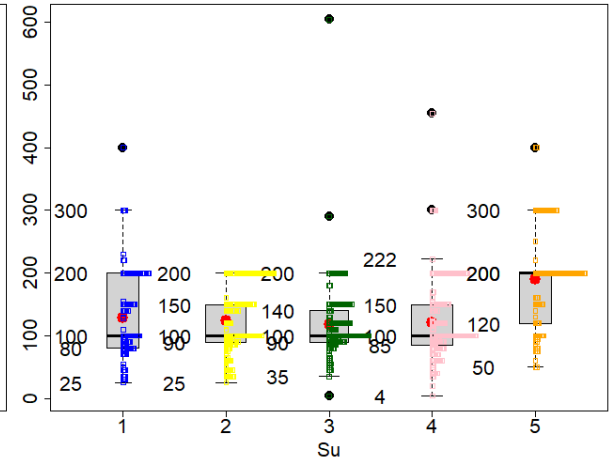
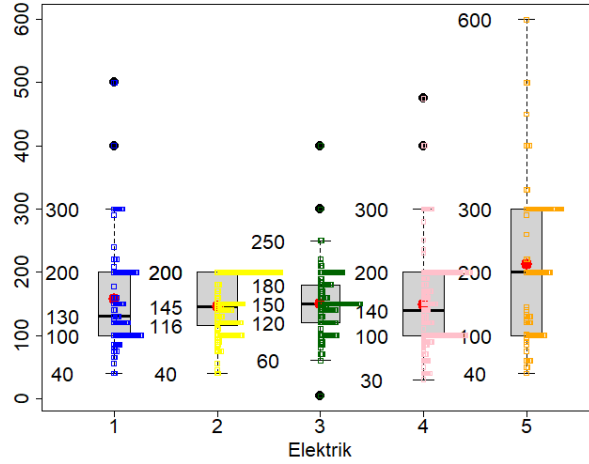
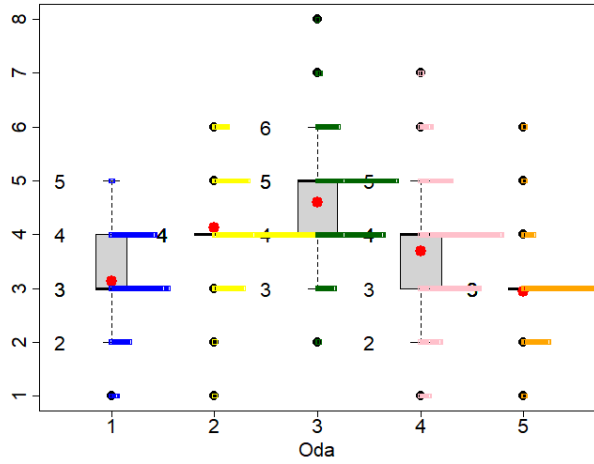
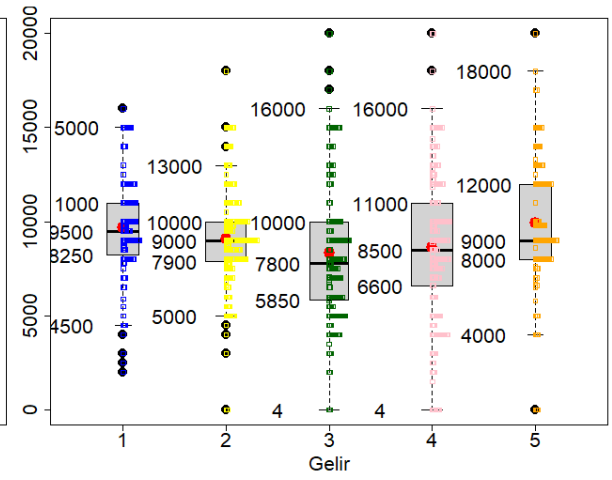
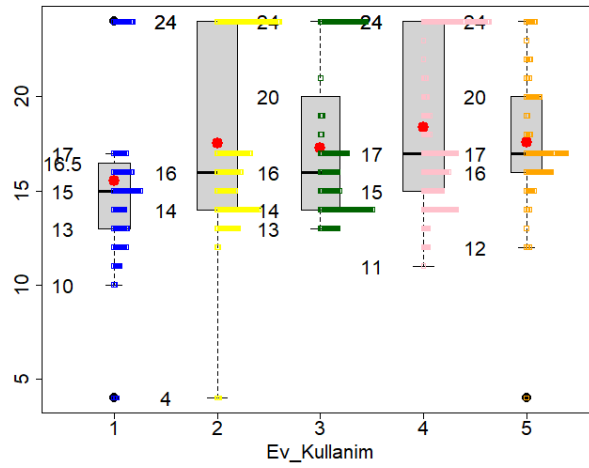
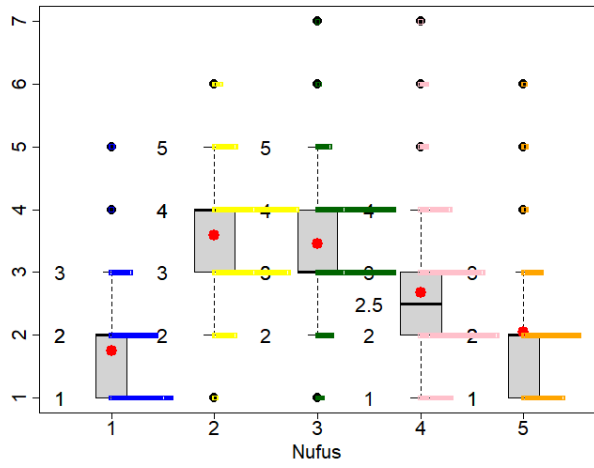
Şekil 88. İkili ve kategorik özneliklerin küme esaslı sütun grafikleri

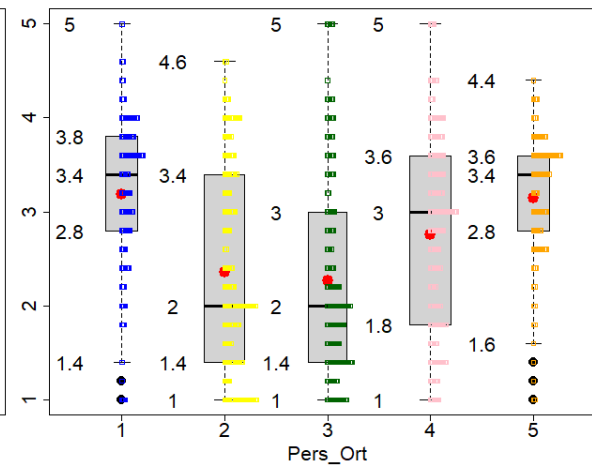
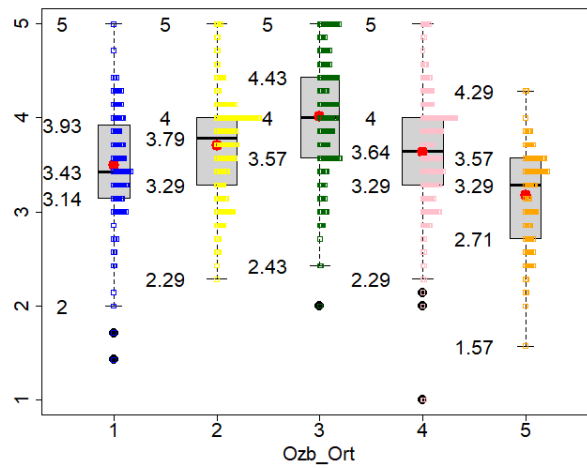
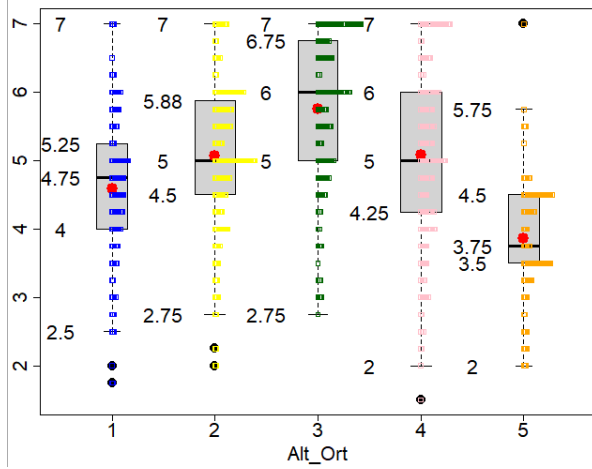
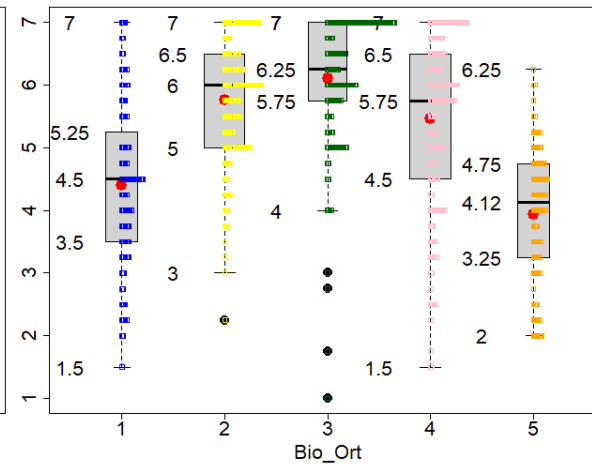
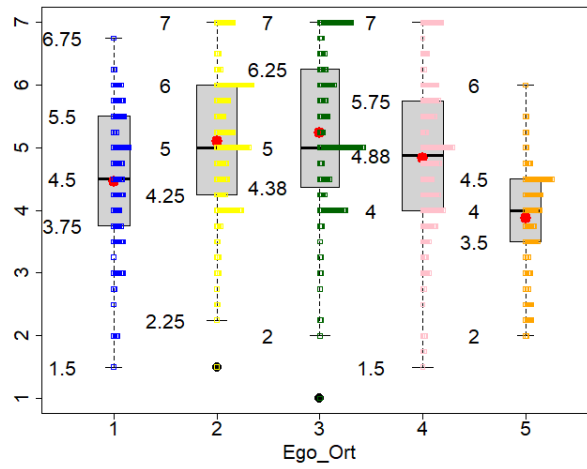
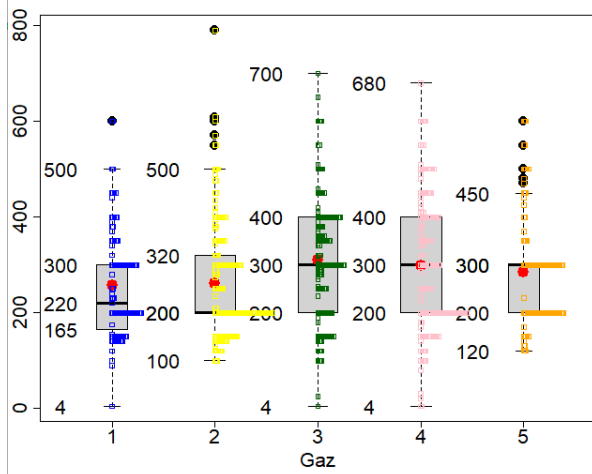


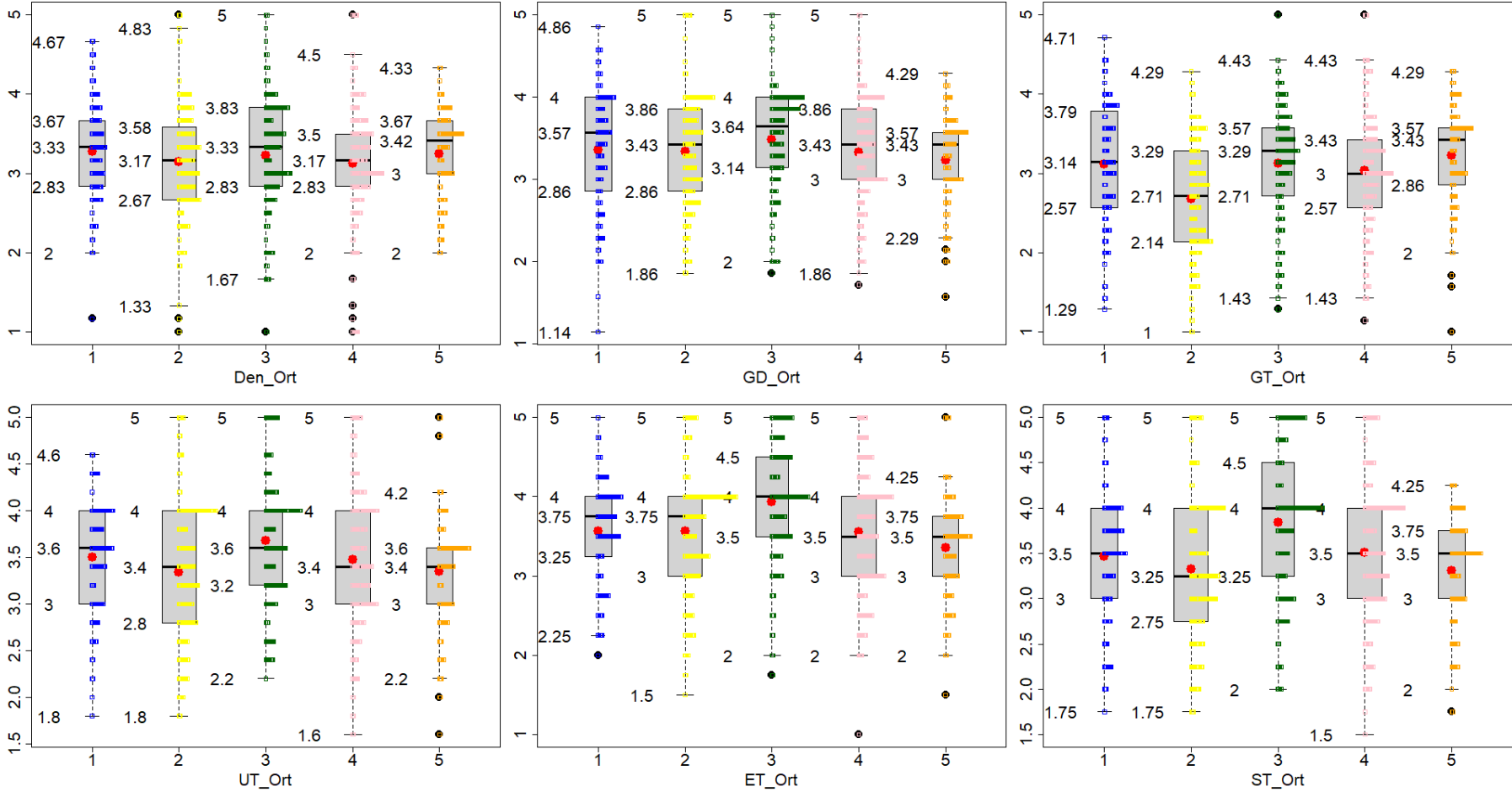
Şekil 89. Demografik-Nümerik özneliklerin küme esaslı sütun grafikleri



Şekil 90. Ölçek-Nümerik özniteliklerin küme esaslı sütun grafikleri







**Şekil 91.** Tüm numerik özneliklerin küme esaslı kutu-bıyık grafikleri (Gelir değişkeni dördüncü küme 40,000 TL değerli bir tane gözleme sahiptir. Küme dağılımlarının daha sağlıklı karşılaştırılabilmesi için bu gözlem sadece bu aşamada veri setinden çıkartılmıştır.)

#### 5.4.2. Karar Ağaçları ile Tanımlama

Betimsel istatistikler, kümelerin geniş bir bakış açısı ile incelenmesini sağlamaktadır. Fakat hangi özniteliklerin kümeleri birbirinden ayırt etmede en yüksek önem derecesine sahip olduğu ya da bir gözlemin hangi genel koşullar altında ve hangi olasılık ile hangi kümede yer alabileceği bilgisi betimsel istatistikler ile elde edilememektedir. Karar ağaçları, bu ve benzeri bilgileri elde etmede başarımı yüksek ve kullanımı ve yorumlaması kolay olan bir sınıflandırma algoritmasıdır. Bu tez çalışmasında, vatandaşları homojen bir şekilde gruplayan öznitelikleri ve her bir vatandaş grubu için öznitelik değerlerini tespit etmek amacı ile karar ağaçları kullanılmıştır. Bu alt aşamadaki analizler de R-Studio yazılımı ile gerçekleştirilmiştir.

Karar ağaçları, veri biliminde sıklıkla kullanılan klasik yöntemlerden olup, bu alanda günümüze kadar çeşitli algoritmalar geliştirilmiştir. Kimlik tanımlama aşamasındaki esas amaç küme analizi sonucunda elde edilen kümeleri anlamlandırmak veya kümeleri, küme bilgilerini kullanarak sınıflandırmak/etiketlemektir.

Bu nedenle bu tez çalışmasında, CART algoritması kullanılmıştır. Veri setini en iyi şekilde (homojen olarak) bölen özneliğin seçilmesinde Gini endeksi dikkate alınmıştır. Karar ağacının, veri setini ezberlemesini engellemek için ise ağacın budanması ve dallanması ile ilgili çeşitli parametreler kullanılmıştır:

- Bir ara (çocuk) düğümün dallanabilmesi için, o düğümde olması gerekli minimum gözlem adedi (minsplit) 20 gözlem,
- Bir düğümün yaprak (terminal) düğüm olabilmesi için, o düğümde olması gerekli olan gözlem sayısı minsplit'in yaklaşık 1/3'ü olacak şekilde 7 gözlem,
- Karmaşıklık (cp) 0.01,
- Çapraz doğrulama için kullanılacak olan alt küme sayısı 10 adet (n-katmanlı çapraz doğrulama) ve
- Ağacın maksimum derinliği 30 olarak belirlenmiştir.

Ayrıca mevcut veri seti ve parametreler dikkate alınarak CART algoritması ile eğitim gerçekleştirildiğinde eğitimin performansını yorumlayabilmek adına mevcut veri setinden %80 – %20 oranlarında eğitim ve test seti olmak üzere 1000 defa örnekleme yapılmıştır. Başka bir ifade ile 1000 farklı ağaç oluşturulmuş, eğitilmiş, yukarıda sayılan

parametreler dikkate alınarak budanmış ve budanmış ağaçların ortalama performansı hesaplanmıştır. Elde edilen 1000 budanmış ağaçtan performansı maksimum olan ağaç ile kural çıkarımı gerçekleştirilmiştir. Elde edilen sonuçlar aşağıdaki gibidir:<sup>175</sup>

- 1000 budanmış ağacın ortalama performansı %89.26 olarak tespit edilmiştir.
- 1000 budanmış ağaçtan maksimum performansa sahip 629 numaralı ağacın performansı %98.48 olup, hata matrisi aşağıdaki gibidir.

**Tablo 53.** Kural çıkarımında kullanılan karar ağacının hata matrisi

Tahmin \ Küme	1	2	3	4	5
1	17	0	0	0	0
2	0	32	0	0	0
3	0	0	33	0	0
4	0	0	1	29	1
5	0	0	0	0	19

- Diğer eğitim parametreleri sabit kalmak üzere, 629 numaralı ağaç  $cp = 0$  koşulu altında tekrar eğitilerek budanmamış haline, tam ağaca “fully grown tree”, ait karmaşıklık tablosu elde edilmiştir, Tablo 54.<sup>176</sup> Budamada hem  $cp = 0.01$  hem de bir standart hata kuralı uygulanmış ve sonuçları karşılaştırılmıştır.

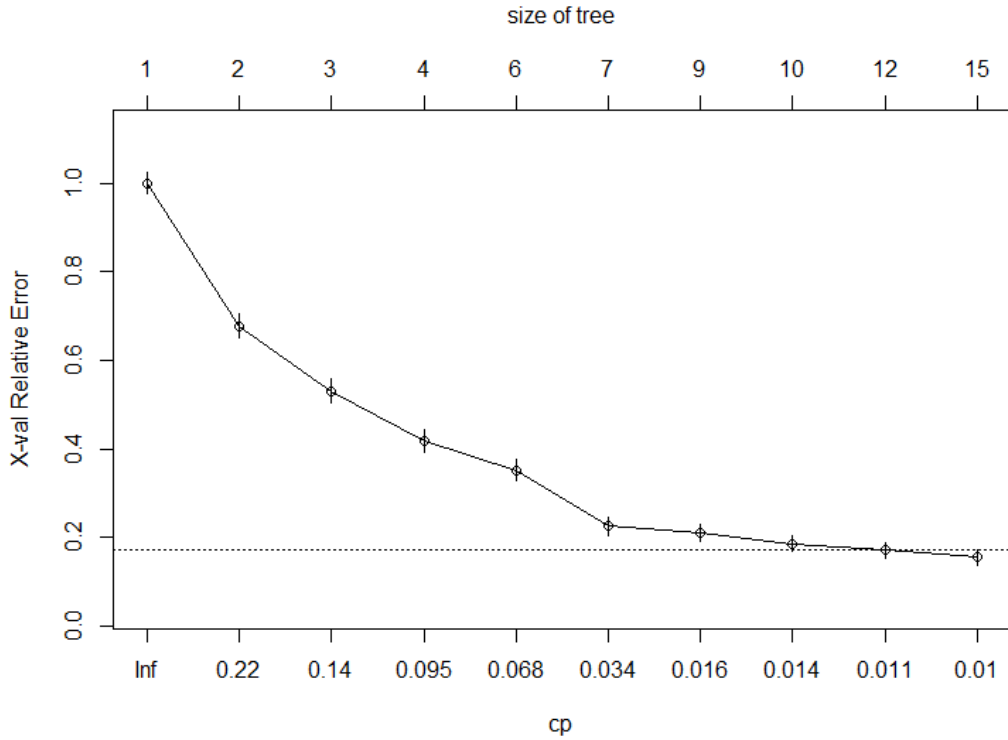
İlk yaklaşıma göre ağaç,  $cp = 0.01$  için 14 dallanmada budanmalıdır. İkinci yaklaşıma göre ise küçük bir hesaplama yapılmalıdır. Tam ağaç için hesaplanan en küçük çapraz doğrulama hatası 18 dallanma ile 0.1244 olarak elde edilmiş ve 18 dallanma için standart hata ise 0.0157’dir. En küçük çapraz doğrulamadan artı bir standart hata üst sınırına tekabül eden değer  $0.1244 + 0.0157 = 0.1401$ ’dir. Dolayısı ile çapraz doğrulama hatası  $< 0.1401$  koşulunu sağlayan en küçük dallanma sayısı

<sup>175</sup> Harita 227’ye rakip olan Harita 132 de ayrıca karar ağacı analizine tabi tutulmuş fakat detaylı raporlanmayı, bu kısımda kısaca özetlenmiştir.  $k = 6$  için ward bağlantılı hiyerarşik küme analizi sonucunda elde edilen küme bölümlenmeleri kullanılarak CART algoritması çalıştırıldığında 1000 budanmış ağacın ortalama performansı %62.44 olarak tespit edilmiştir.  $cp = 0.01$  olmak üzere budanmış ağaçlardan elde edilen maksimum performans %74.41’dir. Maksimum performansı sağlayan 120 numaralı ağacın yaprak sayısı 18’dir. 1-SE kuralı uygulanarak budama yapıldığında ise  $cp = 0.014$ ’e yükselmekte, ağacın performansı %71.31’e düşmekte, yaprak sayısı 9’a inmekte ve dolayısı ile yaprakların safsızlığı artmaktadır. Ayrıca elde edilen kompleksite grafiğinde  $cp = 0.018$  tam 1-SE sınırında kalmıştır.  $cp = 0.018$  için budama yapılırsa eğer, elde edilecek olan ağacın performansı  $cp = 0.014$ ’ten daha kötü olacağı aşikardır.

<sup>176</sup> Tam ağacın görseli Ek 13’te verilmiştir.

referans alınarak 629 numaralı ağaç budanmalıdır. Bu koşulu sağlayan en küçük dallanma sayısı 14 olup, ilk yaklaşımın sağladığı sonucun aynısıdır.

Her iki sonuç, aynı zamanda Şekil 92.'de yer alan karmaşıklık grafiğinden de incelenebilir. Alt x-ekseni karmaşıklık parametresi, üst x-ekseni yaprak sayısı, y-ekseni çapraz doğrulama hatası ve noktalı doğru bir standart hata kuralının sınır noktasını temsil etmektedir. Dolayısı ile eğrinin, doğru altında kalan ilk noktası, budama noktasını temsil etmekte ve bu nokta  $cp = 0.01$  olup, 15 yaprağa denk gelmektedir.



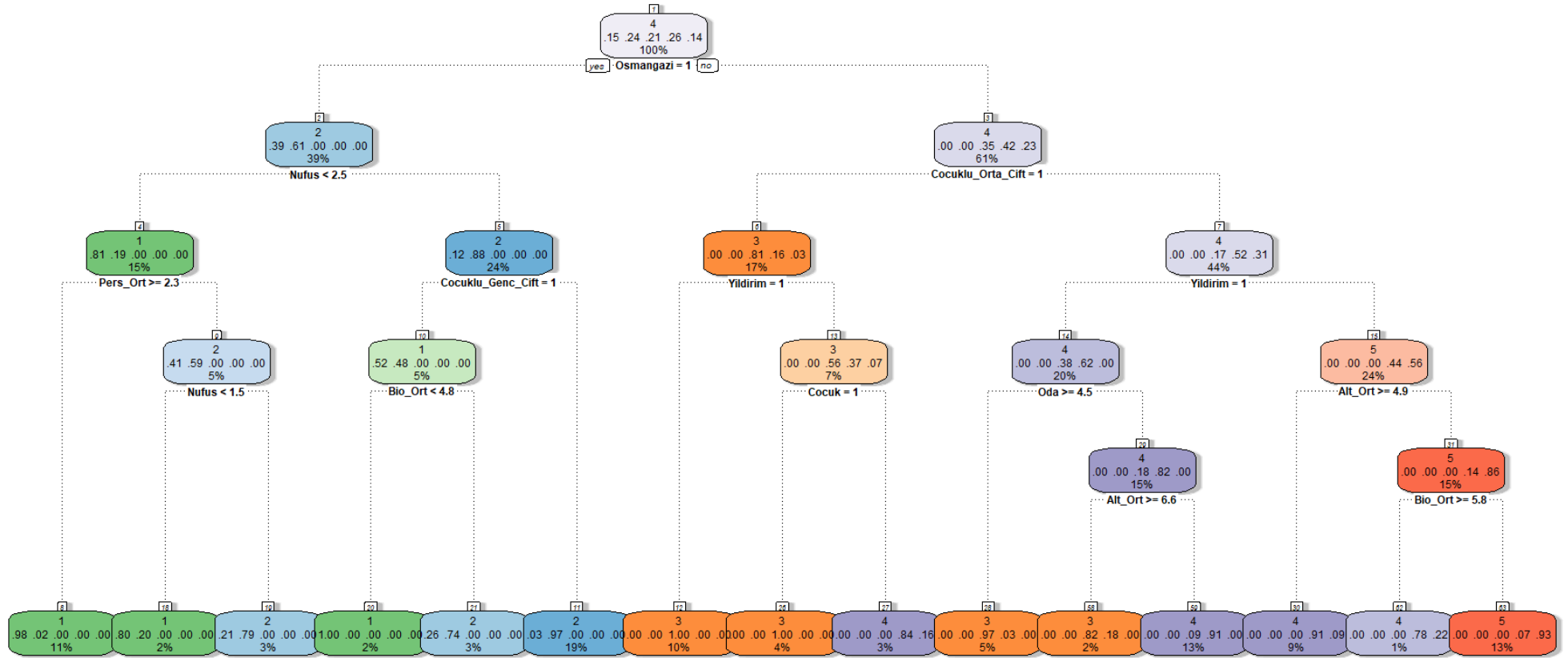
Şekil 92. Karar ağacı karmaşıklık grafiği

Tablo 54. Kural çıkarımda kullanılan 629 numaralı ağacın çapraz doğrulama sonuçları

Karmaşıklık	Dallanma sayısı	Yerine koyma hatası	Çapraz doğrulama hatası	Çapraz doğrulamanın standart hatası
0.3231	0	1	1	0.0236
0.1462	1	0.6768	0.6768	0.0270
0.1266	2	0.5305	0.5305	0.0264
0.0709	3	0.4039	0.4170	0.0250
0.0655	5	0.2620	0.3515	0.0238
0.0174	6	0.1965	0.2248	0.0202
0.0152	8	0.1615	0.2096	0.0196
0.0120	9	0.1462	0.1855	0.0186
0.0109	11	0.1222	0.1703	0.0180
0.0100	14	0.0895	<b>0.1310</b>	0.0160
0	18	0.0851	0.1244	0.0157



Şekil 93.'te yer alan 629 numaralı karar ağacı sonuçları dikkate alındığında, karar ağacının oluşturulmasında kullanılan bölümlene kriterlerinin (özniteliklerin) önem dereceleri Tablo 55.'te özetlenmiştir.



