

**PLAKALI EŐANJÖR ÜRETİM VERİLERİNİN MAKİNE
ÖĞRENMESİ YÖNTEMLERİYLE İNCELENMESİ,
ÖMÜR TEST VERİLERİNİN TAHMİNİ VE
SINIFLANDIRILMASI**

Anıl IŐIĐIÇOK



T.C.
BURSA ULUDAĞ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**PLAKALI EŞANJÖR ÜRETİM VERİLERİNİN MAKİNE ÖĞRENMESİ
YÖNTEMLERİYLE İNCELENMESİ, ÖMÜR TEST VERİLERİNİN TAHMİNİ
VE SINIFLANDIRILMASI**

Anıl İŞİĞİÇOK
0000-0002-4606-0629

Doç. Dr. Fatih KARPAT
(Danışman)
Dr. Öğr. Üyesi. Celalettin YÜCE
(İkinci Danışman)

YÜKSEK LİSANS TEZİ
MAKİNE MÜHENDİSLİĞİ ANABİLİM DALI

BURSA – 2020
Her Hakkı Saklıdır

TEZ ONAYI

Anıl IŞIĞIÇOK tarafından hazırlanan "PLAKALI EŞANJÖR VERİLERİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE İNCELENMESİ, ÖMÜR TEST VERİLERİNİN TAHMİNİ VE SINIFLANDIRIMLARI" adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Bursa Uludağ Üniversitesi Fen Bilimleri Enstitüsü Makine Mühendisliği Anabilim Dalı'nda **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Danışman : Doç. Dr. Fatih KARPAT

Başkan : Doç. Dr. Fatih Karpat
Bursa Uludağ Üniversitesi
Mühendislik Fakültesi,
Makine Mühendisliği Anabilim Dalı
0000-0001-8474-7328

İmza


Üye : Doç. Dr. Ümit ÜNVER
Yalova Üniversitesi
Mühendislik Fakültesi,
Enerji Sistemleri Mühendisliği Bölümü
0000-0002-6968-6181

İmza



Üye : Dr. Öğr. Üyesi. Celalettin YÜCE
Bursa Teknik Üniversitesi
Mühendislik ve Doğa Bilimleri Fakültesi,
Mekatronik Mühendisliği Anabilim Dalı
0000-0003-1387-907X

İmza



Üye : Prof. Dr. Ömer KAYNAKLI
Bursa Uludağ Üniversitesi
Mühendislik Fakültesi,
Makine Mühendisliği Anabilim Dalı
0000-0002-9763-6464

İmza



Üye : Doç. Dr. Nurullah ARSLANOĞLU
Bursa Uludağ Üniversitesi
Mühendislik Fakültesi,
Makine Mühendisliği Anabilim Dalı
0000-0003-4970-4490

İmza



Yukarıdaki sonucu onaylarım



Prof. Dr. Hüseyin Aksel EREN
Enstitü Müdürü

..../..../..

Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

01.10.2022



Anıl IŞIĞIÇOK

ÖZET

Yüksek Lisans Tezi

PLAKALI EŞANJÖR ÜRETİM VERİLERİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE İNCELENMESİ, ÖMÜR TEST VERİLERİNİN TAHMİNİ VE SINIFLANDIRILMASI

Aml İŞİĞİÇOK

Bursa Uludağ Üniversitesi
Fen Bilimleri Enstitüsü
Makine Mühendisliği İmalat Anabilim Dalı

Danışman: Doç. Dr. Fatih KARPAT

Bu çalışmada Manisa Bosch Termoteknik fabrikasının 2018 ve 2019 yıllarında plakalı eşanjör üretiminde gerçekleştirdiği ölçüm ve test sonuçları Python programlama dili ve makine öğrenmesi yöntemleri kullanılarak analiz edilmiştir. Seri üretim koşullarında üretilen eşanjör plakaları, belli periyotlarda kalite kontrol birimine iletilmekte ve çeşitli ölçümler yapılmaktadır. Plakalı eşanjörlerin çalışma sırasında ise; çeşitli nedenlerle, ani basınç dalgalanmaları meydana gelmektedir. Bu davranış su çekici (water hammer) olarak adlandırılır ve plakalı eşanjörün ömründe belirleyici rol oynayan bir kriterdir. Bosch Termoteknik'te su çekici etkisini gözlemlemek için ise ömür testleri gerçekleştirilmektedir. Test standına bağlanan plakalı eşanjörlere, kombinin çalışma koşullarında arka arkaya su çekici etkisi yaratacak basınç uygulanır ve ömür değerleri tespit edilir. Ancak bu test yaklaşık 12 gün sürmektedir ve bu süre günümüz piyasa ve seri üretim koşullarında çok uzundur. Bu çalışmanın amacı geçmiş dönem verilerinden yola çıkarak karar verme sürecini hızlandırmak, kaliteyi güvence altına almak ve verimliliği arttırmaktır. Ölçüm değerleri ve test sonuçlarından oluşan veri kümesi, regresyon ve sınıflandırma yöntemleri kullanılarak analiz edilmiştir. Lineer regresyon yöntemi ile sayısal tahminler yürütülmüştür. Ayrıca lojistik regresyon, destek vektör makineleri, rassal orman ve k en yakın komşu algoritmaları ile de sınıflandırma analizleri yapılmış ve oluşturulan matematiksel modellerin performansları karşılaştırılmıştır.

Anahtar kelimeler: Plakalı Eşanjör, Makine Öğrenmesi, Regresyon, Sınıflandırma

2020, viii + 74 sayfa.

ABSTRACT

MSc Thesis

INVESTIGATION OF PLATE HEAT EXCHANGER PRODUCTION DATA BY
MACHINE LEARNING METHODS, ESTIMATION AND CLASSIFICATION OF
LIFETIME TEST DATA

Amr İŞİĞİÇOK

Bursa Uludağ University
Graduate School of Natural and Applied Sciences
Department of Mechanical Engineering

Supervisor: Assoc. Prof. Dr. Fatih KARPAT

In this study, the measurement and test results of the Manisa Bosch Termoteknik factory in the production of plate heat exchangers in 2018 and 2019 were analyzed using Python programming language and machine learning methods. Heat exchanger plates produced under serial production conditions are transmitted to the quality control unit in certain periods and various measurements are made. During the operation of plate heat exchangers for various reasons, sudden pressure fluctuations occur. This behavior is called water hammer effect and is a criterion that plays a decisive role in the life of the plate heat exchanger. Bosch Termoteknik performs lifetime tests to observe the water hammer effect. The plate heat exchangers connected to the test stand are applied with pressure that will create a water hammer effect similar to the operating conditions of the boiler and the life values are determined. However, this test takes about 12 days and is very long in today's market and mass production conditions. The aim of this study is to accelerate the decision making process, to ensure quality and to increase efficiency based on past data. The data set consisting of measurement values and test results was analyzed using regression and classification methods. Numerical estimates were carried out with the linear regression method. In addition, classification analyzes were carried out with logistic regression, support vector machines, random forest and nearest neighbor algorithms, and the performances of the mathematical models created were compared.

Key words: Plate Heat Exchanger, Machine Learning, Regression, Classification

2020, viii + 74 pages.

ÖNSÖZ VE/VEYA TEŞEKKÜR

Yüksek lisans eğitimim boyunca bilgi ve tecrübelerinden yararlandığım, değerli görüşleri ve ilgisiyle bana yol gösteren Danışman Hocam Sayın Doç. Dr. Fatih KARPAT'a ve tez yazım sırasında deneyimlerini paylaşıp, önerilerini sunan Sayın Dr. Öğr. Üyesi Celalettin YÜCE'ye teşekkür ederim.

Bu çalışma için gerekli altyapıyı sunan ve destek olan BOSCH Termoteknik Isıtma ve Klima San. Tic. A.Ş.'ye ve tezimde bana destek olan tüm çalışma arkadaşlarıma teşekkürlerimi sunarım.

Her konuda bana destek veren ve yanımda olan, emeğini asla esirgemeyen, beni büyüten ve bugünlere getiren, babam Prof. Dr. Erkan IŞIĞIÇOK'a, annem Prof. Dr. Özlem IŞIĞIÇOK'a, emekli öğretmen anneannem Nurhan ZEYBEK'e ve emekli öğretmen, rahmetli dedem İsmail ZEYBEK'e gönülden teşekkür ederim.

Anıl IŞIĞIÇOK
01...108...2020

İÇİNDEKİLER

	Sayfa
1. GİRİŞ	1
2. KURAMSAL TEMELLER VE LİTERATÜR TARAMASI	4
2.1. Eşanjörlerin Sınıflandırılması	4
2.1.1. Isı transfer prosesine göre eşanjörler.....	4
2.1.2. Akışkan sayısına göre eşanjörler.....	5
2.1.3. Yüzey alanı kompaktlığına göre eşanjörler.....	5
2.1.4. Konstrüksiyona göre eşanjörler.....	5
2.1.5. Akış şekillerine göre eşanjörler.....	5
2.1.6 Isı transfer mekanizmalarına göre eşanjörler	6
2.2. Plakalı Eşanjörler	6
2.2.1. Contalı Plakalı Eşanjörler	9
2.2.2. Lehimli Plakalı Eşanjörler.....	11
2.3. Makine Öğrenmesinin Teorik Çerçevesi ve Çeşitli Yöntemler	14
2.3.1. Gözetimli öğrenme.....	16
2.3.2. Gözetimsiz öğrenme.....	17
2.3.3. Pekiştirmeli öğrenme	17
2.4. Makine Öğrenmesi Modelleri	18
2.4.1. Lineer regresyon modeli	19
2.4.2. Lojistik regresyon modeli	21
2.4.3. Destek vektör makineleri modeli	23
2.4.4. Karar ağaçları ve rassal orman algoritmaları	27
2.4.5. K-en yakın komşu algoritması modeli	31
3. MATERYAL VE YÖNTEM	33
3.1. Problemin Tanımlanması	37
3.2. Veriyi Anlamak	39
3.3. Veriyi hazırlama.....	44
3.3.1. Değişken seçimi	44
3.3.2. Normalizasyon	46
3.3.3. Verinin Polinom Hale Getirilmesi	46
3.4. Model Performansının Değerlendirilmesi.....	47
3.5. Aşırı Öğrenme ve Eksik Öğrenme	50
3.6. Çapraz Doğrulama	51

4. BULGULAR VE TARTIŞMA	53
4.1. Lineer Regresyon Algoritması Bulguları	53
4.2. Lojistik Regresyon Algoritması Bulguları	55
4.3. Destek Vektör Makinesi Algoritması Bulguları	56
4.4. Rassal Orman Algoritması Bulguları	57
4.5. K-En Yakın Komşu Algoritması Bulguları	59
5. SONUÇ	62
KAYNAK	64
EKLER	66
EK 1 Numpy, Pandas ve Sklearn Kütüphaneleri	67
EK 2 Lineer Regresyon	68
EK 3 Sınıflandırma Algoritmaları için Kod	69
EK 4 Lojistik Regresyon Algoritması	70
EK 5 Destek Vektör Makineleri Algoritması	71
EK 6 Rassal Orman Algoritması	72
EK 7 K-En Yakın Komşu Algoritması	73
ÖZGEÇMİŞ	74

,

SİMGELER ve KISALTMALAR DİZİNİ

Simgeler

Açıklama

β	Plakalı Eşanjör Kompaktlık Derecesi
θ_0	Bias (Kesme) Terimi
m	Toplam örnek sayısı

Kısaltmalar

Açıklama

HKO	Hata Kareleri Ortalaması
SVM	Support Vector Machines (Destek Vektör Makineleri)
RBF	Radial Basis Function (Radyal Temel Fonksiyon)
k-NN	k-Nearest Neighbours (k En Yakın Komşu)
ID3	Iterative Dichotomiser 3
CART	Classification and Regression Tree (Sınıflandırma ve Regresyon Ağacı)
DP	Doğru Pozitif
DN	Doğru Negatif
YP	Yanlış Pozitif
YN	Yanlış Negatif
EN	European Norms

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 1.1. Kombi Şeması	2
Şekil 2.1. Richard Seligman ve ilk plakalı ısı eşanjörü.....	6
Şekil 2.2. Modern plakalı eşanjör	7
Şekil 2.3. Yıkama tahtası plaka.....	7
Şekil 2.4. Balık kılıcı plakalar	8
Şekil 2.5. Pirinç perçinli plaka ısı eşanjörleri	9
Şekil 2.6. Contalı Plakalı Eşanjör	9
Şekil 2.7. Lehimli plakalı eşanjör	12
Şekil 2.8. Lehimli plakalı eşanjör kesiti.....	12
Şekil 2.9. Plakalı Eşanjör	13
Şekil 2.10. Kanal plaka örneği	14
Şekil 2.11. Lineer Regresyon Grafiği	21
Şekil 2.12. Lojistik fonksiyon örnekleri.....	22
Şekil 2.13. İki boyutlu uzayda ayrılabilir bir sorun örneği	24
Şekil 2.14. Veri kümesi ve karşılık gelen karar ağacı örneği.....	28
Şekil 2.15. Kalp krizi geçirme oranlarının görselleştirildiği karar ağacı	30
Şekil 3.1. Gözetimli makine öğrenim süreci	33
Şekil 3.2. Makine öğrenmesinin girdi ve çıktıları, işlem adımları ve bu adımlar için kullanılan Python kütüphaneleri	35
Şekil 3.3. Scikit-learn algoritma seçimi için yol haritası	36
Şekil 3.4. Lehimli plakalı eşanjör su giriş-çıkış gösterimi.....	37
Şekil 3.5. Lehimli plakalı eşanjör kesit alınmış iç görünümü.....	39
Şekil 3.6. Plakalı eşanjör üretim, kalite kontrol ve test aşamaları	40
Şekil 3.7. Plakalı eşanjör ölçüm bölgeleri şematik gösterimi	43
Şekil 3.8. Plakalı eşanjör ölçüm bölgeleri şematik gösterimi	43
Şekil 3.9. Çevrim sayısı ortalaması (ORT) ve a6 değişkeninin serpilme diyagramı	44
Şekil 3.10. Çevrim sayısı ve F2 değişkeninin serpilme diyagramı	45
Şekil 3.11. Bazı değişkenlerin ile çevrim sayısının çoklu serpilme diyagramları	45
Şekil 3.12. Lineer ve polinom regresyonların görselleştirilmesi	47
Şekil 3.13. Aşırı Öğrenme.....	50
Şekil 3.14. Eksik Öğrenme.....	51
Şekil 3.15. 3-Katlı çapraz doğrulama sistemi	52
Şekil 4.1. Lineer regresyon modeli test sonuçları ile tahminlerin karşılaştırılması	54

ÇİZELGELER DİZİNİ

Sayfa

Çizelge 2.1. Plaka Malzemeleri	10
Çizelge 2.2. Sık kullanılan conta malzemeleri.....	11
Çizelge 2.3.-Bernoulli rassal değişken dağılımı	22
Çizelge 2.4. Çekirdek fonksiyonlar.....	27
Çizelge 3.1. Su Çekici Testi Değerlendirme Kriterleri	41
Çizelge 3.2. Bağımsız değişkenler, açıklamaları ve kısaltmalar.....	42
Çizelge 3.3: Karışıklık Matrisi	48
Çizelge 4.1. Polinom derecelerine ve HKO sonuçları	53
Çizelge 4.2. 2.ci dereceden polinom dağılımında bağımsız değişken katsayıları.....	54
Çizelge 4.4. Lojistik regresyon karışıklık matrisi	55
Çizelge 4.5. Lojistik regresyon sınıflandırma raporu.....	56
Çizelge 4.6. Lojistik regresyon k-kat çapraz doğrulama sonuçları	56
Çizelge 4.7. Destek vektör makinesi karışıklık matrisi.....	56
Çizelge 4.8. Destek vektör makinesi sınıflandırma raporu	57
Çizelge 4.9. Destek vektör makinesi k-kat çapraz doğrulama sonuçları	57
Çizelge 4.10. Rassal orman algoritması parametre optimizasyonu, kırılım analizi.....	58
Çizelge 4.11. Rassal orman algoritması karışıklık matrisi.....	58
Çizelge 4.12. Rassal orman algoritması sınıflandırma raporu	58
Çizelge 4.13. Rassal orman algoritması k-kat çapraz doğrulama sonuçları	59
Çizelge 4.14. K-en yakın komşu algoritması parametre optimizasyonu komşu seçimi .	59
Çizelge 4.15. K-en yakın komşu algoritması karışıklık matrisi.....	60
Çizelge 4.16. K-en yakın komşu algoritması sınıflandırma raporu	60
Çizelge 4.17. K-en yakın komşu algoritması k-kat çapraz doğrulama sonuçları.....	60
Çizelge 4.18. Sınıflandırma modelleri tahmin sonuçları	60

1. GİRİŞ

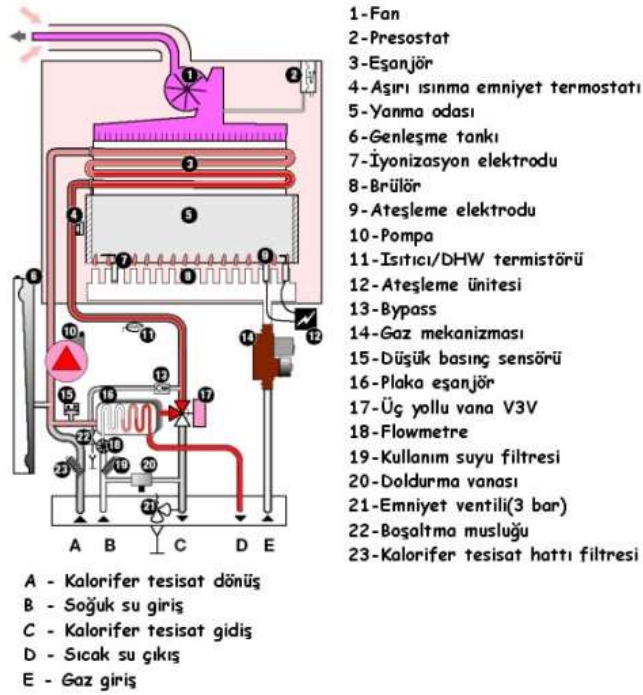
Günümüzde endüstriyel ürünlerin üretimi ve üretim sürecini oluşturan aşamaları, artan nüfusa bağlı olarak artan talep doğrultusunda büyük ölçeklerde planlanmakta ve gerçekleştirilmektedir. Bunun sonucu olarak, endüstriyel faaliyetlerde kalite ve verimlilik gibi kavramlar gün geçtikçe daha fazla önem kazanmaktadır.

Endüstriyel üretim sürecinde verimlilik ve benzeri performans kriterlerini iyileştirmek için üretim sürecinde aktif rol oynayan proses parametrelerinin performans üzerindeki etkiler bilinmeli ve gerek görüldüğü takdirde ilgili parametreler iyileştirilmelidir. Bu aşamada girdi ve çıktı parametreleri arasındaki ilişkiye dair çıkarımlar oluşturulmalı ve matematiksel modeller yardımı ile ifade edilmelidir. Karmaşık girdi-çıkıtı ilişkileri barındıran bunun gibi doğrudan çözemediğimiz problemlerde ise örnek verileri inceleyerek çıkarım yapabiliriz. Bu süreç öğrenmek olarak tanımlanmaktadır. Son yıllarda yaygınlaşan makine öğrenmesi, karmaşık problemlerde girdi ve çıktı arasındaki matematiksel modelin belirlenmesi için örnek verileri veya geçmiş deneyimleri kullanarak bir performans ölçütünü optimize etmek ve bu amaçla bilgisayarları programlamak olarak tanımlanabilir. İnsan uzmanlığının mevcut olmadığı veya insanların uzmanlıkları ile açıklayamadığı durumlarda öğrenmeye ihtiyaç vardır. Makine öğrenmesi bu durumlarda kullanabileceğimiz çok güçlü bir araçtır ve günümüzde yaygın olarak kullanılmaktadır. Konuşma ve el yazısını tanıyan sistemler, müşteri davranışlarını analiz etmek için satış verilerini analiz eden uygulamalar, minimum kaynakları kullanarak bir görevi tamamlamak için davranışlarını optimize eden robotlar akla gelen ilk uygulamalardır. Bilimsel araştırma alanlarının çoğunda, büyük verilerin analizlerine ihtiyaç duyulur ve makine öğrenmesi yaygın olarak kullanılır. Sunduğu avantajlar göz önüne alındığında, makine öğrenmesi ve yapay zeka gibi alanların gelecekte de hayatımızla iç içe olacağı aşikardır.

Ayrıca günümüzde, verimli ve temiz enerji yönetimine artan talep, çok sayıda yenilikçi ısı değiştirici tasarımının geliştirilmesini sağlamıştır. Plakalı eşanjörler, kompakt boyutları ve ısı olarak verimli davranışları nedeniyle bu yenilikçi ürünler arasında büyük öneme sahiptir. Plakalı eşanjör, iki veya daha fazla sıvı arasındaki ısı transferi için

kullanılan bir tasarımdır. Isı transferinin artırılması için transferin gerçekleştiği yüzey alanını arttırmak gerekmektedir. Bu amaçla plakalı eşanjörlerde plakaların kıvrılması gereklidir. Çeşitli yöntemlerle üretilmelerine karşın plakalı eşanjörler çoğunlukla belirli ortak özellikler sergilerler. Genellikle kompakt bir boyuta sahip ve boru şeklindeki eşanjörlerden daha hafif ve daha ucuzdurlar. Bununla birlikte, plakalı eşanjörlerin dayanabildikleri işletme sıcaklık ve basınç değerleri sınırlı olması nedeniyle uygulama tarafında bazı sınırlamaları vardır. Ancak yüksek sıcaklık ve basınç değerleri gerektirmeyen kombi uygulamaları için plakalı eşanjör kullanmak ideal bir çözümdür.

Kombilerin genelinde iki çeşit ısı değiştirici yer almaktadır. İlki brülörün hemen üzerinde yer alıp, doğrudan alev maruz kalan birincil ısı değiştiricidir. İkincisi ise plakalı ısı değiştirici veya plakalı eşanjör olarak isimlendirilir. Plakalı eşanjör birincil eşanjörden beslenir. Birincil eşanjörde ısınan su, ikincil eşanjör olan plakalı eşanjöre aktarılır. Dışarıdan gelen soğuk şebeke suyu plakalı eşanjör içerisinden geçerken sıcaklığa maruz kalır ve ısınır. Bu nedenle plakalı eşanjörlere ikincil eşanjör de denir. Kombi ısıtma sisteminde bulunan bütün parçalar ile plakalı ısı değiştiricisi Şekil 1.1’de verilmiştir.



Şekil 1.1. Kombi Şeması (Anonim, 2017)

Çevrimsel şartlarda çalışan ve anlık basınç yükselmeleri ile birlikte termo-mekanik yüklere maruz kalan plakalı eşanjörlerde yorulma, kritik rol oynamaktadır. Plakalı eşanjörlerde oluşabilecek yapısal problemler merkezi ısıtma suyu ve kullanım suyunun karışmasına neden olmaktadır. Bu yapısal başarısızlık sistemin termal verimliliğinin azalmasıyla birlikte, bazı termo-akışkan problemlere de neden olmaktadır. Ayrıca, kullanım suyuna ağır metal ve yanmış atık kirliliğinin aktarılmasına da neden olabilir.

Bu çalışmada, plakalı eşanjörlerde gerçekleşen yorulma kavramı, makine öğrenmesi yöntemleri kullanılarak analiz edilmiştir. Bu amaçla Manisa Bosch Termoteknik firmasında plakalı eşanjörlerde meydana gelen yorulmayı test etmek için su çekici (water hammer) adı verilen bir test düzeneği ile çalışılmıştır. Bu düzenekte plakalı eşanjör belirli periyotlarda olarak açılıp kapanarak, çalışma şartları simüle edilmekte ve normalde uzun yıllarda meydana gelen yorulma durumu 12 günde test edilebilmektedir.

