T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI
YÖNETİM BİLİŞİM SİSTEMLERİ PROGRAMI
YÜKSEK LİSANS TEZİ

FREN SANAYİNDE TALAŞLI ÜRETİM SÜRECİNDE MAKİNE ÖĞRENMESİ MODELLERİNİN KARŞILAŞTIRILMASI İLE KESTİRİMCİ BAKIM VERİMLİLİĞİNİN DEĞERLENDİRİLMESİ

Burak EVRENTUĞ

Danışman
Doç. Dr. Can AYDIN

2023

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI
YÖNETİM BİLİŞİM SİSTEMLERİ PROGRAMI
YÜKSEK LİSANS TEZİ

FREN SANAYİNDE TALAŞLI ÜRETİM SÜRECİNDE MAKİNE ÖĞRENMESİ MODELLERİNİN KARŞILAŞTIRILMASI İLE KESTİRİMCİ BAKIM VERİMLİLİĞİNİN DEĞERLENDİRİLMESİ

Burak EVRENTUĞ

Danışman
Doç. Dr. Can AYDIN

İZMİR – 2023

TEZ ONAY SAYFASI

YEMİN METNİ

Yüksek Lisans Tezi olarak sunduğum “Fren Sanayinde Talaşlı Üretim Sürecinde Makine Öğrenmesi Modellerinin Karşılaştırılması İle Kestirimci Bakım Verimliliğinin Değerlendirilmesi” adlı çalışmanın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

16.06.2023

Burak EVRENTUĞ

ÖZET

Yüksek Lisans Tezi

Fren Sanayinde Talaşlı Üretim Sürecinde Makine Öğrenmesi Modellerinin Karşılaştırılması İle Kestirimci Bakım Verimliliğinin Değerlendirilmesi

Burak EVRENTUĞ

Dokuz Eylül Üniversitesi
Sosyal Bilimler Enstitüsü
Yönetim Bilişim Sistemleri Anabilim Dalı
Yönetim Bilişim Sistemleri Programı

Endüstriyel sektörlerde, yapay zeka ve makine öğrenme gibi teknolojilerin kullanımı, üretim süreçlerinin verimliliğini ve kararlılığını artırmak için önemli bir gereksinim haline gelmiştir. Üretim süreçlerinde yer alan makinelerin düzenli bakımı ve arızaların önceden tespiti, kesintisiz üretim ve iş sürekliliği açısından kritik öneme sahiptir. Kestirimci bakım uygulamaları, sensörler ve veri analizi yöntemleri ile birleştirilerek, makinelere dair verilerin toplanması, analizi ve anlamlı bilgilere dönüştürülmesi sağlanmaktadır. Böylelikle makinelerde olası arızaların önceden tahmin edilmesi, planlı bakım faaliyetlerinin gerçekleştirilmesi ve beklenmedik durmaların önüne geçilmesi mümkün olmaktadır. Bu yöntemler, üretim süreçlerinde verimliliği artırırken, bakım maliyetlerinin azaltılmasına da yardımcı olmaktadır.

Kestirimci bakım, makine öğrenmesi tekniklerinin kullanıldığı bir yaklaşım olarak endüstride giderek daha fazla ilgi görmektedir. Makine öğrenmesi, büyük miktardaki veriyi analiz ederek modelleri tanımlamak, gelecekteki olayları tahmin etmek ve karar verme süreçlerini otomatikleştirmek için algoritmalar kullanmaktadır. Kestirimci bakım bağlamında, makine öğrenmesi modelleri, makineden gelen verileri analiz ederek ekipman veya sistemlerde olası arızaları tahmin etmek ve önlemek için kullanılmaktadır. Bu modeller, makineye ait verileri kullanarak öğrenme işlemini gerçekleştirir, modelleri tanır ve anormallikleri tespit eder. Bu sayede, beklenmedik arızaların

önceden tahmin edilmesi ve bakım süreçlerinin planlanması verimli bir şekilde sağlanmaktadır.