Şekil 93. Kural çıkarmada kullanılan budanmış karar ağacı

**Tablo 55.** Karar ağacı bölümlenme kriterlerinin önem dereceleri

Osmangazi	128.8658	Tek_Genc	23.9627
Cocuklu_Orta_Cift	74.5137	ST_Ort	17.0066
Nufus	72.4068	Bebek	14.8850
Oda	71.2661	Ergen	14.4274
Yildirim	67.8536	Cocuklu_Genc_Cift	11.2901
Alt_Ort	61.7520	Su	5.3250
Bio_Ort	50.7009	Den_Ort	4.0573
Ozb_Ort	46.4995	Gaz	3.7796
Ego_Ort	34.6264	Gelir	3.1921
Cocuk	34.62086	ET_Ort	2.1343
Pers_Ort	32.7811	GD_Ort	1.8774
Nilufer	31.9502	Tek_Orta	1.8774
Ev_Kullanım	30.9358	Elektrik	1.5619
GT_Ort	25.3897		

Şekil 93.'te yer alan karar ağacı incelendiğinde, beş küme için aşağıda yer alan kural tabanı elde edilmiştir:

- **Birinci küme**, Osmangazi ilçesinde yaşayan,
  - a. Yaprak 8:** Tek veya iki kişilik hanelerden Pers\_Ort boyutu puanı 2.3 ve üzeri olan haneler %98 olasılık ile birinci kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %11'i kadardır.
  - b. Yaprak 18:** Tek kişilik hanelerden Pers\_Ort boyutu puanı 2.3'ten küçük olan haneler %80 olasılık ile birinci kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %2'si kadardır.
  - c. Yaprak 20:** Hane halkı büyüklüğü üç ve üzeri olan çocuklu genç çiftlerden Bio\_Ort boyutu puanı 4.8'den küçük olan haneler kesinlikle birinci kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %2'si kadardır.
- **İkinci küme**, Osmangazi ilçesinde yaşayan,
  - a. Yaprak 19:** İki kişilik hanelerden Pers\_Ort boyutu puanı 2.3'ten küçük olan haneler %79 olasılık ile ikinci kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %3'ü kadardır.
  - b. Yaprak 21:** Hane halkı büyüklüğü üç ve üzeri olan çocuklu genç çiftlerden Bio\_Ort boyutu puanı 4.8'den ve üzeri olan haneler %74 olasılıkla birinci kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %3'ü kadardır.

- c. Yaprak 11:** Hane halkı büyüklüğü üç ve üzeri olan ve çocuklu genç çiftler dışında kalan haneler %97 olasılıkla ikinci kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %19'u kadardır.
- **Üçüncü küme,**
    - a. Yaprak 12:** Çocuklu ve orta yaşlı çiftlerden Yıldırım ilçesinde yaşayanlar kesinlikle üçüncü kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %10'u kadardır.
    - b. Yaprak 26:** Çocuklu ve orta yaşlı çiftlerden Nilüfer ilçesinde yaşayanların 5 – 15 yaş arası çocuk sahibi olanlar kesinlikle üçüncü kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %4'ü kadardır.
    - c. Yaprak 28:** Çocuklu ve orta yaşlı çiftler dışında kalan hanelerden Yıldırım ilçesinde en az beş odalı meskenlerde yaşayanlar %97 olasılıkla üçüncü kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %5'i kadardır.
    - d. Yaprak 58:** Çocuklu ve orta yaşlı çiftler dışında kalan hanelerden Yıldırım ilçesinde en fazla dört odalı meskende yaşayan ve Alt\_Ort boyutundan 6.6 ve üzeri puana sahip olanlar %82 olasılık ile üçüncü kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %2'si kadardır.
  - **Dördüncü küme,**
    - a. Yaprak 27:** Çocuklu ve orta yaşlı çiftlerden Nilüfer ilçesinde yaşayanların 0 – 4 yaş arasında bebeği veya 16 – 19 yaş arasında ergen çocuğu veya 19 yaş üstü yetişkin çocuğu olanlar %84 olasılık ile dördüncü kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %3'ü kadardır.
    - b. Yaprak 59:** Çocuklu ve orta yaşlı çiftler dışında kalan hanelerden Yıldırım ilçesinde en fazla dört odalı meskende yaşayan ve Alt\_Ort boyutundan 6.6'dan az puana sahip olanlar %91 olasılık ile dördüncü kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %13'ü kadardır.

- c. **Yaprak 30:** Çocuklu ve orta yaşlı çiftler dışında kalan hanelerden Nilüferde yaşayan ve Alt\_Ort boyutundan 4.9 ve üzeri puana sahip olan haneler %91 olasılık ile dördüncü kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %9'u kadardır.
  - d. **Yaprak 62:** Çocuklu ve orta yaşlı çiftler dışında kalan hanelerden Nilüferde yaşayan, Alt\_Ort boyutundan 4.9'un altında ve Bio\_Ort boyutundan 5.8 ve üzerinde puan alanlar %78 olasılık ile dördüncü kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %1'i kadardır.
- **Beşinci küme,**
    - a. **Yaprak 63:** Çocuklu ve orta yaşlı çiftler dışında kalan hanelerden Nilüferde yaşayan, Alt\_Ort boyutundan 4.9'un ve Bio\_Ort boyutundan 5.8'in altında puan alanlar %93 olasılık ile beşinci kümedendir. Alt küme büyüklüğü 616 hanelik eğitim setinin %13'ü kadardır.

## 6. GENEL DEĞERLENDİRME

Harita 227.'nin TBA grafiği<sup>177</sup> ve Tablo 55. dikkate alınarak oluşturulan önem grafiği, betimsel istatistikler ile tanımlama kapsamında açıklanan ekolojik kimlik profilleri ile birlikte değerlendirildiğinde Şekil 94,

- Grafikte yer alan vektör uzunlukları (yük değerleri) ile tabloda yer alan önem dereceleri karşılaştırıldığında, varyansı açıklama gücü yüksek olan özneliklerin büyük bir kısmının, karar ağacı analizi ile elde edilen önem derecelerinin de yüksek olduğu görülmektedir.
- Diğer taraftan Nilüfer, GT\_Ort ve Ev\_Kullanım'ın yük değeri büyük olmasına rağmen karar ağacı önem önem dereceleri düşüktür. Bunun nedeni, bu özneliklere ait vektörlerin bileşen düzlemleri ile yaptığı açıdan dolayı tam olarak ayırt edici niteliğe sahip olmamalarıdır.
- Osmangazi özneliği en büyük yük değerine sahip özneliktir.

---

<sup>177</sup> Karar ağacında kullanılan 227 numaralı haritanın TBA analizi sonucunda özneliklerin birinci ve ikinci temel bileşenler üzerindeki yük değerleri Ek 12'de verilmiştir.

- Osmangazi vektörünün, Nilüfer ve Yıldırım vektörleri ile arasındaki açı neredeyse  $180^\circ$  olup negatif korelasyondur. Dolayısı ile Osmangazi ilçesinden bir gözlem ile Nilüfer veya Yıldırım ilçelerinden olan gözlemlerin aynı kümede yer alma ihtimali bulunmamaktadır.

- İkinci bileşendeki faktör puanı pozitif veya birinci bileşendeki faktör puanı (pozitif ya da negatif) yüksek olan bir gözlemin Osmangazi ilçesinden bir gözlem ile aynı kümede yer alma ihtimali çok düşüktür.

- Nilüfer ve Yıldırım vektörleri birinci bileşende ters ve ikinci bileşende neredeyse eş yönlü olduğu için, Yıldırım ve Nilüfer’de oturan haneler bazı ortak özelliklere sahiptir ve aynı kümelerde yer alabilir.

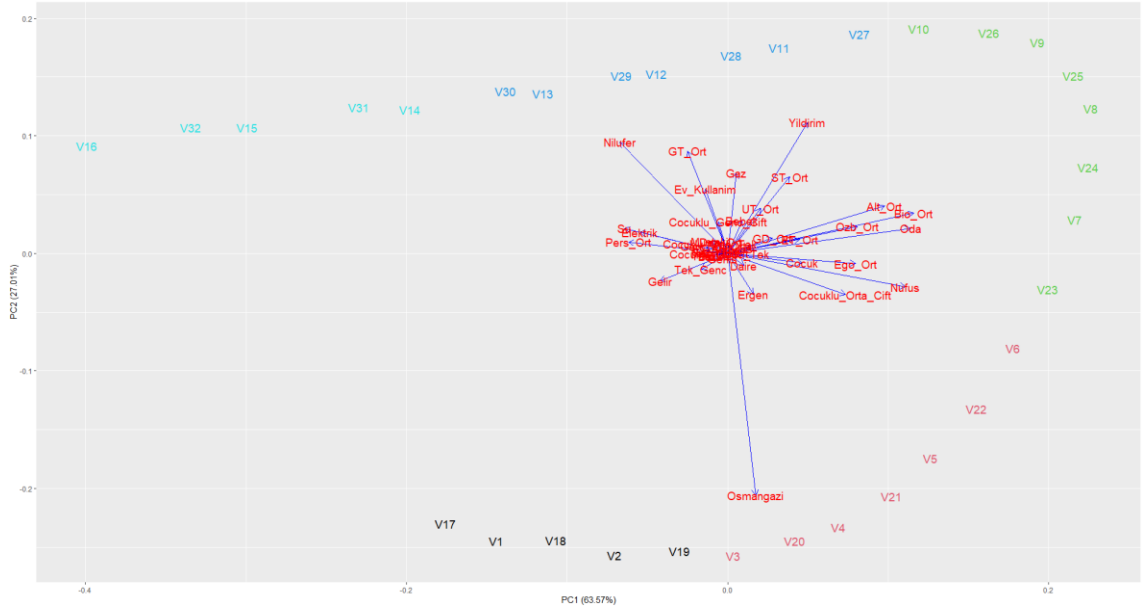
- Osmangazi ile Ergen vektörleri arasındaki açı dar olduğu için bu iki öznitelik birbiri ile pozitif korelasyonludur. Zira sadece Osmangazi ilçesinden hanelerin yer aldığı iki numaralı kümede, ailelerin %51’i ergen çocuğa sahiptir.

- Osmangazi ile Gaz vektörleri arasındaki açı da neredeyse tam  $180^\circ$ ’dir. Zira sadece Osmangazi ilçesinden hanelerin yer aldığı birinci ve ikinci kümeler, diğer kümelere kıyas ile doğalgazı en az tüketen kümelerdir veya Yıldırım ve Nilüfer ilçelerinden hanelerin yer aldığı tüm kümeler doğalgazı fazla tüketen kümelerdir.

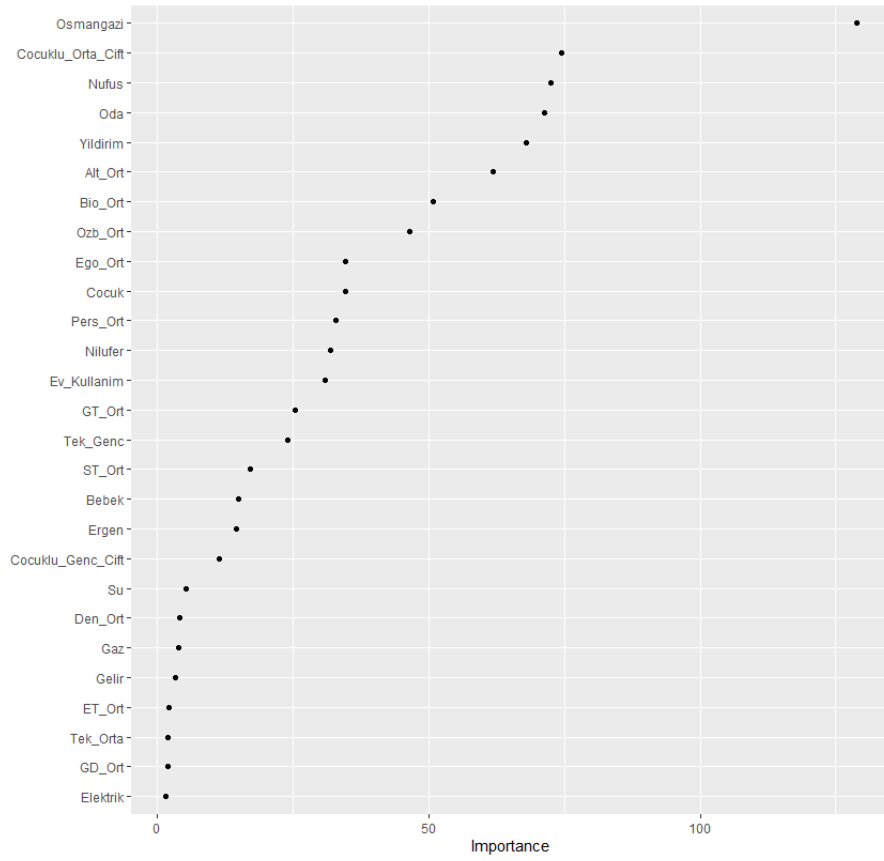
- Gaz vektörü, Nilüfer ve Yıldırım vektörleri ile eş yönlüdür ve Yıldırım vektörüne daha yakındır çünkü Yıldırım’dan hanelerin yoğunluklu olarak yer aldığı üçüncü ve dördüncü kümelerin doğalgaz tüketimleri, Nilüfer’deki hanelerin yer aldığı beşinci kümeden daha fazladır.

- Cocuklu\_Ort\_Cift, Nüfus, Çocuk ve Oda vektörleri arasındaki açılar dardır ve dolayısı ile aralarında pozitif yüksek korelasyon bulunmaktadır.

- Alt\_Ort, Bio\_Ort ve Ozb\_Ort vektörleri arasındaki açılar dardır. Bu sebeple pozitif yüksek korelasyona sahiptirler. Diğer taraftan bu vektörler, olumsuz ifadelerden oluşan Pers\_Ort vektörü ile negatif korelasyonludur çünkü aralarındaki açı nerede ise tam  $180^\circ$ ’dir.



(a)



(b)

Şekil 94. Özniteliklerin (a) TBA faktör yükleri ve (b) karar ağacı önem derecelerinin karşılaştırılması<sup>178</sup>

<sup>178</sup> (a) şıkında yer alan görsel Şekil 62 'deki görselin aynısı olup, bu kısımda paylaşılan karar ağacı analizi sonuçlarının anlaşılabilirliği için burada tekrar paylaşılmıştır.

Karar ağaçları ile elde edilen kural tabanı, betimsel istatistikler ile birlikte değerlendirildiğinde,

- Birinci küme için kural tabanında yer alan en önemli nitelikler betimsel istatistikler ile yapılan tespitler ile uyumlu olarak Osmangazi ilçesinde yaşamak, genellikle 1-2 kişilik hane nüfusuna sahip olmak, daha kalabalık haneler için aile yaşam döngüsünün başında olmak (çocuklu genç çift), düşük biyosferik tutuma ve yüksek perspektife sahip olmaktır.

- İkinci küme için kural tabanında yer alan en önemli nitelikler betimsel istatistikler ile yapılan tespitler ile uyumlu olarak Osmangazi ilçesinde yaşamak, genellikle üç ve üzeri hane nüfusuna sahip olmak, daha kalabalık haneler için aile yaşam döngüsünün ortasında olmak (çocuklu orta yaşlı çift), yüksek biyosferik tutuma ve düşük perspektife sahip olmaktır.

- Üçüncü küme için kural tabanında yer alan en önemli nitelikler betimsel istatistikler ile yapılan tespitler ile uyumlu olarak genellikle Yıldırım ilçesinde yaşamak, genellikle genellikle büyük meskenlerde yaşamak, genellikle aile yaşam döngüsünün ortasında olmak (çocuklu orta yaşlı çift) ve yüksek altruistik tutuma sahip olmaktır.

- Dördüncü küme için kural tabanında yer alan en önemli öznitelikler betimsel istatistikler ile yapılan tespitler ile uyumlu olarak Yıldırım veya Nilüfer ilçelerinde yaşamak, genellikle aile yaşam döngüsünün çocuklu orta yaşlı çift dönemi dışında olmak, orta büyüklükte meskenlerde yaşamak (maksimum dört odalı) ve orta ve üstü altruistik tutuma sahip olmaktır (ortalama 5-6 puan gibi).

- Beşinci küme için kural tabanında yer alan en önemli öznitelikler betimsel istatistikler ile yapılan tespitler ile uyumlu olarak Nilüfer ilçesinde yaşamak, genellikle aile yaşam döngüsünün çocuklu orta yaşlı çift dönemi dışında olmak ve orta altı altruistik ve biyosferik tutuma sahip olmaktır.

Son olarak araştırma modelinin Şekil 49.'da görselleştirilen çıktıları, kimlik tanımlama aşamasının çıktıları ile birlikte değerlendirildiğinde,

- ÇE'nin ekolojik yaşam görüşünü destekleyen biyosferik ve altruistik tutum boyutlarının, neredeyse tüm STD boyutları üzerindeki etkisinde, Dİ'nin özbenlik boyutunun olumlu yönde kısmi aracı etkisi bulunmaktadır. Bu durum, kimlik tanımlama aşaması çıktıları ile uyumaktadır. Nitekim Şekil 62.'de yer alan TBA grafiğinde,



biyosferizm, altruizm ve özbenlik vektörleri arasında çok dar bir açı bulunmakta ve bu durum söz konusu boyutlar arasında yüksek korelasyon olduğu anlamına gelmektedir. Aynı zamanda betimsel istatistikler kısmında yer alan sütun grafikler ve kutu-bıyık diyagramları da incelendiğinde küme analizinde bu boyutları temsil eden özniteliklerden yüksek-orta-düşük puan alan kümelerin (birinci ve ikinci küme hariç) genellikle aynı zamanda tüketimlerinin de düşük-orta-yüksek şeklinde uyduğu gözlemlenmektedir.

- Biyosferizmin, STD'nin elektrik ve gıda tüketimi boyutları üzerindeki etkisinde, Dİ'nin olumsuz ifadelerden oluşan perspektif boyutunun sırası ile olumlu ve olumsuz yönde etkisi bulunmaktadır. Bu durum da kimlik tanımlama aşaması çıktıları ile uyumaktadır. Biyosferik tutum değerinin artışı, olumsuz ifadelerden oluşan perspektif boyutu değerini olumsuz etkilemektedir. Bu etki, beklenen bir sonuçtur. Aynı zamanda, perspektif boyutunun değerindeki artışın elektrik tüketimi boyutunu olumsuz yönde etkilemesi de beklenen bir sonuçtur. Fakat diğer taraftan, biyosferik tutumun gıda tüketimini olumsuz yönde etkilemesi ve olumsuz ifadelerden oluşan perspektif boyutunun gıda tüketimini olumlu yönde etkilemesi beklenen bir sonuç değildir. Söz konusu beklenmeyen durum, ülkemiz özelinde bir durum olarak değerlendirilmiştir. Bu duruma en güzel örnek olan küme ise beş numaralı kümedir. Nitekim bu küme hem ÇE hem de Dİ boyutlarında en kötü puanlara sahip olup ve geliri en yüksek küme olup gıda tüketimi hariç diğer tüm sürdürülebilir tüketim davranışlarında en kötü puana sahip kümedir.

- ÇE'nin egoistik tutum boyutunun, STD'nin geri dönüşüm boyutu üzerindeki etkisinde, Dİ'nin deneyim boyutunun olumlu yönde kısmi aracı etkisi bulunmaktadır. Bu etkileşim kısmi aracılık için alt sınır olan 0.20'ye çok yakındır. Profil çıkarma aşamasında ise elde edilen karar ağacı sonuçları ve betimsel istatistikler ile değerlendirmeler dikkate alındığında her iki boyut da (egoizm ve deneyim) maalesef hiçbir küme için ayırt edici öznitelik olamamıştır. Diğer taraftan doğa ile etkileşim içinde bulunma durumunu temsil eden deneyim boyutunun egoizm boyutu tarafından olumlu etkilenmesi ve aynı zamanda egoizmin, geri dönüşüm boyutunu olumlu yönde etkilemesi beklenmeyen bir bulgudur. Söz konusu bulgunun anlamlandırılmasının ancak yarı-yapılandırılmış görüşmeler ile sağlanabileceği düşünülmektedir.

## SONUÇ

Artan kentleşme ile birlikte insanın doğadan kopuşu, tüketim davranışlarında doğaya duyarsızlaşmaya, enerji tüketiminin artmasına, doğal kaynakların kıtlaşmasına ve doğal dengenin bozulmasına neden olmuştur. Bu durum, kentlerin ve dolayısı ile toplumların geleceğini riske atmıştır. Bu bağlamda geliştirilen sürdürülebilirlik paradigması daha yaşanabilir kentler, verimli ekonomiler, refah toplumlar ve dengenin sağlandığı bir çevre oluşturmayı amaçlamaktadır.

Dünya canlı bir organizma ise, ülkeler, kentler ve sahip olduğu toplumlar onun alt sistemleridir. Bozulan yapının düzelebilmesi ise alt sistemlerin kendi şartları, kaynakları, kısıtları veya sahip olduğu nitelikleri özelinde değerlendirilerek iyileştirilmesine bağlıdır. Bu bütünleşik sistem, sürdürülebilirlik açısından değerlendirildiğinde sürdürülebilirlik çalışmalarının kentler özelinde yerelleştirilmesi, bütün problemlerin çözümü olarak düşünülebilir.

Dünyada çeşitli örgütler yerel yönetimlere yol gösterici nitelikte sürdürülebilir kent modelleri/sistemleri tasarlamıştır. Bunlardan bir tanesi AYB olup, kentimiz Bursa bir süredir bu sisteme adapte olmaya çalışmaktadır. Bursa'nın bir AYB kenti olması ile hem "Yeşil Bursa" markası tescillenmiş olacak hem de yerel halk ve iş çevresi çeşitli sosyal, çevresel ve ekonomik kazanımlar elde edecektir. Fakat Bursa her AYB başvurusunda maalesef ön değerlendirmeden geçememektedir.

Sürdürülebilirlik yolundaki bir kentin tüm bileşenleri ile birlikte yeşil olması gerekmektedir. Örneğin; hane ve binek taşıtlar kaynaklı tüketimin amacının insan konforu olan ve iktisadi bir getirisi olmayan yaşam faaliyetlerinden kaynaklandığı düşünülürse, Bursa'nın enerji tüketiminin yarısına denk gelen 17 milyon MWh'lik tüketimi toplumsal ya da ekonomik bir fayda sağlamamaktadır. Bu durum, başka bir açıdan ele alınır ise bu tüketimin neden olduğu sera gazı salınımı da kentin toplam sera gazı salınımının nerede ise yarısına denk gelmektedir. Dolayısı ile Bursa'nın yeşil olması öncelikle halkının yeşil olmasına bağlıdır. Değişim, vatandaştan başlamalıdır. Değişen tüketim alışkanlıklarından ötürü enerji talebinin yıldan yıla artmaktadır. Fakat bunu değiştirmek veya enerji talebini yönetmek için, fiziki yatırımlar dışında, herhangi bir faaliyet planlandığı dikkati çekmemektedir.

Diğer taraftan gerçekleştirilmesi planlanan faaliyetler incelendiğinde hiçbirisinin insan ile ilgili olmadığı, daha çok kentin alt ve üst yapı sistemleri ile ilgili olduğu görülmektedir. Söz konusu yatırımlar yapay sistemin yeşil ve sürdürülebilir olması için kesinlikle gereklidir. Fakat bakış açısı, sadece fiziki iyileştirme/dönüşüm olmamalıdır. Katı atık dağılımları incelendiğinde, Bursa atıklarının %52'sinin evsel organik atık olduğu dikkati çekmektedir ve katı atık işleme ya da dönüştürme tesislerine yapılacak yatırımlar kesinlikle çok önemlidir. Fakat bu yatırımlar kadar, evsel organik atıkları oluşturan gıda atıklarının azaltılmasına yönelik politikalar geliştirmek ve bu doğrultuda yatırımlar yapmak da önemli olmalıdır. Aslında insana yapılacak olan yatırım öncelikli olmalıdır ki atık oranı ve atık dönüşümü ve işlenmesine gerekli olacak yatırımlar için ayrılacak olan kent bütçesi de azalabilsin.

Katı atıklardaki duruma benzer bir durum ise su yönetiminde görülmektedir. Kişi başı günlük ortalama su tüketimi 2016'da 133 litre olarak belirtilmiş ve su kaynaklarının kıtlığından ve dolayısı ile talebi karşılamak için gerçekleştirilmesi planlanan yatırımlar dikkat çekiyor olsa da bu kapsamda sadece tek bir faaliyetin gerçekleştirildiği ve bunun ilk ve orta öğrenim okullarına verilen eğitimler olduğu görülmektedir. Örneğin bu kapsamda 2013-2017 yılları arasında toplam 153 eğitim gerçekleştirilmiştir. Genç nesile yapılan bilinçlendirme faaliyetleri kesinlikle önemlidir. Bu faaliyetler aslında geleceğe yatırım olarak değerlendirilebilir. Fakat faaliyetlerin etkililiği ya da ailelere erişim hakkında bir bilgi bulunamamıştır.

Dikkati çeken son husus ise Bursa'nın ülkedeki diğer kentler ile karşılaştırıldığında verimli topraklara sahip olduğu ve tarımda iyi bir noktada olduğunun ifade edilmesidir. Fakat örneğin kar amacı gütmeyen bir insiyatif olan Bursa Gıda Topluluğunun maalesef Bursa il sınırları içerisinde ilaçlanmış ya da ticari tohum kullanmayan, atalık tohum kullanan ya da tohumunu kendisi üreten, tarımsal ilaç kullanmayan, ekim alanı sanayiye yakın olmayan, yeşil gübre, hayvan gübresi ya da kompost kullanan bir tane bile onaylı tedarikçisi yoktur. Şarküteri ürünlerinden taze ve kurutulmuş gıdaya ve hatta sürdürülebilir kozmetik ürünlerine kadar tüm ürünler topluluk tarafından Türkiye'nin dört bir yanındaki üreticilerden temin edilmektedir. Kısacası Bursa bir sanayi kenti olduğu kadar bir tarım kentidir fakat sürdürülebilir tarım uygulamaları maalesef yoktur ve BBB'nin de bu perspektifte bir aksiyon planı maalesef bulunmamaktadır.