Ek olarak firma, seri üretim sırasında yaptığı kontrollerde plakaların ölçüm verilerine de sahiptir. Üretim ölçüm verileri ile test verileri birleştirilip çeşitli makine öğrenmesi yöntemleri kullanılarak analizler yapılmıştır. 12 gün ömür testini gerçekleştirmek için makul bir süre olsa da günümüz seri üretim şartları göz önüne alındığında oldukça uzun bir süredir. Bu tez kapsamında makine öğrenmesi yöntemleri, plakalı eşanjörlerin üretimi sırasında etkili olan proses parametrelerinin ortaya çıkan ürünün ömrü üzerindeki etkilerini modelleyebilmek ve sınıflandırabilmek amacıyla kullanılmıştır. İnceleme sırasında farklı makine öğrenmesi yöntemleri irdelenmiş ve sonuçları karşılaştırılmış, incelenen her bir yöntem için parametre optimizasyonu yapılarak oluşturulan modellerin doğruluğu iyileştirilmeye çalışılmıştır. Yapılan çalışmanın amacı, yorulma kavramı hakkında geçmiş dönem verilerinden yola çıkarak öğrenmeler sağlamak, makine öğrenimi yöntemlerinin sonuçları yardımı ile test sürelerini kısaltmak ve bir mühendislik probleminin makine öğrenmesi ile ele alınırken izlenmesi gereken yol haritasını ortaya koymaktır.

2. KURAMSAL TEMELLER VE LİTERATÜR TARAMASI

Sıcaklık farkları olan ve temas halinde bulunan iki veya daha fazla akışkan ya da bir katı ve bir akışkan ya da katı parçacıklar ile akışkan arasında termal enerji (entalpi) geçişini sağlayan cihazlara eşanjör ya da ısı değıştirciler denir. Bir akışkanı ısıtmak, diğeri akışkanı soğutmak ya da her ikisini birden gerçekleştirmek eşanjörün temel amacıdır (Kaçar ve Erbay 2013).

Bu tip cihazlar hem günlük hayatımızda hem de termik santrallerin buhar jeneratörleri, kimya endüstrisinde damıtıcılar, ısıtma, havalandırma, iklimlendirme ve soğutma uygulamalarında buharlaştırıcı ve yoğunlaştırıcılar, elektronik cihazlarda soğutucular, otomobil radyatörleri, gaz türbin motorlardaki rejeneratörler gibi endüstriyel uygulamalarda yaygın kullanıma sahiptir. (Kaçar ve Erbay, 2013). Günlük hayatımızda karşılaşılan en yaygın örnekler ise kombiler, klimalar, araba radyatörleri ve yağ soğutucuları olarak verilebilir. Dolayısıyla eşanjör, adını her gün duymasak da hemen her gün kullandığımız ve hayatımızı kolaylaştıran bir teknolojidir.

2.1. Eşanjörlerin Sınıflandırılması

Kullanım alanları yukarıda bahsedildiği gibi çok geniş olan eşanjörler oldukça farklı çeşitlerde tasarlanmış ve üretilmektedirler. Eşanjörler, üretim yöntemleri, çalışma şartları, kullanım alanları, fiziki ebatları, görünümü, akış şekli ve sayısı, vb. pek çok şekilde sınıflandırılabilirler. Bu sınıflandırmalardan birkaçı aşağıda yer almaktadır.

2.1.1. Isı transfer prosesine göre eşanjörler

Isı transfer prosesine göre sınıflandırma, direk temaslı eşanjörler ve dolaylı temaslı eşanjörler olarak iki ana başlığa ayrılır. İşlem gören akışkanların birbirlerinden katı bir yüzey ile ayrıldıkları tipe direkt temaslı eşanjörler denir. Akışkanların ısı transferini başka bir enerji transfer aracı kullanarak sağladıkları tip ise dolaylı temaslı eşanjör olarak adlandırılır (Shah ve Sekulic 2003).

2.1.2. Akışkan sayısına göre eşanjörler

Isıtma, soğutma, ısı geri kazanımı ve ısı atımı çoğu proseslerde iki akışkan arasındaki ısı transferine dayanır. En yaygın kullanım tipinin iki akışkanlı olma nedeni ise, bu eşanjörlerin çoğu alandaki ihtiyacı karşılayabilmesi sayesinde. Üç akışkanlı eşanjörler ise genellikle kriyojenik ve kimyasal proseslerde (hava ayrıştırma sistemleri, helyum-hava ayrıştırma üniteleri, hidrojenin arıtılması ve sıvılaştırılması, amonyak gaz sentezi) kullanılır. Karmaşık ve oldukça özel kimyasal uygulamalarda ise kullanılan akışkan sayısının on ikiyi bulduğu özel eşanjörler kullanıldığı da bilinmektedir (Shah and Sekulic 2003).

2.1.3. Yüzey alanı kompaktlığına göre eşanjörler

Isı transfer alanının (m^2), eşanjör hacmine (m^3) bölümü eşanjör kompaktlık derecesi (β) olarak ifade edilir. Kompaktlaşma, eşanjörlerde hacmin küçülmesi nedeniyle taşıyıcı yapının küçülmesi, üretim maliyetlerinin azalması, yerleşim planlama ve çalışma koşullarının iyileşmesi gibi ciddi avantajları beraberinde getirir (Shah ve Sekulic 2003).

2.1.4. Konstrüksiyona göre eşanjörler

Eşanjörler çoğunlukla konstrüksiyon tiplerine göre sınıflandırılırlar. Dört ana konstrüksiyon sınıfı; borulu eşanjör, plakalı eşanjör, genişletilmiş yüzeyli eşanjör ve rejenetör tipleridir (Shah ve Sekulic 2003).

2.1.5. Akış şekillerine göre eşanjörler

Eşanjörlerde akış şekillerinin seçiminde belirleyici olan çeşitli kriterler vardır. Verim, basınç düşümü, izin verilen minimum ve maksimum akış hızları, akış yönleri, termal stresler, sıcaklık seviyeleri, borulama işlemleri ve diğer tasarım kriterlerine göz önüne alınmalıdır. Tek geçişli ve çok geçişli eşanjörler (iki veya daha fazla) olmak üzere iki ana başlığa ayrılır (Shah ve Sekulic 2003).

2.1.6 Isı transfer mekanizmalarına göre eşanjörler

Termal enerjiyi, ısı deęiřtiricinin bir tarafındaki akıřkandan ara yüzeye aktarmak için kullanılan temel ısı transfer mekanizmaları; tek faz konveksiyon (zorlanmış ya da doğal), iki faz konveksiyon (zorlanmış ya da doğal konveksiyon aracılığıyla yoęuşma veya buharlaşma) ve birleşik konveksiyon ve radyasyon ısı transfer mekanizmalarıdır (Kaçar ve Erbay 2013).

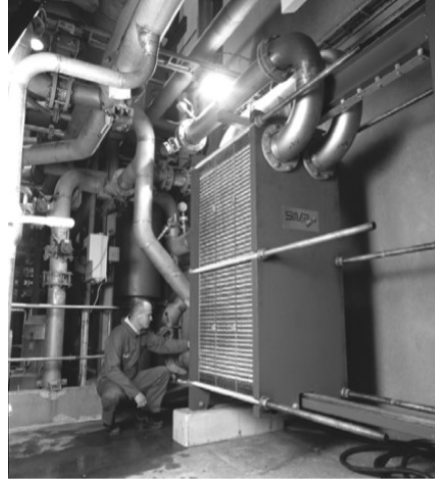
Dięer taraftan, eşanjörlerle ilgili farklı kaynaklarda farklı sınıflandırma sistemleri mevcuttur. Yukarıdaki kısımda ise en yaygın kullanılan sınıflandırma sistemleri özetlenmiştir. Devam eden bölümde ise plakalı eşanjörler ele alınacaktır.

2.2. Plakalı Eşanjörler

Plaka tipi olarak adlandırabileceğimiz ilk ısı eşanjörü için ilk patentler Almanya' da 1800'lerin sonlarında ortaya çıktı. İlk ticari plaka ısı eşanjörü (makine ile üretilmiş pirinç plakalar) 1923' te İngiltere' de Dr. Richard Seligman tarafından APV' de geliştirildi. Şekil 2.1' de yer alan ilk plaka ısı eşanjörü, o zamanlar yaygın ve tehlikeli bir hastalık durumunda olan tüberküloz ile savaşabilmek amacıyla, sütün çabuk soęutulması için geliştirilmişti (Toksoy 2003).

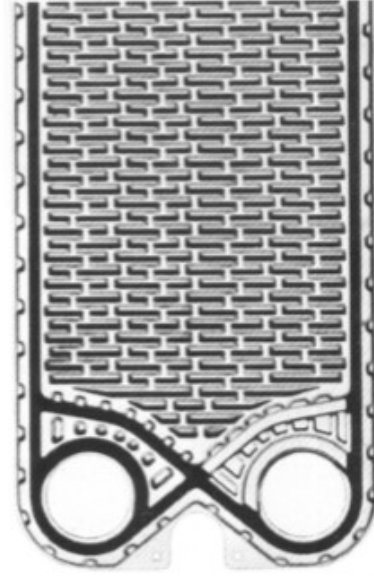


Şekil 2.1. Richard Seligman ve ilk plakalı ısı eşanjörü

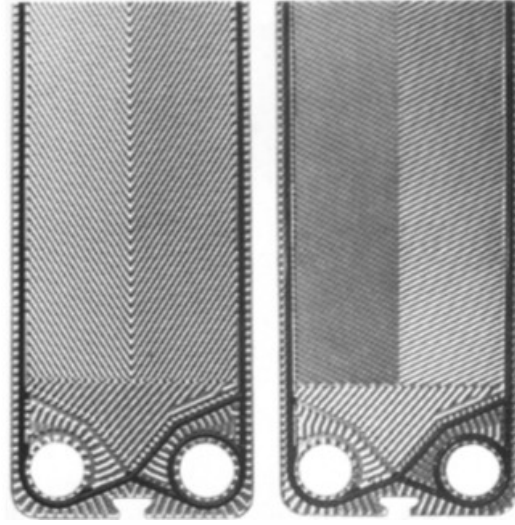


Şekil 2.2. Modern plakalı eşanjör (Toksoy 2003)

Preslenerek üretilen ilk plakalar, 1930'larda geliştirildi. Bunlar yıkama tahtası biçimindeydiler ve öncelikle gıda uygulamalarında halen kullanılmaktadırlar. Yıkama tahtası plakaların az miktarda temas noktaları vardır. Bu nedenle oldukça kalın malzemedен (0,9-1,2 mm) yapılmışlardır ve çok yüksek basınçlarda kullanılamıyorlardı (Toksoy 2003).



Şekil 2.3. Yıkama tahtası plaka (Toksoy 2003)



Şekil 2.4. Balık kılçığı plakalar (Toksoy 2003)

İlk geliştirilen plakalı eşanjörler Şekil 2.3'te gösterildiği biçimde yıkama tahtası biçimindeydiler ve gıda uygulamaları başta olmak üzere hala kullanılmaktadırlar. Şekil 2.4'te gösterilen balık kılçığı plakaları ise 1950 başlarında İsveç şirketi "Rosenblads Patenter" geliştirmiştir. Bu plakalar daha çok sayıda temas noktasına sahip oldukları için daha iyi mekanik stabilite göstermektedir. Bu icat ile plakalar daha ince (0,6 mm) üretilbildi, işletme basıncı daha da yükseldi ve böylece plakalı eşanjörler gıda endüstrisi dışında daha yaygın olarak kullanım alanı buldu. İki plakayı farklı açılarda bir araya getirme yeniliği ise, 1960 sonlarında geliştirilmiştir. Balık kılçığı plakalar son 50 yıl içinde daha da geliştirilmiştir. Ancak çapraz düzen belli başlı tüm sağlayıcılar tarafından hala kullanılmaktadır. Yeni gelişmeler sayesinde plakalar günümüzde 0,4 mm'ye incelebilmektedir (Toksoy 2003).

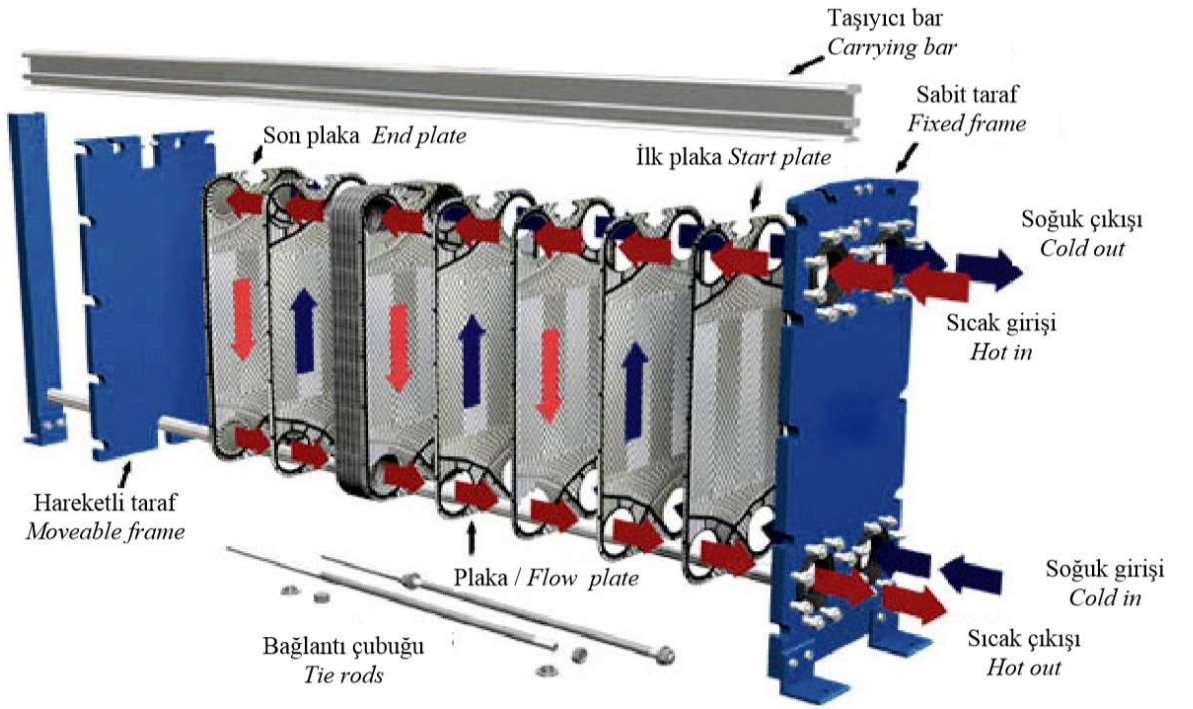
Şekil 2.5'te gösterilen pirinç kaynaklı plakalı eşanjörler, 1980'lerin başlarında geliştirilmiştir. Bunlar, balık kılçığı plakaların tüm temas noktalarında pirinç kaynakla bir araya getirilerek, perçinsiz plaka ısı eşanjörünü oluşturmasından ibarettir. Bu tip eşanjörler özellikle aşırı soğutma uygulamalarında kullanılmakla birlikte temizlemek için açmanın gerekmeyeceği kirlenmeyen uygulamalarda da kullanılmaktadır (Toksoy 2003). Günümüzde ise en yaygın kullanılan plakalı eşanjör tipleri ise contalı ve lehimli plakalı eşanjörlerdir.



Şekil 2.5. Pirinç perçinli plaka ısı eşanjörleri

2.2.1. Contalı Plakalı Eşanjörler

Contalı plakalı eşanjörler, Şekil 2.6’da görüleceği üzere, plakaların iki montaj elemanı arasına cıvata vb. bağlantı elemanları ile sıkıştırılarak sabitlenen ve sızdırmazlığın contalarla sağlandığı plakalı eşanjör uygulamasıdır.



Şekil 2.6. Contalı Plakalı Eşanjör (Kaçar ve Erbay 2013)

Bu tasarımın en önemli avantajları arasında sökölüp kolayca temizlenebilmeleri, sızdırma anında sızdıran alanın kolayca bulunup müdahale edilebilmesi ve ihtiyaç halinde plaka sayısında deęişiklik yaparak uygulamaya esneklik sağlaması yer alır.

Performans kabiliyetleri açısından incelendiklerinde ise kullanılan plaka ve contalar nedeniyle sınırlıdır. Özellikle contalar; çalışma sıcaklıklarında (160°C-250°C), basınçlarında (25-30bar) ve kullanılan akışkanlarda sınırlamalara neden olur. Plaka geometrilerinin kompleks şekillerde tasarlanması ise yüksek kesme gerilmelerine neden olabilir. Ancak, plakalar arası kanal derinliklerinin dar olmasından ve akışkanın hızının düşük olması sayesinde, sürtünme faktörlerinin türbülanslı akışlar için yüksek olmasına rağmen, tek faz akışkan kullanılan uygulamalarda basınç düşümleri emniyetli değerlerde korunabilir (Kakaç ve Liu 2002).

Uygun geometrik şekil verilebilen hemen her malzeme ile plaka üretiminde kullanılabilir. En yaygın malzemeler ise paslanmaz çelik, titanyum, titanyum-palaydum alaşımları, incoloy 825, hastelloy C-276, diabon F100, Monel 400, alüminyum ve alüminyum-pirinç alaşımları olarak sıranabilir ve Çizelge 2.1’de belirtilmiştir Çizelge 2.2’de ise sık kullanılan conta malzemeleri yer almaktadır (Kakaç ve Liu 2002).

Çizelge 2.1. Plaka Malzemeleri (Kakaç ve Liu 2002)

Malzeme	Termal İletkenlik ($\frac{W}{m^2.K}$)
Paslanmaz çelik (316)	16.5
Titanyum	20
Inconel 600	16
Inconel 825	12
Hastelloy C-276	10.6
Monel 400	66
Nikel 200	66
9/10 Nikel bakırı	52
70/30 Nikel bakırı	25

Çizelge 2.2. Sık kullanılan conta malzemeleri (Kakaç ve Liu 2002)

Malzeme	Ticari Adı	Sıcaklık dayanımı (°F)
Styrene-Butadiene	Buna-S	185
Neoprene	Neoprene	250
Acrylonitrile-Butadine	Buna-N	275
Ethylene/Propylene	EPDM	300
Fluorocarbon	Viton	300
Resin-Cured Butyl	Resin-Cured Butyl	300
Compressed Asbestos	Compressed Asbestos	500

Başlıca kullanım alanları ise kimyasal, farmasötik, yemek ve mandıra endüstrileri olarak sıralanabilir. Contalı plakalı eşanjörde, plakaların kolayca sökülebilmesi, temizlenip sterilize edilebilmesi sağlığın korunması ile ilgili gerekliliklerin yerine getirdiğinden, belirtilen endüstrilerde ısıtma, soğutma ve yoğuşurma uygulamalarında vazgeçilmez kılar (Kakaç ve Liu 2002).

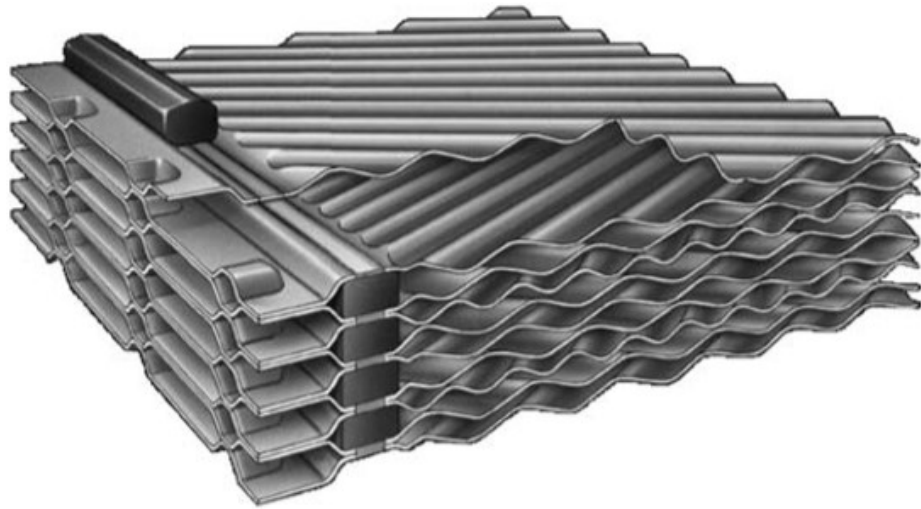
2.2.2. Lehimli Plakalı Eşanjörler

Şekil 2.7’de yer alan lehimli plakalı eşanjörler yüksek sıcaklık ve yüksek basınç uygulamalarında kullanılan ve conta, bağlantı cıvataları, taşıyıcı bar vb. ekipmanlar bulunmayan kompakt plakalı eşanjörlerdir. Paslanmaz çelik iki uç plakası arasında genellikle bakır ile lehimlenmiş plakalardan oluşur. Amonyak üretimi için kullanılan versiyonlarında nikel ile lehimlenen türleri vardır. Plaka boyutları genellikle $0.3m^2$ ile sınırlıdır. Böyle bir braket veya sabitleme ünitesi vb. gerektirmeden direkt olarak tesisata, boruların üzerine monte edilebilir (Shah ve Sekulic 2003).



Şekil 2.7. Lehimli plakalı eşanjör (Anonim, 2020a)

Bu tasarım ve lehim ile bağlantı şekli, lehimli plakalı eşanjörleri, contalı tiplere göre daha kompakt, daha hafif ve daha ucuz kılar. Buna karşılık lehimli bağlantının dezavantajı, plakaların sökülüp değiştirilmesine veya temizlenmesine olanak vermemesidir. Ek olarak kullanım esnasında plaka eklenmesi veya çıkartılması mümkün olmadığından, eşanjör kapasitesinde değişiklik yapılma şansı da yoktur (Rafferty ve Culver 2006).

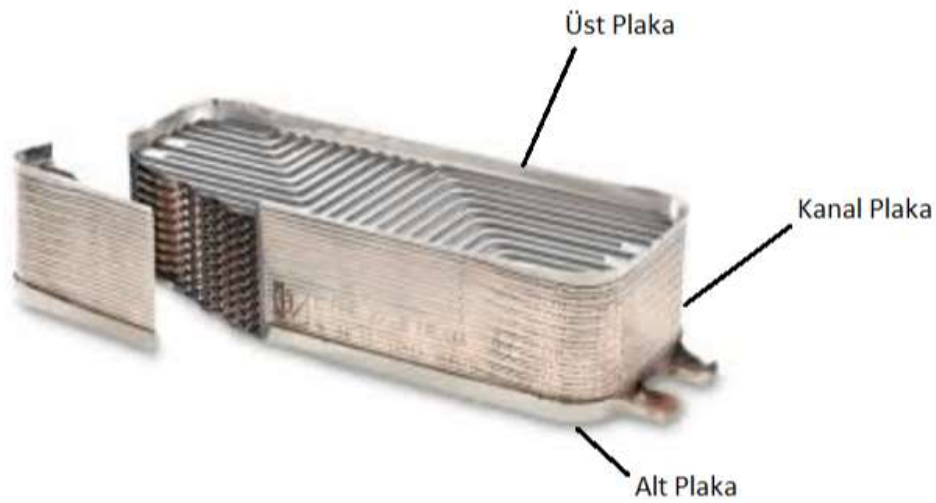


Şekil 2.8. Lehimli plakalı eşanjör kesiti (Shah ve Sekulic 2003)

Plakaların iki tarafından da lehimlenmesi daha yüksek sıcaklık (350°C) ve basınç (4.0MPa) değerlerinde çalışmalarına olanak tanır. Ek olarak uygun plaka malzemesi seçilmesi ile korozif akışkanlar da lehimli plakalı eşanjörlerde kullanılabilir. Lehimli plakalı eşanjörlerde kullanılan plaka malzemeleri ise; paslanmaz çelik, hastelloy alaşımı, nikel bazlı alaşımlar, bakır ve titanyumdur (Shah ve Sekulic 2003).

Lehimli tip plakalı eşanjörler, contalı tiplerin kullanılabilabileceği uygulamalara ek olarak içindeki akışkanın tortu, oksijen ve korozif diğer etkenlerin eşanjörü kullanım dışı bırakmayacak limitlerde olduğu uygulamalarda da kullanılabilir. Kızgın su kullanılan su ısıtma sistemlerinde kapalı devre su sistemleri için uygun bir seçimdir. Özellikle ısı pompalarında ve soğutma sistemlerinde gerek evaporatör gerek ise kondenser olarak yaygın şekilde kullanılırlar. Sanayi tesislerinde ise genellikle yağ soğutma amaçlı tercih edilirler.

Bu çalışmada ele alınan plakalı eşanjör, görsel olarak Şekil 2.7'dekilere benzeyen bir lehimli plakalı eşanjördür. Alt plaka, üst plaka ve kanal plaka olarak adlandırılan üç farklı plakanın birleşiminden meydana gelir. Plakaların gösterimi Şekil 2.9'da yer almaktadır. Bu üç plaka preslerde farklı kalıplarda ayrı ayrı üretilir. Ardından bir adet alt plaka, bir adet üst plaka ve kombi tasarımı ve ihtiyacına göre değişkenlik gösteren 16, 18, 20 vb. adette kanal plaka üst üste eklenir ve preslenerek birleştirilir.



Şekil 2.9. Plakalı Eşanjör

Preslenen eşanjör vakum fırınına yüklenir ve yüksek sıcaklığa maruz bırakılır. Bakırın ergime sıcaklığı çeliğe göre daha düşük olduğundan, kanal plaka içerisinde yer alan ve Şekil 2.10'da gösterilen bakır plakalar erir. Yapıştırıcı görevi gören bakır plakalar, çelik plakaları birbirine bağlar.



Şekil 2.10. Kanal plaka örneği (Anonim, 2020b)

2.3. Makine Öğrenmesinin Teorik Çerçevesi ve Çeşitli Yöntemler

Makine öğrenmesi; bilgisayar sistemleri ile yürütülen, istatistiksel ve matematiksel modelleri kullanan algoritmalarıdır. Belirli bir görev herhangi bir talimat ya da yönerge olmaksızın, veri kümesi içerisindeki çıkarımlar ve kalıplar bulunarak yerine getirilir. Günümüzde popüler araştırma konularından olan yapay zeka sistemlerinde kullanılan yapı taşlarından biridir. Makine öğrenmesi algoritmaları veri kümelerine dayanır. Veri kümeleri öğrenme ve test adı altında iki gruba ayrılır. Öğrenme veri kümesi içerisinde matematiksel model geliştirilir. Daha sonra, test veri kümesinin değerleri tahmin edilerek, gerçek değerlerle karşılaştırılır. Öğrenme denmesinin nedeni algoritmalar çalıştırıldıkça, veri kümelerini ayırtıran matematiksel modellerdeki çeşitli parametrelerin değişiklik göstererek güncellenmesidir. Farklı bir şekilde ifade etmek gerekirse model eğitildikçe tecrübe kazanır ve sonuçlar iyileşir.

Bir makinenin öğrenmesi için davranışını, performansı da göz önünde bulundurarak nasıl değiştireceğini Mitchell (1997) şu şekilde ifade etmiştir; eğer bir bilgisayar programının G görevlerinde, P ile ölçülen performansı, deneyim D ile artıyorsa, o bilgisayar programının bazı G görevlerinin sınıflarına ve performans ölçüsü P'ye göre deneyim D'den öğrendiği söylenmektedir.

Örneğin, dama oynamayı öğrenen bir bilgisayar programı, kendine karşı oyun oynayarak edinilen deneyim sayesinde dama oyunları oynamayı içeren görevler sınıfında kazanma yeteneği ile ölçülen performansını artırabilir. Genel olarak, iyi tanımlanmış bir öğrenme problemine sahip olmak için görev sınıfı, geliştirilecek performansın ölçüsü ve deneyim kaynağı olmak üzere üç özellik tanımlamalıyız.

Bu problemde;

- G görevi, dama oynamak
- P performansı, rakiplere karşı kazandığı oyunların yüzdesi
- D deneyimi, kendi ile oynayarak pratik/idman yapmak

Makine öğrenmesi algoritmaları veri kümesi üzerinde çalıştırıldıklarında makine öğrenmesi modelleri elde edilir. Öğrenme sonrasında modele bir girdi verdiğinizde, model de size bir çıktı verecektir. Kuşkusuz bu çıktı tahmini bir sonuçtur. Örnek vermek gerekirse; bir tahmin algoritması, tahmin modeli üretecektir. Bu tahmin modeline bir girdi verildiğinde çalıştırılan veri kümesine dayalı bir tahmin çıktısı elde edilir. Makine öğrenmesi günümüzde analitik modeller yaratmak için kullanılan ana sistemdir. Çoğu kişi farkında olmadan makine öğrenme uygulamaları ile karşılaşır. E-ticaret sitelerindeki reklamlar bu örneklerden sadece biridir. Siteye girip çeşitli ürünleri incelemeye başladıktan sonra karşınıza benzer ve ilgi çekici ürünler çıkmaya başlar. Bu öneriler yazılımcılar tarafından kodlanmamıştır. Makine öğrenme algoritmaları arka planda internet geçmişiniz üzerinden satın alma ihtimaliniz yüksek olan sistemleri size otomatik olarak sunar (Hurwitz ve Kirsch, 2018).

Değişkenler arasında ilişki olup olmadığını ve varsa bu ilişkinin yapısını incelemek bilimin ilgi alanındadır. Kuşkusuz, günlük yaşamda veya iş hayatında karşılaşılan sorunların çoğunluğu değişkenler arasındaki ilişkilerden kaynaklanmaktadır. Günlük yaşantımızdan verilebilecek örnekler arasında; reklamlar ile satış, stres seviyesi ile verimlilik, Makine mühendisliğinden birkaç örnek ise; malzemelerdeki ısı iletkenlik katsayısı ile ısı kayıpları, çelik alaşımlarındaki karbon oranları ile sertlik ve dayanım ölçüleridir. Kullanılan öğrenme yöntemleri ise; gözetimli öğrenme, gözetimsiz öğrenme ve pekiştirmeli öğrenme olarak üç ana başlıkta incelenmiştir.

2.3.1. Gözetimli öğrenme

Gözetimli öğrenme etiketli ve özellikleri bilinen bir veri kümesinin incelenmesi ile başlar. Gözetimli öğrenme algoritmalarının hedefi veri kümesi içerisindeki benzerlikleri ve örüntüleri bulup analitik yollarla işlemek ve matematiksel bir model oluşturmaktır. Veri kümesindeki özellikler etiketlenmiş olması gerekir. Veri kümesindeki özellikler sürekli (continuous) ise regresyon ve korelasyon analizlerine başvurulur. Regresyon analizi sürekli değişkenler arasındaki fonksiyonel ilişkiyi ortaya koyarken, korelasyon analizi sürekli değişkenler arasındaki ilişkinin yönünü ve şiddetini belirler. Veri kümesindeki özellikler ayrık (discrete) ise sınıflandırma yapılıdır. Örnek olarak hava durumu tahminini ele alalım. Analiz için tarihsel hava durumu verileri (sıcaklık, basınç, nem, rüzgar vb.) kullanılmaktadır. Veri kümesi içerisindeki değişkenler sayısal olduğundan regresyon analizi uygulanabilir ve matematiksel benzerlikler araştırılarak tahmin yapılıdır (Hurwitz ve Kirsch 2018).

Algoritmanın performansını ölçmek amacıyla, çalışmaya başlamadan önce veri kümesi ikiye bölünür. Bunlar öğrenme veri kümesi ve test veri kümesidir. Algoritma, öğrenme veri kümesi kullanılarak çalıştırılır ve matematiksel bir model elde edilir. Ardından oluşturulan matematiksel modele test verileri girildiği sunulur. Bu sayede aşırı uyum gösterme (overfitting) probleminin önüne geçilir. Aşırı uyum gösterme, gözetimli öğrenme algoritmasının oluşturduğu veri kümesinin mevcut veri kümesine fazla uymasıdır. Aynı probleme ait bilinmeyen farklı bir veri matematiksel modele sokulduğunda bu durum modelin başarısını düşürebilir. Ancak yukarıda belirtildiği gibi

veri kümesi ikiye ayrılarak model oluşturulursa, test veri kümesi ile model karşılaştırılarak bu problemin önüne geçilebilir ve model başarısı artırılabilir. Gözetimli öğrenme iş hayatındaki çok çeşitli uygulamalarda kullanılmaktadır. Örnek olarak öneri sistemleri, risk analizleri, ses tanıma sistemleri, dolandırıcılık tespiti vb. verilebilir (Hurwitz ve Kirsch 2018).

2.3.2. Gözetimsiz öğrenme

Gözetimsiz öğrenme ise; etiketlenmemiş çok büyük veri kümelerini içeren problemler için kullanılacak en etkin yöntemlerden biridir. Örnek olarak; “Twitter”, “Instagram” ve “Snapchat” gibi sosyal medya uygulamalarındaki büyük miktardaki etiketsiz veri kümeleri verilebilir. Bu verinin arkasındaki anlamı anlamak, verileri bulduğu kalıplara veya kümelere göre sınıflandırabilmeye dayalı olarak anlamını anlamaya başlayabilecek algoritmalar gerektirir. Kısaca, gözetimsiz öğrenme, insan müdahalesi olmadan verileri analiz etmek için yinelemeli bir süreç yürütür.

Gözetimsiz öğrenme algoritmalarının ana kullanım alanı veri kümesini benzerliklerine göre gruplara bölmektir. Gruplamanın ardından veri kümesi etiketlenebilir ve gözetimli öğrenme algoritmaları için kullanılarak daha detaylı incelenebilir. Gözetimsiz öğrenme işletmelerdeki büyük veri kümelerini anlamak ve problemleri çözmekte aktif olarak kullanılırlar. Örnek olarak sağlık hizmetlerindeki büyük ebatlardaki veri kümelerini ele alalım. Diyabet gibi özel hastalıklara ait semptomları içeren verilerin gruplanması sağlık personellerinin hastalığa ilişkin kavrayışını iyileştirebilir. Bu çözüm ile hızlıca aksiyon alınarak ciddi iyileştirmeler yapılabilir (Hurwitz ve Kirsch 2018).

2.3.3. Pekiştirmeli öğrenme

Pekiştirmeli öğrenme, ne yapılacağını öğrenmedir. Herhangi bir kural belirtilmez, sayısal bir ödül sinyali olup kullanılan algoritma da bu ödül sinyalini maksimize etmeye çalışır. Algoritmanın kendisinin kuralları ve aksiyonları keşfetmesi beklenir. Bu keşifler de ödül sinyalini maksimize edecek şekilde deneme yanılma yoluyla gerçekleşir. Deneme ve yanılma ile araştırma ve ödül sistematiği pekiştirmeli öğrenmenin en belirgin ve ayırt

edici özellikleridir. Pekiştirmeli öğrenmede, tanımlı algoritmalar yoktur, problemler tarafından karakterize edilirler. Problemin en önemli yönleri ele alınır ve bir öğrenme ajanına problemle etkileşebileceği alanda, bir hedef verilir. Öğrenme ajanı, çalıştırıldığı ortamdaki durumu gözlemleyebilmeli, aksiyon alabilmeli ve bu aksiyonlar doğrultusunda ödül döngüsünü çalıştırılabilmelidir (Sutton ve Barto 1998).

Popüler örnekler; çeşitli firmaların geliştirdikleri satranç ve go (çin daması) vb. strateji içeren oyun algoritmalarıdır. Ancak pekiştirmeli öğrenme algoritmaları endüstriyel ve mühendislik uygulamalarında da büyük potansiyele sahiptir. Pekiştirmeli öğrenme günümüzde robotik sistemler, üretim planlama, lojistik uygulamaları, ürün tasarımı, fabrika otomasyonu, kalite kontrol vb. uygulamalarda etkin olarak kullanılmakta ve geliştirilmeye devam edilmektedir.

2.4. Makine Öğrenmesi Modelleri

Algoritmalar, bilgisayarın verilerle nasıl etkileşime geçebileceğini, verileri manipüle ederek değiştirebileceğini ifade eden birtakım talimatlardır. Bir algoritma bir sütuna sayı eklemek gibi çok basit olabileceği gibi bir fotoğraftaki kişinin yüzünü tanımak şeklinde kompleks de olabilir. Bir algoritmanın kullanılabilir olabilmesi için bilgisayarların algılayabileceği bir program şeklinde derlenmesi gerekmektedir. Makine öğrenmesi algoritmaları, genellikle Java, Python ve R programlama dillerinde yazılırlar. Bahsedilen tüm programlama dilleri, çeşitli makine öğrenme algoritmalarını destekleyen kütüphanelere sahiptir. Ek olarak bu diller aktif kullanıcılara olan topluluklara sahiptirler ve kullanıcılar düzenli olarak fikirlerini ve programlarını paylaşarak zorlu problemleri çözmeye çalışırlar (Hurwitz ve Kirsch 2018).

Makine öğrenmesi algoritmaları diğer algoritmalarından farklıdır. Çoğu algoritmada programcı komut dizinlerini oluşturarak işe başlar. Ancak makine öğrenmesi işleminde bu durum tam tersidir. Makine öğrenmesi algoritmalarında veri kümeleri kullanılarak matematiksel modeller oluşturulur. Ne kadar çok veri eklenirse, algoritma o kadar çok gelişmiş ve karmaşık hale gelir. Daha çok veri ile çalıştırılan makine öğrenmesi algoritmaları daha doğru sonuçlar üretir. Doğru algoritmayı seçmek bilimin bir parçası

olduđu kadar sanattır da. Aynı problemi çözmek için uğraşan iki veri bilimcisi, aynı problemle karşılaşmalarına rağmen bambaşka yöntemler seçebilirler. Ancak farklı sınıflara ait makine öğrenme yöntemlerini bilmek, veri bilimcilere en uygun algoritmayı seçmekte yol göstermektedir (Hurwitz ve Kirsch 2018). Bu çalışmada kullanılan makine öğrenmesi yöntemleri ise; lineer regresyon, lojistik regresyon, destek vektör makineleri, rassal orman ve en yakın komşu algoritmasıdır ve aşağıdaki alt başlıklarda detayları yer almaktadır.

2.4.1. Lineer regresyon modeli

Regresyon kavramı hakkında ilk araştırmacılardan bir tanesi olan Sir Francis Galton 1885 yılında babalar ve oğullarının boylarına ilişkin bir araştırma yapmıştır. Çok uzun boylu babaların oğullarının da uzun boylu olmakla birlikte, ortalama olarak babalarından kısa boylu olduklarını ve çok kısa boylu babaların oğullarının ortalama olarak babalarından daha uzun boylu olduklarını bulmuştur. Galton oğulların boylarının ortalamaya doğru yönlendiđini veya ortalamaya yönelme eğilimi gösterdiđini (regressed) belirlemiş ve regresyon (bađlanım) kavramını ortaya atmıştır (McColl, 2020).

Birinci dereceden regresyon modeli lineer regresyon olarak adlandırılır. Lineer regresyonun en basit formu bir bađımsız deđişken ve bir bađımlı deđişkenden oluşan basit lineer regresyon modelidir. Genellikle bađımlı deđişken birden fazla bađımsız deđişkene bađlıdır ve bu duruma da çoklu lineer regresyon modeli denir. Standart çoklu deđişken regresyon modeli matris cebiri ile aşağıdaki şekilde ifade edilir;

$$y = X\theta + \varepsilon \quad (2.1)$$

Eşitlik (2.1)'de y , $(n \times 1)$ boyutunda bađımlı deđişken barındıran bir sütun vektörü, X ise; $(m \times k)$ boyutunda bađımsız deđişkenleri içeren bir matristir. θ , $(k \times 1)$ boyutunda regresyon katsayılarını barındıran bir vektördür ve ε ise $(m \times 1)$ boyutlu hata vektörüdür. Çoklu lineer regresyon modeli birden fazla bađımsız deđişkenin bir bađımlı deđişkeni açıkladıđı durumlarda seçilebilecek bir istatistiksel yöntemdir (Shah 2015).

Değişken sayısının fazla olması bağımlı değişkeni daha iyi açıkladığından çoklu doğrusal regresyon yaygın olarak kullanılmaktadır. Eşitlik (2.1)'de verilen doğrusal regresyon modeli genel olarak Eşitlik (2.2) halinde gösterilir (Montgomery, Peck ve Vining 2012).

$$y_i = \theta_0 + \theta_1 x_{i1} + \theta_2 x_{i2} + \dots + \theta_k x_{ik} + \varepsilon_i, \quad i = 1, \dots, n \quad (2.2)$$

Lineer bir model girdi değişkenlerindeki ağırlıkları hesaplayıp, bias (kesme) terimi olarak ifade edilen θ_0 'ı da modele ekleyerek tahmin yapar. Eğrinin y eksenini kestiği noktaya bias terimi denir ve bias terimi lineer model ile birlikte Şekil2.9'da görselleştirilmiştir. Lineer regresyon modelinin tahmini aşağıdaki şekilde ifade edilebilir;

$$\hat{y} = h_\theta(x) = \theta^T x \quad (2.3)$$

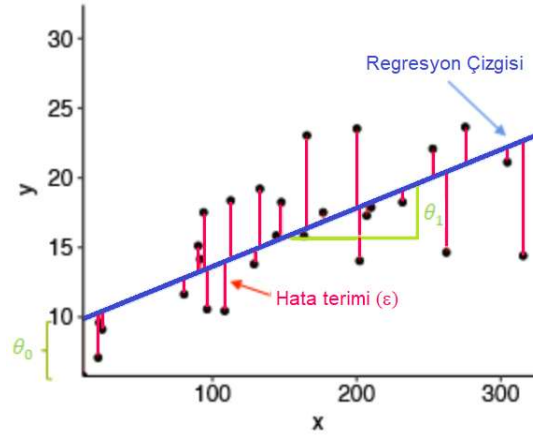
$$h_\theta(x) = [\theta_0 \quad \theta_1 \quad \dots \quad \theta_n] \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_n \end{bmatrix} = \theta^T x \quad (2.4)$$

\hat{y} , tahmin edilen değeri, θ , ağırlık parametresini, θ_0 bias terimini, θ_1 'den θ_n 'e kadar olan değerler regresyon katsayıları/ağırlıkları ifade eder. x girdi vektörü (içerisinde x_0 'dan x_n 'e kadar değişkenleri barındırır ve x_0 her zaman 1'e eşittir) ve h_θ hipotez fonksiyonudur. Lineer regresyon modeline öğretmek veri kümesine en iyi uyan θ parametrelerini bulmak demektir. Hata kareler ortalaması (HKO) ile veri kümesine en uygun θ parametreleri bulunabilir.

$$HKO = \frac{1}{n} \sum_{i=1}^m (y_i - \hat{y})^2 \quad (2.5)$$

Veri kümesindeki veri sayısı m , bağımsız değişkenleri barındıran vektörlerden i . sırasındaki örnek x_i ve x_i bağımsız değişkeninin sonucu olan bağımlı değişken ise;

y_i 'dir. X ise; tüm bağımsız değişkenleri barındıran ve veri kümesinin tamamını kapsayan bir matristir (Wall ve Jonas, 2020).



Şekil 2.11. Lineer Regresyon Grafiği

Yukarıda yer alan Şekil 2.11'de bir bağımsız değişken ve bir bağımlı değişkenden oluşan bir veri kümesine ait dağılım gösterilmektedir. Regresyon çizgisi θ_1 eğimi ile HKO' ya uygun şekilde veri kümesi içerinden geçmektedir.

2.4.2. Lojistik regresyon modeli

Lojistik regresyon, lineer regresyon modeline çok benzeyen makine öğrenme tekniklerinden biridir. Regresyon ismi sayısal bir hesaplama yöntemini çağrıştırırsa da lojistik regresyon, kategorik yapıdaki verileri sınıflandırma için kullanılır ve matematiksel olarak lineer regresyon modeli ile oldukça benzerlik gösterir. Lineer regresyon probleminde, bağımlı değişken sayısal iken, lojistik regresyon modelinde bağımlı değişken, sadece 0 ve 1 gibi kategorik değerler oluşur. Kuşkusuz bu kategorik değişkenin 0 ve 1 gibi iki şıklı olması durumunda ikili lojistik regresyona başvurulurken, 0, 1, 2, ..., vb. durumlarda çoklu lojistik regresyondan yararlanır. Örnek olarak; yarı iletken bir cihazın elektriksel testinin sonucunu ele alalım. Test sonucu ya başarılı yani cihaz çalışıyor veya başarısız yani fonksiyonel bir problem sebebi ile başarısızdır. Modelin aşağıdaki Eşitlik 2.6'daki gibi olsun;

$$y_i = x_i' \theta + \varepsilon_i \quad (2,6)$$

$x'_i = [1, x_{i1}, x_{i2}, \dots, x_{ik}]$, $\theta = [\theta_0, \theta_1, \theta_2, \dots, \theta_k]$ ve bağımlı değişken y_j , sadece 0 veya 1 olabilir. Bu durumda çıktı değişkeni olan y_i , Bernoulli rassal değişkenidir ve dağılım Çizelge 2.3'te gösterilmektedir (Montgomery, Peck ve Vining, 2012).

Çizelge 2.3.-Bernoulli rassal değişken dağılımı

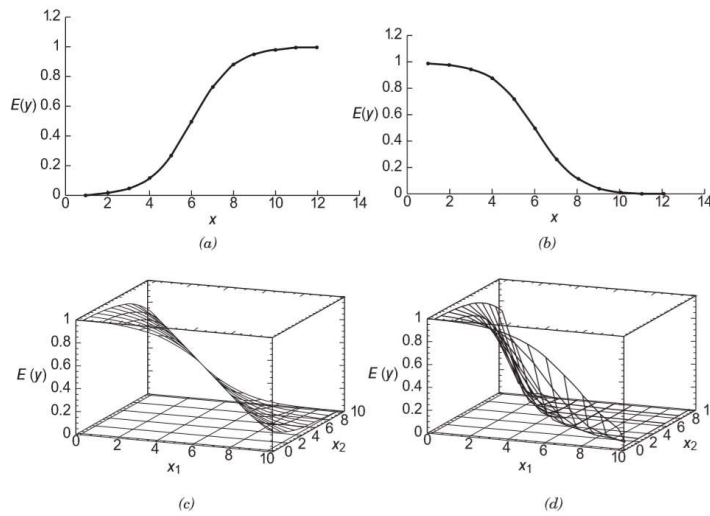
y_i	Olasılık
0	$P(y_i = 0) = 1 - \pi_i$
1	$P(y_i = 1) = \pi_i$

Eşitlik (2.6)'da belirtilen denklemi $E(\varepsilon_i) = 0$ olduğundan aşağıdaki şekilde yazabiliriz.

$$E(y_i) = x'_i \theta = 1(\pi_i) + 0(1 - \pi_i) = \pi_i \quad (2.7)$$

Bağımlı değişkenin ikili değer (binary) olduğu durumlarda deneysel çalışmalar lojistik fonksiyonun (response function) Şekil 2.12'de olduğu gibi artan veya azalan S şeklinde olduğunu göstermektedir. Lojistik fonksiyonun denklemi ise;

$$E(y) = \frac{1}{1+e^{-x}} \quad (2.8)$$



Şekil 2.12. Lojistik fonksiyon örnekleri (Montgomery, Peck ve Vining 2012)

Şekil 2.12’de yer alan (a) grafiğinin denklemi Eşitlik (2.9)’da, (b) grafiğinin denklemi Eşitlik (2.10)’da, (c) grafiğinin denklemi Eşitlik (2.11)’de ve (d) grafiğinin denklemi de Eşitlik (2.12)’de verilmiştir.

$$E(y) = 1/(1 + e^{-6.0+1.0x}) \quad (2.9)$$

$$E(y) = 1/(1 + e^{-6.0+1.0x}) \quad (2.10)$$

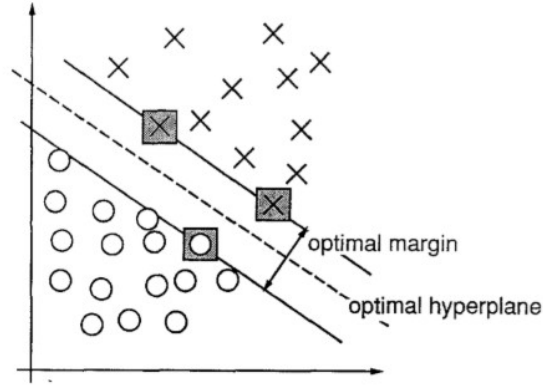
$$E(y) = 1/(1 + e^{-5.0+0.65x_1+0.4x_2}) \quad (2.11)$$

$$E(y) = 1/(1 + e^{-5.0+0.65x_1+0.4x_2}) \quad (2.12)$$

2.4.3. Destek vektör makineleri modeli

Destek vektör makineleri (Support Vector Machines, SVM), Vladimir Vapnik ve Alexey Chervonenkis tarafından geliştirilen, karar vermede bağımlı destek noktaları arasındaki genişliği en yükseğe çıkarmak için kullanılan ve örüntü tanıma, zaman serileri analizi, sınıflandırma gibi uygulamalarda başarıyla kullanılan gözetimli öğrenme algoritmalarıdır. SVM hem doğrusal hem de doğrusal olmayan verilere uygulanabilmektedir. Doğrusal olmayan x veri dizisini, daha yüksek boyuta taşıyacak olan fonksiyon haritalama fonksiyonu olarak tanımlanmakta ve $\Phi(x)$ ile gösterilmektedir. Bu durumda regresyon süreci, çekirdek fonksiyonu $K(x_i, x_j)$ ile yüksek boyutlu alana taşınmaktadır (Altan ve Hacıoğlu 2018).