Özetle, yerel yönetimin sunacağı fiziki yatırımlar son teknoloji olsa bile bu yatırımları veya yatırımların çıktılarını kullanacak olan yine Bursa halkı ve iş çevresidir. Halkın bakış açısı ve dolayısı ile yaşam tarzı değişmedikçe toplu taşıma kullanım oranı düşük kalır, bireysel araçlanma ve araçların kullanımı, hanehalkı genel tüketimi, enerji ve su tüketimi ve atık üretimi, çevre ve gürültü kirliliği artmaya ve kaynaklar kıtlaşmaya devam eder. Bursa'nın bir AYB kenti olabilmesi ve “yeşil” kent markasını kazanabilmesi için kesinlikle ve öncelikle kentin fiziki çevresinden öte, halkının dönüşüme uğrayarak yeşil olması gereklidir. Halkın bakış açısının ve davranışlarının değişimi ise onların duyuşsal, bilişsel, deneyimsel ve davranışsal açıdan analiz edilmesi ve anlaşılmasını gerektirir. Vatandaşlar ancak bu şekilde tanımlandıktan sonra yerel yönetimin tasarlayacağı çeşitli politikalar ve faaliyetler ile vatandaşların dönüşüm süreci başlar ve yeşil kent markası kazanılabilir.

Bu tezde, bir AYB kenti adayı olan Bursa özelinde kent halkının sürdürülebilir tüketim davranışlarını, değerlere bağlı olarak çevresel endişelerini (tutumları) ve doğa ile ilişkilerini dikkate alarak ekolojik kimlik profillerini keşfetme hedeflenmiştir. Bu amaç ile alan yazınında çevresel endişe bağlamında tutumu, insan-doğa ilişkisini ve sürdürülebilir tüketim davranışlarını ölçtüğü belirtilen ve Türkçeye uyarlanmış olan üç ölçme aracı kullanılarak Osmangazi, Yıldırım ve Nilüfer merkez mahallelerinde yaşayan vatandaşlar ile yüzyüze görüşme yöntemi ile veri toplanmıştır. Verilerin analizi iki aşamada gerçekleştirilmiştir. İlk aşamada, Bursa örnekleme için doğa ile ilişkinin tutum ve davranış arasındaki aracı etkisi KEKK-YEM ile keşfedilmiştir. İkinci aşamada ise vatandaşların ekolojik kimlik profilleri çeşitli kümeleme algoritmaları (SÖÖH, ward bağlantılı hiyerarşik ve k-temsili) ile elde edilen küme bölümlenmelerine uygulanan karar ağaçları ve bazı betimsel istatistikler yardımı ile tanımlanmıştır.

Şekil 49'da özetlenmiş olan ilk analiz aşamasını değerlendirmek gerekirse bu tez çalışmasında Nisbet, Zelenski, Murphy (2009) çalışmasında geliştirilmiş olan DİÖ'nün kullanılma nedeni, söz konusu ölçeğin insan-doğa ilişkisini duyuşsal-bilişsel-deneyimsel olarak kapsayıcı bir biçimde ele almasıdır. Çünkü alan yazınında geliştirilmiş olan diğer ölçeklerin hepsi insan doğa-ilişkisinde tek veya iki boyutu dikkate almaktadır. Bu aşamada elde edilen FA sonuçlarına göre Bursa örnekleme için söz konusu bütünsel yapı

özbenlik boyutundaki beş numaralı ifade ve perspektif boyutundaki altı ve yedi numaralı ifadeler hariç olmak üzere geçerli ve güvenilirdir.

Sürdürülebilir tüketim davranışı ve çevresel endişe bağlamındaki tutumun da dahil edilmiş olduğu yapısal modelde ise bireyin doğa ile duygusal bağını temsil eden özbenlik boyutunun, altruistik tutumun sürdürülebilir tüketim davranışı üzerindeki etkisinde olumlu yönde kısmi aracı etkisi bulunmaktadır. Dolayısı ile insana ve içinde yaşadığı topluma değer veren bireylerin doğa ile olan duygusal bağının da artması ile sürdürülebilir tüketim davranışlarını (geri dönüşüm, ulaşım tercihi, su tüketimi ve gıda tüketimi) daha sık sergileyebilecekleri ifade edilebilir.

Bireyin doğa hakkındaki farkındalığını temsil eden bilişsel boyut perspektifin, biyosferik tutumun sürdürülebilir tüketim davranışı üzerindeki etkisinde olumsuz yönde kısmi aracı etkisi vardır. Bu etkinin matematiksel olarak negatif işaretli olması, perspektif boyutunun olumsuz ifadeler içermesinden kaynaklanmaktadır. Dolayısı ile negatif etki ifadeler olumlandığında pozitif işaretli olacaktır. Söz konusu boyutlar arasındaki etkileşimi açıklamak gerekirse, canlı-cansız tüm bileşenleri ile evreni bir bütün olarak gören ve değer veren biyosferik bireylerin çevresel farkındalıklarının da artışı ile sürdürülebilir tüketim davranışlarını (enerji ve su tüketimi) daha sık sergileyebilecekleri ifade edilebilir.

Fakat diğer taraftan, beklenilenin aksine, deneyim boyutu egoistik tutum ve gıda tüketimi arasındaki etkileşim hariç olmak üzere hiçbir ilişkide aracılık rolüne sahip olamamıştır. Sahip olduğu aracılık etkisi ise teorik olarak anlamlandırılmamıştır. Zira söz konusu etkileşim, egoizmin hem deneyim hem de geri dönüşüm davranışı üzerinde olumlu etkisi olduğu ve deneyimin bu etkiyi arttırdığı yönündedir.

İkinci analiz aşamasının çıktılarına göre ise Bursa merkez ilçelerinin merkez mahallelerini temsil eden ve kış aylarında ısınma ihtiyacını doğalgazdan karşılayan hane halkı, SÖÖH'ler kullanılarak kümelendiğinde en uygun küme sayısı beş olarak tespit edilmiştir.

**Birinci kümede yer alan hanelerin**, çevresel endişeleri ve doğa ile ilişkileri düşük olmasına rağmen enerji ve su tüketimleri, beyan ettikleri yüksek sürdürülebilir tüketim davranışları ile uyularak, ortalama-düşük seviyede olan vatandaşlardır. Bu durum teori

ile çatışmaktadır. Düşük çevresel endişe ve doğa ile ilişki puanına sahip bireylerin, düşük STD ve yüksek enerji ve su tüketimleri olmalıdır. Birinci kümede gözlemlenen bu tezatlık üç nedenden kaynaklanabilir:

- İlk neden, bu kümede yer alan hanelerin iktisadi güdüler ile su ve enerji tasarrufunda bulunmalarıdır. Fakat bu olasılık, gelirin düşük olması durumunda daha gerçekçidir. Birinci küme ikinci en yüksek gelire sahip olan küme olduğu için, düşük çevresel endişe ve doğa ile ilişki durumunda, orta-yüksek STD ve orta-düşük enerji ve su tüketiminin iktisadi güdülerden kaynaklandığı ifade edilemez.

- İkinci neden bu durumun aileden aktarılan kültür ile bilinç dışı bir şekilde gelişmiş ve içselleştirilmiş bir yaşam tarzı olmasıdır. Fakat bu olasılık ise STD davranışlarından alınan puanların hem orta-yüksek hem de ranjının dar olduğu durumlarda daha gerçekçidir. Ranjin dar olması, bu kümede yer alan hanelerin tüketim davranışlarının orta-yüksek ve dar bir aralıkta gözlemlenmesi anlamına gelmektedir. Fakat STD boyutlarına ait kutu-bıyık diyagramları incelendiğinde bu kümeye ait dağılımlarının ranjlarının çok geniş olduğu ve dolayısı ile aslında farklı puanlarda STD sergileyen hanelerin bu kümede yer aldığı belirtilebilir.

- Üçüncü neden ise bu kümede yaşayan hanelerin nüfuslarının az ve aile yaşam döngüsünün en başı veya en sonunda olmalarından ötürü ihtiyaçlarının da az olduğu ve bu yüzden enerji ve su tüketimlerinin de az olduğudur. Bu olasılık diğer iki olasılığa nazaran daha gerçekçidir.

Dolayısı ile birinci kümede yer alan hanelerin çevresel bilince ve gelecek endişesine sahip olmaması, STD puanlarının yüksek ve ranjının geniş, beyan edilen tüketim meblağlarının düşük ve gelirlerinin yüksek olmasından ötürü bu küme riskli olarak değerlendirilmiştir. Çünkü bu kümedeki hanelerin davranışları, değişen çevre şartlarında öngörülebilir değildir. Bu küme, “bilinçsiz yeşiller” olarak nitelendirilebilir. Bu kümede yer alan hanelere çevresel bilinçlendirme hakkında eğitimler verilebilir. Aynı zamanda, yarı-yapılandırılmış görüşmeler aracılığı ile bu vatandaşlardan kendi STD uygulamaları ya da deneyimleri hakkında bilgi alarak ve hatta “yeşil” olmayan hane gruplarına verilecek eğitimlerde onları bir eğitmen gibi kullanarak, onların da sürece dahil edilip, süreci sahiplenmeleri sağlanabilir.

**İkinci kümede yer alan hanelerin**, çevresel endişeleri ve doğa ile ilişkileri yüksek ve bununla paralel olarak enerji ve su tüketimleri orta ya da az seviyede fakat beyan ettikleri sürdürülebilir tüketim davranışları diğer kümelere nazaran ortalama ya da en düşük seviyededir. Dolayısı ile bu kümede yer alan vatandaşların, beyan ettikleri davranış bilgisinin yanıltıcı olduğu ve eylemleri ile söylemlerinin uyuşmadığı ifade edilebilir. Bu durum, birinci bölümün ikinci kısmında bahsedilen tutum-davranış boşluğu ile benzeşmek ile birlikte, boşluğun etkisi beklenenin aksine olumludur. Teorideki tutum-davranış boşluğuna göre çevreye duyarlı bir tutuma sahip olan bireylerin bu tutumu davranışlarına yansıtamaması ve enerji tüketimlerinin yüksek olması beklenmektedir. Bu kümede gözlemlenen boşluk ise hanelerin olumlu tutuma sahip olması, beyan ettikleri davranışların sahip oldukları olumlu tutumla örtüşmeyerek, tüketimlerinin az olmasıdır. Bu kümedeki hanelerde gözlemlenen su ve enerji tüketiminin azlığı, yaşadıkları hanelerin ısı yalıtımlı olması ve kullandıkları elektrikli aletlerin ve armatürlerin enerji tasarruflu olmasından kaynaklanabilir. Çevresel endişe ve doğayla ilişkileri fazla olmasına rağmen STD sergileme sıklıklarının az olması ise alışkanlıklar veya rutinlerden vazgeçme zorluğu ile açıklanabilir (Davoudi, Dilley, Crawford, 2014).

Dolayısı ile bu kümede yer alan hanelerin en kalabalık haneler olması, ev kullanım süresinin yüksek olması ve aile yaşam döngüsünde enerji ihtiyacının en fazla olduğu dönemde bulunmalarına rağmen tüketimlerinin az ve çevresel endişelerinin ve doğa ile ilişkilerinin fazla olması, bu hanelerde yaşayan vatandaşların “riskli yeşiller” olarak değerlendirilmesine neden olmaktadır. Çünkü bu kümede yer alan vatandaşların beyan ettikleri STD sergileme sıklıkları gerçeği yansıtıyor ise davranışlarının gerçek tüketim yansıması beyan edilen tüketim değerlerinden çok daha fazla olmalıdır. Bu kümede yer alan haneler için STD hakkında bilgilendirme eğitimleri ve çalıştaylar düzenlenebilir. Ayrıca, bu kümedeki vatandaşlar olumlu tutum ve yüksek çevresel endişeye sahip oldukları için, kendi davranışlarından kaynaklanan tüketimlerin çevresel sonuçlarının kendileri ile paylaşımı da davranış değişimini tetikleyebilir ve tüketim miktarları daha da azaltılabilir.

**Üçüncü kümede yer alan haneler**, en yüksek çevresel endişe, doğayla ilişki ve STD puanlarına sahip olan haneler olduğu için teoriyle uyumlu olarak en düşük enerji ve su tüketimine sahiptir. Fakat bu hanelerin gaz tüketimi çok yüksektir. Bu durum, bu

kümede yer alan hanelerin gelir seviyesi en düşük ve çoğunlukla Yıldırım ilçesinde ikamet eden haneler olmasından ötürü yaşadıkları meskenlerin ısı yalıtımına sahip olmamasından kaynaklandığı düşünülmektedir.<sup>179</sup> Çünkü Yıldırım ilçesi Bursa'nın en fazla göç alan ilçelerinden birisi olup, genellikle tek katlı evlerin birkaç yılda bir yapı izni alınarak, kat çıkmak sureti ile az katlı aile apartmanlarına dönüştürüldüğü yapılara sahiptir.

Ek olarak bu kümede yer alan hanelerin nüfus olarak en kalabalık, mesken olarak en büyük, en düşük gelirli ve aile yaşam döngüsünde enerji ihtiyacının en fazla olduğu haneler olmasına rağmen elektrik ve su tüketimlerinin az ve çevresel endişelerinin, doğa ile ilişkilerinin ve sürdürülebilir tüketim davranışı puanlarının fazla olması, bu hanelerde yaşayan vatandaşların diğer kümelerde yer alan vatandaşlara kıyas ile çok daha sürdürülebilir bir yaşama ve bakış açısına sahip oldukları ifade edilebilir. Fakat bu hanelerin en yüksek doğalgaz tüketimine sahip olması, ekonomik faktörler dolayısı ile bu hanelerin ısı yalıtımsız meskenlerde (aile apartmanlarında) oturmasından kaynaklanıyor ise elektrik ve su tüketiminin az olmasının ardında sadece ekolojik kaygılar değil aynı zamanda ekonomik kaygılar da yatıyor olabilir (Van Raaij, Fred, Verhallen, 1983).

Sonuç olarak bu kümede yer alan vatandaşlar diğer dört kümeden nerede ise sürdürülebilir bir yaşam ve bakış açısı bağlamında en ideal konumda olan kümedir. Fakat, bu kümede yer alan hanelerin en düşük gelir grubundan olması, elektrik ve su tüketimleri en az küme iken, gaz tüketimlerinin en yüksek küme olması, yaşadıkları ilçenin de sosyodemografik durumu göz önüne alındığında, bu kümedeki hanelerin ekonomik şartları değiştiği zaman tüketimlerinin de artabileceği düşünülmektedir. Bu sebeple bu kümede yer alan hanelerin “ekonomik yeşiller” olabileceği düşünülmektedir. Fakat bu kanıya tam olarak varabilmek için mesken türü, ve yalıtım bilgisi ve aynı zamanda yarı-yapılandırılmış görüşmeler ile “soft data” olarak adlandırılan, çeşitli ölçme araçları ile ölçümlenemeyen, gözlemlenerek veya deneyimleyerek elde edilebilen niteliksel soyut bilgi temin edilmelidir.

---

<sup>179</sup> Olası düşük bir ihtimal ise bu kümede yer alan hanelerin büyük bir kısmının bebeğinin olması dolayısı ile sağlık ve konfor ihtiyacının diğer kümelere kıyas ile fazla olmasıdır. Fakat, bu seçenek uzak bir ihtimaldir çünkü sağlık ve konfor ihtiyacı aynı zamanda elektrik ve su tüketiminin de artmasına neden olmaktadır.



**Dördüncü kümede yer alan haneler,** ÇEÖ, DİÖ ve STD ölçeklerinden ortalama puanlara sahip olan, beyan ettikleri tüketim seviyeleri de bu durumla paralel olup, ortalama olan hanelerdir. Dolayısı ile ölçek puanları ve tüketimler birbirine paralel olduğu için, bu hanelerin diğer kümelerdeki hanelere kıyas ile daha tutarlı oldukları ifade edilebilir. Bu yüzden bir risk teşkil etmemektedirler. Her türden aile yapısına sahip olan bu kümenin bir potansiyele sahip olduğu ve bu potansiyelin çeşitli araçlar ile ortaya çıkarılması gerektiği ve davranış değişiminin daha kolay tetiklenebileceği düşünülmektedir. Bu kümede yer alan haneler “potansiyel yeşiller” olarak değerlendirilmektedir.

**Beşinci kümede yer alan haneler,** ÇEÖ, DİÖ ve STD ölçeklerinden en az puanlara sahip olan ve bu durumla paralel olarak (hane nüfusu ile oda sayısı az olmasına rağmen) elektrik ve su tüketimleri çok yüksektir. Diğer taraftan, doğalgaz tüketimleri ortalama seviyededir. Bu durumun nedeni, bu kümede yer alan hanelerin gelir seviyesinin yüksek olmasından ötürü ısı yalıtımlı, merkezi sistemli veya pay ölçerli meskenlerde oturması veya ayrıca yaşadıkları ilçedeki konut stoğunun, diğer ilçelerdekilere nazaran daha yüksek bir kısmının yalıtımlı yeni/tadilatlı olması olabilir. Ayrıca bu kümedeki hanelerin sürdürülebilir gıda tüketimi davranışına ait öznitelik puanının yüksek olmasının sebebi ise ülkemizde hala lüks olan, organik-doğal ve sürdürülebilir gıda tüketiminin bu kümede yer alan hane halkları gibi geliri yüksek olan tüketiciler tarafından tercih edilmesinden kaynaklanabilir.

Dolayısı ile beşinci küme kesinlikle ekolojik ve sürdürülebilir bir dünya görüşüne sahip olmayıp sadece tüketim ve konfor odaklı yaşayan (Fred, Raaij, Verhallen, 1983) hanelerden oluşmaktadır. Doğa için ve gelecek nesiller için en tehlikeli vatandaş grubudur. Perspektiflerinin değişmesi ve davranış değişimlerinin tetiklenmesi için üzerinde en fazla çalışılması gerekli haneler grubudur. Bu kümede yer alan haneler “israf edenler” olarak değerlendirilmektedir.

Özetle, Bursa merkez ilçelerinin merkez mahallelerinde yaşayan halkı temsil ettiği düşünülen veri seti, sürdürülebilirlik bağlamında çevresel endişeleri, doğa ile ilişkileri, tüketim davranışları ve bazı sosyodemografik nitelikleri dikkate alınarak kümelendiğinde beş küme elde edilmektedir. Çalışmanın çıktıları Bursa halkına genellendiğinde tam anlamı ile sürdürülebilir bir bakış açısına ve yaşam tarzına sahip olan bir küme maalesef

tespit edilememiştir. Örneğin, bilinçsiz yeşiller olarak nitelendirilen hanelerin kesinlikle sürdürülebilir bir bakış açısına sahip olmadığı ama tüketim davranışlarının sürdürülebilir olup, tüketimlerinin de az olduğu görülmektedir. Riskli yeşiller olarak nitelendirilen hanelerin sürdürülebilir bir bakış açısına sahip olduğu, tüketimlerinin az ve fakat davranışlarının ise sürdürülebilir olmadığı tespit edilmiştir. Ekonomik yeşiller ise kesinlikle sürdürülebilir bir bakış açısı ve yaşam tarzına sahip iken tüketimlerinin ekonomik kaygılar ile değişkenlik gösterebileceği düşünülmektedir. Potansiyel yeşiller ortalama bir bakış açısı ve yaşam tarzı ile sürdürülebilir olup, tüketimleri de bu duruma paralel olarak ortalamadır. En risksiz ve yönetilebilir grup olarak değerlendirilmektedir. İsrar edenler olarak nitelendirilen hanelerin kesinlikle sürdürülebilir bakış açısına, yaşam tarzına ve tüketime sahip olmadığı görülmektedir. Bu grupta yer alan vatandaşların hem çevre hem de gelecek nesiller için kaygı duymadığı ve önceliklerinin kendi konforları olduğu ifade edilebilir.

Söz konusu beş kümenin nitelikleri dikkate alındığında Bursa halkının genel olarak sürdürülebilir bir bakış açısına sahip olmadığı, az tüketiyor ise bu durum genellikle ekonomik kaygılardan ya da aileden gelen alışkanlıklardan kaynaklandığı ve bu yüzden yerel yönetimin kapsayıcı bir bakış açısı ile vatandaşları sürdürülebilir dönüşüm sürecine dahil ederek bu süreci benimsemelerini sağlamaları gerekmektedir. Mevcut durumda, pek de “yeşil” olmayan vatandaşlara sahip bir kent için “yeşil” markasının elde edilebilirliği gerçekçi ve başarılabilir değildir.

Bu tez çalışmasında ÇEÖ'nün STD üzerindeki etkisinde her ne kadar DİÖ'nün deneyim boyutunun olumlu yönde aracı etkisi tespit edilememiş olsa da, deneyimin STD boyutlarının hepsinin üzerinde olumlu yönde anlamlı etkisi tespit edilmiştir, Tablo 42. Fakat tespit edilen kümeler deneyim boyutu açısından incelendiğinde ise birbirinden çok farklılık göstermediği dikkati çekmektedir, Şekil 91. Dolayısı ile deneyimin STD üzerinde anlamlı etkisi ile birlikte kümeler arasında bariz bir şekilde deneyimsel fark tespit edilebilmiş olsa idi yerel yönetime, insanların daha çok doğa ile iç içe olmalarını sağlayacak yatırımlar yapılması önerilebilirdi. Ayrıca, elde edilen kümelerden bariz bir şekilde diğerlerinden ayrılarak sürdürülebilir olarak nitelendirilecek bir küme olsa idi, bu kümedeki haneler ile yüz yüze görüşme yöntemi ile soyut bilgi toplanarak, elde edilen bilgiler ışığında bir yol haritası oluşturulabilirdi.

Tez çalışması, profil çıkarma aşamasında kullanılan yöntemler açısından değerlendirildiğinde ise, çevrimiçi öğrenme algoritması ile eğitilen ağlar, daha küçük NH sağladığı için yığın öğrenme algoritması ile eğitilen ağlardan daha yüksek performans sağlamıştır.

Ayrıca, topolojik olarak uzun dikdörtgen şekle sahip olan haritaların daha karesel olan haritalara kıyas ile daha iyi küme bölümlenmeleri sağladığı tespit edilmiştir.

Ağların eğitim parametrelerinden komşuluk yarıçapının iterasyon sayısı ilerledikçe sıfırlanmaması ve fakat sıfıra yaklaşması, fine tuning işleminin yapılma gerekliliğini ortadan kaldırmıştır (Dipnot 155-sayfa 223-224).

Ağların eğitim parametrelerinden katman ağırlıkları, küme bütünlüğüne sahip küme bölümlenmeleri elde edilmesini sağlamıştır (Dipnot 154-sayfa 222)

İşsel değerlendirme kriterlerine göre en uygun küme sayısı ya da en iyi performansa sahip algoritma tespit edilemediğinde çok değişkenli varyans analizinin kullanımı, ayırt edici bir kriter olmuştur.

PERMANOVA yönteminin kullanımı, bu tez çalışmasındaki gibi eşit (ya da denk) sayıda gözlem içermeyen kümelerin/grupların var olduğu ve gerek toplam gerek ise gruplardaki gözlem sayısının çok az olduğu durumlarda çok değişkenli varyans analizinin gerçekleştirilebilmesini mümkün kılmıştır.

TBA'ların karar ağaçları ile birlikte kullanımı dallanmada kullanılan ayırt edici özneliklerin çapraz kontrolünün yapılmasını sağlamıştır. Aynı zamanda betimsel istatistiklerin de karar ağaçları ile birlikte kullanımı, hangi boyutların ve davranışların ne tür bir profil oluşturduğu bilgisinin daha etkin bir şekilde keşfedilmesini sağlamıştır.

Ayrıca araştırma modelinin değerlendirilmesine ilişkin kısımda gerçekleştirilen aykırı veri analizinde, SMM ölçütünün Mahalanobis mesafesinden daha iyi bir performans sergilediği görülmüştür. Çünkü SMM ölçütü ile temizlenen veri setlerinin AFA sonuçları, Mahalanobis mesafesi kullanılarak elde edilen veri setlerinin AFA sonuçlarından daha iyidir.

Bu tez çalışmasının devamı niteliğindeki çalışmalarda ise sırası ile

- Vatandaşların yaşadıkları kent hakkındaki algılarının keşfedilmesi ve ihtiyaçlarının tespit edilmesi

- Analitik yöntemler kullanarak STD davranışı ve biyofilik eğilimleri destekleyecek politikalar geliştirilmesi

planlanmaktadır. Yakın gelecek için planlanan çalışmaların bitirilmesi ile Yeşil Bursa kent markalaşmasında vatandaşların katılım boyutu tamamlanmış olacaktır. Bu tez çalışmasında; BBB, AYB ödülü için hazırlanmış olduğu başvuru dosyasını paylaşarak katılımında bulunmuştur. Belediyeden tam destek alınması durumunda ise daha kapsamlı bir proje ekibi ile kent markalaşmasının yerel yönetim ve iş çevresi boyutları da tamamlanıp BBB için uzun vadeli bir yol haritası oluşturulabilir.