Burada optimal bir hiper düzlem, iki sınıfın vektörleri arasında maksimum marj ile doğrusal karar fonksiyonu olarak tanımlanır (Bkz. Şekil 2.13). Gri karelerle işaretlenmiş destek vektörleri, iki sınıf arasındaki en büyük ayrım payını tanımlar.



Şekil 2.13. İki boyutlu uzayda ayrılabilir bir sorun örneği (Vapnik ve Cortes 1995)

Böyle optimal hiper düzlemler (hyper plane) oluşturmak için, kişinin sadece bu marjı belirleyen destek vektörleri denilen az miktarda eğitim verisini dikkate alması gerektiği gözlenmiştir. Eğitim vektörleri hatasız olarak optimal bir hiper düzlemlerle ayrılırsa, bir test örneğinde hata yapma olasılığının beklenen değerinin, destek vektörlerinin sayısının beklenen değeri arasındaki oranla sınırlandırıldığı gösterilmiştir (Vapnik ve Cortes 1995).

Giriş parametrelerinin $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_m, y_m)$ veri noktalarını içeren bir veri dizini olduğunu varsayalım. Burada x_i normalleştirilmiş giriş parametreleri (i bir girdi örneğini temsil eder) ve y_i değeri de i . girdi örneği altındaki çıktıdır. Toplam örnek sayısı m 'dir. SVM, Eşitlik (2.13)'teki formu kullanarak fonksiyona yaklaşır (Cao ve Tay 2003).

$$f(x) = \omega \cdot \phi(x) + b \quad (2.13)$$

Burada $\phi(x)$, x düşük boyutlu uzayından doğrusal olmayan bir şekilde eşlenen yüksek boyutlu bir özellik alanını temsil eder. ω ve b ise; Eşitlik (2.14)'teki düzenlenmiş risk fonksiyonu (regularized risk function) minimize edilerek hesaplanan katsayılarıdır.

$$\frac{1}{2} \|\omega\|^2 + C \frac{1}{m} \sum_{i=1}^m L_\epsilon(y_i, f(x_i)) \quad (2.14)$$

Burada $\|\omega\|^2$, düzenlenmiş (regularized) terimdir. $\|\omega\|^2$ 'nin en aza indirgenmesi fonksiyonu olabildiğince düz (flat) yapar. $\frac{1}{m} \sum_{i=1}^m L_\varepsilon(y_i, f(x_i))$ ise; aşağıdaki gibi tanımlanan ε duyarsız kayıp fonksiyonu (insensitive loss function) tarafından ölçülen ampirik hatadır (Vapnik, 1995):

$$L_\varepsilon(y_i, f(x_i)) = \begin{cases} 0, & y_i - f(x_i) \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon, & \text{aksi takdirde} \end{cases} \quad (2.15)$$

Bu, ε değerleri aralığını tanımlar; böylece tahmin edilen değer aralık dahilindeyse, kayıp sıfırdır ve tahmin edilen nokta aralığın dışındaysa, kayıp tahmin edilen değer ile mesafe arasındaki farkın büyüklü kadardır. C 'ye ceza faktörü (penalty factor) olarak da adlandırılan bir düzenleyici sabit (regularization coefficient) denir. Belirtilen katsayıların tamamı kullanıcı tarafından belirlenir.

ω ve b kestirimini elde etmek için, Eşitlik (2.14)'e ξ_i^* pozitif gevşeklik değişkenleri (positive slack variable) eklenerek orijinal objektif fonksiyon (original objective function) olan Eşitlik (2.16)'ya dönüştürülür.

$$\left. \begin{aligned} & \text{Minimize } \frac{1}{2} \omega^2 + C \frac{1}{m} \sum_{i=1}^m L_\varepsilon(y_i, f(x_i)) \\ & \text{Subject to } y_i - \omega \cdot \phi(x_i) - b \leq \varepsilon + \xi_i \\ & \omega \cdot \phi(x) + b - y_i \leq \varepsilon + \xi_i, \quad i = 1, \dots, m \\ & \xi_i^* \geq 0 \end{aligned} \right\} \quad (2.16)$$

Eşitlik (2.16)'nın çözülmesinin zor olduğu göz önüne alındığında, Eşitlik (2.17) eyer noktasını (saddle point) önceki eşitsizlik kısıtlamaları altında bulabiliriz. Dört Lagrange çarpanı, $a_i, a_i^*, \eta_i, \eta_i^*$, getirerek, Lagrange fonksiyonuna dönüştürebiliriz. Böylece fonksiyon Eşitlik (2.17)'de gösterilen hale gelir.

$$L = \frac{1}{2}\omega^2 + C \sum_{i=1}^m (\xi_i + \xi_i^*) - \sum_{i=1}^m (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^m a_i (\varepsilon + \xi_i - y_i + \omega \cdot \phi(x) + b) - \sum_{i=1}^m a_i^* (\varepsilon + \xi_i^* - y_i - \omega \cdot \phi(x) + b) \quad (2.17)$$

Dört Lagrange çarpanı kısıtlamaları $a_i, a_i^*, \eta_i, \eta_i^* \geq 0, (i = 1, 2, \dots, m.)$ aralığında sağlıyor ve eğer ilişkiler Eşitlik (2.18)'ü sağlıyorsa, Eşitlik (2.19)'i elde ederiz, bu da bizi Eşitlik (2.20)'ye ulaştırır

$$\frac{\partial L}{\partial \omega} = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial \xi^*} = \frac{\partial L}{\partial \xi} \quad (2.18)$$

$$\omega - \sum_{i=1}^m (a_i - a_i^*) x_i = 0 \quad (2.19)$$

$$f(x) = \sum_{i=1}^m (a_i - a_i^*) \phi(x_i) \cdot \phi(x) + b \quad (2.20)$$

Kernel (çekirdek) fonksiyonları; $K(x_i, x_j)$ 'yi kullanarak Eşitlik (2.20)'yi son hali olan Eşitlik (2.21)'ye dönüştürebiliriz

$$f(x) = \sum_{i=1}^m (a_i - a_i^*) K(x_i, x_j) + b \quad (2.21)$$

$K(x_i, x_j)$ 'nin değeri $\phi(x_i)$ ve $\phi(x_j)$ özellik alanındaki iki x_i ve x_j vektörünün iç çarpımına eşittir. Yani $K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$. Çekirdek fonksiyonları kullanılarak, gerekli tüm hesaplamalar $\phi(x)$ 'i hesaplamak zorunda kalmadan doğrudan giriş alanında yapılabilir. Popüler fonksiyonlar Çizelge 2.4'te ifade edilmiştir.

Çizelge 2.4. Çekirdek fonksiyonlar

Çekirdek fonksiyonlar	Çekirdek Foksiyon
Lineer	$K(x_i, x_j) = x_i \cdot x_j$
Polinomal	$K(x_i, x_j) = (x_i \cdot x_j + 1)^d$
Radyal temel (radial basis function, RBF)	$K(x_i, x_j) = \exp(-\gamma \ x_i - x_j\ ^2), \gamma > 0$

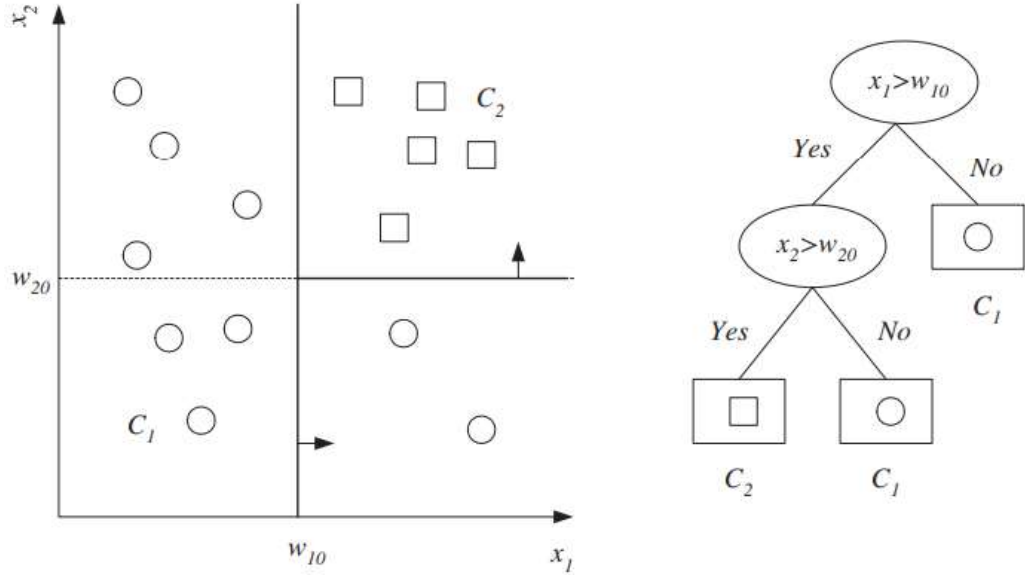
Çizelge 2.4’te belirtilen Polinomal ve RBF fonksiyonlarındaki d ve γ değerleri çekirdek parametreleridir. Farklı türde çekirdek fonksiyonları kullanarak, farklı türde karar yüzeylerine sahip çoklu makine öğrenme modelleri inşa edilebilir. RBF fonksiyonu SVM algoritmaları arasında en yaygın kullanılan çeşittir. Bu çalışmadaki analizlerde de RBF çekirdek fonksiyonu seçilmiştir (Shao, Wang X., Bu, Chen ve Wang Y. 2020).

2.4.4. Karar ağaçları ve rassal orman algoritmaları

Karar ağacı (decision tree), böl ve ele geçir stratejisini uygulayan hiyerarşik bir veri yapısıdır. Hem sınıflandırma hem de regresyon için kullanılabilen verimli ve parametrik olmayan bir yöntemdir. Belirli bir etiketli veri kümesi ile karar ağacı oluşturulabileceği gibi, anlaşılması kolay bir dizi basit kurala da dönüştürülebilirler (Alpaydın 2010).

Girdi alanı (input space), Öklid normu gibi bir mesafe ölçüsü ile tanımlanan yerel bölgelere (local regions) bölünür ve her girdi için, o bölgedeki eğitim verilerinden hesaplanan karşılık gelen yerel model kullanılır. Karar ağacı, denetimli öğrenme için hiyerarşik bir modeldir; burada yerel bölge, daha az sayıda adımda yinelemeli bölünmeler sırasında tanımlanır. Bir karar ağacı, iç karar düğümleri (internal decision nodes), karar düğümü (decision node) ve terminal yapraklarından (terminal leaves) oluşur (bkz. Şekil 2.14). Oval düğümler karar düğümleri iken, dikdörtgenler yaprak düğümlerdir. Tek değişkenli karar düğümü bir eksen boyunca ayrılır ve birbirini izleyen bölümler her birine diktir. Her karar düğümü m , dalları etiketleyen ayrı sonuçlara sahip bir $f_m(x)$ test fonksiyonunu uygular. Bir girdi verildiğinde her düğümde bir test uygulanır ve sonuca bağlı olarak dallardan (branch) biri alınır. Bu işlem kökten (root) başlar ve bir yaprak

(leaf) düğümü bulunana kadar tekrarlanır. Bu noktada yaprakta yazılan değer çıktıyı oluşturur. Ağaç yapısı bir öngörü olarak sabitlenmemiştir. Öğrenme sırasında verilere özgü sorunun karmaşıklığına bağlı olarak ağaç büyür, dallar ve yapraklar eklenir (Alpaydın 2010).



Şekil 2.14. Veri kümesi ve karşılık gelen karar ağacı örneği.

Literatürde farklı türde karar ağacı algoritmaları yer alsa da karar ağaçları, entropi (entropy) ve bilgi kazancı (information gain) olmak üzere iki ana karakteristik doğrultusunda hesaplama yapar (Mitchell 1997).

$$Entropy(S) = \sum_{i=1}^C -p_i \log_2 p_i \quad (2.22)$$

$$p_i = \frac{|c_{i,S}|}{|S|} \quad (2.23)$$

Burada p_i , i sınıfına ait S oranıdır ve Eşitlik 2.23'de gösterilmektedir. Eşitlik 2.22'de belirtilen logaritmanın tabanı ise 2'dir. Bunun nedeni entropinin, bit cinsinden ölçülen beklenen kodlama uzunluğunun bir ölçüsü olmasıdır. Hedef öznelik (attribute) olası değerleri alabilirse, entropinin log kadar büyük olabileceği unutulmamalıdır. Bir özelliğin eğitim verilerinin sınıflandırılmasındaki etkinliğinin ölçüsünü tanımlamak için Eşitlik 2.24'de belirtilen bilgi kazancı kullanılır (Mitchell 1997).

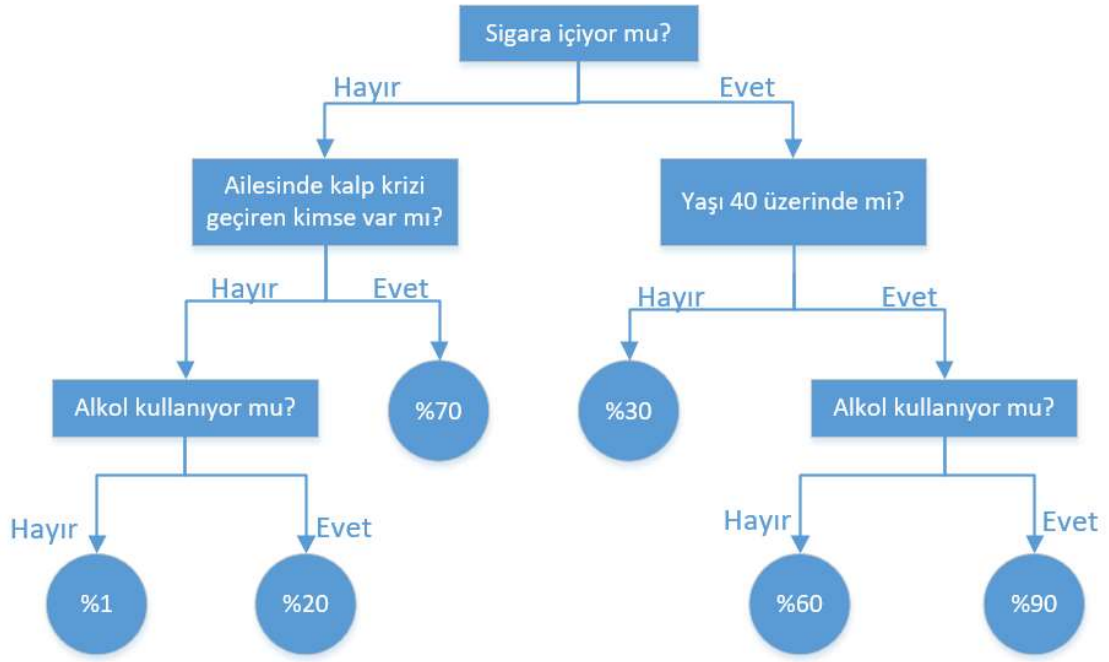
$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2.24)$$

Eşitlik 2.24’de Values(A), A niteliği için tüm olası değerlerin kümesidir. S_v ise; v özelliğinin S değerine sahip olduğu A’lerin alt kümesidir (Yani $s_v = \{s \in S | A(s) = v\}$) (Mitchell 1997).

Rassal orman algoritmaları (Random forest algorithm) çalıştırıldıklarında S kümesine bağlı tüm değişkenlerin bilgi kazancı ölçülür ve en yüksek bilgi kazancına sahip değişken seçilir. En yüksek kazanca sahip değişken kökü oluşturur ve dallar bu kök üzerinden ilerler. Yeni bir öznitelik seçme ve eğitim örneklerini bölümlere ayırma işlemi arta kalan her düğüm için tekrarlanır. Bu işlem her iki yaprak düğümü için aşağıdaki iki koşuldan biri karşılanana kadar devam eder:

- Her özelliğin ağaç boyunca bu yol üzerine eklenmesi ve
- Yaprak düğümü ile ilişkili eğitim örneklerinin tümünün aynı hedef değere sahip olması (yani entropinin sıfır)

Şekil 2.15’de belirtilen ağaç örneği yukarıda anlatılan rassal orman algoritmasının görselleştirilmiş halidir. İnsanların çeşitli özellikleri ele alınarak kalp krizi geçirme olasılıkları ele alınmaktadır (Belirtilen yüzdeler örnek amaçlı verilmiş tahmin değerleridir.). Tüm öznitelikler için entropi hesaplanır ve en yüksek öznitelik kök seçilir. Bu örnekte sigara içmek en yüksek entropiyi oluşturduğundan kök seçilmiştir. Diğer öznitelikler ise; alt düğümleri oluşturur ve entropi hesaplama işlemleri yukarıda açıklanan detaylar doğrultusunda tekrarlanır. En yaygın kullanılan karar ağacı algoritmalarına (ID3, C4.5, C5.0 ve CART) ise; bir sonraki sayfada değinilmiştir.



Şekil 2.15. Kalp krizi geçirme oranlarının görselleştirildiği karar ağacı (Şeker, 2012)

ID3 (Iterative Dichotomiser 3) algoritması, 1986 yılında Ross Quinlan tarafından geliştirilmiştir. Algoritma her bir düğüm için açgözlü (greedy) bir şekilde kategorik hedeflerde en büyük bilgi kazanımını sağlayacak özelliği bulan çok yollu bir ağaç oluşturur. Ağaçlar maksimum boyutlarına kadar büyütülür. Ardından da görünmeyen verileri (unseen data) genelleme yeteneğini geliştirme için budama (pruning) adımı uygulanır.

C4.5 eğitilmiş ağaçları (yani ID3 algoritmasının çıktısı) ise eğer-sonra (if-then) kural kümelerine dönüştürür ID3'ün ayrık aralık kümesi (discrete set of intervals) ve sürekli öznitelik değerini (continuous attribute value), bir aralık kümesine ayıran ayrık bir özniteliği (discrete attribute value) dinamik olarak tanımlayarak özelliklerin kategorik olma kısıtlamasını kaldırmıştır. Daha sonra her bir kuralın bu doğruluğu uygulanmaları gereken sırayı belirlemek için değerlendirilir. Budama, kuralın doğruluğu onsuz iyileşirse, bir kuralın ön koşulunun kaldırılması ile yapılır. C5.0 tescilli bir lisans altında Quinlan'ın en son sürüm algoritmasıdır. Daha az bellek kullanırken C4.5'te n daha küçük kural kümeleri oluşturur ve daha doğru sonuçlar verir.

CART (Classification and regression trees, sınıflandırma ve regresyon ağaçları) algoritması, C4.5'e çok benzeyen ancak sayısal hedef değişkenleri (regresyon) desteklemesi ve kural kümelerini hesaplamaması bakımında farklılık gösterir. CART, her düğümde en büyük bilgi kazanımını sağlayan özellik ve eşiği kullanarak ikili ağaçlar oluşturur. Scikit-learn, CART algoritmasının optimize edilmiş bir sürümünü kullanır. Ancak, scikit-learn uygulaması şimdilik kategorik değişkenleri desteklememektedir. (Bu çalışma Python dili kullanılarak hazırlanmıştır ve scikit-learn Python'da kullanılabilir en yaygın makine öğrenmesi kütüphanelerinden biridir. Scikit-learn'ün desteklediği algoritma olduğundan bu çalışmada CART kullanılmıştır.

2.4.5. K-en yakın komşu algoritması modeli

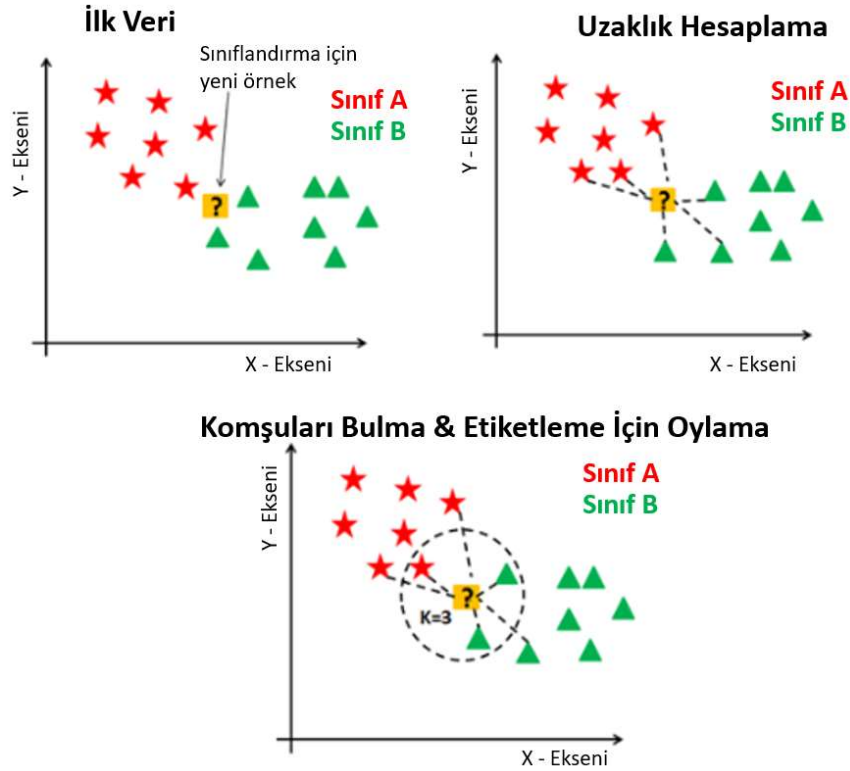
K-en yakın komşu (k-Nearest Neighbour, k-NN) algoritması ilk olarak 1950'lerin başında tanımlanmıştır. Bu algoritma, büyük eğitim kümeleri verildiğinde emek yoğunudur ve artan hesaplama gücünün mevcut olduğu 1960'lara kadar popülerlik kazanmamıştır. O zamandan beri, örüntü tanıma alanında yaygın olarak kullanılmaktadır. k-NN sınıflandırıcıları (classifiers); analogi yoluyla öğrenmeye, yani belirli bir test grubunu, ona benzeyen eğitim gruplarıyla karşılaştırmaya dayanırlar. Eğitim (Training) demetleri (tuple-çok ögeli değişken grup) n adet özellikle tanımlanır. Her bir demet n-boyutlu bir alanda bir noktayı temsil eder. Bu şekilde tüm eğitim demetleri n boyutlu bir desen alanında saklanır. Bilinmeyen bir demet verildiğinde en yakın komşu sınıfçı bilinmeyen gruba en yakın olan k eğitim demetlerinin desen alanını araştırır. Bu k eğitim demetleri, bilinmeyen demetin k “en yakın komşuları”dır (Han, Kamber ve Pei 2012). “Yakınlık” bir ölçü metriği olan Öklid mesafe ölçüsü ile tanımlanır. İki nokta veya iki demet arasındaki Öklid mesafesi, $x_1 = (x_{11}, x_{12}, \dots, x_{1n})$ ve $x_2 = (x_{21}, x_{22}, \dots, x_{2n})$ için denklem, Eşitlik 2.25 de ifade edilmiştir.

$$dist(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad (2.25)$$

Diğer bir deyişle, her sayısal özellik için, x_1 ve x_2 gruplarında bu özelliğin karşılık gelen değerleri arasındaki farkı n karelerinin toplamının karekökü hesaplanır. Farkların

karelerinin toplamı biriken mesafe sayısından alınır (Han, Kamber ve Pei 2012). Farkların bazılarının negatif bazılarının pozitif olması nedeniyle kareleri alınır. Kare alma işlemi fark değerlerini büyütür. En sonunda bu büyümenin etkisini gidermek üzere, karekök alınır.

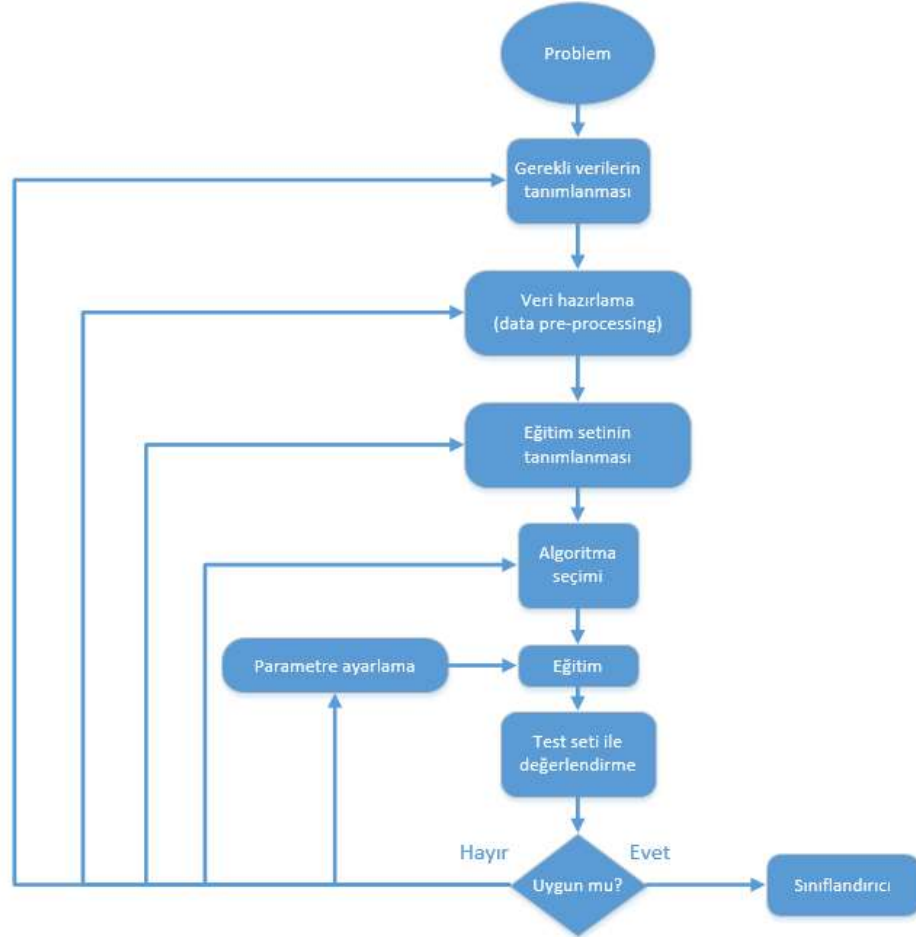
Şekil 2.16’da k-NN algoritması için iki sınıf görselleştirilmiştir; A sınıfı kırmızı yıldızlar ile ve B sınıfı yeşil üçgenler ile grafikte yer almaktadır. Bu noktalar etiketlenmiş ve bilinmektedir. Bilinmeyen bir noktanın hangi sınıfa ait olduğunu anlamak için K-NN algoritması kullanıldığında ise; bilinmeyen noktaya en yakın noktalar hesaplanır. Seçilen k değerine göre hesaplama yapılır. Görseldeki durumda $k=3$ ’tür. Yani bilinmeyen noktaya en yakın 3 noktanın Öklid mesafesi hesaplanır. Örnekte bilinmeyen nokta $k=3$ iken B sınıfına daha yakın olduğundan B sınıfına dahil edilir. Algoritmanın mantığı k parametresine bağlı olduğundan k değişkeni değiştirildiğinde çok farklı sonuçlar elde edilebilir.



Şekil 2.16. K-NN Algoritması Görselleştirme (Deepthi, 2019)

3. MATERYAL VE YÖNTEM

Makine öğrenmesi, örneklerden (eğitim kümesindeki örnekler) veya daha genel bir anlatımla, yeni örneklerden genelleştirmek için kullanılabilir bir sınıflandırıcı oluşturma kurallarını öğrenme sürecidir. Gözetimli makine öğrenmesi gerçek dünya problemlerine uygulanan akış şeması, Şekil 3.1’de görselleştirilmiştir (Konsiantis 2007).



Şekil 3.1. Gözetimli makine öğrenim süreci (Konsiantis 2007)

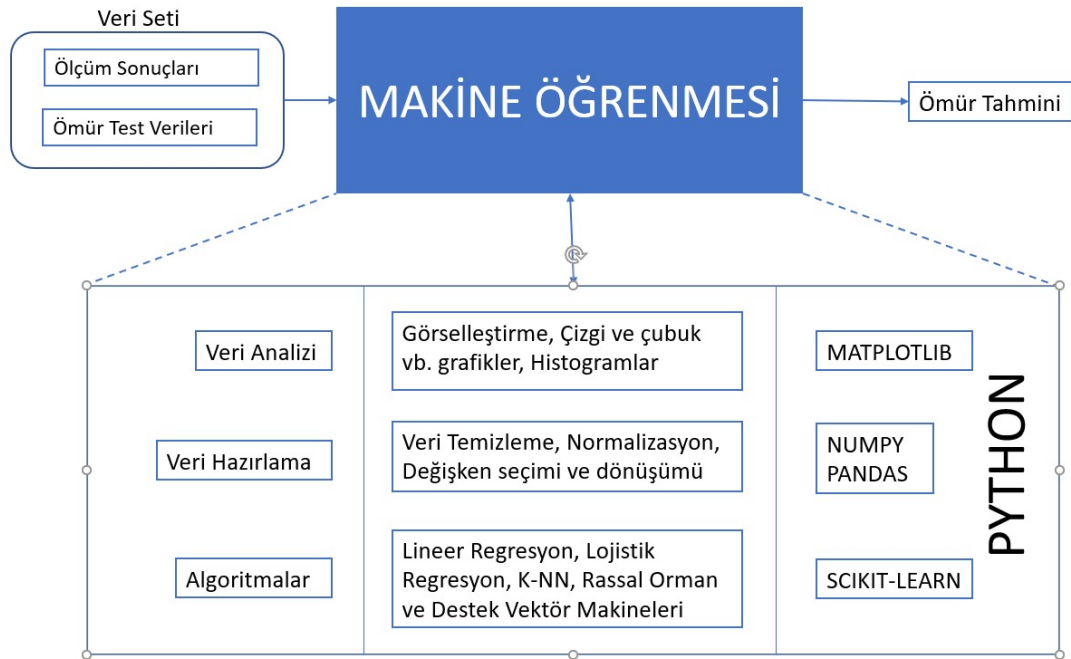
Şekil 3.1’de de görüldüğü gibi makine öğrenmesinin ilk adım veri kümesini toplamaktır. Problem hakkında tecrübeli bir uzman varsa, hangi alanların (nitelikler, özellikler) en bilgilendirici olduğunu önerebilir. İkinci adım ise veriyi hazırlamaktır. Koşullara bağlı olarak, araştırmacılar eksik verilerle karşılaşabilir ve eksik verileri işlemek için seçebilecekleri çeşitli yöntemleri vardır (Batista ve Monard, 2003). Hodge ve Austin (2004) ise aykırı (gürültü) verilerin tespiti için çağdaş tekniklerin bir araştırmasını

sunmuşlardır. Değişken seçimi yalnızca gürültüyü (noise) işlemek için değil, aynı zamanda çok büyük veri kümelerinden öğrenmenin olanaksızlığıyla baş etmek için de kullanılır. Veriler azaltılarak algoritmanın büyük veri kümeleriyle etkili bir şekilde çalışması sağlanır. Değişken seçimi, mümkün olduğunca sonuç üzerinde etkisi çok az veya etkisiz özelliklerin tanımlanması ve kaldırılması işlemidir (Yu ve Liu, 2004). Bu, verilerin boyutsallığını azaltır ve algoritmaların daha hızlı ve daha etkin çalışmasını sağlar. Birçok özelliğin birbirine bağlı olması, genellikle gözetimli makine öğrenmesi sınıflandırma modellerinin doğruluğunu (accuracy) etkiler. Bu sorun, temel özellik kümesinde yeni özellikler oluşturularak giderilebilir (Markovitch ve Rosenstein 2002). Bu tekniğe özellik yapımı/dönüşümü denir. Bu yeni oluşturulan özellikler sayesinde daha özlü ve doğru sınıflandırıcılar oluşturulabilir. Buna ek olarak, anlamlı özelliklerin bulunması, üretilen sınıflandırıcının daha iyi anlaşılabilirliğine ve öğrenilen kavramın daha iyi anlaşılmasına katkıda bulunur (Konsiantis 2007).

Şekil 3.1'deki Konsiantis'in yaklaşımı üzerinden ilerlersek, bu çalışmadaki problem, su çekici testine giren plakalı eşanjörlerin çevrim sonuçlarını tahmin etmektir. Çalışma, 2018 ve 2019 yıllarında yapılmış ölçüm sonuçları ve test verileri ile sürdürülmüştür. Hazırlanan veri görselleştirmeleri ile plakalı eşanjör ömrü üzerinde etkisi olmayan bağımsız değişkenler analizlerden çıkartılmıştır. Veri hazırlamaya ilişkin detaylar Bölüm 3.3'te belirtilmiştir. Eğitim kümesi ise Python dili için yazılmış, "scikit-learn" adlı makine öğrenmesi kütüphanesi ile hazırlanmıştır. Veri kümesi "train_test_split()" fonksiyonu ile %80 eğitim kümesi ve %20 test kümesi olacak şekilde ikiye ayrılmıştır. Sonraki aşama ise algoritma seçimidir. Seçilen algoritmalar, yapılan analizler ve tahmin edilen sonuçlar doğrultusunda ikiye ayrılmaktadır. Sayısal bir tahmin elde etmek için regresyon yöntemlerinden biri olan lineer regresyon yöntemi seçilerek analiz edilmiştir. Ancak Bölüm 3.2'de de belirtilen test koşulları nedeniyle, analizler doğruluk oranları limitlidir. Tahmini farklı bir yaklaşımla gerçekleştirmek için nümerik olan bağımlı değişken, kategorik hale dönüştürülmüştür. Böylece test sonuçları başarılı (1) ve başarısız (0) olarak ayrılmış ve mevcut veriler ile sınıflandırma analizleri yapılmıştır.

Literatürde çok çeşitli sınıflandırma algoritmaları yer almaktadır. Veri kümesindeki değişkenler, değişkenlerin çeşitleri (nümerik, kategorik, vb.) ve veri kümesinin

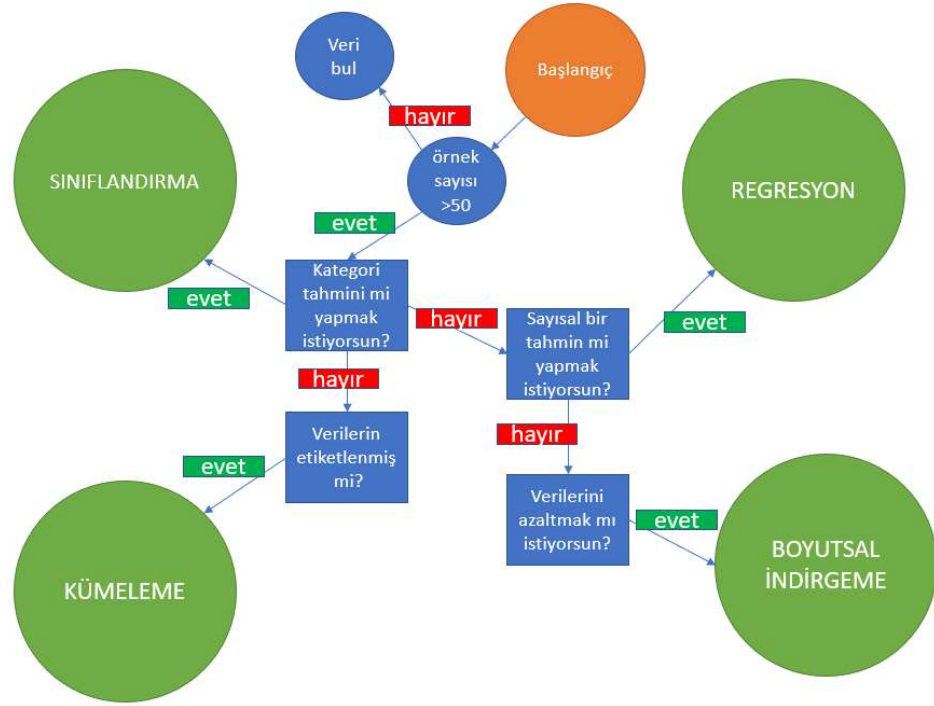
büyüklüğü göz önüne alınarak; lojistik regresyon, destek vektör makineleri, rassal orman ve k en yakın komşu algoritmaları seçilerek analizler yürütülmüştür. Seçilen algoritmalar ile sınıflandırma analizleri yapılmış ve test sonuçları değerlendirilmiştir. Test sonuçları doğrultusunda ise parametreler değiştirilerek incelemelere devam edilmiştir. Kuşkusuz algoritmaların matematiksel modelleri birbirlerinden farklı olduğundan parametreleri de birbirlerinden farklıdır. Bulgular ve tartışma bölümünde parametre değişikliklerinden detaylı bir şekilde bahsedilmiştir. Parametreler optimize edildikten sonra da algoritmaların başarıları karşılaştırılarak en başarılı algoritma seçilmiştir. Doğruluk oranları ve seçilen algoritmaya ilişkin bulgular sonuç bölümünde detaylı şekilde açıklanmıştır. Şekil 3.2’de ise makine öğrenmesine ait girdi ve çıktılar, işlem adımları, adımların açıklamaları ve bu adımları gerçekleştirmek için gerekli Python kütüphaneleri yer almaktadır.



Şekil 3.2. Makine öğrenmesinin girdi ve çıktıları, işlem adımları ve bu adımlar için kullanılan Python kütüphaneleri

Veri bilimcilerin problemleri ele alırken gerçekleştirdikleri ilk adım veri görselleştirmedir. Veri kümesindeki değişkenlerin dağılımları grafiklere dökülerek; değişken seçimi ve algoritma seçimi gibi belirleyici kararların alınması kolaylaşır.

Matplotlib görselleştirme için bu çalışmada kullanılan Python kütüphanesidir. Numpy ve Pandas kütüphaneleri ise veri hazırlama adımı kullanılmıştır. Makine öğrenmesi algoritmaları, kompleks hesaplamaları daha hızlı sonuçlandırabilmek için matrisleri kullanır ve Numpy, veri kümesini matris formatına dönüştürmektedir. Pandas'ın ana kullanım alanı ise veri çerçevesi (framework) oluşturmaktır. Veri işleme, dönüştürme, temizleme ve düzenleme gibi çeşitli adımlarda yaygın olarak kullanılan bir kütüphanedir. Scikit-learn ise çeşitli makine öğrenme algoritmalarını barındıran açık kaynak kodlu bir kütüphanedir. Şekil 3.3'te algoritma seçiminde kullanılan yol haritası yer almaktadır. Scikit-learn web sitesinde yer alan görselden esinlenerek seçim kriterleri görselleştirilmiştir. Web sitesinde ek detaylar yer almaktadır.



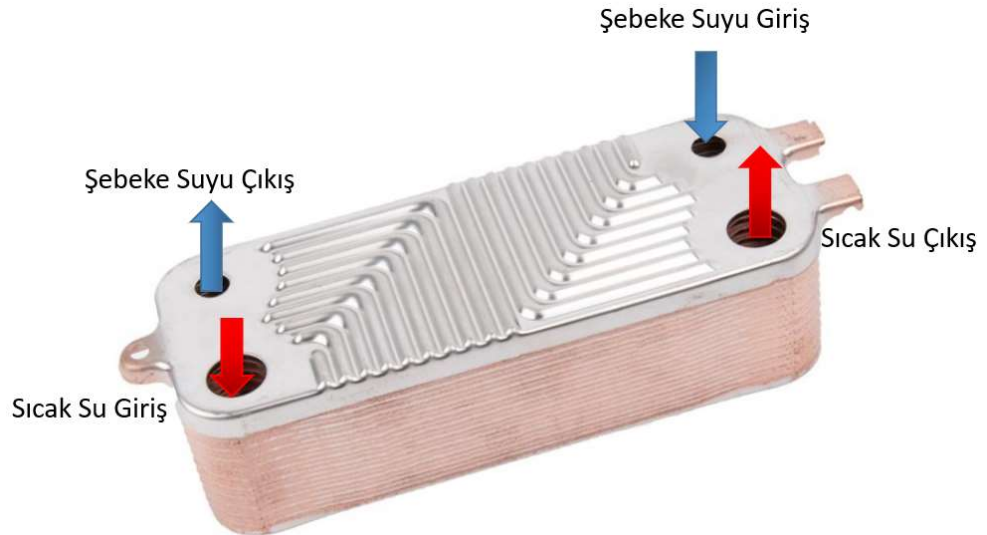
Şekil 3.3. Scikit-learn algoritma seçimi için yol haritası (Anonim, 2020c)

Şekil 3.3'teki görsel karmaşık görünse de regresyon, sınıflandırma, kümeleme ve boyutsal indirgeme olmak üzere dört ana başlıktan meydana gelmektedir. Regresyon analizi sayısal tahmin yapmak için kullanılır ve görselde veri kümesinin büyüklüğüne ve değişkenlerin özelliklerine göre kısımlar belirtilmiştir. Sınıflandırma ise etiketlenmiş veri kümesi üzerinde sınıflara ayırma için kullanılan ve Bölüm 2 literatür taramasında

detaylıca açıklanmış seçenektir. Regresyon analizinde olduğu gibi veri kümesinin büyüklüğü, dağılımı ve çeşitli özellikleri algoritma seçiminde farklılıklar yaratmaktadır. Kümeleme, etiketi olmayan ham veri olarak adlandırılabilir veri kümelerine uygulanan bir yöntemdir. Genellikle gözetimsiz öğrenme algoritmaları ise veri kümesinin farklı kümelerine yani etiketlere ayrılır. Boyutsal indirgeme işlemleri ise; büyük veri kümelerinde sonuç üzerinde etkisi düşük olan bağımsız değişkenleri bulmak ve veri kümesini sadeleştirmek için kullanılan bir yöntemdir.

3.1. Problemin Tanımlanması

Plakalı eşanjörler, evsel tesisat hattına (mutfak, banyo, vb.) sıcak su sağlamak için kombilerde ikincil bir ısı eşanjörü olarak bulunur. Nispeten yüksek basınçlarda çalışırlar. Plakalı eşanjörler, ısı çıkış gereksinimlerini karşılamak için birkaç sayıda kanal plakasından oluşur. Plakalı eşanjörler merkezi ısıtma ve kullanım suyu olmak üzere iki farklı su kanalından oluşur. Merkezi ısıtma kanalı, ısı hücresi veya yanma odası olarak adlandırılan birincil ısı eşanjöründen gelen sıcak suyu taşır. Kullanım suyu ise şehrin ana su sisteminden gelen soğuk musluk suyunu taşır. Şekil 3.4'de giriş ve çıkış bölgeleri belirtilmiştir:

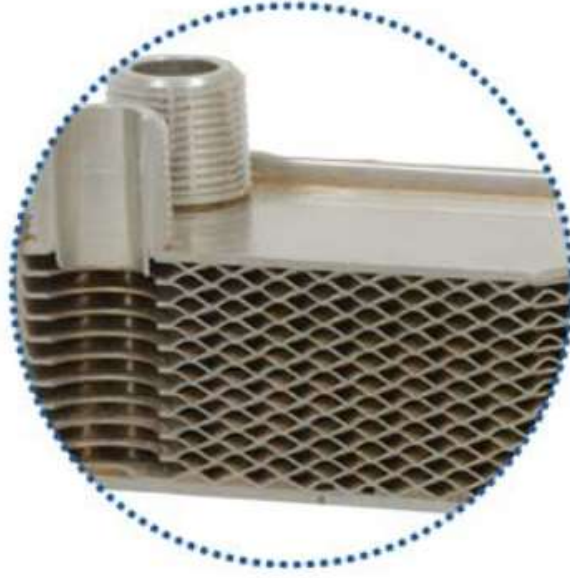


Şekil 3.4. Lehimli plakalı eşanjör su giriş-çıkış gösterimi (Anonim, 2020d)

Kullanım sıcak suyu devresindeki plakalı eşanjörlerde kullanıcı alışkanlıkları nedeniyle, kanal basıncında ani dalgalanmalar meydana gelebilir. Öte yandan; bu durum mutfak, banyo veya benzeri evsel kullanım suyu hattındaki musluklardan birinin aniden kapanmasıyla da meydana gelebilir. Bu davranış su çekici (water hammer) olarak adlandırılır. Su çekici basıncının büyüklüğü, şehir ve ülkenin ana su sistemi ile doğrudan ilişkilidir. Bosch Termoteknik tarafından yapılan çalışmalarda, çeşitli ülkeler üzerindeki minimum su darbesi basınç değerlerinin ortalaması 16 bar olarak gözlemlenmiştir. Bu değer ek olarak EN (European Norms) standartlarına da uygundur. Su çekici plakalı eşanjörlerin ömürleri boyunca ortaya çıktığı için yorulma, değerlendirilirken göz önüne alınması gereken önemli bir parametredir.

Bosch Termoteknik'te plakalı eşanjörlerde oluşan yorulmayı belirlemek için su çekici testi kullanılmaktadır. Test standına bağlanan plakalı eşanjörlere kombinin çalışma koşullarında arka arkaya su çekici etkisi yaratacak basınç uygulanır. Bu koşullar eşanjörde bir yırtılma dolayısıyla kaçak oluşana kadar sürdürülür. Plakalı eşanjörlerin ömrü 15 yıl olarak kabul edilmektedir ve bu da 300.000 kez su çekici darbesine maruz kalma olarak öngörülmektedir. Test standı bu fenomeni test etmekte başarılı olsa da 300.000 çevrime ulaşma süresi ortalama olarak 12 gündür.

Kanal plakaları oldukça incedir ve yumuşak bir malzeme olan bakırdan üretilmektedir. Şekil 3.5'de lehimli plakalı eşanjörün kesit alınmış hali yer almaktadır. Şekilde de görüldüğü üzere, tasarım geometrisi karmaşık olup üretim de oldukça hassastır. Üretim aşamalarında oluşabilecek bir sapma yüzünden üst üste konan iki plakada örtüşmeyen noktalar oluşabilir. Örtüşmeme durumu firmada lehim boşluğu adı verilen hata türüne neden olabilir.



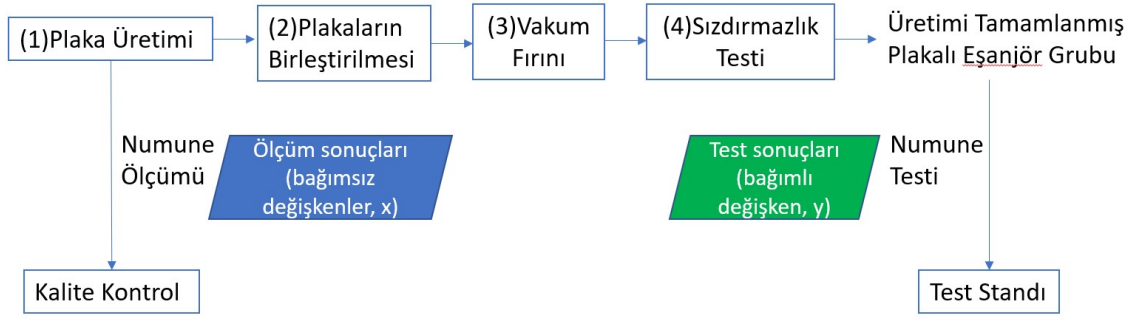
Şekil 3.5. Lehimli plakalı eşanjör kesit alınmış iç görünümü (Anonim, 2020e)

Bu boşluklar plakalı eşanjör içerisindeki sıcak ve soğuk su hatlarının karışmasına yol açıp eşanjörün verimini düşürebilir. Ek olarak su çekici fenomeni en zayıf halka olan bu boşluklarda zaman içerisinde yırtılmalara ve eşanjörün planlanan ömründen daha kısa sürede kullanılamaz hale gelmesine neden olabilir. Seri üretimde yaşanan işlem parametre sapmalarının plakalı eşanjörün ömründeki etkisini gözlemlemenin en etkili yöntemi su çekici testidir. Ancak belirtildiği gibi, bu test yaklaşık olarak 12 gün sürmektedir. Günümüzdeki seri üretim koşullarında bu bekleme süresi kabul edilemez. Bu çalışmanın amacı karar verme sürecini hızlandırmak ve kaliteyi güvence altına almaktır. Bu amaçla geçmiş dönem verileri, makine öğrenmesi algoritmaları ile incelenerek matematiksel modeller oluşturulmuştur. Bu modeller sayesinde teste girmemiş plakalı eşanjörlerin test sonuçları saniyeler içerisinde tahmin edilebilmektedir.