## KAYNAKLAR

- AAKER David A., 1996, *Building Strong Brands*, NY: The Free Press.
- ABRAHAMSE Wokje, Linda STEG, 2009, "How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings?", *Journal of Economic Psychology*, doi:10.1016/j.joep.2009.05.006.
- ACUNA Edgar, Caroline RODRIGUEZ, 2004, "On detection of outliers and their effect in supervised classification", *IPSI 2004*, 2004, ss. 1-14.
- AHMAD Amir, Lipika DEY, 2007, "A k-mean clustering algorithm for mixed numeric and categorical data", *Data and Knowledge Engineering*, C. 63, S. 2, ss. 503-27, doi:10.1016/j.datak.2007.03.016.
- AHMAD Amir, Sarosh HASHMI, 2016, "K-Harmonic means type clustering algorithm for mixed datasets", *Applied Soft Computing Journal*, C. 48, S. November, ss. 39-49, doi:10.1016/j.asoc.2016.06.019.
- AHMAD Amir, Shehroz KHAN, 2016, "Survey of state-of-the-art mixed data clustering algorithms", *IEEE Access*, C. 4, ss. 31883-902, doi:10.1109/ACCESS.2017.DOI.
- AKINDUKO Ayodeji A, Evgeny M MIRKES, Alexander N GORBAN, 2015, "SOM: Stochastic initialization versus principal components", *Information Sciences*, C. 364-365, ss. 213-21, doi:10.1016/j.ins.2015.10.013.
- AKKÜÇÜK Ulaş, 2011, *Veri Madenciliği-Sınıflama ve Kümeleme Algoritmaları*, İstanbul: Yalın Yayıncılık.
- AKPINAR Haldun, 2018, *DATA Veri Madenciliği Veri Analizi*. İkinci Bas, İstanbul: Papatya Yayıncılık Eğitim.
- ALAHAKOON Damminda, Saman K. HALGAMUGE, Bala SRINIVASAN, 2000, "Dynamic self-organizing maps with controlled growth for knowledge discovery", *IEEE Transactions on Neural Networks*, C. 11, S. 3, ss. 601-14, doi:10.1109/72.846732.
- ALTUNKAYNAK Bülent, 2017, *Veri Madenciliği Yöntemleri ve R Uygulamaları*, Ankara: Seçkin Yayıncılık.
- AMIGÓ Enrique vd., 2009, "A comparison of extrinsic clustering evaluation metrics based on formal constraints", *Information Retrieval*, C. 12, S. 4, Springer Netherlands, ss. 461-86, doi:10.1007/s10791-008-9066-8.
- AMRUTHNATH Nagdev, Tarun GUPTA, 2019, "Fault diagnosis using clustering. What statistical test to use for hypothesis testing?", *Machine Learning and Applications: An International Journal (MLAIJ)*, C. 6, S. 1, doi:10.5121/mlaij.2019.6102.
- ANDERSON Marti, Ray N GORLEY, K Robert CLARKE, 2008, "PERMANOVA + for PRIMER user manual", Plymouth, UK: PRIMER-e (Quest Research Limited).
- ANDERSON Marti J., 2001, "A new method for non-parametric multivariate analysis of variance", *Austral Ecology*, C. 2001, S. 26, ss. 32-46, doi:10.1080/13645700903062353.
- , 2017, "Permutational Multivariate Analysis of Variance (PERMANOVA)",

- Wiley StatsRef: Statistics Reference Online*, doi:10.1002/9781118445112.stat07841.
- , 2017, “Some solutions to the multivariate Behrens-Fisher problem for dissimilarity-based analyses”, *Australian & New Zealand Journal of Statistics*, C. 59, S. 1, Blackwell Publishing Ltd, ss. 57-79, doi:10.1111/anzs.12176.
- ANHOLT Simon, 2002, “Foreword to the special issue on place branding”, *Brand Management*, C. 9, S. 4-5, ss. 229-39.
- , 2006, “The Anholt-GMI city brands index: How the world sees the world’s cities”, *Place Branding*, C. 2, S. 1, ss. 18-31, doi:10.1057/palgrave.pb.5990042.
- ASTRACHAN Claudia Binz, Vijay K. PATEL, Gabrielle WANZENRIED, 2014, “A comparative study of CB-SEM and PLS-SEM for theory development in family firm research”, *Journal of Family Business Strategy*, C. 5, ss. 116-28, doi:10.1016/j.jfbs.2013.12.002.
- AYDIN KESKIN Gülşen, 2008, *Hata Türü ve Etkileri Analizinde Bulanık Adaptif Rezonans Teorisi Yaklaşımı ile Bir Model Önerisi*, Kocaeli Üniversitesi.
- AZEVEDO Ana, Manuel Filipe DOS SANTOS, 2008, “KDD, SEMMA and CRISP-DM: A parallel overview”, *IADIS European Conf. Data Mining*, 2008, ss. 182-85.
- BABAHAN Ali, Zuhul GÜLER, 2015, “Kentsel dönüşüm, sürdürülebilir kalkınma ve sosyal çeşitlilik/karışım”, *Neden, Nasıl ve Kim İçin Kentsel Dönüşüm*, ed. Betül Duman, Rafet Coşkun, İstanbul, ss. 471-96.
- BAHAR Fatma, Elvan SAHİN, 2017, “An associational research on Turkish children’s environmentally responsible behaviors, nature relatedness, and motive concerns”, *Science Education International*, C. 28, S. 2, ss. 111-18.
- BALAJI K, K LAVANYA, 2018, “Clustering Algorithms for mixed datasets: A review”, *International Journal of Pure and Applied Mathematics*, C. 118, S. 7, ss. 547-56, <http://www.ijpam.eu>.
- BARON Reuben M., David A. KENNY, 1986, “The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations”, *Journal of Personality and Social Psychology*, C. 51, S. 6, ss. 1173-77, doi:10.1037/0022-3514.51.6.1173.
- BAŞAR Enes Emre, 2016, *Hane Halklarının Sürdürülebilir Tüketim Davranışlarının Ölçülmesine Yönelik Bir Ölçek Geliştirme Çalışması*, Atatürk Üniversitesi.
- BECCALI M. vd., 2004, “Forecasting daily urban electric load profiles using artificial neural networks”, *Energy Conversion and Management*, doi:10.1016/j.enconman.2004.01.006.
- BECKER Lawrence J. vd., 1981, “Relating attitudes to residential energy use”, *Environment and Behavior*, C. 13, S. 5, ss. 590-609.
- BEN-ISRAEL Adi vd., 2007, “Probabilistic D-Clustering”, *Journal of Classification*, C. 25, doi:10.1007/s00357-007-0021-y.
- BIBRI Simon Elias, John KROGSTIE, 2017, “Smart sustainable cities of the future: An extensive interdisciplinary literature review”, *Sustainable Cities and Society*, C. 31, Elsevier B.V., ss. 183-212, doi:10.1016/j.scs.2017.02.016.

- BLINN Jim, 2003, *Notation, Notation, Notation*, San Francisco, CA: Morgan Kaufmann.
- BOEHMKE Brad, Brandon GREENWELL, 2020, *Hands-On Machine Learning with R*, Boca Raton, FL: CRC Press.
- BORER Michael Ian, 2006, "The location of culture: The urban culturalist perspective", *City & Community*, C. 5, S. 2, ss. 173-97, [http://www.urbanlab.org/articles/urb\\_loc\\_culture.pdf](http://www.urbanlab.org/articles/urb_loc_culture.pdf).
- BREARD Gregory T, 2017, *Evaluating Self-Organizing Map Quality Measures as Convergence Criteria*, University of Rhode Island.
- BREIMAN Leo vd., 1984, *Classification and Regression Trees*, Boca Raton, FL: Chapman & Hall/CRC.
- BRUNI Coral M., Randie C. CHANCE, P. Wesley SCHULTZ, 2012, "Measuring values-based environmental concerns in children: An environmental motives scale", *Journal of Environmental Education*, C. 43, S. 1, ss. 1-15, doi:10.1080/00958964.2011.583945.
- BRUNTLAND Gro Harlem, 1987, "Our Common Future", Oslo, doi:10.1016/0022-2364(91)90424-R.
- BURDENSKI Tom, 2000, "Evaluating univariate, bivariate, and multivariate normality using graphical and statistical procedures", *Multiple Linear Regression Viewpoints*, C. 26, S. 2, ss. 15-25.
- BUSCH Henner, Stefan ANDERBERG, 2015, "Green attraction-Transnational municipal climate networks and green city branding", *Journal of Management and Sustainability*, C. 5, S. 4, ss. 1-16, doi:10.3868/s050-004-015-0003-8.
- BÜYÜKÖZTÜRK Şener, 2005, *Sosyal Bilimler İçin Veri Analizi El Kitabı*, 5. b., Ankara: Pegem Yayıncılık.
- BÜYÜKŞEHİR BELEDİYESİ Bursa, 2018, "Bursa, 'Avrupa Yeşil Başkenti' olmaya aday", *Basın Açıklaması*, <https://www.bursa.bel.tr/bursa---avrupa-yesil-baskenti--olmaya-aday/haber/25636>.
- C.S.B. Grimmond vd., 2015, "Urban energy budget models", *Understanding Urban Metabolism: A Tool for Urban Metabolism*, ed. Nektarios Chrysoulakis, Eduardo Anselmo de Castro, Eddy J. Moors, Abingdon, Oxon: Routledge, ss. 91-105.
- ÇAKIR Birgül vd., 2015, "Adaptation of nature relatedness scale to Turkish", *Elementary Education Online İlköğretim Online*, C. 14, S. 144, ss. 1370-83, doi:10.17051/ieo.2015.95299.
- CARUANA Robert, Michal J. CARRINGTON, Andreas CHATZIDAKIS, 2016, "Beyond the Attitude-Behaviour Gap: Novel Perspectives in Consumer Ethics': Introduction to the Thematic Symposium", *Journal of Business Ethics*, 2016, doi:10.1007/s10551-014-2444-9.
- CATENI Silvia vd., 2013, "Variable Selection and Feature Extraction Through Artificial Intelligence Techniques", *Multivariate Analysis in Management, Engineering and the Sciences*, ed. Leandro Valim de Freitas, Ana Paula Barbosa Rodrigues de Freitas, INTECH, ss. 103-18, doi:10.5772/53862.

- CHABANE Ahmed Nait, Naveed ISLAM, Benoit ZERR, 2017, “Incremental clustering of sonar images using self-organizing maps combined with fuzzy adaptive resonance theory”, *Ocean Engineering*, C. 142, S. 2017, Elsevier Ltd, ss. 133-44, doi:10.1016/j.oceaneng.2017.06.061.
- CHAIR Sara, Malika CHARRAD, 2016, “A new R package for Multi-SOM clustering”, *Conférences Conjointes Francophones sur la Sciences des Données AAFD & SFC*, 2016, s. .
- CHAN Chung-Shing, Lawal M MARAFA, 2014, “Rebranding Hong Kong ‘Green’: The potential for connecting city branding with green resources”, *World Leisure Journal*, C. 56, S. 1, ss. 62-80, doi:10.1080/04419057.2013.876587.
- CHAO Hsin Chieh vd., 2011, “Modified adaptive resonance theory for alarm correlation based on distance hierarchy in mobile networks”, *2011 13th Asia-Pacific Network Operations and Management Symposium*, 2011, ss. 1-4, doi:10.1109/APNOMS.2011.6076958.
- CHARRAD Malika vd., 2014, “NbClust: An R package for determining the relevant number of clusters in a data set”, *Journal of Statistical Software*, C. 61, S. 6, ss. 1-36.
- CHECKER Melissa, 2011, “Wiped out by the ‘Greenwave’: Environmental gentrification and the paradoxical politics of urban sustainability”, *City and Society*, C. 23, S. 2, ss. 210-29, doi:10.1111/j.1548-744X.2011.01063.x.
- CHEN Ning, Nuno C. MARQUES, 2005, “An extension of self-organizing maps to categorical data”, *Progress in Artificial Intelligence: 12th Portuguese Conference on Artificial Intelligence*, ed. Carlos Bento, Amilcar Cardoso, Gael Dias, 2005, ss. 304-13, doi:10.1007/11595014\_31.
- CHERNICK Michael R., Robert A. LABUDDE, 2011, *An Introduction to Bootstrap Method with Applications in R*, New Jersey: John Wiley&Sons, Inc.
- CHICCO Gianfranco vd., 2004, “Load pattern-based classification of electricity customers”, *IEEE Transactions on Power Systems*, doi:10.1109/TPWRS.2004.826810.
- CHIESURA Anna, 2004, “The role of urban parks for the sustainable city”, *Landscape and Urban Planning*, C. 68, S. 1, Elsevier, ss. 129-38, doi:10.1016/J.LANDURBPLAN.2003.08.003.
- CIAMPI Antonio, Yves LECHEVALLIER, 2000, “Clustering large, multi-level data sets: An approach based on kohonen self organizing maps”, *Principles of Data Mining and Knowledge Discovery, 4th European Conference, PKDD*, 2000, ss. 353-58, doi:10.1007/3-540-45372-5\_36.
- CIVELEK Mustafa Emre, 2018, *Yapısal Eşitlik Modellemesi Metodolojisi*, İstanbul: Beta.
- CLAYTON Susan, 2003, “Environmental identity: A conceptual and an operational definition”, *Identity and the Natural Environment*, ed. Susan Clayton, Susan Opatow, Cambridge, Massachusetts: The MIT Press, ss. 45-66, doi:10.7551/mitpress/3644.001.0001.



- COŞKUN Recai vd., 2015, *Sosyal Bilimlerde Araştırma Yöntemleri SPSS Uygulamalı*, 8. b., Sakarya: Sakarya Yayıncılık.
- COSO Carmelo DEL vd., 2015, “Mixing numerical and categorical data in a self-organizing map by means of frequency neurons”, *Applied Soft Computing Journal*, S. 36, ss. 246-54, doi:10.1016/j.asoc.2015.06.058.
- COSTA Ivan G, Francisco De A T DE CARVALHO, Marcílio C P DE SOUTO, 2002, “Comparative study on proximity indices for cluster analysis of gene expression time series”, *Journal of Intelligent & Fuzzy Systems*2, C. 2002/2003, S. 13, ss. 133-42.
- , 2004, “Comparative analysis of clustering methods for gene expression time course data”, *Genetics and Molecular Biology*, C. 27, S. 4, ss. 623-31, www.sbg.org.br.
- COTTRELL Marie vd., 2016, “Theoretical and applied aspects of the self-organizing maps”, *11th International Workshop WSOM 2016*, ed. Erzsebet Merenyi, Michael J. Mendenhall, Patrick O’Driscoll, 2016, ss. 3-26, doi:10.1007/978-3-319-28518-4\_1.
- CROCKETT Shane R., Leiza J. WOOD, 2004, “Western Australia: Building a state brand”, *Destination Branding: Creating the Unique Destination Proposition*, ed. Nigel Morgan, Annette Pritchard, R. Pride, 2nd b., Burlington, MA, ss. 185-206.
- DAVOUDI Simin, Luke DILLEY, Jenny CRAWFORD, 2013, ...“...57.Davoudi.Dilley.Crawford.2014”.
- , 2014, “Energy consumption behaviour: rational or habitual?”, *disP - The Planning Review*, C. 50, S. 3, ss. 11-19, doi:10.1080/02513625.2014.979039.
- DEAN Julie H. vd., 2018, “Is nature relatedness associated with better mental and physical health?”, *International Journal of Environmental Research and Public Health*, C. 15, S. 7, ss. 9-11, doi:10.3390/ijerph15071371.
- DEMIRUTKU Kürşad, Elis GÜNGÖR, 2017, “Bilimsellik değerlerinin Schwartz’ın değer kuramı çerçevesinde incelenmesi”, 2. *Sosyal Psikoloji Kongresi*, 2017, s. .
- DEPARTMENT FOR ENVIRONMENT Food and Rural Affairs, 2011, “A framework for pro-environmental behaviours”, www.defra.gov.uktel:02072386000Website:www.defra.gov.ukhttp://www.opsi.gov.uk/click-use/value-added-licence-information/index.htm.
- DEVELLIS Robert F, 2014, *Ölçek Geliştirme: Kuram ve Uygulamalar*, ed. Tarık Totan, 3. b., Nobel.
- DIAMANTIDIS N A, D KARLIS, E A GIAKOUMAKIS, 2000, “Unsupervised stratification of cross-validation for accuracy estimation”, *Artificial Intelligence*, C. 116, S. 2000, ss. 1-16.
- DIAMANTOPOULOS Adamantios, Judy A. SIGUAW, 2006, “Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration”, *British Journal of Management*, C. 17, S. 4, ss. 263-82, doi:10.1111/j.1467-8551.2006.00500.x.
- DINNIE Keith, 2011, *City Branding-Theory and Cases*, 1st b., Basingstoke, Hampshire:

Palgrave MacMillan.

- DRUCKMAN A., T. JACKSON, 2008, "Household energy consumption in the UK: A highly geographically and socio-economically disaggregated model", *Energy Policy*, doi:10.1016/j.enpol.2008.03.021.
- DUDA Richard O., Peter E. HART, David G. STORK, 2000, *Pattern Classification*. Second, Wiley Interscience.
- DURMUŞ Beril, Serra YURTKORU, Murat ÇINKO, 2016, *Sosyal Bilimlerde Spss'le Veri Analizi*, 6. b., İstanbul: Beta Basım.
- DUTCHER Daniel D. vd., 2007, "Connectivity with nature as a measure of environmental values", *Environment and Behavior*, C. 39, S. 4, ss. 474-93, doi:10.1177/0013916506298794.
- EFRON Bradley, Robert J. TIBSHIRANI, 1994, *An Introduction to the Bootstrap*, London: Chapman & Hall.
- EGGER Steve, 2006, "Determining a sustainable city model", *Environmental Modelling and Software*, C. 21, S. 9, ss. 1235-46, doi:10.1016/j.envsoft.2005.04.012.
- ERGAN Senem, Ayse AKYOL, Derman KUCUKALTAN, 2014, "The effect of city branding and its components on resident satisfaction", *Journal of Management, Marketing & Logistics-JMML*, C. 1, S. 2, Kucukaltan, ss. 82-97.
- ERSÖZ Filiz, 2019, *Veri Madenciliği Teknikleri ve Uygulamaları-Kavram, Teori, Modeller-Yöntem*. Üçüncü Bas, Ankara: Seçkin Yayıncılık.
- ESHUIS Jasper, Erik Hans KLIJN, Erik BRAUN, 2014, "Place marketing and citizen participation: Branding as strategy to address the emotional dimension of policy making?", *International Review of Administrative Sciences*, C. 80, S. 1, ss. 151-71, doi:10.1177/0020852313513872.
- ESIN Alptekin, Müslim EKNI, Hamza GAMGAM, 2006, *Istatistik*, Ankara: Gazi Kitabevi.
- EUROPEAN COMMISSION DIRECTORATE-GENERAL FOR ENERGY, 2019, "Energy Efficiency Directive", <https://ec.europa.eu/energy/en/topics/energy-efficiency/energy-efficiency-directive>.
- EUROPEAN ENVIRONMENTAL AGENCY, 2013, "Final electricity consumption by sector , EU-27", European Environment Agency.
- EUROPEAN UNION, 2018, "Indicators for Sustainable Cities", doi:10.2779/121865.
- EVERITT B. S., 1988, "A finite mixture model for the clustering of mixed-mode data", *Statistics and Probability Letters*, C. 6, S. 5, ss. 305-9, doi:10.1016/0167-7152(88)90004-1.
- EVERITT Brian S vd., 2011, *Cluster Analysis*. Fifth edit, West Sussex, UK: John Wiley&Sons, Inc.
- FAIERS Adam, Matt COOK, Charles NEAME, 2007, "Towards a contemporary approach for understanding consumer behaviour in the context of domestic energy use", *Energy Policy*, doi:10.1016/j.enpol.2007.01.003.

- FAYYAD Usama, Gregory PIATETSKY-SHAPIRO, Padhraic SMYTH, 1996a, "From data mining to knowledge discovery in databases", *AI Magazine*, C. 17, S. 3, ss. 37-53, doi:10.1609/aimag.v17i3.1230.
- , 1996b, "The KDD process for extracting useful knowledge from volumes of data", *Communications of the ACM*, C. 39, S. 11, ss. 27-34, doi:10.1145/240455.240464.
- FAYYAD Usama, Paul STOLORZ, 1997, "Data mining and KDD: Promise and challenges", *Future Generation Computer Systems*, C. 1997, S. 13, ss. 99-115.
- FENG Changyong vd., 2014, "Log-transformation and its implications for data analysis", *Shanghai Archives of Psychiatry*, C. 26, S. 2, ss. 105-9.
- FERAL Crystal-Helen, 1998, "The connectedness model and optimal development: Is ecopsychology the answer to emotional well-being?", *The Humanistic Psychologist*, C. 26, S. 1-3, ss. 243-74, doi:10.1080/08873267.1998.9976975.
- FILZMOSER Peter, Robert G GARRETT, Clemens REIMANN, 2005, "Multivariate outlier detection in exploration geochemistry", *Computers & Geosciences*, C. 31, S. 2005, ss. 579-87, doi:10.1016/j.cageo.2004.11.013.
- FLANNER Mark G., 2009, "Integrating anthropogenic heat flux with global climate models", *Geophysical Research Letters*, C. 36, S. 2, ss. 1-5, doi:10.1029/2008GL036465.
- FORNELL Claes, David F. LARCKER, 1981, "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, C. 18, ss. 39-50, doi:10.2307/3151312.
- FORT Jean-Claude, Patrick LETREMY, Marie COTTRELL, 2002, "Advantages and drawbacks of the batch kohonen algorithm", *10th European Symposium on Artificial Neural Networks, ESANN*, ed. M. Verleysen, 2002, ss. 223-30.
- FOSS Alex vd., 2016, "A semiparametric method for clustering mixed data", *Machine Learning*, C. 105, ss. 419-58, doi:10.1007/s10994-016-5575-7.
- FREDERIKS Elisha R., Karen STENNER, Elizabeth V. HOBMAN, 2015a, "Household energy use: Applying behavioural economics to understand consumer decision-making and behaviour", *Renewable and Sustainable Energy Reviews*, doi:10.1016/j.rser.2014.09.026.
- , 2015b, "The socio-demographic and psychological predictors of residential energy consumption: A comprehensive review", *Energies*, C. 8, ss. 573-609, doi:10.3390/en8010573.
- FRIEDMAN Jerome, Trevor HASTIE, Robert TIBSHIRANI, 2008, *The Elements of Statistical Learning*. Second edi, Springer.
- FRITZSCHE David J., 1981, "An analysis of energy consumption patterns by stage of family life cycle", *Journal of Marketing Research*, C. 18, S. May, ss. 227-32.
- GABRIELSSON Sam, Stefan GABRIELSSON, 2006, *The Use of Self-Organizing Maps in Recommender Systems*, Uppsala University.
- GAMA J, C PINTO, 2006, "Discretization from Data Streams: Applications to

- Histograms and Data Mining”, *Proceedings of the 2006 ACM symposium on Applied computing*, 2006, ss. 23-27, doi:10.1145/1141277.1141429.
- GARSON G. David, 2016, *Partial Least Squares: Regression & Structural Equation Models*, G. David Garson and Statistical Associates Publishing.
- GATERSLEBEN Birgitta, 2013, “Measuring environmental behavior”, *Environmental Psychology, An Introduction*, ed. Linda Steg, Agnes E. van den Berg, Judith I.M. de Groot, West Sussex: BPS Blackwell, ss. 131-40.
- GATERSLEBEN Birgitta, Linda STEG, 2013, “Affective and symbolic aspects of environmental behaviour”, *Environmental Psychology: An Introduction*, ed. Linda Steg, Agnes E. van den Berg, Judith I.M. De Groot, West Sussex: BPS Blackwell, ss. 165-74.
- GHOSH J K, Partha P MAJUMDER, 1994, “Letter to the editor: P.C. Mahalanobis’ contributions to biometry”, *Annals of Human Biology*, C. 21, S. 3, ss. 287-89, doi:10.1080/03014469400003292.
- GOOD Philip, 2000, *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. Second Edi, Springer.
- GRANTHAM Stephanie, 2010, “Household energy consumption, conservation & efficiency”.
- GROOT Judith I. M. DE, John THOGERSTEN, 2013, “Values and pro-environmental behavior”, *Environmental Psychology: An Introduction*, ed. Linda Steg, Agnes E. van den Berg, Judith I. M. de Groot, West Sussex: BPS Blackwell, ss. 141-52.
- GUERIN Denise A., Becky L. YUST, Julie G. COOPET, 2000, “Occupant predictors of household energy behavior and consumption change as found in energy studies since 1975”, *Family and Consumer Sciences Research Journal*, doi:10.1177/1077727X00291003.
- GÜLOĞLU Bülent, Emre AKIN, 2014, “Türkiye’de hane halkları elektrik talebinin belirleyicileri: Sıralı Logit yaklaşımı”, *Siyaset, Ekonomi ve Yönetim Araştırmaları Dergisi*, C. 2, S. 3, ss. 1-20.
- HAGAN Martin T. vd., 1996, *Neural Network Design*, 2. b., Boulder, Colorado.
- HAIR J. F. vd., 2006, *Multivariate Data Analysis*. Sixth edit, Uppersaddle River, N.J: Pearson Prentice Hall.
- HAIR Joseph F. vd., 1998, *Multivariate Data Analysis*, 5. b., Upper Saddle River, New Jersey: Prentice-Hall, Inc.
- HAIR JR. Joseph F. vd., 2014, *Multivariate Data Analysis*. Internatio, Essex: Pearson.
- HAIR JR Joe F. vd., 2014, “Partial least squares structural equation modeling (PLS-SEM) An emerging tool in business research”, *European Business Review*, C. 26, S. 2, ss. 106-21.
- , 2017, *A Primer on Partial Least Squares Structural Equation Modeling, A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. Second, Los Angeles: SAGE Publications Ltd.

- HALKIDI Maria, Yannis BATISTAKIS, Michalis VAZIRGIANNIS, 2001, "On clustering validation techniques", *Journal of Intelligent Information Systems*, C. 17, S. 2-3, ss. 107-45, doi:10.1023/A:1012801612483.
- HAN Jiawei, Micheline KAMBER, Jian PEI, 2012, *Data Mining: Concepts and Techniques*, 3rd b., Waltham, MA: Morgan Kaufmann.
- HASSAN Abbas M., Hyowon LEE, 2015, "The paradox of the sustainable city: definitions and examples", *Environment, Development and Sustainability*, C. 17, S. 6, ss. 1267-85, doi:10.1007/s10668-014-9604-z.
- HAYKIN Simon, 2009, *Neural Networks and Learning Machines*, 3. b., Upper Saddle River, New Jersey: Pearson.
- HAYTON James C., David G. ALLEN, Vida SCARPELLO, 2004, "Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis", *Organizational Research Methods*, doi:10.1177/1094428104263675.
- HENNIG Christian, Marina MEILA, 2016, "Cluster analysis: An overview", *Handbook of Cluster Analysis*, ed. Christian Hennig vd., Boca Raton, FL: CRC Press, ss. 1-19.
- HENSELER Jörg, Geoffrey HUBONA, Pauline Ash RAY, 2016, "Using PLS path modeling in new technology research: updated guidelines", *Industrial Management & Data Systems*, C. 16, S. 1, ss. 2-20, doi:10.1108/IMDS-09-2015-0382.
- HENSELER Jörg, Christian M. RINGLE, Marko SARSTEDT, 2015, "A new criterion for assessing discriminant validity in variance-based structural equation modeling", *Journal of the Academy of Marketing Science*, C. 43, ss. 115-35, doi:10.1007/s11747-014-0403-8.
- HIDALGO Anne, 2017, "What Happened in 2017?", C. 9, <http://dblp.uni-trier.de/db/journals/isjgp/isjgp9.html#Kuchta00b>.
- HILLER Carolina, 2014, "Factors influencing residents' energy use - A study of energy-related behaviour in 57 Swedish homes", *Energy and Buildings*, doi:10.1016/j.enbuild.2014.11.013.
- HSU Chung-Chian, 2006, "Generalizing self-organizing map for categorical data", *IEEE Transactionson neural network*, C. 17, S. 2, ss. 294-304, doi:10.1109/TNN.2005.863415.
- HSU Chung-Chian, Yan-Ping HUANG, Chieh-Ming HSIAO, 2006, "Modified adaptive resonance theory network for mixed data based on distance hierarchy", *6th International Conference on Computational Science – ICCS 2006*, ed. Vassil N. Alexandrov vd., 2006, ss. 757-64.
- HSU Chung Chian, Yu Cheng CHEN, 2007a, "Mining of mixed data with application to catalog marketing", *Expert Systems with Applications*, C. 32, S. 2007, ss. 12-23, doi:10.1016/j.eswa.2005.11.017.
- , 2007b, "Mining of mixed data with application to catalog marketing", *Expert Systems with Applications*, C. 32, S. 1, ss. 12-23, doi:10.1016/j.eswa.2005.11.017.
- HSU Chung Chian, Yan Ping HUANG, 2008, "Incremental clustering of mixed data based on distance hierarchy", *Expert Systems with Applications*, C. 35, S. 3, ss. 1177-

85, doi:10.1016/j.eswa.2007.08.049.