3.2. Veriyi Anlamak

Çalışmada kullanılan veri kümesi Bosch Termoteknik Manisa fabrikasından temin edilmiştir. Bu veri kümesi 2018 ve 2019 yıllarına ait ölçüm ve test sonuçlarını kapsamaktadır. Veri analizinde bağımsız değişkenler olarak ele alınan ölçüm sonuçları,

İPK (İstatistiksel Proses Kontrol) kapsamında kalite kontrol biriminin seri üretim içerisinde gerçekleştirdiği ölçümleri kapsamaktadır. Test sonuçları ise, detayları Bölüm 3.1’de belirtilen su çekici test standından elde edilmiş ve bağımlı değişken olarak ele alınmıştır. Firmaya ait gizli bilgilerin korunması amacıyla, malzeme bilgisi, üretim parametresi, teknik resim veya ölçüm sonucu paylaşılmamıştır. Analizde kullanılan ölçüm verilerinin buldukları konumlar kabaca gösterilmiş, kesin konumları belirtilmemiştir. Analizde kullanılan tüm veriler Bosch Termoteknik’e aittir ve bu çalışmada ürün ve üretim süreçleri gibi firmaya ait hassas bilgilerin gizliliği esas alınmıştır. Üretim, ölçüm ve test aşamaları şematik olarak Şekil 3.6’te görselleştirilmiştir.



Şekil 3.6. Plakalı eşanjör üretim, kalite kontrol ve test aşamaları

İlk adım plakalı eşanjörlerin plakalarının üretilmesidir. Ardından preslerin yardımı ile plakalar birleştirilir ve vakum fırınlarına gönderilirler. Vakum fırınlarında bakır plakalar erir ve çelik plakalarla birleştirilerek lehimli plakalı eşanjör üretimi tamamlanır. Birleştirilmiş plakalı eşanjörler, vakum fırınına gruplar halinde yerleştirildiğinden, uygulanan üretim biçimi grup bazlı üretimdir (batch production) ve fırından çıkan grup üretim partisi olarak adlandırılır. Plaka üretim aşamasında kalite kontrol birimi düzenli olarak kanal plakalardan numune alarak ölçüm yapmaktadır. Veri kümesindeki bağımsız değişkenler bu ölçüm sonuçlarıdır. Plakalı eşanjörün kontrolü ise sızdırmazlık testi ile gerçekleştirilir. Ancak sızdırmazlık testi basınçlı hava ile kaçakları kontrol ettiğinden, ürünün ömrü hakkında bilgi sağlamamaktadır. Ömür hakkında kontrolleri sağlayabilmek için su çekici testi yapılır.

Test standına toplamda 8 adet plakalı eşanjör bağlanabilmektedir. Üretilen partiden 8 adet numune rastgele seçilir ve test düzeneklerine bağlanarak test edilirler. Kalite kontrol biriminin yaptığı numune ölçümleri de parti başına tek plaka ile gerçekleştirildiğinden, verileri eşleştirmek için 8 eşanjörün çevrim sayısı ortalaması alınır. Farklı eşanjörler, test standında farklı performanslar gösterebildiğinden üretim partisinin kalitesi ve ömrünü kontrol etmek için ortalama ömürler alınmıştır.

Teste giren eşanjörlerin tamamının 300.000 limitine ulaşması durumunda ise test başarılı kabul edilmekte ve test durdurulmaktadır. Hedeflenen ömür süresine ulaşıldığı için teste devam etmek ek zaman ve gereksiz maliyet anlamına gelir. Ancak bu durum ömrü daha uzun plakalı eşanjörlerin ömrünün ölçülememesi anlamına da gelir. Firmada geçmiş dönemlerde yapılan denemelerde teste devam edildiğinde 300.000 limitinin oldukça üzerinde çevrim sonrasında bile çalışmaya devam edebilen eşanjör numuneleri gözlemlenmiştir. Hedeflenen aralığa ulaşıldığında testin durdurulması oldukça mantıklı olsa da bu durum veri analizinde ne yazık ki tutarsızlıklara yol açmaktadır. Yapılan ömür testlerindeki kabul kriterleri ise Çizelge 3.1’de belirtilmiştir.

Çizelge 3.1. Su Çekici Testi Değerlendirme Kriterleri

Senaryo	Sonuç	Yorum
Önceden belirlenen A adet eşanjörde, minimum çevrim değeri olan önceden belirlenen B limiti öncesi sızdırma	Başarısız (0)	Test durdurulur, eşanjörler bloke alanına alınır.
8 eşanjörün tamamı 300.000 çevrime ulaşmasa da hepsi B çevrimi barajını geçmesi	Şartlı Kabul (1)	Test durdurulur, eşanjörlerde ek inceleme ve risk değerlendirmesi yapılır.
8 eşanjörün tamamının 300.000 çevrime ulaşması	Başarılı (1)	Test durdurulur, eşanjörlere serbesti verilir.

Çizelge 3.1’de A ve B olarak belirtilmiş değerler yer almaktadır. Bu değerler aslında nümerik sayılardır ancak firma gizli bilgilerini saklamak için çalışmada bu şekilde ifade edilmiştir. Test yapılan 8 plakalı eşanjör içerisinde en fazla A tanesinde B çevrim öncesinde sızdırma yaşanırsa test başarısız olarak kabul edilir ve üretilen parti blokeye alınır. B çevrim limiti Bosch Termoteknik tarafından geçmiş yıllardaki üretim

değişkenlikleri ve saha şikayetleri göz önüne alınarak belirlenmiş bir alt sınırdır. Başarılı ve başarısız kabul kriterlerine ek olarak şartlı kabul senaryosu da vardır. Örneğin test yapılan sekiz eşanjörden ilki 235.000, üçü 220.000, geriye kalan dördü ise 300.000 çevrim ile testi tamamlamış olsun. Eşanjörlerin tamamı 300.000 sınırına ulaşamamış olsa da eğer B limitini aşmışlarsa firmadaki risk değerlendirme ekibi bir araya gelir. Üretim proses parametreleri, ölçüm ve test sonuçlarını analiz ederek karar verir. Bu senaryo şartlı kabul olarak tanımlanmaktadır.

Ömür test sonuçlarındaki 300.000 çevrim limiti nedeniyle, nümerik ömür tahminleri sayısal olarak belirli bir doğruluğa kadar tahmin edilebilmektedir. Ancak veri kümesindeki nümerik olan çevrim sayısı Çizelge 3.1'deki değerlendirme kriterleri kullanılarak başarılı (1) ve başarısız (0) olarak kategorik hale getirilmiştir. Bu sayede nümerik tahminleri içeren regresyon analizlerine ek olarak çeşitli yöntemler ile sınıflandırma analizleri de yapılabilir.

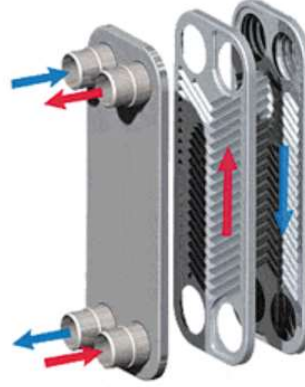
Veri analizinde bulunan bağımlı ve bağımsız değişkenlere ait açıklamalar ve kısaltmalar ise Çizelge 3.2'de yer almaktadır. Plakalı eşanjör içerisinde kanal plakalar sıcak ve soğuk plaka olarak adlandırılmakta ve ikiye ayrılmaktadır. Merkezi ısıtma kanalının temas ettiği plaka sıcak plaka, şebekeden gelen kullanım suyunun temas ettiği plaka ise soğuk plakadır

Çizelge 3.2. Bağımsız değişkenler, açıklamaları ve kısaltmalar

SICAK PLAKA BÖLGESEL (mm)						
Arka Sıra 1 (AS1)	Arka Sıra 2 (AS2)	Arka Sıra 3 (AS3)	Orta (O1)	Ön Sıra 3 (ÖS3)	Ön Sıra 2 (ÖS2)	Ön Sıra 1 (ÖS1)
SOĞUK PLAKA BÖLGESEL (mm)						
Arka Sıra 1 (AS11)	Arka Sıra 2 (AS21)	Arka Sıra 3 (AS31)	Orta (O11)	Ön Sıra 3 (ÖS31)	Ön Sıra 2 (Ö21)	Ön Sıra 1 (ÖS11)
SICAK-SOĞUK PLAKA AÇI UYUMU/FARKI (mm)				DÜZLEMSELLİK		
a3 (Sıcak Ön- Soğuk Ön)	a4 (Sıcak Sağ- Soğuk Sol)	a5 (Sıcak Arka- Soğuk Arka)	a6 (Sıcak Sol- Soğuk Sağ)	F1	F2	

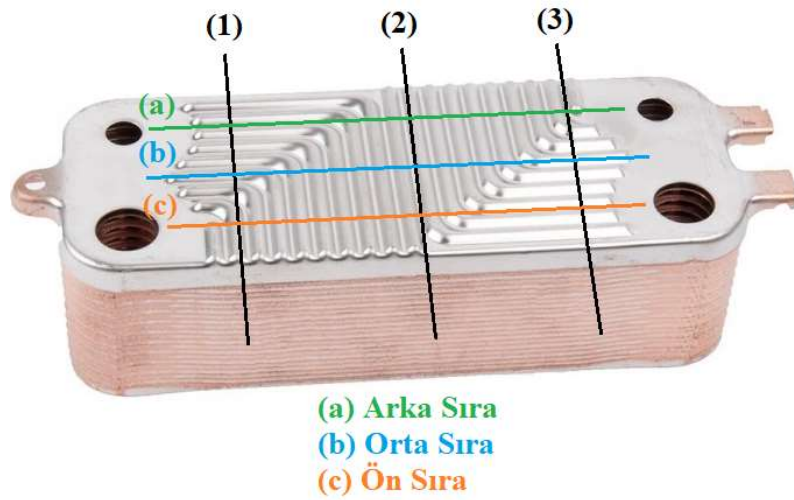
Firmanın gizli bilgisi olması nedeniyle ile plakalarla ilgili görsel Şekil 3.7, örnek vermek amacıyla internetten alınarak kullanılmıştır. Sıcak ve soğuk iki plaka arasındaki temel fark, görselde de yer aldığı gibi plaka iç geometridir. Plakaların iç yapılarındaki bu

farklılık sayesinde, üst üste konduklarında temas noktası artmakta ve daha iyi lehim yapılmaktadır. Bu durum da dayanım değerlerini iyileştirmektedir.



Şekil 3.7. Plakalı eşanjör ölçüm bölgeleri şematik gösterimi (Anonim, 2020f)

Çalışmada bilgi güvenliği nedeniyle, plakalar üzerindeki ölçüm yapılan noktalar da paylaşılammıştır. Şekil 3.8 deki görsel, internetten alınmış ve ölçüm yapılan bölgeler sembolik olarak belirtilmiştir. Örneğin; Arka Sıra 3 olarak belirtilen ölçüm bölgesi; (a) çizgisi ile (3) çizgisinin kesişim noktasıdır. Çizelge3.2'deki düzlemsellik değerleri, plakalar üzerinde çeşitli noktaların ölçümleri yapılarak hesaplanan bir ölçüm değeridir. Plakaların katlanan kenar bölgeleri 90° üretilseydi birbirleri üzerine konamazlardı. Bu nedenle bu kenarlar preslerde belli bir açıda katlanmaktadır. Çizelge3.2'de yer alan a3, a4, a5 ve a6 değişkenleri ise sıcak ve soğuk plaka kenarlarındaki açılal farklardır.



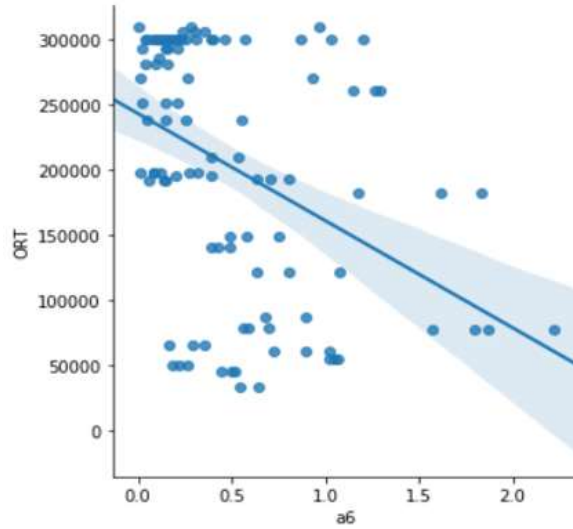
Şekil 3.8. Plakalı eşanjör ölçüm bölgeleri şematik gösterimi (Anonim, 2020g)

3.3. Veriyi hazırlama

Veri analizine geçilmeden önce, verilerin düzenlenmesi ve hazır hale getirilmesi oluşturulan modelin doğruluğu için son derece önemlidir. Çeşitli kaynaklarda, farklı temizleme, birleştirme, dönüştürme, indirgeme vb. işlemler dikkate alınmaktadır. Ancak veri hazırlama süreci standardı olmayan, mevcut veri kümesine göre değişkenlik gösteren bir süreçtir. Bu çalışmada kullanılan işlemlere aşağıda değinilmiştir.

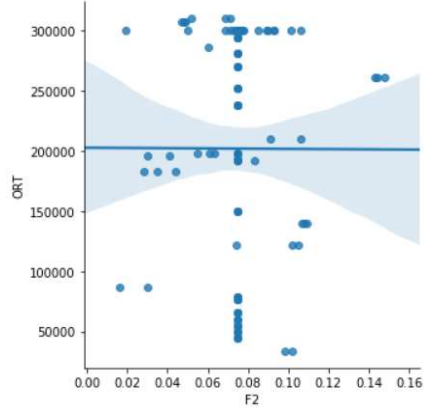
3.3.1. Değişken seçimi

Veriler incelenmeye başlandığında, toplamda 20 değişken vardı. Tüm değişkenler ile ortalama çevrim sayıları serpilme diyagramı (scatter plot) ile görselleştirilmiştir. Şekil3.9'daki grafik, değişkenlerden a6 ile ortalama çevrim sayısına aittir. (Bu veri kümesinden geçen doğru lineer regresyon ile hesaplanmıştır) Grafikte de görüldüğü gibi a6 değişkeninin değeri arttıkça ortalama çevrim düşmektedir.



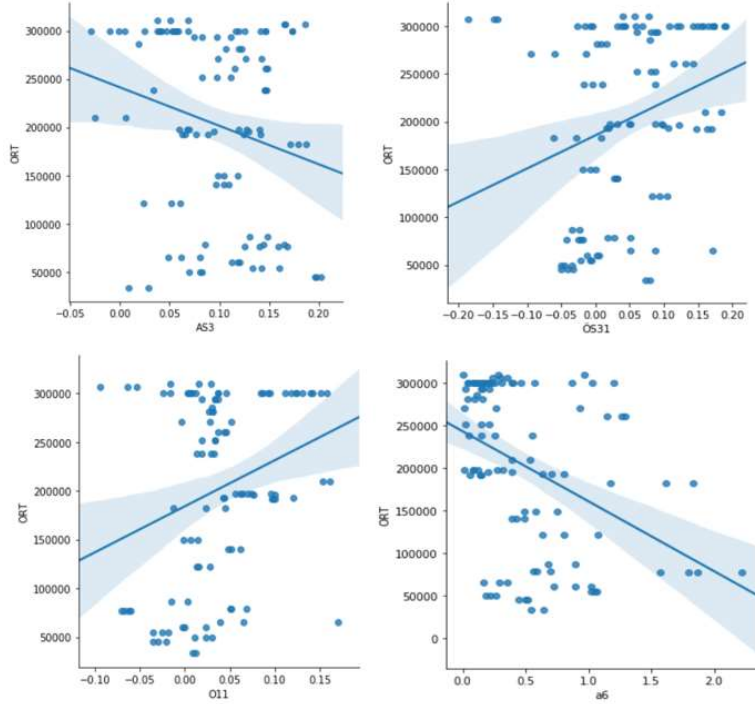
Şekil 3.9. Çevrim sayısı ortalaması (ORT) ve a6 değişkeninin serpilme diyagramı

Ancak Şekil3.10'daki grafikte ise F2 değişkeni ile yine ortalama çevrim sayıları incelendiğinde, görüldüğü üzere veri kümesinden geçen optimum doğru yatay olarak seyir etmektedir. Diğer bir deyişle, eğrinin sıfır olması nedeniyle, F2 parametresindeki değişimim çevrim sayısı üzerinde herhangi bir etkisi olmadığı söylenebilir.



Şekil 3.10. Çevrim sayısı ve F2 değişkeninin serpilme diyagramı

Bazı değişkenlerin (AS3, ÖS31, O11 ve a6), çevrim sayısı ile ilişkileri Şekil 3.11’de görselleştirilmiştir. Eğimleri pozitif olan bazı değişkenlerin artması çevrim sayısında olumlu etki yaratırken eğitimi negatif olan bazı değişkenler olumsuz etki yaratmıştır. Ancak bu iki kategoriye girmeyen ve eğimleri sifıra yakın olan ve nötr olarak adlandırabileceğimiz 2 değişken (a5 ve F2) olduğu gözlemlenmiştir. Bu değişkenler, çevrim sayısı üzerinde herhangi bir etkileri olmadığından veri analizinden çıkartılmışlardır.



Şekil 3.11. Bazı değişkenlerin ile çevrim sayısının çoklu serpilme diyagramları

3.3.2. Normalizasyon

Çalışmada kullanılan ölçüm birimi veri analizini etkileyebilir. Örneğin, ölçü birimlerini yükseklik için metreden inç'e veya kilo için kilogramdan pound'a değiştirmek, çok farklı sonuçlara yol açabilir. Genelde bir özellik küçük birimlerle ifade edildiğinde, bu özellik için tanımlanan katsayı daha büyük olur. Böylece aslında kritik olmayan bir özellik model içerisinde daha önemli rol oynayan bir değişkene dönüşebilir. Ölçüm birimleri seçiminde bağımlılığı önlemek için, veriler normalleştirilmeli (normalisation) veya standartlaştırılmalıdır (standardization). Bu, verilerin $[-1,1]$ veya $[0.0, 1.0]$ gibi daha küçük veya ortak bir aralığa dönüştürülmesini içerir. Terimler standartlaştır ve normalize, veri ön işlemede birbirinin yerine kullanılır. Bu amaçla kullanılan bazı formüller aşağıda yer almaktadır (Han, Kamber ve Pei 2012).

Min-max:

$$v'_i = \frac{v_i - \min_A}{\max_A - \min_A} (\text{new_max}_A - \text{new_min}_A) + \text{new_min}_A \quad (3.1)$$

Z-score:

$$v' = \frac{v - \bar{A}}{\sigma_A} \quad (3.2)$$

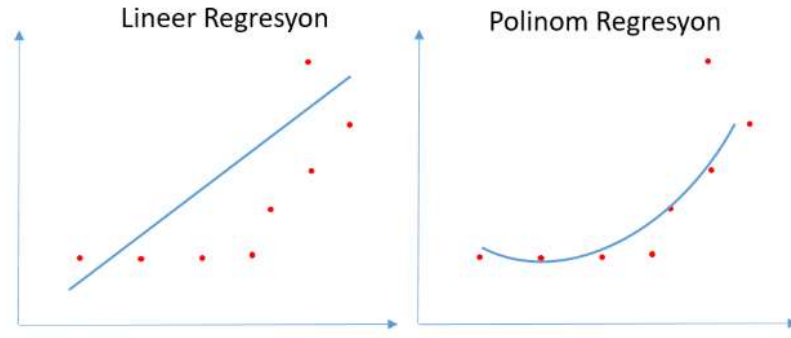
Ondalık ölçekleme:

$$v'_i = \frac{v_i}{10^j} \quad (3.3)$$

$\text{Max}(|v'|) < 1$ koşulunu sağlayan en küçük tam sayı j olmak üzere

3.3.3. Verinin Polinom Hale Getirilmesi

Regresyon problemlerinde birinci dereceden denklemler çözüm için her zaman en ideal durum olmayabilir. Şekil 3.12'de da görüldüğü üzere polinomik bir regresyon bazı veri kümeleri için daha uygun olabilir



Şekil 3.12. Lineer ve polinom regresyonların görselleştirilmesi (Şirin, 2017)

Modelin başarısını arttırmak için bu çalışmada da veriler polinom hale çevrilmişlerdir. Scikit-learn kütüphanesi içerisinde “PolynomialFeatures” adında bir fonksiyon bulunmaktadır. Bu fonksiyon sayesinde, verilen dereceye göre denklem tüm polinom kombinasyonlarını oluşturacak şekilde bağımsız değişken matrisi oluşturmaktadır. Örneğin bağımsız değişkenlerin yer aldığı girdi iki boyutlu ve $[a, b,]$ biçiminde olsun. Bu girdiyi ikinci dereceye yükseltirsek $[1, a, b, a^2, axb, b^2]$ halini alır. Daha çok bağımsız değişkenin olması model başarısını yükseltebilir ancak beraberinde farklı aşırı öğrenme durumunu getirebilir. Aşırı öğrenmeyi engellemek için yapılabilecek yöntemlerden bir tanesi ise çapraz doğrulamadır (cross-validation).

3.4. Model Performansının Değerlendirilmesi

Algoritma seçilip veriler analiz edildikten sonraki adım, oluşturulan modelin performansının değerlendirilmesidir. Bölüm 3.2’de de değinildiği gibi; bu çalışmada regresyon analizi ve sınıflandırma olmak üzere iki ayrı analiz yapılmıştır ve bu analizlerin değerlendirme ölçüleri birbirlerinden farklıdır. Regresyon analizi, Bölüm 2.4.1’de ele alınmıştır ve modelin performans değerlendirme kriteri olarak hata kareleri ortalaması seçilmiştir. Bu bölümde ise sınıflandırma modellerinin performans değerlendirme kriterleri ele alınacaktır.

Hangi sınıflandırıcının daha iyi çalıştığını bulmak için çeşitli yöntemler geliştirilmiştir. Karışıklık matrisi (Confusion matrix) bu yöntemler arasında sıkça kullanılan yöntemlerden bir tanesidir. Bu çalışmada değerlendirme yöntemi olarak seçilmesinin

nedeni ise veri kümemizin ikili kategorik (bias) olmasıdır. Bu çalışmadaki sınıflandırma analizi başarılı ve başarısız olmak üzere iki seçenek üzerinden yürütüldüğünden karışıklık matrisi performans değerlendirmesinde kullanılabilir basit ve hızlı bir yöntemdir. Literatürde ikiden fazla sınıflandırma üzerine yapılan çalışmalar da mevcuttur ve bu tür sınıflandırmalarda farklı değerlendirme kriterleri seçilmesi daha uygun olacaktır.

Çizelge 3.3: Karışıklık Matrisi (Han, Kamber ve Pei 2012)

		Tahmin		
		Pozitif	Negatif	Toplam
Gerçek	Pozitif	Doğru pozitif (DP)	Yanlış Negatif (YN)	P
	Negatif	Yanlış Pozitif (YP)	Doğru Negatif (DN)	N
	Toplam	P'	N'	P+N

Karışıklık matrisi ise Çizelge 3.3'te belirtilmiştir. Karışıklık matrisi ile doğruluk, hata oranı gibi hesaplamalar yapılabilmektedir. Hesaplamaların tamamı ise karışıklık matrisi içerisinde yer alan ana dört öge ile gerçekleştirilmektedir. Bu dört ögeye ait açıklamalar ise aşağıda belirtilmiştir (Han, Kamber ve Pei 2012);

Doğru Pozitif (DP): Gerçek değeri pozitif olup, sınıflandırıcı tarafından da pozitif olarak etiketlenmiş veri demetleridir. (tuple)

Doğru Negatif (DN): Gerçek değeri negatif olup, sınıflandırıcı tarafından da negatif olarak etiketlenmiş veri demetleridir.

Yanlış Pozitif (YP): Gerçek değeri negatif olan ancak sınıflandırıcının yaptığı hata sonucu pozitif olarak etiketlenmiş veri demetleridir.

Yanlış Negatif (YN): Gerçek değeri pozitif olan ancak sınıflandırıcının yaptığı hata sonucu negatif olarak etiketlenmiş veri demetleridir.

Doğrulukla başlayarak değerlendirme ölçütlerine bakacak olursak; belirli bir veri kümesindeki sınıflandırıcının doğruluğu, sınıflandırıcı tarafından doğru bir şekilde

sınıflandırılan test kümesi gruplarının yüzdesidir ve Eşitlik (3.1)'de denklemi yer almaktadır.

$$\text{Doğruluk (ACC)} = \frac{DP+DN}{P+N} \quad (3.4)$$

Sınıflandırıcının pozitif sınıf etiketlerini tahmin etmedeki etkililiği duyarlılık (sensitivity, true positive rate) olarak adlandırılmaktadır.