- HSU Chung Chian, Kuo M. WANG, Sheng Hsuan WANG, 2006, “GViSOM for multivariate mixed data projection and structure visualization”, *IEEE International Conference on Neural Networks*, 2006, ss. 3300-3305, doi:10.1109/ijcnn.2006.247327.
- HSU Chung Chian, Sheng Hsuan WANG, 2006, “An Integrated Framework for Visualized and Exploratory Pattern Discovery in Mixed Data”, *IEEE Transactions on Knowledge and Data Engineering*, C. 18, S. 2, ss. 161-73, doi:10.1109/TKDE.2006.23.
- HUANG S vd., 2014, “A formal algorithm for verifying the validity of clustering results based on model checking”, *PLoS ONE*, C. 9, S. 3, ss. 1-14, doi:10.1371/journal.pone.0090109.
- HUANG Zhexue, 1998, “Extensions to the k-means algorithm for clustering large data sets with categorical values”, *Data Mining and Knowledge Discovery*, C. 2, ss. 283-304.
- HUBERT Lawrence, Phipps ARABIE, 1985, “Comparing partitions”, *Journal of Classification*, C. 1985, S. 2, ss. 193-218.
- HUBERT Mia, Michiel DEBRUYNE, 2010, “Minimum covariance determinant”, *WIREs Computational Statistics*, C. 2, S. January/February, ss. 36-43, doi:10.1002/wics.61.
- HUEBNER Gesche M. vd., 2015, “Explaining domestic energy consumption - The comparative contribution of building factors, socio-demographics, behaviours and attitudes”, *Applied Energy*, doi:10.1016/j.apenergy.2015.09.028.
- İLGÜNER Muhterem, Christer ASPLUND, 2011, *Marka Şehir*, İstanbul: Bilnet.
- İŞÇİ GÜNERİ Öznur, Dursun AYDIN, 2017, “Grup üyelerini belirlemede istatistiksel sınıflandırma yöntemleri: Karşılaştırmalı bir çalışma”, *Türkiye Klinikleri Journal of Biostatistics*, C. 9, S. 1, ss. 45-67, doi:10.5336/biostatic.2016-52498.
- İŞİĞİÇOK Erkan, 2011, *Altı Sigma Kara Kuşaklar İçin Hipotez Testleri Yol Haritası*. İkinci, Bursa: Marmara Kitabevi.
- JAIN Anil K., Richard C. DUBES, 1988, *Algorithms for Clustering Data*, New Jersey: Prentice-Hall, Inc.
- JAJUGA K., M. WALESIAK, 2000, “Standardisation of data set under different measurement scales”, *Classification and Information Processing at the Turn of the Millennium, 23rd Annual Conference of the Gesellschaft für Klassifikation e.Y.*, ed. Reinhold Decker, Wolfgang Gaul, 2000, ss. 105-12, doi:10.1007/978-3-642-57280-7.
- JÄRVISALO Susanna, 2012, *How to Build Successful City Brands?-Case Munich, Berlin & Hamburg*, HAAGA-HELIA University of Applied Sciences.
- JI Jinchao vd., 2012, “A fuzzy k-prototype clustering algorithm for mixed numeric and categorical data”, *Knowledge-Based Systems*, C. 30, S. 2012, Elsevier B.V., ss. 129-35, doi:10.1016/j.knosys.2012.01.006.
- , 2015, “An initialization method for clustering mixed numeric and categorical

- data based on the density and distance”, *International Journal of Pattern Recognition and Artificial Intelligence*, C. 29, S. 7, ss. 1-16, doi:10.1142/S021800141550024X.
- JINYIN Chen vd., 2017, “A novel cluster center fast determination clustering algorithm”, *Applied Soft Computing*, C. 57, S. August, ss. 539-55, doi:10.1016/j.asoc.2017.04.031.
- JOHANSON George A., Gordon P. BROOKS, 2010, “Initial scale development: Sample size for pilot studies”, *Educational and Psychological Measurement*, doi:10.1177/0013164409355692.
- JONG Martin DE vd., 2018, “Explaining city branding practices in China’s three mega-city regions: The role of ecological modernization”, *Journal of Cleaner Production*, C. 179, S. (2018), Elsevier, ss. 527-43, doi:10.1016/J.JCLEPRO.2018.01.098.
- JOSHI Ameet V., 2020, *Machine Learning and Artificial Intelligence*, Springer.
- KAISER Florian G., Mark WILSON, 2004, “Goal-directed conservation behavior: The specific composition of a general performance”, *Personality and Individual Differences*, C. 36, S. 7, ss. 1531-44, doi:10.1016/j.paid.2003.06.003.
- KALAYCI Şeref, 2010, *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*, 5. b., Ankara: Asil Yayın Dağıtım.
- KARAKURT TOSUN Elif, 2009, “Sürdürülebilirlik olgusu ve kentsel yapıya etkileri”, *PARADOKS, Ekonomi, Sosyoloji ve Politika Dergisi*, C. 5, S. 2.
- KARDES Frank R., Maria L. CRONLEY, Thomas W. CLINE, 2011, *Consumer Behavior*, Mason, OH: South-Western, Cengage Learning.
- KARVOUNIS Alexandros, 2015, “Urban metabolism”, *Understanding Urban Metabolism: A Tool for Urban Metabolism*, ed. Nektarios Chrysoulakis, Eduardo Anselmo de Castro, Eddy J. Moors, Abingdon, Oxon: Routledge, ss. 3-12, doi:10.4324/9781315765846.
- KASSAMBARA Alboukadel, 2015, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*, 1st b., STHDA.
- , 2017, *Principal Component Methods in R: Practical Guide*, STHDA.
- KAUFMAN Leonard, Peter J. ROUSSEEUW, 2005, *Finding Groups in Data*. Second, New Jersey: Wiley Interscience.
- KAVARATZIS Mihalis, 2009, “Cities and their brands: Lessons from corporate branding”, *Place Branding and Public Diplomacy*, C. 5, S. 1, ss. 26-37, doi:10.1057/pb.2008.3.
- KAWAMOTO Tatsuro, Yoshiyuki KABASHIMA, 2017, “Cross-validation estimate of the number of clusters in a network”, *Scientific Reports*, C. 7, S. 1, Nature Publishing Group, doi:10.1038/s41598-017-03623-x.
- KELLERT Stephen R., 2005, *Building for Life: Designing and Understanding the Human-Nature Connection*, Washington, DC: Island Press.
- KHALID Samina, Tehmina KHALIL, Shamila NASREEN, 2014, “A survey of feature

- selection and feature extraction techniques in machine learning”, *Proceedings of Science and Information Conference, SAI 2014*, 2014, ss. 372-78, doi:10.1109/SAI.2014.6918213.
- KHANCHOUCHI., M. CHARRAD, M. LIMAM, 2015, “A Comparative Study of Multi-SOM Algorithms for Determining the Optimal Number of Clusters”, *International Journal of Future Computer and Communication*, C. 4, S. 3, ss. 198-202, doi:10.7763/ijfcc.2015.v4.384.
- KHANNA Nina Zheng, Jin GUO, Xinye ZHENG, 2016, “Effects of demand side management on Chinese household electricity consumption: Empirical findings from Chinese household survey”, *Energy Policy*, doi:10.1016/j.enpol.2016.04.049.
- KIANG Melody Y., 2003, “A comparative assessment of classification methods”, *Decision Support Systems*, C. 35, S. 4, ss. 441-54, doi:10.1016/S0167-9236(02)00110-0.
- KIPPING A., E. TRØMBORG, 2016, “Modeling and disaggregating hourly electricity consumption in Norwegian dwellings based on smart meter data”, *Energy and Buildings*, doi:10.1016/j.enbuild.2016.02.042.
- KIRCHNER Kathrin, Jelena ZEC, Boris DELIBAŠIĆ, 2016, “Facilitating data preprocessing by a generic framework: a proposal for clustering”, *Artificial Intelligence Review*, doi:10.1007/s10462-015-9446-6.
- KLINER, Rex B., 2011, *Principles and Practice of Structural Equation Modeling*. The Third, New York: The Guilford Press.
- KLINER Paul, 1994, *An Easy Guide to Factor Analysis*, Oxon: Routledge.
- KOÇ Erdoğan, 2016, “Benlik ve kişilik”, *Tüketici Davranışı ve Pazarlama Stratejileri: Global ve Yerel Yaklaşım*, Yedinci, Ankara: Seçkin Yayıncılık, s. 535.
- KOÇAK Duygu, Ömay ÇOKLUK, Murat KAYRI, 2016, “Faktör sayısının belirlenmesinde MAP testi, paralel analiz, K1 ve yamaç birikinti grafiği yöntemlerinin karşılaştırılması”, *YYÜ Eğitim Fakültesi Dergisi*, C. XIII, S. I, ss. 330-59.
- KOHONEN Teuvo, 1989, *Self-Organization and Associative Memory*, 3. b., Springer.
- , 2001, *Self-Organizing Maps*, 3rd b., Springer.
- , 2013, “Essentials of the self-organizing map”, *Neural Networks*, C. 37, ss. 52-65, doi:10.1016/j.neunet.2012.09.018.
- KORKMAZ Selçuk, Gökmen GÖKSÜLÜK, DİNÇER, ZARARSIZ, 2014, “MVN: An R package for assessing multivariate normality”, *The R Journal*, C. 6, S. 2, ss. 151-62, <https://journal.r-project.org/archive/2014-2/korkmaz-goksuluk-zararsiz.pdf>.
- KÖSE İlker, 2018, *Veri Madenciliği-Teori, Uygulama ve Felsefesi*, ed. Gökhan Silahtaroğlu. Birinci Ba, İstanbul: Papatya Yayıncılık Eğitim.
- KOTLER Philip vd., 1999, *Marketing Places Europe: How to Attract Investments, Industries, Residents and Visitors to Cities, Communities, Regions, and Nations in Europe*, Financial Times.



- LAM Dao, Mingzhen WEI, Donald WUNSCH, 2015, "Clustering data of mixed categorical and numerical type with unsupervised feature learning", *IEEE Access*, C. 3, S. 2015, ss. 1605-13, doi:10.1109/ACCESS.2015.2477216 Clustering.
- LEBBAH Mustapha, Khalid BENABDESLEM, 2010, "Visualization and clustering of categorical data with probabilistic self-organizing map", *Neural Computing and Applications*, C. 19, S. 3, ss. 393-404, doi:10.1007/s00521-009-0299-2.
- LEE Gyeong Suk, Young Hee YOM, 2013, "Structural equation modeling on life-world integration in people with severe burns", *Asian Nursing Research*, C. 7, S. 3, Elsevier, ss. 112-19, doi:10.1016/j.anr.2013.06.003.
- LEISCH Friedrich, 2016, "Resampling methods for exploring cluster stability", *Handbook of Cluster Analysis*, ed. Christian Hennig vd., Boca Raton, FL: CRC Press, ss. 637-52.
- LEYS Christopher vd., 2018, "Detecting multivariate outliers: Use a robust variant of the Mahalanobis distance", *Journal of Experimental Social Psychology*, C. 74, ss. 150-56, doi:10.1016/j.jesp.2017.09.011.
- LIANG Jiye vd., 2012, "Determining the number of clusters using information entropy for mixed data", *Pattern Recognition*, C. 45, S. 6, ss. 2251-65, doi:10.1016/j.patcog.2011.12.017.
- LIM Jongwoo vd., 2012, "A framework for clustering mixed attribute type datasets", *Proceeding of the fourth International Conference on Emerging Databases -EDB 2012*, 2012, ss. 1-12.
- LIU Hongyan vd., 2011, "Advances in Data Mining Applications and Theoretical Aspects 12th Industrial Conference, ICDM", *Electricity consumption time series profiling: A data mining application in energy industry*, ed. Randy Goebel, Yuzuru Tanaka, Wolfgang Wahlster, 2011, ss. 52-66, doi:10.1007/978-3-642-34182-3.
- LIU Huan, Hiroshi MOTODA, 1998, *Feature Extraction, Construction and Selection: A Data Mining Perspective*, New York: Springer.
- LONGHI Simonetta, 2015, "Residential energy expenditures and the relevance of changes in household circumstances", *Energy Economics*, doi:10.1016/j.eneco.2015.03.018.
- LOUV Richard, 2008, *Last Child in the Woods\_ Saving Our Children From Nature-Deficit Disorder*, North Carolina: Algonquin Books.
- LU Shen, Richard S. SEGALL, 2013, "Multi-SOM: an algorithm for high-dimensional, small size datasets", *Journal of Systemics, Cybernetics and Informatics*, C. 11, S. 2, ss. 41-46.
- MAESSCHALCK R. DE, D. JOUAN-RIMBAUD, D. L. MASSART, 2000, "The mahalanobis distance", *Chemometrics and Intelligent Laboratory Systems*, C. 50, S. 1, ss. 1-18, doi:10.1016/S0169-7439(99)00047-7.
- MAHALANOBIS Prasanta Chandra, 1936, "On the generalized distance in statistics", *Proceedings of the National Institute of Sciences of India*, C. 2, S. 1, ss. 49-55.
- MANNING Christie, 2009, "The Psychology of Sustainable Behavior: Tips for

Empowering People to Take Environmentally Positive Action”, St. Paul, MN, <http://www.pca.state.mn.us/index.php/view-document.html?gid=12949%5Cnhttp://www.pca.state.mn.us/index.php/topics/preventing-waste-and-pollution/sustainability/sustainable-communities/psychology-of-sustainable-behavior.html>.

- MARIN-AGUILAR Jose, Natalia VILA-LÓPEZ, 2014, “How can mega events and ecological orientation improve city brand attitudes?”, *International Journal of Contemporary Hospitality Management*, C. 26, S. 4, ss. 629-52, doi:10.1108/IJCHM-03-2013-0117.
- MATSUSHITA Haruna, Yoshifumi NISHIO, 2010, “Batch-learning self-organizing map with weighted connections avoiding false-neighbor effects”, *IEEE World Congress on Computational Intelligence, WCCI 2010*, 2010, ss. 1075-80.
- MAYER F Stephan, Cynthia MCPHERSON FRANTZ, 2004, “The connectedness to nature scale : A measure of individuals ’ feeling in community with nature \$”, *Journal of Environmental Psychology*, C. 24, S. 2004, ss. 503-15, doi:10.1016/j.jenvp.2004.10.001.
- MCARDLE Brian H., Marti J. ANDERSON, 2001, “Fitting multivariate models to community data : A comment on distance-based redundancy analysis”, *Ecological Society of America*, C. 82, S. 1, ss. 290-97.
- MCCUNE Bruce, James B. GRACE, 2002, “Distance measures”, *Analysis of Ecological Communities*, Gleneden Beach, Oregon: MJM Software Design, ss. 45-58.
- MCIVER John P., 2004, “Ratio scale”, *Encyclopedia of Social Science Research Methods*, SAGE Publications Ltd, ss. 925-26.
- MEILA Marina, 2005, “Comparing Clusterings-An Axiomatic View”, *the 22nd International Conference on Machine Learning*, 2005, ss. 577-84.
- , 2016, “Criteria for comparing clusterings”, *Handbook of Cluster Analysis*, ed. Christian Hennig vd., Boca Raton, FL: CRC Press, ss. 619-35.
- MERCHANT Carolyn, 2005, *Radical Ecology: The Search for a Livable World*, 2nd b., Abingdon Oxon: Routledge.
- MICHELI-TZANAKOU Evangelia, 2000, *Supervised and Unsupervised Pattern Recognition: Feature Extraction and Computational Intelligence*, Florida: CRC.
- MICHIE D. vd., 1994, *Machine learning, neural and statistical classification*, *Journal of the American Statistical Association*, Ellis Horwood.
- MILJKOVIC Dubravko, 2017, “Brief review of self-organizing maps”, *40th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2017*, 2017, ss. 1061-66, doi:10.23919/MIPRO.2017.7973581.
- MOILANEN Teemu, Seppo RAINISTO, 2009, *How to Brand Nations, Cities and Destinations*, *How to Brand Nations, Cities and Destinations*, Trowbridge, Wiltshire: PALGRAVE MACMILLAN.
- MORGAN Nigel, Annette PRITCHARD, 2004, “Meeting the destination branding

- challenge”, *Destination Branding: Creating the Unique Destination Proposition*, ed. Nigel Morgan, Annette Pritchard, Roger Pride, 2nd b., Burlington, MA: Elsevier Butterworth-Heinemann, ss. 59-78.
- MOUSTAKI Iriini, Ioulia PAPAGEORGIU, 2005, “Latent class models for mixed variables with applications in Archaeometry”, *Computational Statistics and Data Analysis*, C. 48, S. 3, ss. 659-75, doi:10.1016/j.csda.2004.03.001.
- MUCHA Hans-joachim, Hans-georg BARTEL, 2017, “Validation of k -means Clustering: Why is bootstrapping better than subsampling?”, *Archives of Data Science, Series A*, C. 2, S. 1, ss. 1-14, doi:10.5445/KSP/1000058749/27.
- MURTAGH Fionn, 2016, “A brief history of cluster analysis”, *Handbook of Cluster Analysis*, ed. Christian Hennig vd., Boca Raton, FL: CRC Press, ss. 21-32.
- NISBET Elizabeth K., John M. ZELENSKI, Steven A. MURPHY, 2009, “The nature relatedness scale: Linking individuals’ connection with nature to environmental concern and behavior”, *Environment and Behavior*, doi:10.1177/0013916508318748.
- NOORBEHBAHANI Fakhroddin vd., 2015, “An incremental mixed data clustering method using a new distance measure”, *Soft Computing*, C. 19, ss. 731-43, doi:10.1007/s00500-014-1296-7.
- NORDLUND Annika M., Jörgen GARVILL, 2002, “Value structures behind proenvironmental behavior”, *Environment and Behavior*, doi:10.1177/001391602237244.
- NOVA David, Pablo A. ESTEVEZ, 2012, “Online visualization of prototypes and receptive fields produced by LVQ algorithms”, *Advances in Self-Organizing Maps-9th International Workshop, WSOM 2012*, ed. Pablo A. Estevez, Jose C. Principe, Pablo Zegers, 2012, ss. 173-82, doi:10.1016/j.neunet.2006.05.011.
- O’SHAUGHNESSY John, Nicholas Jackson O’SHAUGHNESSY, 2003, *The Marketing Power of Emotion*, Journal of Macromarketing, NY: Oxford University Press.
- ÖNDER Hasan, 2005, *Permütasyon Testlerinin Doğrusal Modellerde Uygulanması ve Karşılaştırılması*, Çukurova Üniversitesi.
- ÖNDER Hasan, Zeynel CEBECİ, 2009, “Use and comparison of permutation tests in linear models”, *Anadolu J. Agric. Sci.*, C. 24, S. 2, ss. 93-97.
- OSKAMP Stuart, 2000, “A sustainable future for humanity? How can psychology help?”, *American Psychologist*, doi:10.1037/0003-066X.55.5.496.
- ÖZDAMAR Kazım, 2013, *Paket Programlar ile İstatistiksel Veri Analizi - Cilt 2*, 9. b., Eskişehir: Nisan Kitabevi.
- ÖZKAN Yalçın, 2008, *Veri Madenciliği Yöntemleri*, İstanbul: Papatya Yayıncılık Eğitim.
- ÖZTEMEL Ercan, 2012, *Yapay Sinir Ağları*, 3. b., İstanbul: Papatya Yayıncılık.
- PANAPAKIDIS Ioannis P. vd., 2014, “Pattern recognition algorithms for electricity load curve analysis of buildings”, *Energy and Buildings*, doi:10.1016/j.enbuild.2014.01.002.

- PASIN Özge vd., 2016, “Parametrik olmayan çok değişkenli varyans analizi ve sağlık alanında bir uygulaması”, *Bilişim Teknolojileri Dergisi*, C. 9, S. 1, ss. 13-20, doi:10.17671/btd.47787.
- PATHAK Arkanath, Nikhil R PAL, 2016, “Clustering of mixed data by integrating fuzzy, probabilistic and collaborative clustering framework”, *International Journal of Fuzzy Systems*, C. 18, S. 3, ss. 339-48, doi:10.1007/s40815-016-0168-y.
- PESCHIERA Gabriel, John E. TAYLOR, 2012, “The impact of peer network position on electricity consumption in building occupant networks utilizing energy feedback systems”, *Energy and Buildings*, doi:10.1016/j.enbuild.2012.03.011.
- POLANCO Xavier, Claire FRANÇOIS, Jean-Charles LAMIREL, 2001, “Using artificial neural networks for mapping of science and technology: a multi self-organizing maps approach”, *Scientometrics*, C. 51, S. 1, ss. 267–292.
- PÖLZLBAUER Georg, 2004, “Survey and comparison of quality measures for self-organizing maps”, *Workshop on Data Analysis, WDA2004*, 2004, ss. 67-82, <http://www.cis.hut.fi/projects/somtoolbox>.
- PRASAD D.Hari, M. PUNITHAVALLI, 2012, “An integrated framework for mixed data clustering using growing hierarchical self-organizing map (GHSOM)”, *International Conference, ICMMSC 2012*, ed. P. Balasubramaniam, R. Uthayakumar, 2012, ss. 471-79, doi:10.1007/978-3-642-28926-2.
- RAAIJ, FRED W. VAN, Theo M. M. VERHALLEN, 1983, “A behavioral modal of residential energy use”, *Journal of Economic Psychology*, C. 3, ss. 39-63.
- RAAIJ W. Fred VAN, Theo M M VERHALLEN, 1983, “Patterns of residential energy behavior”, *Journal of Economic Psychology*, C. 4, ss. 85-106.
- RÄSÄNEN Teemu, Juhani RUUSKANEN, Mikko KOLEHMAINEN, 2008, “Reducing energy consumption by using self-organizing maps to create more personalized electricity use information”, *Applied Energy*, doi:10.1016/j.apenergy.2007.10.012.
- RIBEIRO Caio Eduardo, Luis Enrique ZÁRATE, 2019, “Classifying longevity profiles through longitudinal data mining”, *Expert Systems with Applications*, C. 117, Mart, Elsevier Ltd, ss. 75-89, doi:10.1016/j.eswa.2018.09.035.
- RIGDON Edward E., Marko SARSTEDT, Christian M. RINGLE, 2017, “On comparing results from CB-SEM and PLS-SEM: Five perspectives and five recommendations”, *Marketing ZFP*, C. 39, S. 3, ss. 4-16, doi:10.15358/0344-1369-2017-3-4.
- ROGERS Richard, 1997, *Cities for a Small Planet*, ed. Philip Gumuchdjian, London: Faber and Faber Limited.
- ROGOVSCHI Nicoleta, Mustapha LEBBAH, Younès BENNANI, 2008, “Probabilistic mixed topological map for categorical and continuous data”, *7th International Conference on Machine Learning and Applications, ICMLA 2008*, 2008, ss. 224-31, doi:10.1109/ICMLA.2008.13.
- ROSENBERG Andrew, Julia HIRSCHBERG, 2007, “V-Measure: A conditional entropy-based external cluster evaluation measure”, *Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2007, ss. 410-20.

- ROUSSEEUW Peter, 1985, "Multivariate estimation with high breakdown", *Mathematical Statistics and Applications*, ed. W. Grossmann vd., Dordrecht, Holland: Reidel Publishing Company, ss. 283-97.
- ROUSSEEUW Peter J, 1984, "Least median of squares regression", *Journal of the American Statistical Association*, C. 79, S. 388, ss. 871-80.
- , 1987, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis", *Journal of Computational and Applied Mathematics*, C. 20.
- ROUSSEEUW Peter J, Katrien Van DRIESSEN, 1999, "A fast algorithm for the minimum covariance determinant estimator", *Technometrics*, C. 41, S. 3, ss. 212-23.
- SALA Antonello vd., 2019, "Trawling in the Mediterranean: An exploration of empirical relations connecting fishing gears, otterboards and propulsive characteristics of fishing vessels", *Frontiers in Marine Science*, C. 6, ss. 1-16, doi:10.3389/fmars.2019.00534.
- SAVAŞ Serkan, 2020, "Veri madenciliği süreci", *Medium*, (17.09.2020), <https://medium.com/veri-madenciliği/veri-madenciliği-süreci-11243e7966fc>.
- SCHMUCK Peter vd., 2005, "Values and their relationship to environmental concern and conservation behavior", *Journal of Cross-Cultural Psychology*, C. 36, S. 4, ss. 457-75, doi:10.1177/0022022105275962.
- SCHULTZ P. W., 1999, "Empathizing with nature: The effects of perspective taking on concern for environmental issues", *Journal of social issues*, C. 56, S. 3, ss. 391-406, doi:10.1111/0022-4537.00174.
- SCHULTZ P. Wesley, Lynnette C. ZELEZNY, 2007, "Values and Proenvironmental Behavior: A five country survey", *Journal of Cross-Cultural Psychology*, C. 29, S. 4, ss. 540-58, doi:10.1177/0022022198294003.
- SCHULTZ P Wesley, 2000, "Empathizing With Nature: The Effects of Perspective Taking on Concern for Environmental Education. *Journal of Social Issues*, 56(3), 391-406.", *Journal of Social Issues*, C. 56, S. 3, ss. 391-406.
- SCHULTZ Wesley P., 2001, "The structure of environmental concern: Concern for self, other people, and the biosphere", *Journal of Environmental Psychology*, C. 21, ss. 327-39, doi:10.1006/jevp.2001.0227.
- , 2002, "Inclusion with nature: The psychology of human-nature relations", *Psychology of Sustainable Development*, ed. Peter Schmuck, Wesley P. Schultz, Kulwer Academic Publishers, ss. 61-78, doi:10.1007/978-1-4615-0995-0.
- SCHULTZ Wesley P., Lynnette ZELEZNY, 1999, "Values as predictors of environmental attitudes: Evidence for consistency across 14 counties", *Journal of Environmental Psychology*, C. 19, S. 3, ss. 255-65, doi:10.1006/jevp.1999.0129.
- SCHWARTZ Shalom H., 2013, "An Overview of the Schwartz Theory of Basic Values", *Online Readings in Psychology and Culture*, C. 2, S. 1, ss. 1-20, doi:10.9707/2307-0919.1116.
- ŞEKER Şadi Evren, 2013, *İş Zekası ve Veri Madenciliği*, İstanbul: Cinius Yayınları.

- ŞENESEN Ümit, 2013, *Istatistik: Sayıların Arkasını Anlamak*, 4. b., İstanbul: Literatür.
- SENI Giovanni, John F. ELDER, 2010, “Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions”, *Synthesis Lectures on Data Mining and Knowledge Discovery*, doi:10.2200/s00240ed1v01y200912dmk002.
- ŞENTÜRK Aysan, 2006, *Veri Madenciliği-Kavram ve Teknikler*, Bursa: Ekin Basım Yayın Dağıtım.
- SERPER Özer, 2014, *Uygulamalı İstatistik*, 7. b., Bursa: Ezgi Kitabevi.
- SERPER Özer, Mustafa AYTAÇ, Nuran BAYRAM, 2016, *Örnekleme*. Dördüncü, Ankara: Pegem Yayıncılık.
- SHEN Furao, Osamu HASEGAWA, 2008, “A fast nearest neighbor classifier based on self-organizing incremental neural network”, *Neural Networks*, C. 21, S. 10, Elsevier Ltd, ss. 1537-47, doi:10.1016/j.neunet.2008.07.001.
- SHIH Ming-Yi, Wen JHENG, Lien-Fu LAI, 2010, “A two-step method for clustering mixed categorical and numeric data”, *Tamkang Journal of Science and Engineering*, C. 13, S. 1, ss. 11-19.
- SHIRKHORSHIDI Ali Seyed, Saeed AGHABOZORGI, Teh Ying WAH, 2015, “A comparison study on similarity and dissimilarity measures in clustering continuous data”, *PLoS ONE*, C. 10, S. 12, ss. 1-20, doi:10.1371/journal.pone.0144059.
- ŞİMŞEK GÜRSOY Umman Tuğba, 2012, *Uygulamalı Veri Madenciliği-Sektörel Analizler*. Üçüncü bas, Ankara: Pegem Yayıncılık.
- SOLORIO-FERNÁNDEZ Saúl, J. Ariel CARRASCO-OCHOA, José Fco MARTÍNEZ-TRINIDAD, 2019, “A review of unsupervised feature selection methods”, *Artificial Intelligence Review*, Springer Netherlands, 1-42, doi:10.1007/s10462-019-09682-y.
- STAŃCZYK Urszula, Lakhmi C JAIN, 2015, “An Introduction”, *Feature Selection for Data and Pattern Recognition*, ed. Urszula Stańczyk, Lakhmi C Jain, Springer, ss. 1-5, <http://www.springer.com/series/7092>.
- STEG Linda, Agnes E. van den BERG, Judith I.M. DE GROOT, 2013, “Environmental psychology: History, scope and methods”, *Environmental Psychology: An Introduction*, ed. Linda Steg, Agnes E. van den Berg, Judith I.M. de Groot, West Sussex: BPS Blackwell, ss. 1-11.
- STEG Linda, Annika NORDLUND, 2013, “Models to explain environmental behaviour”, *Environmental Psychology: An Introduction*, ed. Linda Steg, Agnes E. van den Berg, Judith I. M. de Groot, West Sussex: BPS Blackwell, ss. 185-96.
- STEG Linda, Charles VLEK, 2009, “Encouraging pro-environmental behaviour: An integrative review and research agenda”, *Journal of Environmental Psychology*, doi:10.1016/j.jenvp.2008.10.004.
- STERN Paul C, Thomas DIETZ, 1994, “The value basis of environmental psychology”, *Journal of Social Issues*, C. 50, S. 3, ss. 65-84.
- STEVENS Stanley Smith, 1946, “On the theory of scales of measurement”, *Science*, C. 103, S. 2684, ss. 677-80.