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (3.5)$$

Sınıflandırıcının negatif sınıf etiketlerini tahmin etmedeki etkililiği belirleyicilik (specificity, true negative rate) olarak adlandırılmaktadır.

$$\text{Belirleyicilik} = \frac{DN}{DN+YP} \quad (3.6)$$

Doğru sınıflandırılan pozitif örneklerin toplam pozitif tahmin edilen örneklere oranına hassasiyet ya da pozitif öngörü değeri (precision / positive predictive value) denir. Doğru sınıflandırılan negatif sınıf etiketine sahip örneklerin toplam negatif tahmin edilen örneklere oranının ise geri çağırma ya da negatif öngörü değeri (recall / negative predictive value) olarak adlandırılır (Han, Kamber ve Pei 2012).

$$\text{Hassasiyet (precision)} = \frac{DP}{DP+} \quad (3.7)$$

$$\text{Geri çağırma (recall)} = \frac{DN}{DN+YN} \quad (3.8)$$

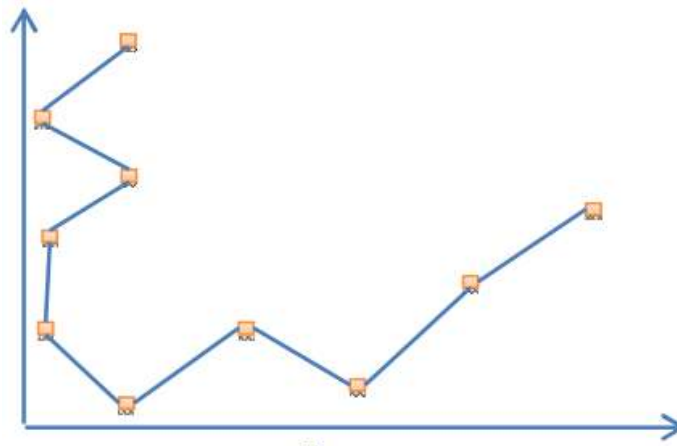
Bu çalışma kapsamında değerlendirmeye alınacak son ölçüt, F puanıdır (F Score). Hassasiyet ve geri çağırma performans değerlendirme ölçülerinin harmonik ortalaması alınarak hesaplandığından, her iki ölçütün birlikte ele alınmasını sağlamaktadır.

$$F \text{ puanı} = \frac{2 \times \text{hassasiyet} \times \text{ç} \text{ağırma}}{\text{hassasiyet} + \text{ç} \text{ağırma}} \quad (3.9)$$

Sınıflandırıcının performansını değerlendirirken ilk akla gelen performans değerlendirme ölçütü olsa da, her zaman kullanılacak en iyi metrik olmayabilir. Örneğin, 100 kez gerçekleşen bir olayda yalnızca 1 kez oluşan bir durumu tahmin etmeye çalıştığımızı varsayalım. Etkinliğin hiç gerçekleşmediğini söyleyerek %99 doğruluk sağlayan bir model oluşturabiliriz. Ancak bu şekilde ilerlersek önem verdiğimiz olayların %0'ını yakalarız. Hassasiyet (precision) ve geri çağırma (recall) metrikleri bu nedenle ile kritiktir. Hassasiyet ve geri çağırma metriklerini birlikte değerlendirmemize olanak sağladığı için F puanı, bu çalışmada ana değerlendirme metriği olarak seçilmiştir.

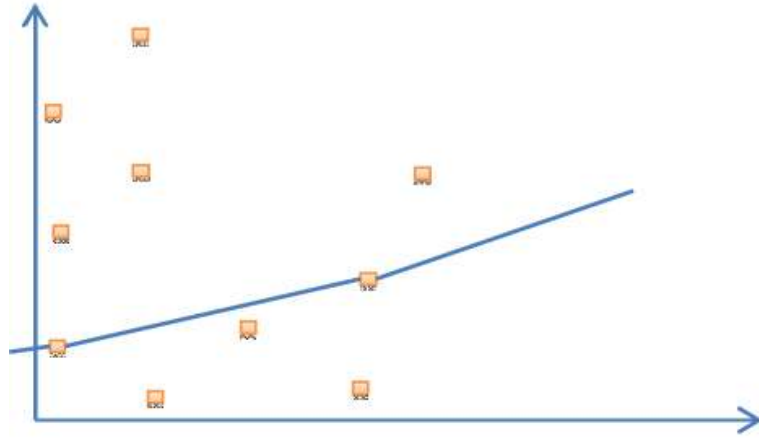
3.5. Aşırı Öğrenme ve Eksik Öğrenme

Gözetimli makine öğrenmesinde anlamsızlıklara yol açabilecek bir sorun vardır. Model analiz edilen veri kümesinde çok iyi çalışmasına rağmen bilinmeyen verilerde kötü performans sergileyebilir. Bu duruma aşırı öğrenme (Overfitting) denir. Aynı durum analiz edilen veriler içerisindeki eğitim kümesinde mükemmel performans sergilerken test kümesine kötü uyum sağlama olarak da karşımıza çıkabilir. Bunun nedeni eğitim ve test olarak ayrılan veri kümelerinin birbirlerinden farklı olması olabilir. (Ying, 2019) Bu durum modelin eğitim süresi boyunca belirli bir zamanda sorun çözme yeteneğini geliştirmediği anlamına gelir. Modelin öğrendiği şey eğitim kümesindeki bazı rastgele düzenliliklerdir. Bu da modelin genelleme yeteneğini düşürerek, farklı veriler ile test edildiğinde zayıf performans göstermesine neden olur (Jabbar ve Khan 2015). Aşırı öğrenme durumu Şekil 3.13'de görselleştirilmiştir.



Şekil 3.13. Aşırı Öğrenme (Jabbar ve Khan 2015)

Görüldüğü üzere model veri kümesi ile birebir eşleşmektedir. Doğal olarak bu model bilinmeyen farklı veriler ile zayıf performans gösterecektir. Aşırı öğrenme durumunun tam tersi ise eksik öğrenmedir (underfitting). Bu, modelin değişkenliği yakalayamadığı durumlarda ortaya çıkar. Eksik öğrenme, Şekil 3.14’de görselleştirilmiştir. Örneğin, parabolik yapıya daha uygun bir veri kümesi üzerinde doğrusal ($y = ax + b$ polinom değil, a ve b sabit) bir model oluşturalım. Sonuçta ortaya çıkan sınıflandırıcının tahmini gücü zayıf olacak ve eğitim verilerini düzgün şekilde haritalayamayacaktır. Bu, belirli bir veri kümesini tanımlamak için çok basit olan bir modeli anlamamanın veya kullanmaya çalışmanın sonucudur (Jabbar ve Khan 2015).



Şekil 3.14. Eksik Öğrenme (Jabbar ve Khan 2015)

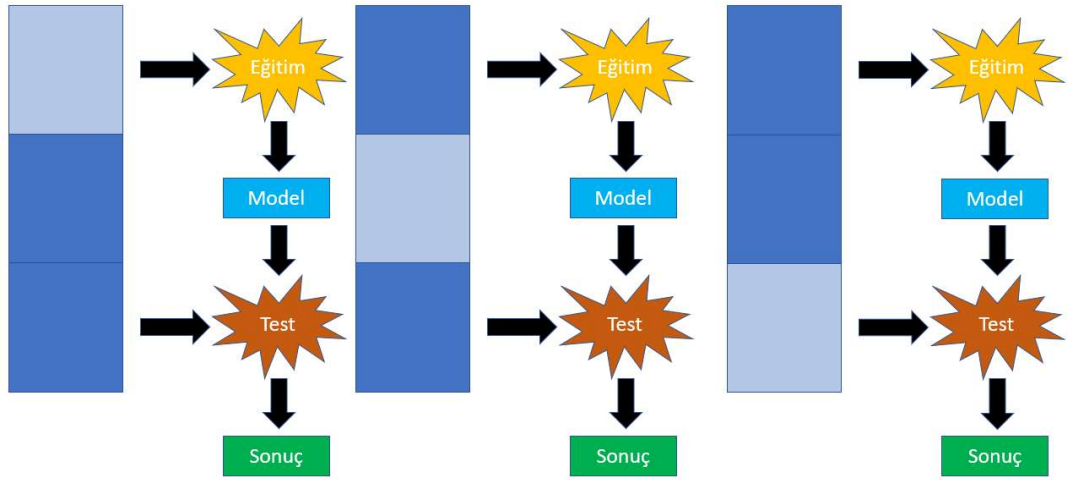
Eksik öğrenmeyi tahmin etmek kolaydır. Farklı makine öğrenmesi yöntemleri ve parametre optimizasyonları ile model başarıları tekrar değerlendirilebilir. Yeni modellerin performansları göz önüne alınarak hızlıca değişikliklere gidilebilir. Ancak aşırı öğrenmeyi yakalamak bu kadar kolay değildir. Veri bilimcileri bu problemi ele almak için çeşitli modeller geliştirmişlerdir. Bu çalışmada ise aşırı öğrenme durumunu incelemek için detayları Bölüm 3.6’da belirtilen çapraz doğrulama adı verilen bir yöntem kullanılacaktır.

3.6. Çapraz Doğrulama

Aşırı öğrenme, istatistiksel modelin iki sorunu çözmek zorunda kalmasının bir sonucu olarak ortaya çıkar. Eğitim sırasında performans hatasını en aza indirmelidir. Ancak test

sırasında bilinmeyen veriler üzerinde doğru tahminler yapmak için becerilerini en üst düzeye çıkartılması da gerekir. Bu çift hedef modelin ana görevi olması gereken veri trendini öğrenmek yerine, eğitim veri kümesini öğrenmesine yol açabilir (Chicco 2017). Çapraz doğrulama (cross validation), bu çift hedefli problemi kontrol edip modelin veri trendini öğrendiğinden emin olmak için kullanılan yöntemlerden bir tanesidir.

Çapraz doğrulama, verileri iki gruba bölerek öğrenme algoritmalarını değerlendirmek ve karşılaştırmak için kullanılan istatistiksel bir yöntemdir. Gruplardan biri bir modeli öğrenmek ya da eğitmek için kullanılır iken diğeri modeli doğrulamak için kullanılır. Çapraz doğrulamanın temel şekli k-kat çapraz doğrulamadır. k-kat çapraz doğrulamada, veriler önce k eşit (veya neredeyse eşit) boyutlu gruplara bölünür. Bölünen k grubu test için, geri kalan veriler ise eğitim için kullanılır. Ardından farklı bir k grubu eğitim için geri kalan veriler ise tekrar eğitim için kullanılarak iterasyonlar gerçekleştirilir. Şekil 3.15'te k=3 olan bir örneği göstermektedir. Verilerin daha koyu olan bölümü eğitim için kullanılırken açık renkli kısımlar test için kullanılır. Veri madenciliği ve makine öğreniminde 5 kat çapraz doğrulama (k=5) yaygın seçimlerden biridir.



Şekil 3.15. 3-Katlı çapraz doğrulama sistemi (Refaeilzadeh 2009)

Çapraz doğrulamanın olası iki hedefi vardır. İlki, bir algoritma kullanarak öğrenilen modelin mevcut verilerden performansını tahmin etmek, ikincisi ise; iki veya daha fazla algoritmanın performansını karşılaştırmak ve mevcut veriler için en iyi algoritmayı bulmaktır.

4. BULGULAR VE TARTIŞMA

Çalışmanın bu bölümünde lineer regresyon, lojistik regresyon, lojistik regresyon, destek vektör makineleri, k-en yakın komşu algoritması ve rassal orman algoritmalarının veri kümesine uygulanması ve elde edilen bulgular yer almaktadır. Analizlerde veri kümesindeki aynı bağımsız değişkenler kullanılmıştır. Ancak problem regresyon ve sınıflandırma olarak iki farklı şekilde ele alındığından bağımlı değişkenler birbirinden farklıdır. Regresyon analizlerinde bağımlı değişkenler sayısal değerler içeriyorken, sınıflandırma analizlerinde Bölüm 3.2’de belirtilen başarı kriterlerine göre başarılı veya başarısız yani 0 veya 1 olarak düzenlenmişlerdir. Analizlerin tamamı %80’e %20 olacak şekilde öğrenme ve test veri kümesi olarak ikiye ayrılmıştır. Kullanılan yöntemlerin tamamında 5 kat çapraz doğrulama uygulanmıştır. Parametre seçimi içeren yöntemlerde ise seçim kriterleri hakkında bilgi verilmiştir. Ekler bölümünde analizlerde kullanılan Python (versiyon 3.7) kodu yer almaktadır. “#” ile başlayan satırlar kodu daha iyi anlayabilmek için açıklamalar barındırmaktadır.

4.1. Linear Regresyon Algoritması Bulguları

Lineer regresyon analizi için Python’un scikit-learn kütüphanesi içerisindeki LinearRegression() fonksiyonu kullanılmıştır. Ayrıca lineer regresyon analizinde bağımsız değişkenler polinomlaştırılarak incelenmiştir. Çizelge 4.1’de polinom derecesine göre HKO (hata kareleri ortalama) sonuçları yer almaktadır.

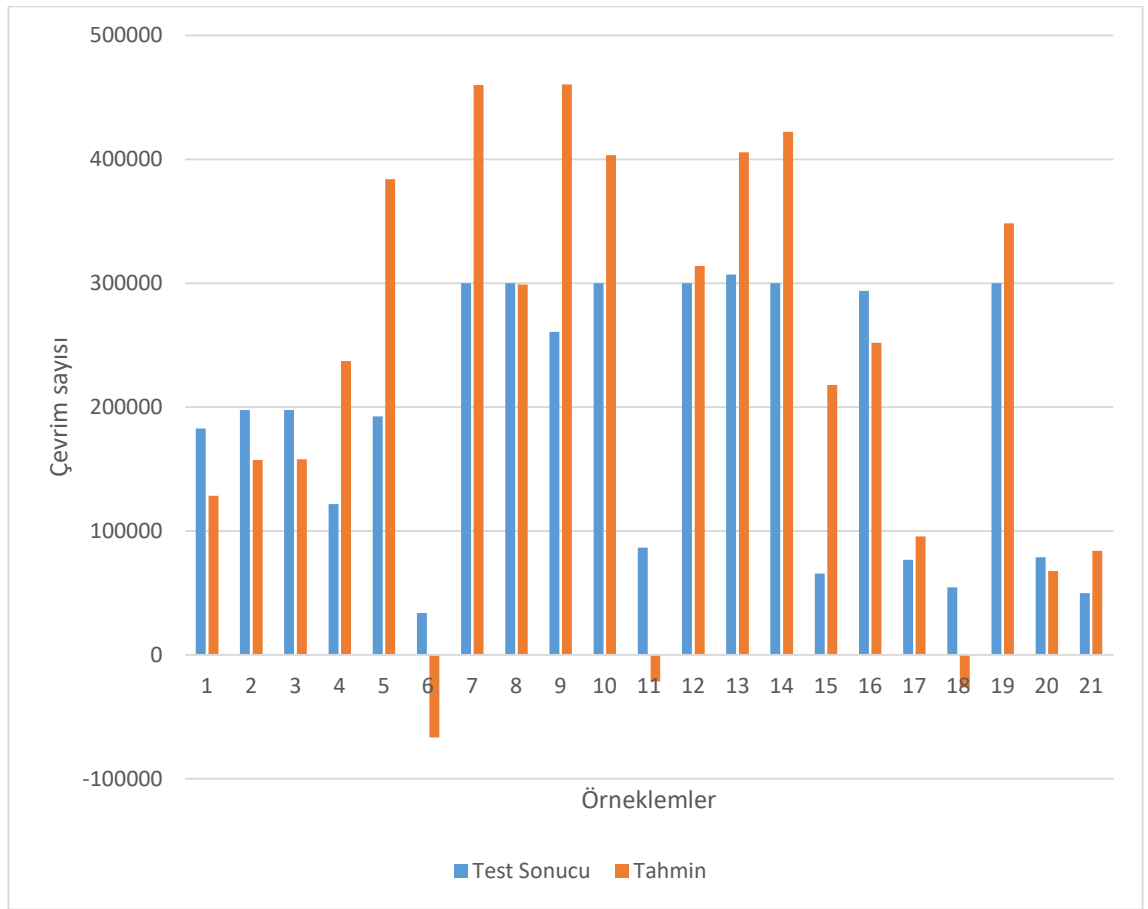
Çizelge 4.1. Polinom derecelerine ve HKO sonuçları

Polinom Derecesi	1	2	3	4	5
HKO Sonucu	-0.61	0.61	-0.36	0.11	0.02

HKO performans değerlendirme ölçütüne göre en başarılı derece 2’dir ve modelin performansı %61’dir Çizelge 4.2’de, 2.ci dereceden dağılımda değişken katsayıları gösterilmiştir.

Çizelge 4.2. 2.ci dereceden polinom dağılımında bağımsız değişken katsayıları

	Bağımsız Değişken	Tahmin Edilen Katsayı		Bağımsız Değişken	Tahmin Edilen Katsayı		Bağımsız Değişken	Tahmin Edilen Katsayı
0	SacK	-6,7291	7	ÖS1	-1,0851	14	ÖS11	-4,7352
1	AS1	8,7859	8	AS11	1,9355	15	a3	-5,3715
2	AS2	3,6156	9	AS21	1,5616	16	a4	7,1256
3	AS3	-3,8199	10	AS31	8,5626	17	a6	-1,1896
4	O1	-1,0283	11	O11	-4,9713	18	F1	-5,7612
5	ÖS3	1,0088	12	ÖS31	2,9771			
6	ÖS2	-2,5423	13	ÖS21	2,0271			



Şekil 4.1. Lineer regresyon modeli test sonuçları ile tahminlerin karşılaştırılması

Şekil 4.1'deki grafik, öğrenme verilerinden oluşturulan matematiksel modelin test verilerindeki tahminlerini içermektedir. Mavi çizgiler gerçek sonucu, turuncu çizgiler ise tahminleri göstermektedir.

2.ci dereceden polinom formatındayken HKO değerlendirme kriterine göre model başarısı %61 olsa da 1.ci derece regresyonda HKO -0,61'dir. 0 ile 1 arasında bir değer alması gereken hata kareler ortalamasının eksi bir değer olması modelin veya veri kümesinin tutarsız olduğunu göstermektedir. Aynı şekilde 3.cü, 4.cü ve 5.ci derecelerde HKO yine eksiye veya 0'a yakın değerlerde sonuçlanmıştır. Yaşanan tutarsızlığın veri kümesinden kaynaklandığı düşünülmektedir. Bağımlı değişken olan çevrim sayısı verisi test düzeneğinden elde edilmektedir. Ancak plakalı eşanjör 300.000 çevrime ulaştığında test durdurulmaktadır. Bu durum test süreleri açısından mantıklı olsa da, ömrü daha yüksek eşanjörlerin ömrünün 300.000 olarak kabul edilmesine sebebiyet vermiştir. Bu nedenle de bağımlı değişken kategorik hale getirilmiş ve sınıflandırma analizleri yapılmıştır.

4.2. Lojistik Regresyon Algoritması Bulguları

Lojistik Regresyon analizi için Python'un Scikit-learn kütüphanesindeki "LogisticRegression()" fonksiyonu kullanılmıştır. Bu fonksiyon içerisinde herhangi bir parametre seçimi yer almamaktadır. Fonksiyon öğrenme verileri ile çalıştırılarak bir model oluşturulmuştur. Ardından oluşturulan modelde test verileri denenmiştir. Çizelge 4.4'te ise lojistik regresyon modeline ait karışıklık matrisi yer almaktadır.

Çizelge 4.4. Lojistik regresyon karışıklık matrisi

		Tahmin		
		Pozitif	Negatif	Toplam
Gerçek	Pozitif	4	2	6
	Negatif	4	12	16

Oluşturulan model 6 pozitif (başarılı) değerden 4 tanesini, 16 negatif (başarısız) değerden ise 12 tanesini doğru bilmiştir. Yine Scikit-learn içerisindeki "classification_report()" fonksiyonu kullanıldığında ise hassasiyet, geri çağırma ve F puanı elde edilebilmektedir. Değerler Çizelge 4.5'te yer almaktadır.

Çizelge 4.5. Lojistik regresyon sınıflandırma raporu

	Hassasiyet	Geri çağırma	F puanı	Bağımlı değişken adedi
0	0.50	0.67	0.57	6
1	0.86	0.75	0.80	16
Ortalama/Toplam	0.76	0.73	0.74	22

Toplamda 22 testin 16'sını doğru tahmin eden lojistik regresyon modelinin doğruluğu %72,7'dir. 5-katlı çapraz doğrulama yapıldığında modelin doğruluk sonuçları Çizelge 4.6'da yer almaktadır. Lojistik regresyon modelinin 5 kat çapraz-doğrulama ile gerçekleştirilen analizinde doğruluk değerlerinin ortalaması 0,6 ve standart sapması $\pm 0,4$ 'tür.

Çizelge 4.6. Lojistik regresyon k-kat çapraz doğrulama sonuçları

1-kat çapraz doğrulama	2-kat çapraz doğrulama	3-kat çapraz doğrulama	4-kat çapraz doğrulama	5-kat çapraz doğrulama
0.5	0.5	1.0	0.5	0.5

4.3. Destek Vektör Makinesi Algoritması Bulguları

Destek vektör makineleri analizi için Scikit-learn kütüphanesindeki "SVC()" (support vector classifier) fonksiyonu kullanılmıştır. Çekirdek fonksiyon seçiminde varsayılan fonksiyon olan radyal temel (radial basis function, rbf) kullanılmıştır. Ek bir parametre seçimi yapılmamıştır. Fonksiyon öğrenme verileri ile çalıştırılarak bir model oluşturulmuştur. Ardından oluşturulan modelde test verileri denenmiştir. Çizelge 4.7'de ise destek vektör makinesi modeline ait karışıklık matrisi yer almaktadır.

Çizelge 4.7. Destek vektör makinesi karışıklık matrisi

		Tahmin		
		Pozitif	Negatif	Toplam
Gerçek	Pozitif	3	3	6
	Negatif	4	12	16

Oluşturulan model 6 pozitif (başarılı) değerden 3 tanesini, 16 negatif (başarısız) değerden, 12 tanesini doğru bilmiştir. Yine sklearn içerisindeki “classification_report()” fonksiyonu kullanıldığında ise; hassasiyet, geri çağırma ve F puanı elde edilebilmektedir. Değerler Çizelge 4.8’de yer almaktadır.

Çizelge 4.8. Destek vektör makinesi sınıflandırma raporu

	Hassasiyet	Geri çağırma	F puanı	Bağımlı değişken adedi
0	0.43	0.50	0.46	6
1	0.80	0.75	0.77	16
Ortalama/Toplam	0.70	0.68	0.69	22

Toplamda 22 testin 15’ini doğru tahmin eden lojistik regresyon modelinin doğruluğu %68,1’dir. 5-katlı çapraz doğrulama yapıldığında modelin doğruluk sonuçları Çizelge 4.9’da yer almaktadır. Destek vektör makinesi modelinin 5 kat çapraz-doğrulama ile gerçekleştirilen analizinde doğruluk değerlerinin ortalaması 0,73 ve standart sapması $\pm 0,07$ ’tür.

Çizelge 4.9. Destek vektör makinesi k-kat çapraz doğrulama sonuçları

1-kat çapraz doğrulama	2-kat çapraz doğrulama	3-kat çapraz doğrulama	4-kat çapraz doğrulama	5-kat çapraz doğrulama
0.66	0.75	0.75	0.75	0.75

4.4. Rassal Orman Algoritması Bulguları

Rassal orman algoritması ile analiz için Python’un scikit-learn kütüphanesindeki “RandomForestClassifier()” fonksiyonu kullanılmıştır. Bu fonksiyon içerisinde “n_estimators” olarak belirtilen 100 varsayılan olarak tanımlanmış bir ağacın dallarını ifade eden bir parametre yer almaktadır. Bu parametre ile rassal orman algoritmasının kaç ağaca yani kırılıma dağılacağı belirlenir. Kırılım sayısı 10’dan 200’e kadar 10’arlı adetlerle test edilerek 20 farklı analiz yapılmıştır. Çizelge 4.10’da da görüldüğü üzere ulaşılan maksimum doğruluk oranı ve standart sapma oranı 0,73 ile $\pm 0,07$ ’dir. Bu değerlere ulaşılan minimum sayı 100 olduğundan n_estimators parametresi 100 seçilmiştir.

Çizelge 4.10. Rassal orman algoritması parametre optimizasyonu, kırılım analizi

Kırılım	Doğruluk	Standart Sapma(\pm)	Kırılım	Doğruluk	Standart Sapma(\pm)
10	0.57	0.41	110	0.68	0.19
20	0.70	0.20	120	0.73	0.07
30	0.68	0.19	130	0.68	0.19
40	0.68	0.19	140	0.73	0.07
50	0.73	0.07	150	0.73	0.07
60	0.65	0.24	160	0.73	0.07
70	0.65	0.24	170	0.68	0.19
80	0.70	0.20	180	0.68	0.19
90	0.68	0.19	190	0.73	0.07
100	0.73	0.07	200	0.73	0.07

Fonksiyon öğrenme verileri üzerinde çalıştırılarak bir model yaratılmış ve daha sonra oluşturulan modelde test verileri denenmiştir. Çizelge 4.11’de ise rassal orman algoritması ile oluşturulan modele ait karışıklık matrisi yer almaktadır.

Çizelge 4.11. Rassal orman algoritması karışıklık matrisi

		Tahmin		
		Pozitif	Negatif	Toplam
Gerçek	Pozitif	2	4	6
	Negatif	4	12	16

Oluşturulan model 6 pozitif (başarılı) değerden 2 tanesini, 16 negatif (başarısız) değerden, 12 tanesini doğru bilmiştir. Yine Scikit-learn içerisindeki “classification_report()” fonksiyonu kullanıldığında ise; hassasiyet, geri çağırma ve F puanı elde edilmiş ve Çizelge4.12’de paylaşılmıştır.

Çizelge 4.12. Rassal orman algoritması sınıflandırma raporu

	Hassasiyet	Geri çağırma	F puanı	Bağımlı değişken adedi
0	0.33	0.33	0.33	6
1	0.75	0.75	0.75	16
Ortalama/Toplam	0.64	0.64	0.64	22

Toplamda 22 testin 14'ünü doğru tahmin eden lojistik regresyon modelinin doğruluğu %63,6'dır. 5-katlı çapraz doğrulama yapıldığında modelin doğruluk sonuçları Çizelge 4.11'de yer almaktadır. Destek vektör makinesi modelinin 5 kat çapraz doğrulama ile gerçekleştirilen analizinde doğruluk değerlerinin ortalaması 0,68 ve standart sapması $\pm 0,19$ 'tür.

Çizelge 4.13. Rassal orman algoritması k-kat çapraz doğrulama sonuçları

1-kat çapraz doğrulama	2-kat çapraz doğrulama	3-kat çapraz doğrulama	4-kat çapraz doğrulama	5-kat çapraz doğrulama
0.66	0.5	0.75	0.75	0.75

4.5. K-En Yakın Komşu Algoritması Bulguları

K-en yakın komşu algoritması ile yapılan analizlerde Scikit-learn kütüphanesindeki “KNeighborsClassifier()” fonksiyonu kullanılmıştır. Bu fonksiyon içerisinde “n_neighbours” olarak belirtilen varsayılan değer 5 olarak tanımlanmış komşu sayısını belirleyen bir parametre yer almaktadır. “n_neighbours” parametresi 1’den 10’a kadar test edilerek 10 farklı analiz yapılmıştır. Yapılan analizlerdeki doğruluk ve standart sapma değerleri Çizelge 4.14’te yer almaktadır. Komşu sayısı 5 seçildiğinde en yüksek doğruluk olan 0,78 sonucuna ulaşılmaktadır, standart sapma ise; $\pm 0,23$ 'tür. Çizelge 4.15’te, 5 komşu parametresi ile yapılan analize ait karışıklık matrisi yer almaktadır.

Çizelge 4.14. K-en yakın komşu algoritması parametre optimizasyonu komşu seçimi

Komşu	Doğruluk	Standart Sapma(\pm)
1	0.73	0.32
2	0.58	0.59
3	0.73	0.32
4	0.52	0.27
5	0.78	0.23
6	0.58	0.21
7	0.68	0.19
8	0.55	0.37
9	0.73	0.07
10	0.68	0.19

Çizelge 4.15. K-en yakın komşu algoritması karışıklık matrisi

		Tahmin		
		Pozitif	Negatif	Toplam
Gerçek	Pozitif	3	3	6
	Negatif	4	12	16

Oluşturulan model 6 pozitif (başarılı) değerden 2 tanesini, 16 negatif (başarısız) değerden ise 12 tanesini doğru bilmiştir. Çizelge 4.16’da ise hassasiyet, geri çağırma ve F puanı yer almaktadır.

Çizelge 4.16. K-en yakın komşu algoritması sınıflandırma raporu

	Hassasiyet	Geri çağırma	F puanı	Bağımlı değişken adedi
0	0.43	0.50	0.46	6
1	0.80	0.75	0.77	16
Ortalama/Toplam	0.70	0.68	0.69	22

Toplamda 22 testin 15’ini doğru tahmin eden lojistik regresyon modelinin doğruluğu %68,1’dir. Destek vektör makinesi modelinin 5 kat çapraz doğrulama ile gerçekleştirilen analizinde doğruluk değerlerinin ortalaması 0,78 ve standart sapması $\pm 0,23$ ’tür, 5 kat çapraz doğrulama sonuçları ise Çizelge 4.17’de yer almaktadır.

Çizelge 4.17. K-en yakın komşu algoritması k-kat çapraz doğrulama sonuçları

1-kat çapraz doğrulama	2-kat çapraz doğrulama	3-kat çapraz doğrulama	4-kat çapraz doğrulama	5-kat çapraz doğrulama
0.66	1	0.75	0.75	0.75

Çizelge 4.18. Sınıflandırma modelleri tahmin sonuçları

	Doğruluk	Çapraz doğrulama	Standart sapma(\pm)
Lojistik Regresyon	72,7	0,6	0,4
Destek Vektör Makineleri	68,1	0,73	0,07
Rassal Orman	63,6	0,68	0,19
k-en yakın komşu	68,1	0,78	0,23

Yapılan tüm sınıflandırma analizlerinin doğruluk, çapraz doğrulama ve çapraz doğrulamaya ait standart sapma verileri Çizelge 4.18’de yer almaktadır. Analizlerde geri çağırma, hassasiyet, F puanı gibi değerler hesaplanmış olsa da bu metrikler, nadir gerçekleşen durumların yaşandığı analizlerde daha önem arz etmektedir. Bu çalışmada ve kullanılan veri kümesinde ise böyle nadir bir durum yaşanmadığından çapraz doğrulama sonuçları ana performans değerlendirme kriteri olarak seçilmiştir.

5. SONUÇ

Günümüzde üretim teknolojileri, her gün ilerlemekte ve çok çeşitli endüstri 4.0 uygulamaları ile karmaşıklaşmakta ve yaygınlaşmaktadır. Bu teknolojik gelişmeler ve edinilen tecrübeleri uygun metotlar kullanılarak kayıt edilmesi sayesinde geçmişte kimsenin sahip olmadığı kadar veri elimizde mevcuttur. Karmaşık veri kümeleri analiz edilerek ve süreçler iyileştirilerek karlılık, kalite ve müşteri memnuniyeti gibi konularda kazanımlar elde edilebilir. Ayrıca firmaların amacı, ürünlerinin kalitesini ve ömür sürelerini arttırarak daha cazip ürünler üretmektir. Ürünlerin ömürlerini ve kalitelerini arttırmak için ise; kalite kontrol süreçlerini detaylandırır ve çeşitli analizler yaparlar. Kaliteli ürünler üretmek, artan talep ve üretim kapasiteleri göz önüne alındığında günümüz piyasa koşullarında sektörel firmalar arası rekabet açısından kritiktir.

Bu çalışmada, Bosch Termoteknik Manisa fabrikasındaki plakalı eşanjör üretim verileri ile ömür test verileri makine öğrenmesi teknikleri ile analiz edilmiştir. Regresyon yöntemi ile ömür değerleri, test sonuçlarını başarılı ve başarısız olarak iki kategoriye ayırarak da sınıflandırma tahminleri yürütülmüştür.

Çalışmanın ilk aşamasında test sonuçlarını sayısal olarak tahmin edebilmek için regresyon yöntemi uygulanmıştır ancak yapılan regresyon analizleri sonucu model başarısı oldukça zayıftır. Plakalı eşanjörün yorulma ömrünün belirlenmesi için hazırlanan test düzeneğinde, hedeflenen çevrim 300.000'dir ve bu eşik değere ulaşıldığında test durdurulmaktadır. Ancak geçmiş yıllardaki denemelerde 750.000 çevrim sonrasında dahi çalışmaya devam edebilen eşanjörler gözlemlenmiştir. Eğer yorulma ömür testlerinde limit çevrim değeri olan 300.000 çevrimin sonrasında testler devam ettirilseydi, bağımlı değişken değerler daha farklı olup girdi parametreleri ve bağımlı değişkenler arasındaki ilişkiye müdahale edilmemiş olurdu. Bu durum nedeniyle mevcut durumdaki veri kümesi regresyon analizi için uygun olmayıp, yapılan regresyon analizine ek olarak sınıflandırma analizleri gerçekleştirilmiştir.

Çalışma kapsamındaki ömür sonuçlarını sınıflandırmak için ise; lojistik regresyon, destek vektör makineleri, rassal orman ve k-en yakın komşu algoritmaları kullanılmış ve model

parametreleri optimize edilmiştir. Dört sınıflandırma modeli, doğruluk oranları ve 5-kat çapraz doğrulama kriteri üzerinden karşılaştırılmıştır. Çizelge 4.18’de de görüldüğü üzere doğruluk değerleri ile çapraz doğrulama değerleri arasında farklılıklar mevcuttur. Bu fark veri sayısının az olduğu çalışmalarda çapraz doğrulama adımının kritik olduğunu kanıtlamaktadır. Tekrarlı analizler içerdiğinden nihai sonuç çapraz doğrulama yapılmış analizlerden alınmaktadır. Çapraz doğrulama değerleri incelendiğinde ise k-en yakın komşu algoritması 0,78 değeri ile en başarılı modeldir. Ancak standart sapma değerleri de göz önüne alındığında destek vektör makinelerinin daha tutarlı sonuç verdiği gözlemlenmiştir. 0,73 doğruluk değerine ulaşabilen destek vektör modeli, k-en yakın komşu algoritmasına nazaran 0,05 düşük olsa da $\pm 0,07$ standart sapma değerine sahip olduğundan daha tutarlı sonuç vermektedir. Tüm bu değerlendirmeler ve analizler göze alınarak en uygun sonucu veren sınıflandırma yöntemi destek vektör makineleridir.

Veri kümesinin büyüklüğü arttığında makine öğrenmesinin başarısı ve doğru tahmin yapabilme ihtimali de artmaktadır. İlerleyen çalışmalarda oluşturulan matematiksel modellere yeni veriler eklenerek, doğruluk artırılabilir ve endüstriyel üretim sürecinde makine öğreniminden edinilen bilgiler doğrultusunda verim artırılabilir. Literatürde endüstriyel üretim süreçlerini ele alan makine öğrenimi çalışmalarındaki eksiklik göz önüne alındığında, bu çalışma devam ettirilerek farklı modellerin karşılaştırıldığı bir doğrulama (benchmark) çalışması olarak literatüre katkı sağlayan bir yayın oluşturulabilir. Bu tez çalışmasından ve ileride yapılacak muhtemel akademik yayınlardan yola çıkılarak farklı endüstriyel ürünlerin üretim süreçlerinde de benzer iyileştirmeler yapılarak endüstriyel üretimde verimlilik artışı sağlanarak ülke ekonomisine doğrudan katkı sağlanabilir.

KAYNAK

- Anonim, 2017.** Kombi şeması. <https://www.koseogluisi.com/projeler/dogalgaz-kombi/kombi-semasi/> (Erişim tarihi: 05.07.2020)
- Anonim, 2020a.** Lehimli plakalı eşanjör. <https://ktk.com.tr/brazed-plate-heat-exchangers/> (Erişim tarihi: 04.07.2020)
- Anonim, 2020b.** Kanal plaka örneği. <http://rahyaab.co/en/component/k2/item/93-brazed-plate-heat-exchangers> (Erişim tarihi: 04.07.2020)
- Anonim, 2020c.** Scikit-learn algoritma seçimi için yol haritası https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html (Erişim tarihi: 23.05.2020)
- Anonim, 2020d.** Lehimli plakalı eşanjör su giriş-çıkış gösterimi <https://www.sogutmamerkezi.com/bosch-kombi-plakali-esanjor-21-p.html> (Erişim tarihi: 03.07.2020)
- Anonim, 2020e.** Lehimli plakalı eşanjör kesit alınmış iç görünümü <http://www.emmegi.co.uk/wb-series-brazed-plate-oil-coolers/wb-series-cutaway-emmegi/> (Erişim tarihi: 04.07.2020)
- Anonim, 2020f.** Plakalı eşanjör ölçüm bölgeleri şematik gösterimi <https://www.southwestthermal.com/brazed-plate-exchanger.html> (04.07.2020 tarihli görünüm)
- Anonim, 2020g.** Plakalı eşanjör ölçüm bölgeleri şematik gösterimi <https://www.southwestthermal.com/brazed-plate-exchanger.html> (Erişim tarihi: 04.07.2020)
- Alpaydın, E. 2010.** Introduction to Machine Learning. The MIT Press, Cambridge, Massachusetts, London, England, 527 pp.ÇA
- Altan, A., Hacıoğlu, R., 2018.** Prediction of Bitcoin prices with machine learning methods using time series data, 26th Signal Processing and Communications Applications Conference, İzmir.
- Chicco, D., 2017.** Ten quick tips for machine learning in computational biology, *BioDataMining*, 10(35):1-17.
- Deepthi, A.R., 2019.** KNN visualization in just 13 lines of code. <https://towardsdatascience.com/knn-visualization-in-just-13-lines-of-code-32820d72c6b6> (Erişim tarihi: 01.05.2020)
- Han, J., Kamber, M., Pei, J., 2012.** Data mining concepts and techniques. Elsevier Inc., USA, 744 pp.
- Hurwitz, J., Kirsch, D., 2018.** Machine learning for dummies. John Wiley & Sons Inc., USA, 68 pp.
- Jabbar, H., Khan, R.Z., 2015.** Methods to avoid over-fitting and under-fitting in supervised machine learning (comparative study), *Computer Science, Communication and Instrumentation Devices/Researchgate*.
- Kaçar, E.N., Erbay, L.B., 2013.** Isı değiştiricilerin tasarımına bir bakış, *Mühendis ve Makine*, 54(644):14-413.
- Kakaç, S., Liu, H., 2002.** Heat exchangers selection rating and thermal design, CRC Press, Florida, USA, 522 pp.
- Konsiantis, S.B., 2007.** Supervised machine learning: a review of classification techniques, *Informatica*, 31(2007): 249-268.

- McCull, L., 2020.** Regression to the mean: as relevant today as it was in the 1900s. <https://select-statistics.co.uk/blog/regression-to-the-mean-as-relevant-today-as-it-was-in-the-1900s/> (Eriřim tarihi: 06.07.2020)
- Mitchell, M.T., 1997.** Machine learning. McGraw-Hill Science/Engineering/Math, USA, 432 pp.
- Montgomery, D.C., Peck, A.E., Vining, G.G., 2012.** Introduction to linear regression analysis. John Wiley & Sons, Inc., New Jersey, USA, 672 pp.
- Rafferty, K.D., Culver, G., 2006.** Heat exchangers. Geo-Heat Center Oregon Institute of Technology, Oregon, USA, 277 pp.
- Refaeilzadeh, P., Tang, L., Liu, H., 2009.** Cross-validation, encyclopedia of database systems. Springer, Boston, USA, 546 pp.
- Shah, R.K., Sekulic, D.P., 2003.** Fundamentals of heat exchanger design. John Wiley & Sons Inc., USA, 976 pp.
- Shah, S.N., 2015.** Comparison of some improved estimators for linear regression model under different conditions. *Master's Thesis*, Florida International University, Statistics, USA.
- Shao, M., Xang, X., Xin, B.Z., Chen, X., Wang, Y., 2020.** Prediction of energy consumption in hotel buildings via support vector machines, *Elsevier Sustainable Cities and Society*, 57(June 2020):102128.
- Sutton, R.S., Barto, A.J., 1998.** Reinforcement learning: an introduction, MIT Press, USA, 344 pp.
- řeker, S.E., 2012.** Karar ağacı öğrenmesi. <http://bilgisayarkavramlari.sadievrenseker.com/2012/04/11/karar-agaci-ogrenmesi-decision-tree-learning/> (Eriřim tarihi: 03.05.2020)
- řirin, E., 2017.** Polinom regresyon: python ile uygulama-1. <https://www.veribilimiokulu.com/polinom-regresyon-python-uygulama-1/> (Eriřim tarihi: 26.05.2020)
- Toksoy, M., 2003.** Doğrudan ısıtma sistemleri: temelleri ve tasarım seminer kitabı. TMMOB Makine Mühendisleri Odası, İzmir, Türkiye, 392 ss.
- Vapnik, V., Cortes, C., 1995.** Support vector machines, *Machine Learning*, 20:273-293.
- Wall, D.H.P., Fontenot, D.J., 2020.** Application and comparison of machine learning models for predicting quality assurance outcomes in radiation therapy treatment planning, *Elsevier Informatics in Medicine Unlocked*, 18(2020):100292.
- Ying, X. 2019.** An overview of overfitting and its solutions, *IOP Conf. Series: Journal of Physics*, 1168(2019): 022022.

EKLER

- EK 1** Numpy, Pandas ve Sklearn Kütüphaneleri
- EK 2** Lineer Regresyon
- EK 3** Sınıflandırma Algoritmaları için Kod
- EK 4** Lojistik Regresyon Algoritması
- EK 5** Destek Vektör Makinesi Algoritması
- EK 6** Rassal Orman Algoritması
- EK 7** K-En Yakın Komşu Algoritması

EK 1 Numpy, Pandas ve Sklearn Kütüphaneleri

```
import numpy as np
import pandas as pd
from pandas import Series,DataFrame
import sklearn
#Regresyon fonksiyonu
from sklearn.linear_model import LinearRegression
#Sınıflandırma fonksiyonları
from sklearn.ensemble import RandomForestClassifier
#Model değerlendirme fonksiyonları
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn import metrics
from sklearn.metrics import r2_score, accuracy_score
from sklearn.metrics import confusion_matrix, classification_report
```

EK 2 Lineer Regresyon

```
xlsfile = pd.ExcelFile('Tez-veri-set.03.xlsx')
#Veriler excelde, veri seti xlsfile adında bir değişken olarak kaydedildi
PHE_df = xlsfile.parse('WH-18-19-Sade3')
#xlsfile excel dosyası içindeki WH-18-19-Sade3 sayfasındaki veriler ile
PHE_df adında bir dataframe yaratıldı.
X = PHE_df.drop(columns = ['ORT','Status','F2','KG','a5'])
Y = PHE_df.ORT
#X bağımsız değişkenleri, Y bağımlı değişkenleri ifade ediyor.
#Bağımlı değişkenler içinde anlamlı olmayan Status, F2, KG ve a5
değişkenleri çıkartıldı.
poly_reg = PolynomialFeatures(degree=2)
X_poly = poly_reg.fit_transform(X_norm)
#X değeri PolynomialFeatures fonksiyonu ile 2.ci dereceden bir
değişkene türetildi ve X_poly olarak kaydedildi.
X_train, X_test, Y_train, Y_test = train_test_split(X_poly, Y, test_size
= 0.2, random_state = 0)
lreg = LinearRegression()
lreg.fit(X_train,Y_train)
#X_train ve Y_train veri seti kullanılarak lreg adında lineer regresyon
kullanan matematiksel bir model yaratıldı.
Y_pred = lreg.predict(X_test),
#X_test için ayrılmış veriler lreg matematiksel modeline yerleştirildi ve
Y_pred adında tahmin değişkeni yaratıldı.
r2_score(Y_pred, Y_test)
#Y_pred yani tahmin edilen değerlerle Y_test yani gerçek değerlerin r2
yöntemi ile performansı değerlendirildi.
scores = cross_val_score(lreg,X_poly,Y,cv = 5)
#Modele çapraz doğrulama yapıldı ve öğrenme-test verileri tekrar
tekrar analiz edildi.
```

EK 3 Sınıflandırma Algoritmaları için Kod

```
PHE_df2= xlsfile.parse('WH-18-19-Sade2');
PHE_df2= xlsfile.parse('WH-18-19-Sade2');
#Sınıflandırma problemi için PHE_df2 adında ikinci dataframe
yaratıldı.
X2 = PHE_df2.drop(columns = ['ORT','Status','F2','KG','a5'])
Y2 = PHE_df2.Status
#Değişkenler ilk analizle karışmasın diye X2, Y2 olarak isimlendirildi
#Aradaki tek fark bağımlı değişken artık Status, yani sayısal bir değer
yerine 0 ya da 1.
#Aynı mantıkla etkisiz değişkenler kaldırıldı.
#Aradaki tek fark bağımlı değişken artık Status, yani sayısal bir değer
yerine 0 ya da 1.
X2_norm = sklearn.preprocessing.normalize(X2, norm='l2')
X2_train, X2_test, Y2_train, Y2_test = train_test_split(X2_norm, Y2,
test_size = 0.2, random_state=0)
```

EK 4 Lojistik Regresyon Algoritması

```
LogReg = LogisticRegression()
LogReg.fit(X2_train, Y2_train),
pred_LogReg = LogReg.predict(X2_test)
print(classification_report(Y2_test, pred_LogReg))
print(confusion_matrix(Y2_test, pred_LogReg))
print(accuracy_score(Y2_test, pred_LogReg))
scores4 = cross_val_score(LogReg,X2_test,Y2_test,cv=5)
print("Accuracy:  %0.2f  (+/-  %0.2f)"  %  (scores4.mean(),
scores4.std()*2))
```

EK 5 Destek Vektör Makineleri Algoritması

```
#Destek Vektör Makineleri
clf = svm.SVC()
clf.fit(X2_train, Y2_train)
pred_clf = clf.predict(X2_test)
print(classification_report(Y2_test, pred_clf))
print(confusion_matrix(Y2_test, pred_clf))
print(accuracy_score(Y2_test, pred_clf))
scores3 = cross_val_score(clf,X2_test,Y2_test,cv=5)
print("Accuracy:  %0.2f  (+/-  %0.2f)"  %  (scores3.mean(),
scores3.std()*2))
```

EK 6 Rassal Orman Algoritması

```
rfc = RandomForestClassifier(n_estimators = 200)
rfc.fit(X2_train, Y2_train)
pred_rfc = rfc.predict(X2_test)
scores2 = cross_val_score(rfc,X2_test,Y2_test,cv=5)
#Çapraz doğrulama yöntemi
print("Accuracy:  %0.2f  (+/-  %0.2f)"  %  (scores2.mean(),
scores2.std()*2))
#Çapraz doğrulama yöntemi ile elde edilen sonuçların ortalaması ve
standart sapmasının hesaplanması
```


EK 7 K-En Yakın Komşu Algoritması

```
kNN = KNeighborsClassifier(5)
kNN.fit(X2_train, Y2_train)
pred_kNN = kNN.predict(X2_test)
print(classification_report(Y2_test, pred_kNN))
print(confusion_matrix(Y2_test, pred_kNN))
print(accuracy_score(Y2_test, pred_kNN))
scores5 = cross_val_score(kNN,X2_test,Y2_test,cv=5)
print("Accuracy:  %0.2f  (+/-  %0.2f)"  %  (scores5.mean(),
scores5.std()*2))
#Uygulanan tüm yöntemlerin arka arkaya doğruluk ortalamalarının ve
standart sapmalarının yazdırılması
print("LinReg Accuracy:  %0.2f  (+/-  %0.2f)"  %  (scores1.mean(),
scores1.std()*2))
print("RFC Accuracy:  %0.2f  (+/-  %0.2f)"  %  (scores2.mean(),
scores2.std()*2))
print("SVM Accuracy:  %0.2f  (+/-  %0.2f)"  %  (scores3.mean(),
scores3.std()*2))
print("LogReg Accuracy:  %0.2f  (+/-  %0.2f)"  %  (scores4.mean(),
scores4.std()*2))
print("K-NN Accuracy:  %0.2f  (+/-  %0.2f)"  %  (scores5.mean(),
scores5.std()*2))
```

ÖZGEÇMİŞ

Adı Soyadı : Anıl Işığışok
Doğum Yeri ve Tarihi : 31.08.1994/BURSA
Yabancı Dil : İngilizce

Eğitim Durumu
Lise : Bursa Milli Piyango Anadolu Lisesi
Lisans : Yıldız Teknik Üniversitesi, Makine Mühendisliği
Yüksek Lisans : Uludağ Üniversitesi, Makine Mühendisliği

Çalıştığı Kurum/Kurumlar : Bosch Sanayi ve Tic. A.Ş, Bursa
Tasarım Mühendisi (2016-2019)

Bosch Termoteknik Isıtma ve Klima A.Ş, Manisa
Üretim ve Proses Mühendisi (2019,-)

İletişim (e-posta) : isigicok94@gmail.com

Yayımları :