- STRANG Gilbert, 2016, *Introduction to Linear Algebra*. Fifth Edit, Wellesley-Cambridge Press.
- STRICKLAND Nicole, 2014, *Nature Relatedness as Predictor of Environmental Behaviour of Recreational Anglers in False Bay*, University of the Western Cape.
- SU Mu-Chun, Ta-Kang LIU, Hsiao-Te CHANG, 2002, "Improving the self-organizing feature map algorithm using an efficient initialization scheme", *Tamkang Journal of Science and Engineering*, C. 5, S. 1, ss. 35-48.
- SUPRIYAMENON M., P. RAJARAJESWARI, 2017, "A review on association rule mining techniques with respect to their privacy preserving capabilities", *International Journal of Applied Engineering Research*, C. 12, S. 24, ss. 15484-88.
- TABACHNICK Barbara G., Linda S. FIDELL, 2013, *Using Multivariate Statistics.*, *PsycCRITIQUES*, Pearson.
- TAI Wei Shen, Chung Chian HSU, 2012, "Growing self-organizing map with cross insert for mixed-type data clustering", *Applied Soft Computing Journal*, C. 12, S. 9, ss. 2856-66, doi:10.1016/j.asoc.2012.04.004.
- TAI Wei Shen, Chung Chian HSU, Jong Chen CHEN, 2010, "A mixed-type self-organizing map with a dynamic structure", *International Joint Conference on Neural Networks*, 2010, ss. 1-8, doi:10.1109/IJCNN.2010.5596823.
- TAM Kim Pong, 2013, "Concepts and measures related to connection to nature: Similarities and differences", *Journal of Environmental Psychology*, C. 34, Elsevier Ltd, ss. 64-78, doi:10.1016/j.jenvp.2013.01.004.
- TAN Pang-Ning, Michael STEINBACH, Vipin KUMAR, 2014, *Introduction to Data Mining*, Pearson.
- TEGHTSOONIAN R., 2001, "Stevens, Stanley Smith (1906-73)", *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier, ss. 15104-8, <http://web.mit.edu/epl/StevensBiography.pdf>.
- TEZBAŞARAN A. Ata, 2008, "Likert Tipi Ölçek Hazırlama Kılavuzu", *Likert Tipi Ölçek Hazırlama Kılavuzu Üçüncü Sürüm e-kitap*, 1-65, doi:10.1017/CBO9781107415324.004.
- THE EUROPEAN COMMISSION DIRECTORATE GENERAL FOR ENERGY, 2013, "World Energy Scenarios".
- THRUN Michael Christoph, 2018, *Projection-Based Clustering through Self-Organization and Swarm Intelligence*, SpringerVieweg.
- TIAN Jing, Michael H AZARIAN, Michael PECHT, 2014, "Anomaly detection using self-organizing maps-based K -nearest neighbor algorithm", *European Conference of the Prognostics and Health Management Society*, 2014, ss. 1-9.
- TIWARI Piyush, 2000, "Architectural, Demographic, and Economic Causes of Electricity Consumption in Bombay", *Journal of Policy Modeling*, doi:10.1016/S0161-8938(98)00003-9.
- TRUEMAN Myfanwy M., Nelarine CORNELIUS, Alison J. KILLINGBECK-WIDDUP, 2007, "Urban corridors and the lost city: Overcoming negative

- perceptions to reposition city brands”, *Brand Management*, C. 15, S. 1, ss. 20-31, doi:10.1057/palgrave.bm.2550107.
- TUNG Anthony T.H. vd., 2001, “Constraint-based clustering in large databases”, *8th International Conference on Database Theory – ICDT 2001*, ed. Jan Van den Bussche, Victor Vianu, 2001, ss. 405-19, doi:10.1108/ir.1999.04926fae.001.
- TURE Mevlut, Imran KURT OMURLU, 2018, “Determining of complexity parameter for recursive partitioning trees by simulation of survival data and an application on breast cancer data”, *Journal of Statistics and Management Systems*, C. 21, S. 1, Informa UK Limited, ss. 125-38, doi:10.1080/09720510.2017.1386878.
- TZOULAS Konstantinos vd., 2007, “Promoting ecosystem and human health in urban areas using Green Infrastructure: A literature review”, *Landscape and Urban Planning*, C. 81, S. 3, Elsevier, ss. 167-78, doi:10.1016/J.LANDURBPLAN.2007.02.001.
- UN, 2018, “68% of the world population projected to live in urban areas by 2050, says UN”, <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html>.
- UN HABITAT, 2018, “Energy”, <https://unhabitat.org/urban-themes/energy/>.
- URIARTE E.Arsuaga, F. Diaz MARTIN, 2008, “Topology preservation in SOM”, *International Journal of Computer and Information Engineering*, C. 2, S. 9, ss. 3192-95.
- VERMA Inder Swaroop, 2015, “Knowledge data discovery and its issues”, *10th Biyani International Conference (BICON-15)*, 2015, ss. 88-91, [http://biyaniconference.com/proceedings-2015/IT-BOOK-\(Bicon-2015\).pdf#page=88](http://biyaniconference.com/proceedings-2015/IT-BOOK-(Bicon-2015).pdf#page=88).
- VIRGO Ben, Leslie DE CHERNATONY, 2006, “Delphic brand visioning to align stakeholder buy-in to the City of Birmingham brand”, *Journal of Brand Management*, C. 13, S. 6, ss. 379-92, doi:10.1057/palgrave.bm.2540280.
- WALESIAK Marek, Andrzej DUDEK, 2019, “Types of variable normalization”, R Studio, <http://www.benetz Korn.com/2011/11/data-normalization-and-standardization/> (1.06.2014).
- WANG Richard Y., Veda C. STOREY, Christopher P. FIRTH, 1995, “A Framework for analysis of data quality research”, *IEEE Transactions on Knowledge and Data Engineering*, C. 7, S. 4, ss. 623-40, doi:10.1109/69.404034.
- WASKE Björn, Jón Atli BENEDIKTSSON, 2014, “PATTERN RECOGNITION AND CLASSIFICATION”, *Encyclopedia of Remote Sensing*, E.G. Njoku, New York: Springer Science+Business Media, ss. 503-9, doi:10.1007/978-0-387-36699-9.
- WEHRENS Ron, Johannes KRUISSELBRINK, 2018a, “Flexible self-organizing maps in kohonen 3.0”, *Journal of Statistical Software*, C. 87, S. 7, ss. 1-18, doi:10.18637/jss.v087.i07.
- , 2018b, “Package ‘kohonen’”, R Studio.
- WIEDENHOFER Dominik, 2011, “Spatial and Socio-economic Drivers of Direct and



- Indirect Household Energy Consumption in Australia”, , 133, Social Ecology, Vienna, www.aau.at/socec.
- WILMINK F.W., H.T. UYTTERSCHAUT, 1984, “Cluster analysis, history, theory and applications”, *Multivariate Statistical Methods in Physical Anthropology*, ed. G.N. Van Vark, W.W. Howells, Dordrecht, Holland: D. Reidel Publishing Company, ss. 135-76.
- WILSON Edward O., 1993, “Biophilia and the conservation ethic”, *The Biophilia Hypothesis*, ed. Stephen R. Kellert, Edward O. Wilson, Washington D.C., California, s. .
- XIE Qing, Hongbing OUYANG, Xiang GAO, 2016, “Estimation of electricity demand in the residential buildings of China based on household survey data”, *International Journal of Hydrogen Energy*, doi:10.1016/j.ijhydene.2016.03.152.
- XU Dongkuan, Yingjie TIAN, 2015, “A Comprehensive Survey of Clustering Algorithms”, *Annals of Data Science*, C. 2, S. 2, ss. 165-93, doi:10.1007/s40745-015-0040-1.
- YAMANE Taro, 2009, *Temel Örneklem Yöntemleri*, ed. Alptekin Esin vd., Ankara: Literatür.
- YIN Hujun, 2002, “ViSOM-a novel method for multivariate data projection and structure visualization”, *IEEE Transactions on Neural Networks*, C. 13, S. 1, ss. 237-43, doi:10.1109/72.977314.
- YOON Jiyoung, 2010, *A Conceptual Model for City Branding Based on Semiotics*, Brunel University.
- YOUNG William vd., 2010, “Sustainable consumption: Green consumer behaviour when purchasing products”, *Sustainable Development*, doi:10.1002/sd.394.
- ZHANG Cha, Yunqian MA, 2012, *Ensemble machine learning: Methods and applications*, *Ensemble Machine Learning: Methods and Applications*.
- ZHANG Chengqi, Shichao ZHANG, 2002, *Association Rule Mining: Models and Algorithms*, Springer.
- ZHANG Tao, Peer Olaf SIEBERS, Uwe AICKELIN, 2012, “A three-dimensional model of residential energy consumer archetypes for local energy policy design in the UK”, *Energy Policy*, doi:10.1016/j.enpol.2012.04.027.
- ZHANG Tian, Raghu RAMAKRISHNAN, Miron LIVNY, 1997, “BIRCH: A new data clustering algorithm and its applications”, *Data Mining and Knowledge Discovery*, C. 1997, S. 1, ss. 141-82.
- ZHOU Zhi Hua, 2012, *Ensemble methods: Foundations and algorithms*.

## EKLER

### Ek 1

$x_{ij}(x_{ij}')$   $i$ 'inci gözlemin  $j$ 'inci öznelik için ölçüm değerini (normalize edilmiş  $x_{ij}$ 'yi),  $\bar{x}_j(\sigma_j)$   $j$ 'inci özneliğe ait tüm ölçüm değerlerinin ortalamasını (standart sapmasını),  $med_j$   $j$ 'inci özneliğe ait tüm ölçüm değerlerinin medyanını,  $mad_j$   $j$ 'inci özneliğe ait tüm ölçüm değerlerinin medyandan mutlak sapmasını,  $r_j$   $j$ 'inci özneliğin ranjını/genişliğini ( $\max_i\{x_{ij}\} - \min_i\{x_{ij}\}$ ),  $m_j$   $j$ 'inci özneliğin ranjının orta değerini ( $\frac{\max_i\{x_{ij}\} + \min_i\{x_{ij}\}}{2}$ ) temsil etmek üzere [cluster.Sim](#) kütüphanesindeki `data.Normalization()` fonksiyonunda yer alan veri normalleştirme/ölçeklendirme yöntemleri aşağıdaki gibidir (Jajuga, Walesiak, 2000; Walesiak, Dudek, 2019).

**Tablo 56.** Veri normalleştirme yöntemleri

İsim	Formülasyon	Dönüşüm
n1: Standardization	$x_{ij}' = \frac{x_{ij} - \bar{x}_j}{\sigma_j}$	Oranlı → Eşit aralıklı Eşit aralıklı → Eşit aralıklı
n2: Positional standardization	$x_{ij}' = \frac{x_{ij} - med_j}{mad_j}$	Oranlı → Eşit aralıklı Eşit aralıklı → Eşit aralıklı
n3: Unitization	$x_{ij}' = \frac{x_{ij} - \bar{x}_j}{r_j}$	Oranlı → Eşit aralıklı Eşit aralıklı → Eşit aralıklı
n3a: Positional unitization	$x_{ij}' = \frac{x_{ij} - med_j}{r_j}$	Oranlı → Eşit aralıklı Eşit aralıklı → Eşit aralıklı
n4: Unitization with zero minimum	$x_{ij}' = \frac{x_{ij} - \min_i\{x_{ij}\}}{r_j}$	Oranlı → Eşit aralıklı Eşit aralıklı → Eşit aralıklı
n5: Normalization in range [-1, 1]	$x_{ij}' = \frac{x_{ij} - \bar{x}_j}{\max_i  x_{ij} - \bar{x}_j }$	Oranlı → Eşit aralıklı Eşit aralıklı → Eşit aralıklı
n5a: Positional normalization in range [-1, 1]	$x_{ij}' = \frac{x_{ij} - med_j}{\max_i  x_{ij} - med_j }$	Oranlı → Eşit aralıklı Eşit aralıklı → Eşit aralıklı
n6: Quotient transformation	$x_{ij}' = \frac{x_{ij}}{\sigma_j}$	Oranlı → Oranlı
n6a: Positional quotient transformation	$x_{ij}' = \frac{x_{ij}}{mad_j}$	Oranlı → Oranlı
n7: Quotient transformation	$x_{ij}' = \frac{x_{ij}}{r_j}$	Oranlı → Oranlı
n8: Quotient transformation	$x_{ij}' = \frac{x_{ij}}{\max_i\{x_{ij}\}}$	Oranlı → Oranlı
n9: Quotient transformation	$x_{ij}' = \frac{x_{ij}}{\bar{x}_j}$	Oranlı → Oranlı

n9a: Positional quotient transformation	$x_{ij}' = \frac{x_{ij}}{med_j}$	Oranlı → Oranlı
n10: Quotient transformation	$x_{ij}' = \frac{x_{ij}}{\sum_{i=1}^n x_{ij}}$	Oranlı → Oranlı
n11: Quotient transformation	$x_{ij}' = \frac{x_{ij}}{\sqrt{\sum_{i=1}^n x_{ij}^2}}$	Oranlı → Oranlı
n12: Normalizationon	$x_{ij}' = \frac{x_{ij} - \bar{x}_j}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}}$	Oranlı → Eşit aralıklı Eşit aralıklı → Eşit aralıklı
n12a: Positional normalization	$x_{ij}' = \frac{x_{ij} - med_j}{\sqrt{\sum_{i=1}^n (x_{ij} - med_j)^2}}$	Oranlı → Eşit aralıklı Eşit aralıklı → Eşit aralıklı
n13: Normalization with zero being the central point	$x_{ij}' = \frac{x_{ij} - m_j}{r_j/2}$	Oranlı → Eşit aralıklı Eşit aralıklı → Eşit aralıklı

## Ek 2

### Soru formu

#### ÇEVRE DOSTU DAVRANIŞLARIN TESPİTİ ANKETİ: BURSA KENT MERKEZİ ARAŞTIRMASI

Değerli Katılımcı,

Bu anket çalışmasının amacı; Bursa'nın Osmangazi, Yıldırım ve Nilüfer ilçelerinin merkez mahallelerinde yaşayan vatandaşların çevre dostu tüketim davranışlarını tespit etmektir. Ankette bulunan sorulara vereceğiniz cevaplar tarafımızca tamamen saklı tutulup, bilimsel amaçlı kullanılacaktır.

Anket iki bölümden oluşmaktadır. Birinci bölüm size, birlikte yaşadığınız hane halkına ve meskene ait bilgilerin elde edilmesi amacıyla hazırlanan sorulardan meydana gelmektedir. İkinci bölüm ise sizin çevre / doğa hakkındaki hisleriniz ve düşünceleriniz ile çevre dostu tüketim davranışlarınızı ve alışkanlıklarınızı belirlemek amacıyla hazırlanan sorulardan oluşmaktadır.

Anket sonuçlarının sağlıklı olabilmesi için soruları samimi ve doğru bir şekilde yanıtlamanız önemlidir.

Söz konusu ankete ait etik kurul izni Bursa Uludağ Üniversitesi Sosyal ve Beşeri Bilimler Araştırma ve Yayın Etik Kurulu'ndan alınmış olup, anket yaklaşık 25 dakika sürecektir.

İlgi ve yardımlarınız için şimdiden teşekkür ederiz.

Doç.Dr. Gül EMEL  
Arş.Gör. Gülcan PETRİÇLİ

#### BİRİNCİ BÖLÜM

Katılımcının ikamet ettiği ilçe:

#### KATILIMCI, HANE HALKI ve MESKEN BİLGİLERİ

Cinsiyet	<input type="checkbox"/> Kadın	<input type="checkbox"/> Erkek	
Yaş	<input type="checkbox"/> 18-24	<input type="checkbox"/> 25-34	<input type="checkbox"/> 35-44
	<input type="checkbox"/> 45-54	<input type="checkbox"/> 55-64	<input type="checkbox"/> 65-74
	<input type="checkbox"/> 75 ve üstü		
Eğitim durumunuz nedir?	<input type="checkbox"/> İlkokul	<input type="checkbox"/> Ortaokul	<input type="checkbox"/> Lise
	<input type="checkbox"/> Ön lisans	<input type="checkbox"/> Lisans	<input type="checkbox"/> Lisansüstü
Çalışma durumunuzu hangisi en iyi şekilde temsil etmektedir?	<input type="checkbox"/> Tam zamanlı çalışan (45 saat)		
	<input type="checkbox"/> Yarı zamanlı çalışan		
	<input type="checkbox"/> Emekli		
	<input type="checkbox"/> Üniversite öğrencisi		
	<input type="checkbox"/> Öğrenci		
	<input type="checkbox"/> İşsiz (iş arıyor)		
Evinizde, bakıma muhtaç (hasta, engelli, yaşlı) kaç birey bulunmaktadır?	<input type="checkbox"/> Çalışmıyor (Ev hanımı veya yetişkin ve sağlıklı fakat çalışmayan birey)		
Siz dâhil olmak üzere, hangi yaş gruplarından kaç kişi evinizde yaşamaktadır?	Yaş grubu	Kişi sayısı	
	0-4		
	5-11		
	12-15		
	16-19		
	18-64		
	65-74		
75+			

Hane halkını hangisi en iyi şekilde tanımlamaktadır? <b>(Genç: 15-29, Orta: 30-64, Yaşlı: 65 ve üstü)</b>	<input type="checkbox"/> Tek başına yaşayan genç <input type="checkbox"/> Çocuksuz evli genç çift <input type="checkbox"/> Çocuklu evli genç çift <input type="checkbox"/> Tek başına yaşayan çocuklu genç ebeveyn <input type="checkbox"/> Tek başına yaşayan orta yaşlı birey <input type="checkbox"/> Çocuklu evli orta yaşlı çift <input type="checkbox"/> Çocuksuz evli orta yaşlı çift <input type="checkbox"/> Tek başına yaşayan çocuklu orta yaşlı ebeveyn <input type="checkbox"/> Tek başına yaşayan yaşlı birey <input type="checkbox"/> Yaşlı çift <input type="checkbox"/> Ev arkadaşı <input type="checkbox"/> Geniş aile (Karı koca, çocuklardan oluşan çekirdek aileye ek olarak, büyük baba, büyük anne, hala, teyze, dayı gibi akrabaların veya var ise evli çocukların ve eşlerinin de oluşturduğu aile çeşidi) <input type="checkbox"/> Diğer ise aşağıda belirtiniz
Evinizin uyku dışı günlük kullanım süresi yaklaşık kaç saattir?	(.....) saat veya <input type="checkbox"/> Ev tüm gün kullanılmakta
Hanenin aylık toplam geliri yaklaşık ne kadardır?	
Yaşadığınız meskenin türü nedir?	<input type="checkbox"/> Apartman dairesi <input type="checkbox"/> Müstakil ev
Evinizde kullanılan toplam oda sayısı nedir (salon ve mutfak dahil)?	
Son bir yıl içinde aylık ödediğiniz maksimum elektrik faturası tutarı yaklaşık ne kadardır?	
Son bir yıl içinde aylık ödediğiniz maksimum su faturası tutarı yaklaşık ne kadardır?	
Son bir yıl içinde aylık ödediğiniz maksimum doğal gaz faturası tutarı yaklaşık ne kadardır?	

İKİNCİ BÖLÜM								
Doğaya verilen zarar, çeşitli çevre sorunlarına neden olmaktadır. Meydana gelen çevre sorunları ise tüm canlıları olumsuz etkilemektedir. İnsanlar çevre sorunları ve bu sorunların canlılar üzerindeki zararlı etkileri hakkında endişe duymaktadır. Fakat herkesin hakkında endişe duyduğu varlık farklı olabilmektedir. Lütfen aşağıda yer alan varlıklar hakkında ne kadar endişe duyduğunuzu, <b>hiç endişe duymuyorsanız 1 puan, son derece endişeliyseniz 7 puan</b> olmak üzere 1 ile 7 arasında değerlendirin.								
ÇEVRESEL ENDİŞE-1								
NO	( ... ) üzerindeki olumsuz etkilerinden dolayı çevresel sorunlardan endişe duyuyorum.	1	2	3	4	5	6	7
BT 1	Bitkiler							
BT 2	Deniz yaşamı							
BT 3	Kuşlar							
BT 4	Hayvanlar							
ÇEVRESEL ENDİŞE-2								
ET 1	Kendim							
ET 2	Kendi yaşam tarzım							
ET 3	Kendi sağlığım							
ET 4	Kendi geleceğim							
ÇEVRESEL ENDİŞE-3								
AT 1	Türkiye'de yaşayan insanlar							
AT 2	Tüm insanlar							
AT 3	Dünyadaki tüm çocuklar							
AT 4	Kendi / gelecekteki çocuklarım							

Bu bölümde çevre/doğa hakkındaki duygularınızı, düşüncelerinizi ve alışkanlıklarınızı öğrenmek amacıyla çeşitli ifadeler verilmiştir. Lütfen bu ifadeleri kendi düşünceleriniz doğrultusunda; **1-Kesinlikle Katılmıyorum, 2-Katılmıyorum, 3-Kararsızım, 4-Katılıyorum, 5-Kesinlikle Katılıyorum**, seçeneklerinden sadece birini seçerek cevaplandırınız.

<b>DOĞAYLA İLİŞKİ-1</b>						
NO	SORULAR	1.Kesinlikle Katılmıyorum	2.Katılmıyorum	3.Kararsızım	4.Katılıyorum	5.Kesinlikle Katılıyorum
Ö 1	Davranışlarımın çevreyi nasıl etkilediğini düşünürüm.					
Ö 2	Doğaya ve çevreye bağlılığım ruhumun bir parçasıdır.					
Ö 3	Çevre sorunlarının farkındayım.					
Ö 4	Kendimi doğanın bir parçası olarak görüyorum.					
Ö 5	Doğa ile ilgili hislerim yaşam tarzımı etkilemez.					
Ö 6	Şehrin ortasında bile etrafımdaki doğayı fark ederim.					
Ö 7	Doğa ile ilişkim benliğimin önemli bir parçasıdır.					
Ö 8	Dünyaya ve canlıların tümüne oldukça bağlıyım.					
<b>DOĞAYLA İLİŞKİ-2</b>						
P 1	Bazı türlerin nesli devam etmese de olur.					
P 2	İnsanların doğal kaynakları istedikleri gibi kullanma hakkı vardır.					
P 3	Çevre konusunda ne yaparsam yapayım dünyanın öteki yerlerindeki problemlere çözüm olmayacaktır.					
P 4	Hayvanlar ve bitkiler insanlardan daha az haklara sahip olmalıdır.					
P 5	Doğayı koruma çalışmaları gereksizdir, çünkü doğa insanların yol açtığı sorunlara karşı kendini yenileyecek kadar güçlüdür.					
P 6	İnsan dışındaki canlıların durumu, insanoğlunun geleceğinin bir göstergesidir.					
P 7	Hayvanların çektiği acıları umursarırım.					
<b>DOĞAYLA İLİŞKİ-3</b>						
D 1	Yağışlı/karlı/rüzgarlı gibi kötü havada bile dışarıda olmayı severim.					
D 2	İdeal tatil yerim uzak, el değmemiş bir doğa alanıdır.					
D 3	Toprakla uğraşmaktan ve ellerimi kirletmekten hoşlanırım.					
D 4	Nerede olursam olayım yabani hayat ilgimi çeker.					
D 5	Doğal alanlara sık gitmem.					
D 6	Medeniyetten uzak, ormanın derinliklerinde olma düşüncesi beni korkutur.					

Bu bölümde geri dönüşüm, ulaşım, enerji, su ve gıda tüketimi alışkanlıklarınızı ve çevreye duyarlı vatandaş olma eğiliminizi öğrenmek amacıyla çeşitli ifadeler verilmiştir. Lütfen bu ifadeleri kendi düşünceleriniz doğrultusunda; **1-Hiçbir zaman , 2-Nadiren, 3-Arasıra, 4-Sıklıkla, 5-Her zaman**, seçeneklerinden sadece birini seçerek cevaplandırınız.

<b>ÇEVRESEL DAVRANIŞ-ATIK OLUŞTURMA ve GERİ DÖNÜŞÜM</b>						
NO	SORULAR	1.Hiçbir zaman	2.Nadiren	3.Arasıra	4.Sıklıkla	5.Her zaman
GD 1	Plastik, metal, cam ve kâğıt atıkları geri dönüşüm için ayırırım.					
GD 2	Evsel atıkları azaltmak için çaba sarf ederim.					
GD 3	Alışverişte geri dönüşümlü ve çevre etiketli ambalajlarda satılan ürünleri tercih ederim.					

GD 4	Gıda atıklarını azaltmak için çabuk bozulan gıda ürünlerini az miktarlarda satın alırım.					
GD 5	Çevreye daha az zarar verdiğini bildiğim ürünleri satın almak için daha fazla para ederim.					
GD 6	Eşyalarını kullanılabilecek durumda oldukları sürece değiştirmem.					
GD 7	İmkânlar dâhilinde geri dönüşümden elde edilen ürünleri satın almayı tercih ederim.					
<b>ÇEVRESEL DAVRANIŞ -KİŞİSEL ULAŞIM TERCİHİ</b>						
UT 1	Yakıt tasarrufu sağlamak için çaba sarf ederim.					
UT 2	Yeni bir taşıt satın alacak olsam, öncelikle yakıt tasarrufu sağlayan ve düşük CO2 emisyonuna sahip bir araç tercih ederim.					
UT 3	Varmak istediğim yere toplu taşıma imkânı varsa, toplu taşıma araçlarını tercih ederim.					
UT 4	Kısa mesafeli varış yerlerine taşıtsız gitmeyi tercih ederim (yürüme, bisikletle vb.).					
UT 5	Birlikte bir yere gideceğimiz zaman, diğer aile fertleriyle aynı taşıtı kullanırız.					
<b>ÇEVRESEL DAVRANIŞ -EVSEL ENERJİ TALEBİ</b>						
ET 1	Enerji tasarrufu sağlamak için çaba sarf ederim.					
ET 2	Evin sadece kullandığım bölümlerinin ışıklarını açarım.					
ET 3	Yeni bir elektronik eşya satın aldığımda enerji verimliliği sağlayanları tercih ederim.					
ET 4	Elektrikle çalışan cihazları sadece kullandığım zaman fişe takarım.					
<b>ÇEVRESEL DAVRANIŞ -EVSEL SU TÜKETİMİ</b>						
ST 1	Su tasarrufu sağlamak için çaba sarf ederim.					
ST 2	Damlayan muslukları hemen tamir ederim.					
ST 3	Duş yaparken gerekli olmadığında musluğu kapatırım.					
ST 4	Yeni musluk aldığımda su tasarrufu sağlayan modelleri tercih ederim.					
<b>ÇEVRESEL DAVRANIŞ -ÇEVREYE DUYARLI GIDA TERCİHİ</b>						
GT 1	Organik olarak yetiştirilen gıdalara daha fazla para ödeyebilirim.					
GT 2	GDO'lu (Genetiği değiştirilmiş) gıdaları tüketmekten kaçınırım.					
GT 3	Çevre etiketli (eco-label) gıdaları tüketmeyi tercih ederim.					
GT 4	Büyükbaş hayvanların yaydıkları sera gazları hava kirliliğinde yüksek derecede etkili olduğu için çok fazla kırmızı et tüketmemeye özen gösteririm.					
GT 5	Uzak yerlerden taşınan süt ve süt ürünlerinin nakliyesinin yarattığı kirlilik nedeni ile bu ürünleri yerel işletmelerden almayı tercih ederim.					
GT 6	Çeşitli egzotik meyveler gibi ithal edilen gıdaların nakliyesinin yarattığı kirlilik nedeni ile bu tür gıdaları tüketmekten kaçınırım.					
GT 7	Konserve hazır gıdaları, nakliye ve ambalaj atıklarının yarattığı kirlilik nedeni ile tüketmekten kaçınırım.					

Katılımınız ve samimiyetiniz için teşekkür ederiz...

Doç.Dr. Gül EMEL  
Arş.Gör. Gülcan PETRİÇLİ

### Ek 3

Katılımcılar, aileleri ve meskenlerine ait bazı betimsel istatistikler.

*i.* Katılımcıların 361'i Osmangazi, 303'ü Yıldırım ve 328'i Nilüfer ilçelerindedir.

*ii.* Katılımcıların 218'i 18-24, 86'sı 25-34, 300'ü 35-44, 189'u 45-54, 132'si 55-64, 49'u 65-74 ve 17'si 75 ve üstü yaş aralığındadır.

*iii.* Katılımcıları 42'si ilkokul, 143'ü ortaokul, 319'u lise, 161'i ön lisans, 256'sı lisans ve 69'u lisans üstü eğitime sahiptir.

*iv.* İlk ve orta öğrenime sahip olanların 232'si Osmangazi, 133'ü Yıldırım ve 139'u Nilüfer ilçelerinde, lisans ve lisans üstü öğrenime sahip olanların 129'u Osmangazi, 169'u Yıldırım ve 188'i Nilüfer ilçelerinde yaşamaktadır.

*v.* Katılımcıların 441'i tam zamanlı çalışan, 147'si yarı zamanlı çalışan, 150'si emekli, 55'i üniversite öğrencisi, 46'sı öğrenci, 107'si işsiz (iş arayan) ve 46'sı çalışmamaktadır.

*vi.* Hanelerin 163'ünde bir tane, ikisinde ise iki tane bakıma muhtaç hasta, engelli veya yaşlı bulunmaktadır.

*vii.* Hanelerin 115'i bir, 12'si iki ve üçü üç tane 0-4 yaş arası bebeğe sahiptir.

*viii.* Hanelerin 264'ü bir ve 26'sı iki tane 5-15 yaş arası çocuğa sahiptir.

*ix.* Hanelerin 188'i bir, 56'sı iki ve biri üç tane 16-19 yaş arası ergen çocuğa sahiptir.

*x.* Hanelerde 18-64 yaş arası 1594, 65-74 yaş arası 221 ve 75 yaş ve üzeri 83 yetişkin yaşamaktadır.

*xi.* Hanelerin 110'u tek başına yaşayan genç birey, 73'ü çocuksuz genç çift, 159'u çocuklu genç çift, 28'i tek başına yaşayan çocuklu genç ebeveyn, 64'ü tek başına yaşayan orta yaşlı birey, 4'ü çocuksuz orta yaşlı çift, 324'ü çocuklu orta yaşlı çift, 51'i tek başına yaşayan çocuklu orta yaşlı ebeveyn, 29'u tek başına yaşayan yaşlı birey, 71'i çocuksuz yaşlı çift, 18'i ev arkadaşı, 56'sı geniş ailedir.

*xii.* Meskenlerin 817'si apartman dairesi ve 175'i müstakil evdir.

*xiii.* Meskenlerin 29'unda tek, 72'sinde iki, 273'ünde üç, 338'inde dört, 212'sinde beş, 57'sinde altı, 10'unda yedi ve birinde sekiz tane salon dahil oda bulunmaktadır.



#### Ek 4

Birinci pilot veri seti (200 gözlemlik) için Oblimin eğik döndürme yöntemi ile elde edilen FA sonuçları aşağıda yer aldığı gibidir.

**ÇEO'ye ait analiz sonuçları:** AFA sonucunda ÇEO'ye ait ifadelerin tümü, tek boyut altında toplanmıştır. Bu yüzden faktör sayısının tespitinde Kaiser kriteri kullanılmadan ve faktör sayısı üç olacak şekilde AFA tekrar edildiğinde ise ortaya çıkan faktör grupları, faktör gruplarına karşılık gelen ifadeler, faktör yükleri ve Cronbach Alfa katsayıları Tablo 57'de özetlenmiştir. AFA sonucunda, her bir ifadenin ilgili faktördeki yükünün 0.50 ve üstü olduğu görülmektedir. Ayrıca her bir ifadenin ilgili faktördeki yükü ile diğer faktörlerdeki yükleri arasındaki fark, 0.10'dan büyüktür. Elde edilen bu değerler ile üç faktörlü ölçeğin yapı ve ayırsama geçerliklerinin sağlandığı ifade edilebilir. Fakat dikkat edileceği üzere Ego1 ifadesinin faktör yükü birden büyüktür. Bu şekildeki bir sonucun nedeni, faktörler arası korelasyondan kaynaklanmaktadır. Eğik döndürme, faktörler arası korelasyona izin veren bir yöntem olduğu için, örneğin üç numaralı faktöre ait faktör yüklerinin oluşumunda diğer faktörlerin de etkisi bulunmaktadır. Bu yüzden hem ifadelerin faktör yüklerinin anlamlandırılması (yorumlanması) zorlaşmakta ve Ego1 örneğindeki gibi birden büyük yük değerleri ile karşılaşılabilen hem de açıklanan toplam varyans değeri gibi bir değer hesaplanamamaktadır.

**Tablo 57.** 200 örneklemlik pilot veri setinin ÇE ölçeği için faktör yükleri

İfadeler	Faktör yükleri		
	1	2	3
Bio1	0.332	0.450	
Bio2		0.602	
Bio3		0.965	
Bio4		0.867	
Ego1			-1.001
Ego2			-0.685
Ego3			-0.831
Ego4			-0.930
Alt1	0.601		
Alt2	0.833		
Alt3	0.908		
Alt4	0.892		

Eğik döndürme yaklaşımlarında dik döndürme yaklaşımlarından farklı olarak faktörler arası korelasyon bilgisi elde edilebilmektedir. Birinci ve ikinci faktörler arası korelasyon 0.713, birinci ve üçüncü faktörler arası korelasyon  $-0.839$  ve son olarak ikinci ve üçüncü faktörler arası korelasyon  $-0.732$  olarak hesaplanmıştır. Bu değerler P. W. Schultz (1999) ve Schmuck, vd. (2005) çalışmalarında sırası ile (0.37, 0.39, 0.25) ve (0.42, 0.69, 0.20) olarak tespit edilmiştir.

***DİÖ'ye Ait Analiz Sonuçları:*** AFA sonucunda ifadeler, Tablo 58'den de inceleneceği üzere beş boyut altında toplanmıştır. P6 ve P7 ifadeleri beklenildiği üzere perspektif boyutunda yüklenmemiştir. Deneyim boyutu altında yer alan D4 ve D5 ifadeleri neredeyse eşit yükler ile iki faktöre yüklenmiş ve D6 ifadesi ise başka bir faktöre güçlü bir şekilde yüklenmiştir.

**Tablo 58.** 200 örneklemlik pilot veri setinin Dİ ölçeği için birinci faktör analizi faktör yükleri

İfadeler	Faktör yükleri				
	1	2	3	4	5
Ö1		.747			
Ö2		.833			
Ö3		.744			
Ö4		.768			
Ö5		.775			
Ö6		.842			
Ö7		.599			
Ö8		.556			
P1	-.825				
P2	-.735				
P3	-.858				
P4	-.757				
P5	-.734				
<b>P6</b>					
<b>P7</b>				<b>-.746</b>	
D1					.727
D2					.579
D3					.660
<b>D4</b>			<b>.550</b>		<b>.463</b>
<b>D5</b>			<b>.405</b>	<b>.308</b>	
<b>D6</b>			<b>.928</b>		

Faktör sayısının tespitinde Kaiser kriteri kullanılmadan ve faktör sayısı üç olacak şekilde AFA tekrar edildiğinde P6 ve P7 ifadeleri hariç olmak üzere her bir ifadenin ilgili

faktördeki yükünün 0.50 ve üstü olduğu görülmektedir, Tablo 59. Ayrıca her bir ifadenin ilgili faktördeki yükü ile diğer faktörlerdeki yükleri arasındaki fark, 0.10'dan büyüktür. Elde edilen bu değerler ile üç faktörlü ölçeğin (P6 ve P7 ifadeleri hariç olmak üzere) yapı ve ayırsama geçerliklerini sağladığı ifade edilebilir.

**Tablo 59.** 200 örneklemlik pilot veri setinin Dİ ölçeği için ikinci faktör analizi faktör yükleri

İfadeler	Faktör yükleri		
	1	2	3
Ö1		.745	
Ö2		.719	
Ö3		.841	
Ö4		.795	
Ö5		.748	
Ö6		.734	
Ö7		.717	
Ö8		.677	
P1	-.862		
P2	-.732		
P3	-.873		
P4	-.776		
P5	-.631		
<b>P6</b>		<b>.316</b>	
<b>P7</b>		<b>.355</b>	
D1	.304		.500
D2			.550
D3			.640
D4			.946
D5			.541
D6			.627

Faktörler arası korelasyon ise birinci ve ikinci faktörler arası 0.460, birinci ve üçüncü faktörler arası korelasyon 0.573 ve son olarak ikinci ve üçüncü faktörler arası korelasyon 0.347 olarak hesaplanmıştır. Bu değerler (Nisbet, Zelenski, Murphy, 2009) ve Schmuck, vd. (2005) çalışmalarında sırası ile (0.47, 0.58, 0.34) ve (0.34, 0.69, 0.31) olarak tespit edilmiştir.

**STD Ölçeğine Ait Analiz Sonuçları:** Söz konusu ölçeğe eğik döndürme yöntemi kullanılarak AFA gerçekleştirilmeye çalışılmış fakat başarısız olmuştur. Çünkü yedi faktörlü bir yapı keşfedilmeye çalışılmış ve ölçek ifadelerinden bir tanesinin ortak

varyansı<sup>180</sup> birden büyük olarak hesaplanmıştır. Mevcut veri seti kullanılarak Kaiser kriteri dikkate alınmadan, faktör sayısı beş olacak şekilde AFA tekrarlanmıştır, Tablo 60. Faktörler ve faktör yükleri incelendiğinde geri dönüşüm, ulaştırma tercihi ve su tüketimi ifadelerinin birden fazla faktöre yüklendiği, ulaştırma tercihinde UT2 ve UT3'ün hiçbir boyuta yüklenmediği ve ayrıca GD1, GD6 ve ST1'in aynı anda iki farklı boyuta zayıf bir şekilde yüklendiği görülmektedir. Sonuç olarak, mevcut veri seti ile STD ölçeğinin tamamen problemlili bir yapıya sahip olduğu gözlemlenmiştir.

**Tablo 60.** 200 örneklemlilik pilot veri setinin STD ölçeği için ikinci faktör analizi faktör yükleri

İfadeler	Faktör yükleri				
	1	2	3	4	5
<b>GD1</b>	<b>0.519</b>			<b>-0.476</b>	
GD2	0.643				
GD3				-0.718	
GD4	0.502				
GD5				-0.655	
<b>GD6</b>	<b>0.366</b>			<b>-0.429</b>	
GD7				-0.625	
UT1	0.623				
<b>UT2</b>					
<b>UT3</b>					
UT4	0.448				
UT5	0.596				
ET1			0.729		
ET2			0.789		
ET3			0.716		
ET4			0.761		
<b>ST1</b>	<b>0.454</b>				<b>-0.472</b>
ST2	0.340				-0.502
ST3					-0.768
ST4					-0.597
GT1		0.549			
GT2		0.607			
GT3		0.547			
GT4		0.395			
GT5		0.857			

<sup>180</sup> Ortak varyans, bir değişkenin (ölçekte yer alan bir ifadenin) varyansının faktörler tarafından ne kadarlık kısmının açıklandığıdır. Bu değer birden büyük olması durumu Heywood vakalarındandır. Bu durumun olası sebepleri, aykırı veriler, örneklem büyüklüğünün 100 – 150'den küçük olması, bir faktörde sadece iki değişkenin yer alması ve mevcut veri setinin çok yüksek veya çok düşük korelasyonlu olmasıdır (Kline, 2011).

İfadeler	Faktör yükleri				
	1	2	3	4	5
GT6		0.731			
GT7		0.916			

Özetle, eğik döndürme yöntemi ile hem Kaiser kriteri dikkate alınarak hem de alınmayarak gerçekleştirilen AFA sonuçları, dik döndürme yöntemi ile elde edilen ve tezin uygulama bölümünde kısım 2.2’de raporlanan sonuçlar ile aynıdır. İkinci bölüm kısım 2.3.4.1’de de ifade edildiği üzere eğik döndürme yöntemi, ölçek maddelerini (ifadeleri) tek boyutluluktan uzaklaştırmaktadır ve bu yüzden boyutlar bağımsız bir yapıya sahip olmayıp, anlamlandırılmaya çalışılan teorik yapı basitlikten uzaklaşmaktadır (DeVellis, 2014). Eğik döndürme yaklaşımı (özellikle sosyal bilimlerde) her ne kadar daha gerçekçi bir yaklaşım olarak kabul edilse de DeVellis (2014) çalışmasında da önerildiği üzere, ölçeğe eğik döndürme yöntemine ek olarak dik döndürme yöntemi de uygulanmalı ve her iki yöntem ile elde edilen sonuçlar karşılaştırılmalı, basit ve anlamlandırılabilir yapılar sağlayan döndürme yöntemi benimsenmelidir.

Faktör döndürme yönteminin seçimindeki bir diğer önemli husus ise örneklem hacminin, gözlemlenebilir değişken (madde/ifade) sayısının veya ortak faktör varyansının (communality) yüksek olduğu durumlarda farklı döndürme yöntemlerinin çok benzer sonuçlar sağladığıdır (Tabachnick, Fidell, 2013). Eğer farklı döndürme yöntemleri birbirine yakın yapılar ve değerler sağlar ise faktörler arasında korelasyona (yani doğrusal ilişkiye) izin vermeyen dik döndürme yöntemi ile elde edilen sonuçların kullanılması tavsiye edilmektedir.

Birinci pilot veri setine ait dik ve eğik döndürme yöntemleri ile gerçekleştirilen AFA sonuçları birbirinin aynısı olduğu için ikinci pilot ve araştırma veri setleri eğik döndürme yöntemi ile AFA’ya tabi tutulmamış ve raporlanmamıştır.

## Ek 5

İkinci pilot veri seti Mahalanobis mesafesi kullanılarak temizlendiğinde, gözlem sayısı 193'e düşmüştür. Mahalanobis mesafesi ile veri setinin yaklaşık 2/3'ü kullanılabilir olmasına rağmen, Tablo 61'den de inceleneceği üzere, söz konusu "temizlenmiş" veri setinin AFA sonuçları ve alfa değerleri SMM ile temizlenmiş veri setine ait sonuçlardan daha kötüdür.

**Tablo 61.** Mahalanobis ve SMM ölçütleri ile temizlenen ikinci pilot veri setine ait FA sonuçları ve Cronbach alfa değerleri

Ölçüt	N	Ölçek	Cronbach Alfa	Kaiser kriteri dikkate alınarak gerçekleştirilen AFA			Kaiser kriteri dikkate alınmadan gerçekleştirilen AFA		
				Problemler Sayısı	Boyut	Varyans %	Problemler Sayısı	Boyut	Varyans %
Mahalanobis	193	ÇE	0.968	8	1	100	-	3	83
		Dİ	0.874	3	4	60	3	3	57
		STD	0.921	6	6	59	6	5	56
SMM	133	ÇE	0.970	4	2	81	-	3	<b>86</b>
		Dİ	0.895	4	4	64	4	3	<b>61</b>
		STD	0.933	5	5	<b>63</b>	Analize gerek kalmamıştır		

**Not:** ÇEÖ ve DİÖ teorik olarak üç ve STD ölçeği beş boyuta sahiptir.

## Ek 6

Ana araştırma veri seti için çok değişkenli normal dağılımın Mardia katsayıları (çok değişkenli basıklık ve çarpıklık) ile kontrolü.

```
> mvn(endise700,mvnTest = "mardia", desc = FALSE)
$multivariateNormality
      Test      Statistic      p value Result
1 Mardia Skewness 736.020010241951 2.12337115250728e-27    NO
2 Mardia Kurtosis 9.08448743773706          0      NO

> mvn(iliski700,mvnTest = "mardia", desc = FALSE)
$multivariateNormality
      Test      Statistic      p value Result
1 Mardia Skewness 3653.74770004033 5.94012679435023e-133    NO
2 Mardia Kurtosis 10.8035127835581          0      NO

> mvn(davranis700,mvnTest = "mardia", desc = FALSE)
$multivariateNormality
      Test      Statistic      p value Result
1 Mardia Skewness 5902.05831423057 2.92126858920833e-110    NO
2 Mardia Kurtosis 12.3272409453307          0      NO
```

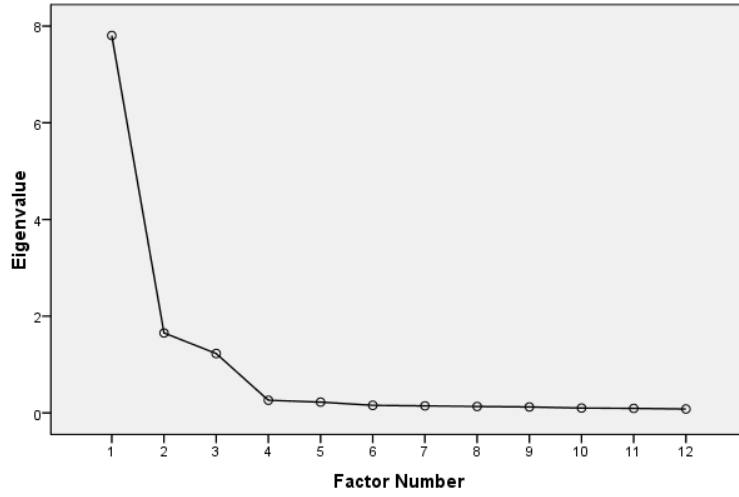
## Ek 7

Ölçme modellerinin SPSS yazılımı kullanılarak elde edilen güvenilirlik ve geçerlik analizleri sonuçları aşağıda yer aldığı gibidir.

**ÇEÖ'ye ait analiz sonuçları** Tablo 62'de yer almaktadır. AFA sonucunda ÇEÖ'ye ait ifadeler, varyansın %85.65'ini açıklayacak şekilde üç boyut altında toplanmıştır. Tablo 62'den ve Şekil 95'te yer alan yamaç-birikinti grafiğinden inceleneceği üzere, birinci faktör toplam varyansın % 29.205'ini, ikinci faktör % 28.757'sini ve üçüncü faktör % 27.503'ünü açıklamaktadır.

**Tablo 62.** Gözlem sayısı 352 olan ana araştırmaya ait veri setinin ÇEÖ için varyans açıklama yüzdeleri

Faktörler	Özdeğerler	Varimax döndürme sonrası	
		Açıklanan varyans	Birikimli yüzde
1	3.505	29.205	29.205
2	3.451	28.757	57.962
3	3.300	27.503	86.465
4			
⋮	⋮	⋮	⋮
12			



**Şekil 95.** ÇEÖ'ye ait yamaç-birikinti grafiği

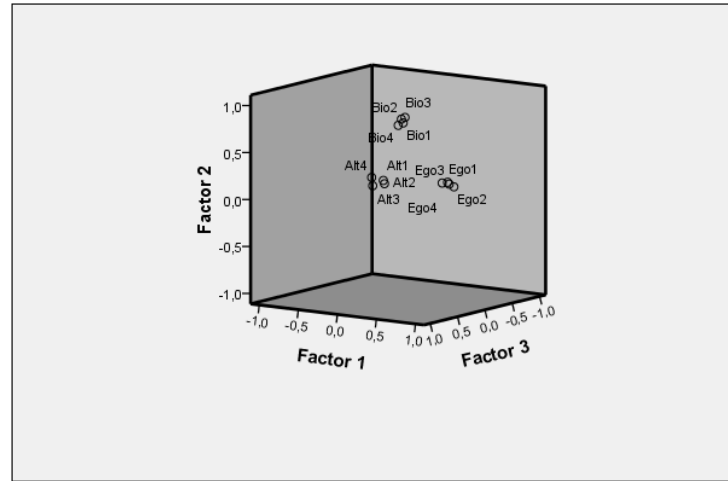
AFA sonucu ortaya çıkan faktör grupları, faktör gruplarına karşılık gelen ifadeler, faktör yükleri ve Cronbach Alfa katsayıları Tablo 63'te özetlenmiştir. İnceleneceği üzere, her bir ifadenin ilgili faktördeki yükünün 0.50 ve üstü olduğu görülmektedir. Ayrıca her bir ifadenin ilgili faktördeki yükü ile diğer faktörlerdeki yükleri arasındaki fark, 0.10'dan



büyükür. Elde edilen bu deęerlerle üç faktörlü ölçeęin yapı ve ayırsama geęerliklerinin saęlandıęı ifade edilebilir, Şekil 96. Güvenirlik analizi sonuçlarına göre Cronbach alfa katsayıları incelendięinde bütün faktörlerin alfa deęerlerinin 0.90'dan büyük olduęu görölmektedir. Ölçeęin genel alfa deęeri ise 0.951 olarak tespit edilmiřtir.

**Tablo 63.** Gözlem sayısı 352 olan ana arařtırmaya ait veri setinin ÇEÖ için faktör yükleri

İfadeler	Faktör yükleri			Cronbach alfa
	1	2	3	
Bio1		0.82		0.954
Bio2		0.87		
Bio3		0.89		
Bio4		0.81	.308	
Ego1	0.86			0.963
Ego2	0.87			
Ego3	0.84		.302	
Ego4	0.82		.362	
Alt1	.347	.300	0.77	0.956
Alt2	.384		0.8	
Alt3			0.86	
Alt4		.332	0.83	

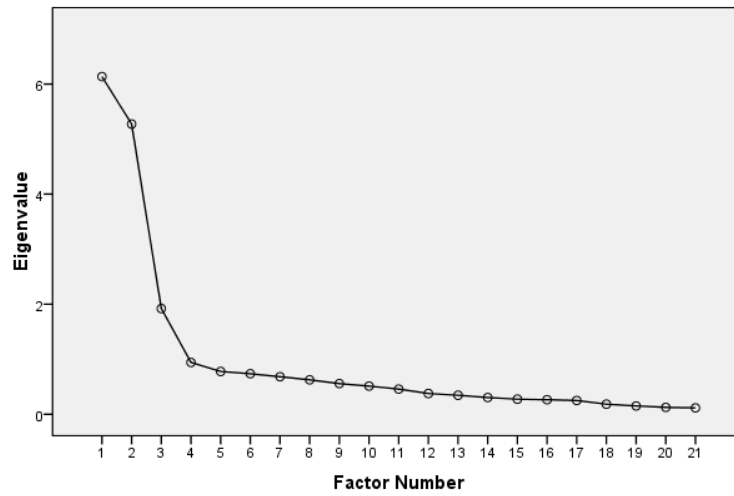


**Şekil 96.** ÇEÖ'ye ait faktörlerin üç boyutlu uzayda temsili

**DİÖ'ye ait analiz sonuçları** Tablo 64'te yer almaktadır. AFA sonucunda DİÖ'ye ait ifadeler, varyansın %58.164'ünü açıklayacak şekilde üç boyut altında toplanmıştır. Tablo 64 ve Şekil 97'de yer alan yamaç-birikinti grafięinden inceleneceęi üzere, birinci faktör toplam varyansın % 22.945'ini, ikinci faktör % 18.967'sini ve üçüncü faktör % 16.252'sini açıklamaktadır.

**Tablo 64.** Gözlem sayısı 352 olan ana araştırmaya ait veri setinin ÇE ölçeği için varyans açıklama yüzdeleri

Faktörler	Özdeğerler	Varimax döndürme sonrası	
		Açıklanan varyans	Birikimli yüzde
1	4.819	22.945	22.945
2	3.983	18.967	41.913
3	3.413	16.252	58.164
4			
⋮	⋮	⋮	⋮
21			



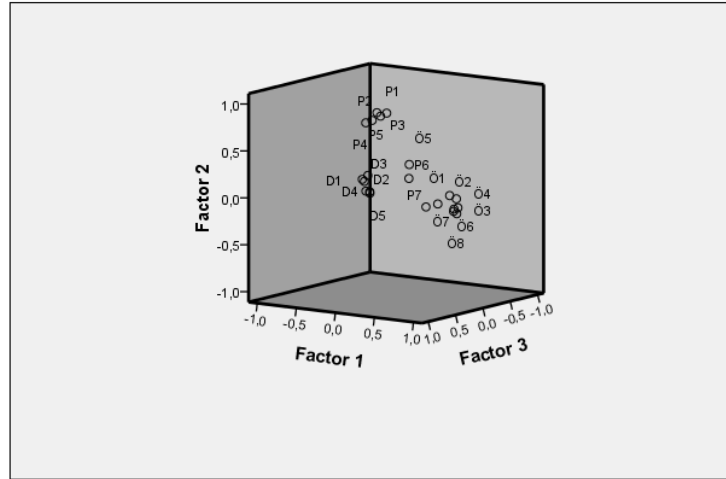
**Şekil 97.** DİÖ'ye ait yamaç-birikinti grafiği

AFA sonucu ortaya çıkan faktör grupları, faktör gruplarına karşılık gelen ifadeler, faktör yükleri ve Cronbach Alfa katsayıları Tablo 65'te özetlenmiştir. AFA sonucunda, P6 ve P7 ifadelerinin orijinal ve uyarlama ölçeklerde olduğu gibi özbenlik boyutunun altında yüklendiği fakat diğer taraftan Ö5 ifadesinin de zayıf bir şekilde perspektif boyutunda yüklendiği görülmektedir. P6, P7 ve Ö5 ifadeleri hariç olmak üzere her bir ifadenin ilgili faktördeki yükünün 0.50 ve üstüdür. Elde edilen bu değerlerle üç faktörlü ölçeğin yapı ve ayırsama geçerliklerinin sağlandığı ifade edilebilir, Şekil 98. Güvenirlik analizi sonuçlarına göre Cronbach alfa katsayıları incelendiğinde bütün faktörlerin alfa değerlerinin 0.90'a yakın olduğu görülmektedir. Ö5 ifadesinin özbenlik boyutundan çıkartılması durumunda özbenlik boyutunun alfa değeri 0.915'e yükselmektedir. Perspektif boyutunun alfa değeri 0.830 olup, P6 ve P7 ifadelerinin perspektif boyutundan çıkartılması ile alfa değeri 0.939'a yükselmektedir. Ölçeğin genel

alfa değeri ise 0.876 olarak tespit edilmiştir. P6, P7 ve Ö5 ifadelerinin ölçekten çıkartılması ile genel alfa 0.865'e düşmektedir. Dolayısı ile P6, P7 ve Ö5 ifadeleri ölçekten çıkartıldığı zaman ölçeğin genel alfa değeri düştüğü için yapısal modelin analizi için gerçekleştirilecek olan KEKK YEM analizinde Dİ ölçeğinin orijinal hali ile kullanılması uygun görülmüştür.

**Tablo 65.** Gözlem sayısı 352 olan ana araştırmaya ait veri setinin DİÖ için faktör yükleri

İfadeler	Faktör yükleri			Cronbach alfa
	1	2	3	
Ö1	.632			0.891
Ö2	.743			
Ö3	.806			
Ö4	.838			
<b>Ö5</b>		<b>.357</b>		
Ö6	.787			
Ö7	.791			
Ö8	.814			
P1		.843		0.830
P2		.879		
P3		.878		
P4		.802		
P5		.771		
<b>P6</b>	<b>.314</b>			
<b>P7</b>	<b>.525</b>			
D1			.740	0.881
D2			.742	
D3			.669	
D4			.734	
D5			.714	
D6			.641	

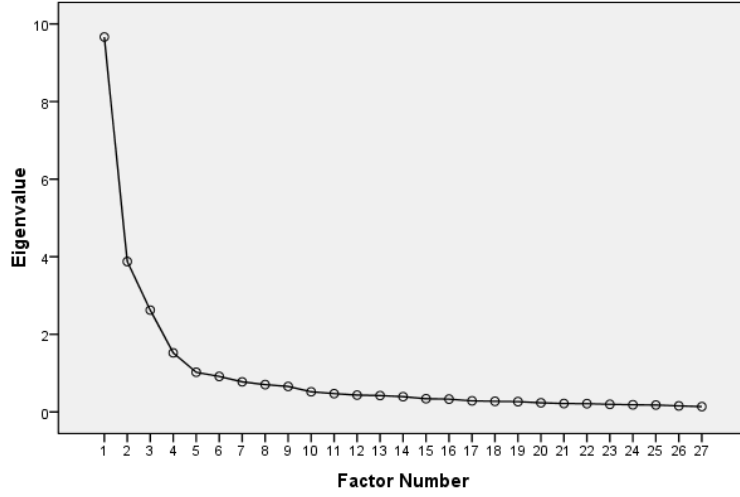


Şekil 98. DİÖ'ye ait faktörlerin üç boyutlu uzayda temsili

*STD ölçeğine ait analiz sonuçları* Tablo 66'da yer almaktadır. AFA sonucunda STD ölçeğine ait ifadeler, varyansın %62.997'sini açıklayacak şekilde beş boyut altında toplanmıştır. Tablo 66 ve Şekil incelendiğinde beş faktör toplam varyansın sırası ile % 16.156, %16.084, % 10.568, %10.134 ve %10.053'ünü açıklamaktadır.

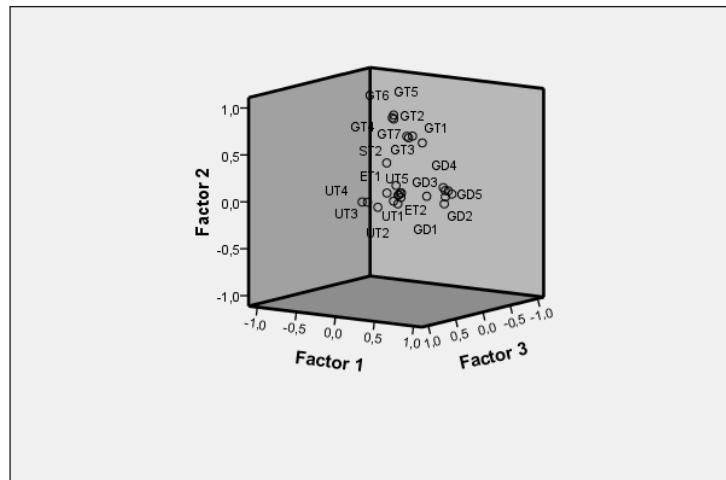
**Tablo 66.** Gözlem sayısı 352 olan ana araştırmaya ait veri setinin STD ölçeği için varyans açıklama yüzdeleri

Faktörler	Özdeğerler	Varimax döndürme sonrası	
		Açıklanan varyans	Birikimli yüzde
1	4.362	16.156	16.156
2	4.343	16.084	32.241
3	2.853	10.568	42.809
4	2.736	10.134	52.943
5	2.714	10.053	62.997
⋮	⋮	⋮	⋮
27			



Şekil 99. STD ölçeğine ait yamaç-birikinti grafiği

AFA sonucu ortaya çıkan faktör grupları, faktör gruplarına karşılık gelen ifadeler, faktör yükleri ve Cronbach Alfa katsayıları Tablo 67’de özetlenmiştir. Her bir ifadenin ilgili faktördeki yükünün, UT5 hariç, 0.50 ve üstü olduğu görülmektedir. Ayrıca her bir ifadenin ilgili faktördeki yükü ile diğer faktörlerdeki yükleri arasındaki fark, 0.10’dan büyüktür. Elde edilen bu değerlerle Şekil 100’den de inceleneceği üzere beş faktörlü ölçeğin yapı ve ayırsama geçerliklerinin sağlandığı ifade edilebilir. Güvenirlilik analizi sonuçlarına göre Cronbach alfa katsayıları incelendiğinde bütün faktörlerin alfa değerlerinin 0.90 civarı olduğu görülmektedir. Ölçeğin genel alfa değeri ise 0.929 olarak tespit edilmiştir.



Şekil 100. STD ölçeğine ait faktörlerin üç boyutlu uzayda temsili

**Tablo 67.** Gözlem sayısı 352 olan ana araştırmaya ait veri setinin STD ölçeği için faktör yükleri

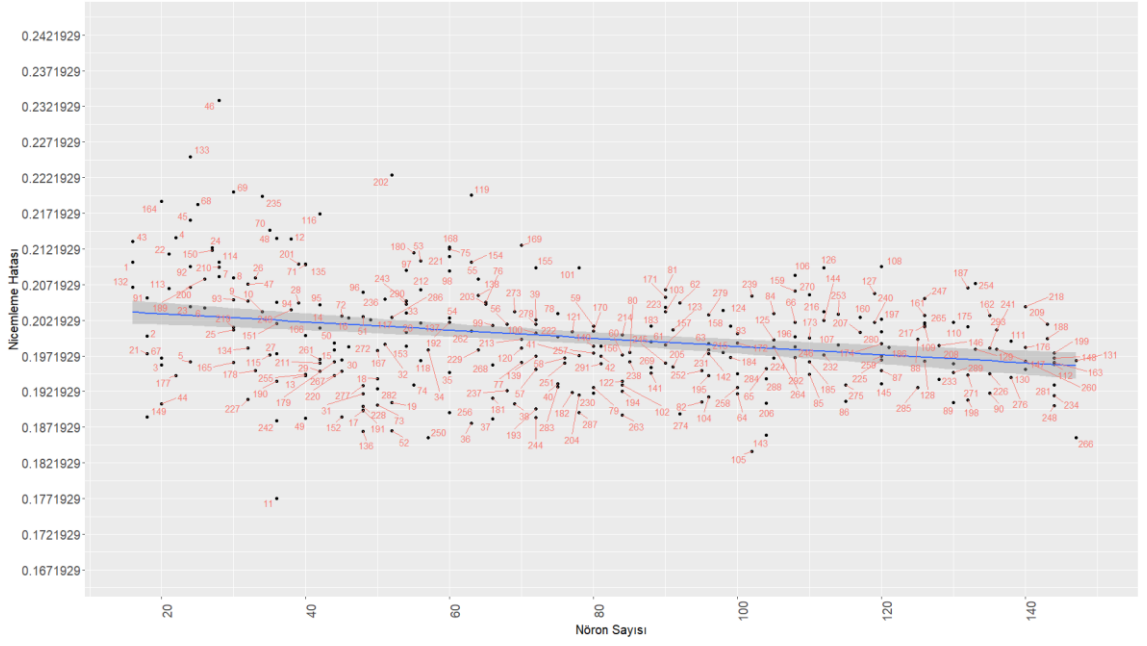
İfadeler	Faktör yükleri					Cronbach alfa
	1	2	3	4	5	
GD1	.546					0.902
GD2	.727					
GD3	.751					
GD4	.699					
GD5	.789					
GD6	.739					
GD7	.707					
UT1	.354		.555		.319	0.875
UT2			.666			
UT3			.847			
UT4			.754			
UT5			.466	.338	.354	
ET1				.571	.326	0.886
ET2				.825		
ET3				.685		
ET4				.673		
ST1					.770	0.866
ST2		.376			.509	
ST3					.786	
ST4				.326	.644	
GT1	.367	.624				0.907
GT2		.676				
GT3		.651				
GT4		.884				
GT5		.856				
GT6		.857				
GT7		.673				

Araştırma modelinin analizinde kullanılan Smart-PLS yazılımı, yapısal modeli dikkate alarak birleşme ve ayrışma geçerliklerine bağlı bir şekilde faktör yapısını tespit etmektedir. SPSS yazılımı ise ölçme modellerini ayrı ayrı ele alarak faktör yapısını tespit etmektedir. Bu kısımda SPSS programı kullanılarak, ölçme modellerine uygulanan güvenilirlik ve geçerlik analizleri sonuçları incelendiğinde, dördüncü bölüm, kısım 3.2.1’de elde edilen sonuçlar ile aynı olduğu görülmektedir. Dolayısı ile söz konusu araştırmada kullanılan ölçeklerin faktör yapılarının her iki yaklaşım ile doğrulandığı ifade edilebilir.

Faktör yapılarının doğrulanmasından ötürü, SPSS yazılımı kullanılarak ve Ö5, P6 ve P7 ifadeleri Dİ ölçeğinden çıkartılarak, her bir gözlem için ilgili boyuttaki ortalama

puanları hesaplanmıştır. Bu işlem ile veri setinin ölçeklere ait 60 özneliđi, 11 özneliđe indirgenmiştir.

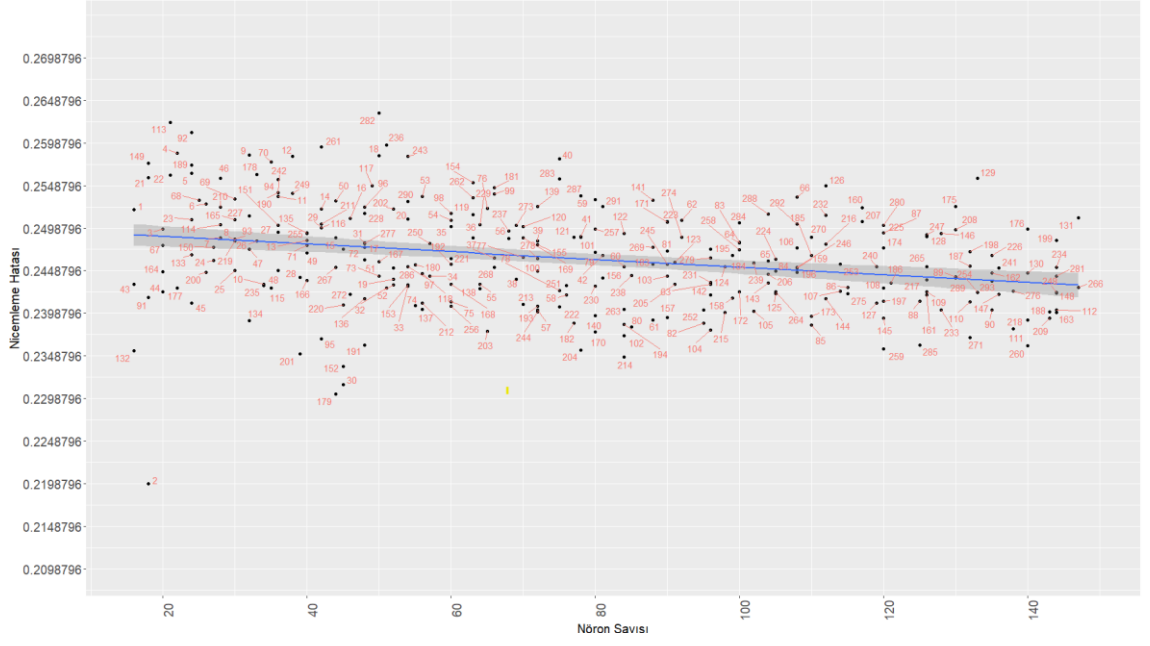
## Ek 8



Şekil 101. Çevrimiçi öğrenme algoritması ile eğitilmiş olan 293 haritanın nicemeleme hatasının nöron sayısına göre değişimi

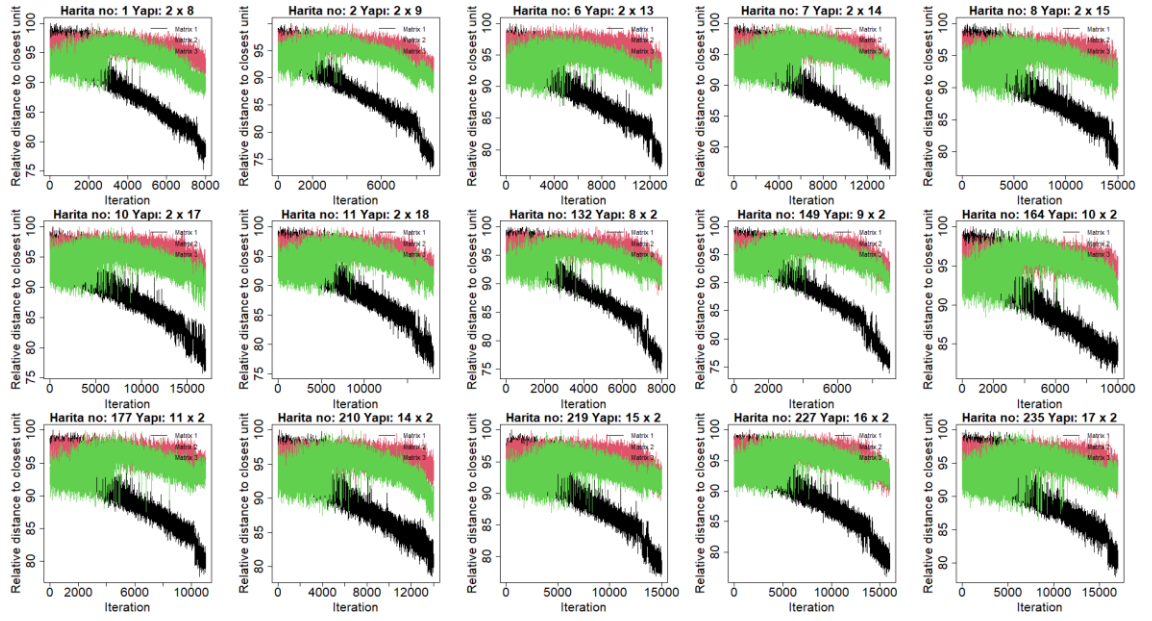


## Ek 9



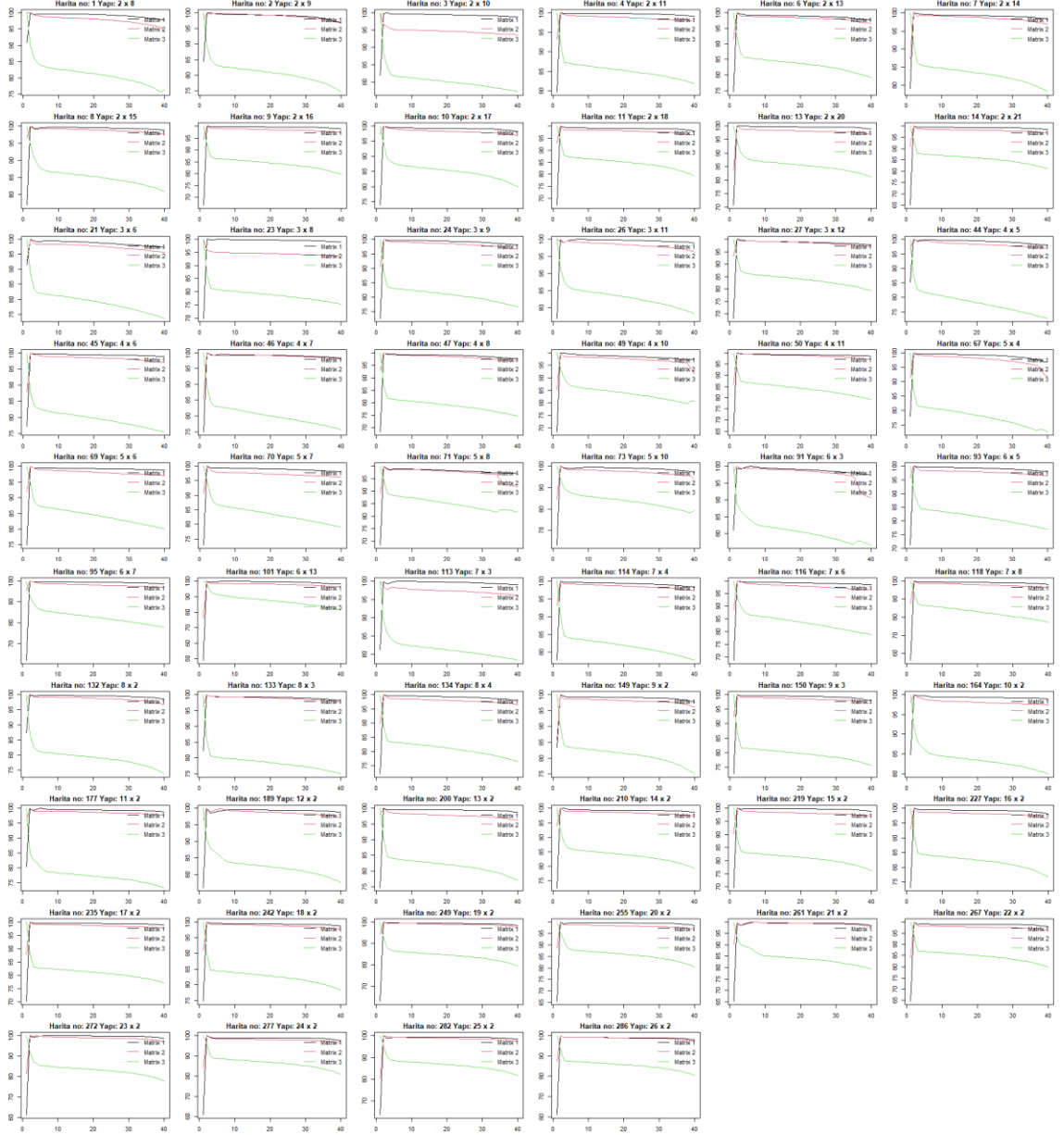
Şekil 102. Yığın öğrenme algoritması ile eğitilmiş olan 293 haritanın nicemeleme hatasının nöron sayısına göre değişimi

## EK 10



Şekil 103 Çevrimiçi öğrenme algoritması ile eğitilmiş olan 15 haritanın iterasyon esnaslı yakınsama grafiği

## Ek 11



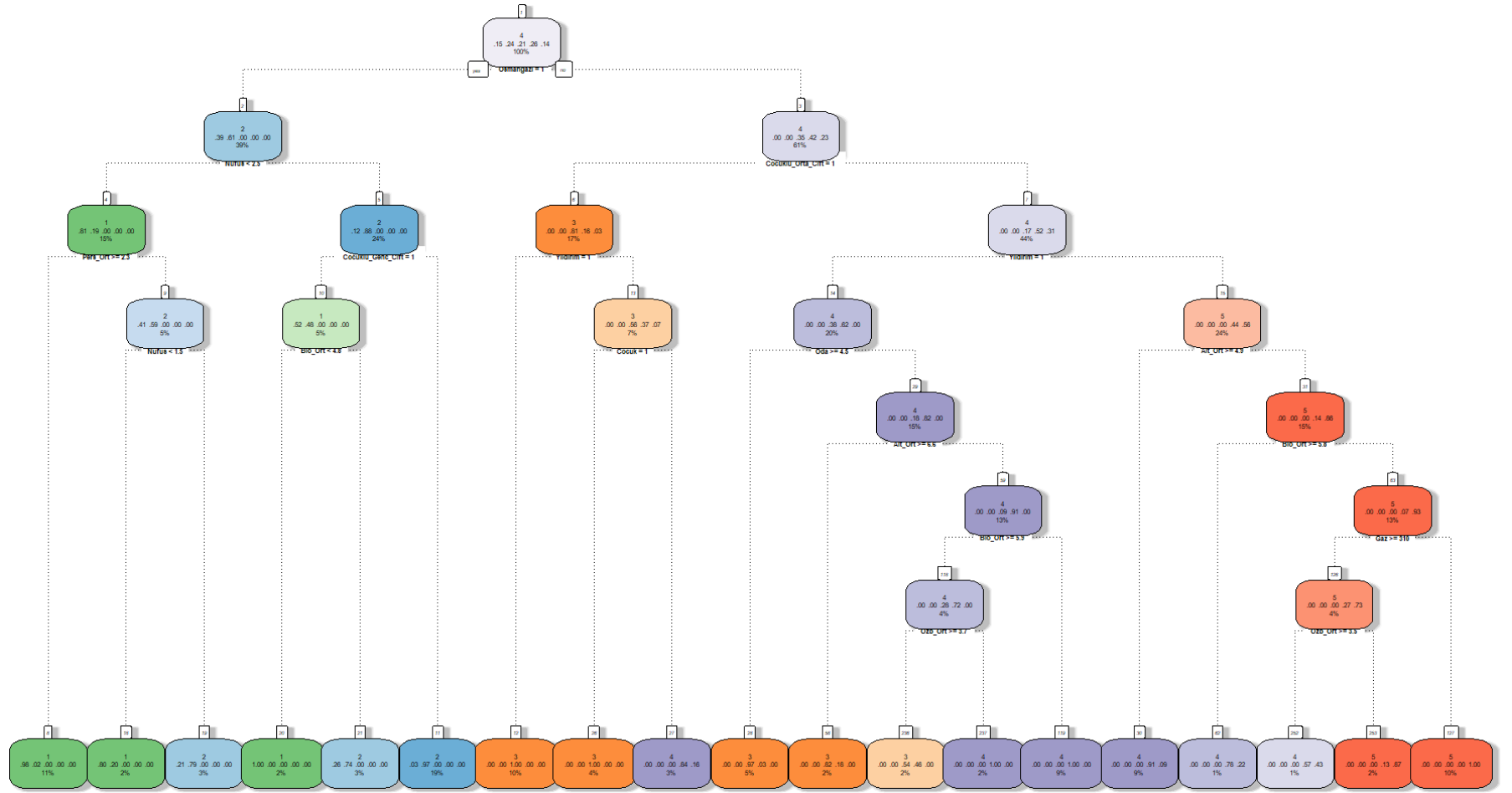
Şekil 104. Yiğın öğrenme algoritması ile eğitilmiş olan 15 haritanın iterasyon esaslı yakınsama grafiği

## Ek 12

**Tablo 68.** Karar ağacında kullanılan 227 numaralı haritanın TBA analizinin sonucunda öz niteliklerin birinci ve ikinci temel bileşenler üzerindeki yükleri


Öznitelik	Birinci temel bileşen	İkinci temel bileşen
Nilufer	-0.21708	0.30545
Su	-0.20741	0.06908
Pers_Ort	-0.19940	0.03112
Elektrik	-0.17644	0.05800
Gelir	-0.13637	-0.07557
GT_Ort	-0.08063	0.28106
Tek_Genc	-0.05357	-0.04401
Yasli	-0.04834	-0.00665
Ev_Kullanım	-0.04492	0.17683
Genc_Cift	-0.04414	0.01997
Cocuklu_Orta_Tek	-0.03498	0.02609
Mustakil	-0.03320	0.03338
Bakım	-0.02815	-0.00730
Tek_Yasli	-0.02713	0.00238
Yasli_Cift	-0.02179	0.00153
Cocuklu_Genc_Cift	-0.01918	0.08694
Cocuklu_Genc_Tek	-0.01664	-0.00107
Den_Ort	-0.01418	0.03198
Genis	-0.01029	-0.01347
Ev_Arkadasi	-0.00851	0.01295
Orta_Cift	-0.00320	0.00066
Tek_Orta	0.00243	0.02190
Gaz	0.01757	0.22049
Bebek	0.02933	0.09023
Daire	0.03320	-0.03338
Ergen	0.05139	-0.11301
Osmangazi	0.05616	-0.66550
UT_Ort	0.06588	0.12318
GD_Ort	0.09039	0.04198
ST_Ort	0.12424	0.21047
ET_Ort	0.14485	0.03844
Cocuk	0.15035	-0.02620
Yildirim	0.16092	0.36004
Cocuklu_Orta_Cift	0.23700	-0.11386
Ego_Ort	0.25687	-0.02818
Ozb_Ort	0.26102	0.07368
Alt_Ort	0.31571	0.13001
Nufus	0.35730	-0.09214
Oda	0.36886	0.06844
Bio_Ort	0.37439	0.11024

Ek 13



Şekil 105. Kural çıkarmada kullanılan karar ağacının budanmamış versiyonu

## ÖZGEÇMİŞ

ÖZGEÇMİŞ			
<b>Adı, Soyadı</b>	Gülcan		PETRİÇLİ
<b>Doğum Yeri ve Yılı</b>	Ayvalık		1982
<b>Bildiği Yabancı Diller</b>	İngilizce		---
<b>ve Düzeyi</b>	TOEFL-IBT (103)		YÖKDİL: 92,5
<b>Eğitim Durumu</b>	<b>Başlama - Bitirme Yılı</b>	<b>Kurum Adı</b>	
<b>Lise</b>	1996	1999	İzmir Kız Lisesi (Burslu Yatılı)
<b>Lisans</b>	1999	2003	Balıkesir Üniversitesi – Makine Mühendisliği
<b>Yüksek Lisans</b>	2009	2013	Uludağ Üniversitesi – İşletme (Tezsiz)
<b>Doktora</b>	2013	2020	Bursa Uludağ Üniversitesi – İşletme
<b>Çalıştığı Kurum (lar)</b>	<b>Başlama - Ayrılma Yılı</b>	<b>Çalışılan Kurumun Adı</b>	
1.	2004	2004	BICO Otomotiv Yan Sanayisi Atoyle Üretim Sorumlusu
2.	2009	2011	SBB Orta Doğu ve Kuzey Afrika Müşteri Bölge Yöneticisi
3.	2011	2019	Bursa Uludağ Üniversitesi SBE Araştırma Görevlisi
<b>Üye Olduğu Bilimsel ve Mesleki Kuruluşlar</b>	MMO, YAD, EURO		
<b>Katıldığı Proje ve Toplantılar</b>	<a href="https://akademik.yok.gov.tr/AkademikArama/view/viewAutor.jsp">https://akademik.yok.gov.tr/AkademikArama/view/viewAutor.jsp</a> linkinden ad-soyad ile arama yapılarak erişilebilir.		
<b>Yayımlar:</b>	<a href="https://akademik.yok.gov.tr/AkademikArama/view/viewAutor.jsp">https://akademik.yok.gov.tr/AkademikArama/view/viewAutor.jsp</a> linkinden ad-soyad ile arama yapılarak erişilebilir		
<b>Diğer:</b>	---		
<b>İletişim (e-posta):</b>	gulcanbp@gmail.com		
<b>Tarih</b> 30.09/2020			
<b>İmza</b>			
<b>Adı Soyadı</b>	Gülcan PETRİÇLİ		