Bu çalışmanın amacı, fren kaldırma bileşeni üreten endüstriyel bir şirketin makinede oluşabilecek arızaların veri analizi yöntemleriyle tahmin edilmesini ve bu tahminlerin doğruluk performansının artırılmasını sağlamaktır. Çalışmada, PyCharm IDE ortamı kullanılarak, veri analizi için Pandas, Numpy, Seaborn, Matplotlib ve Sklearn modüllerinden yararlanılmıştır. Veri analizi yöntemleriyle, makinedeki arızaların belirtilerini ve ilişkilerini anlamak için verilerin kapsamlı bir şekilde incelenmesi ve analiz edilmesi hedeflenmiştir.

Anahtar Kelimeler: Makine Öğrenmesi, Kestirimci Bakım, Arıza Tahmini.

ABSTRACT

Master's Thesis

Evaluation of Predictive Maintenance Efficiency with the Comparison of Machine Learning Models in Machining Production Process in Brake Industry

Burak EVRENTUĞ

Dokuz Eylül University

Graduate School of Social Sciences

Department of Management Information Systems

Management Information Systems Program

The use of technologies such as artificial intelligence and machine learning has become a crucial requirement in industrial sectors to enhance the efficiency and stability of production processes. Regular maintenance of machines involved in the production processes and early detection of failures are of critical importance for uninterrupted production and business continuity. Predictive maintenance applications combined with sensors and data analysis methods, enable the collection, analysis, and transformation of machine-related data into meaningful insights. This allows for the early prediction of potential failures, the execution of planned maintenance activities, and the prevention of unexpected downtime.

Predictive maintenance, as an approach that utilizes machine learning techniques, is gaining increasing interest in the industry. Machine learning uses algorithms to analyze large amounts of data, identify patterns, predict future events, and automate decision-making processes. In the context of predictive maintenance, machine learning models are used to analyze machine-generated data and predict and prevent potential failures in equipment or systems.

The purpose of this study is to predict failures in the machine used for producing brake lifting components in an industrial company through data analysis methods and enhance the accuracy performance of these predictions. PyCharm IDE was utilized in this study, along with modules namely Pandas, Numpy, Seaborn, Matplotlib, and Sklearn for data analysis. The comprehensive

analysis and examination of data were aimed at understanding the symptoms and relationships of failures in the machine through data analysis methods.

Keywords: Machine Learning, Predictive Maintenance, Failure Prediction.

FREN SANAYİNDE TALAŞLI ÜRETİM SÜRECİNDE MAKİNE ÖĞRENMESİ MODELLERİNİN KARŞILAŞTIRILMASI İLE KESTİRİMCİ BAKIM VERİMLİLİĞİNİN DEĞERLENDİRİLMESİ

İÇİNDEKİLER

TEZ ONAY SAYFASI ii
YEMİN METNİ iii
ÖZET iv
ABSTRACT vi
İÇİNDEKİLER viii
KISALTMALAR xi
TABLOLAR LİSTESİ xiii
ŞEKİLLER LİSTESİ xiv

GİRİŞ 1

BİRİNCİ BÖLÜM
ENDÜSTRİDE MAKİNE ÖĞRENMESİ İLE KESTİRİMCİ BAKIM: UYGULAMA VE AKADEMİK ÇALIŞMALAR

1.1. ENDÜSTRİ SEKTÖRÜNÜN YENİ İHTİYAÇLARI VE OTOMASYONUN ARTAN ÖNEMİ 7
1.2. KESTİRİMCİ BAKIM 7
1.3. LİTERATÜR TARAMASI 10
1.3.1. Veri Toplama Yöntemleri Açısından Kestirimci Bakım Çalışmaları 10
1.3.2. Veri Analizi Yöntemleri açısından Bakım Çalışmaları 11

İKİNCİ BÖLÜM
ENDÜSTRİ 4.0 VE MAKİNE ÖĞRENMESİ

2.1. YAPAY ZEKA 14
2.2. YAPAY ZEKANIN KARAR VERME ÜZERİNE ETKİSİ 15
2.3.YAPAY ZEKA UYGULAMALARINDA VERİ ANALİTİĞİ 16
2.3.1. Betimleyici Analitik 16
2.3.2. Teşhis/Tanı Analitiği 17
2.3.3. Tahminsellik Analitiği 17
2.3.4. Yönergeli Analitik 17
2.4. CRISP-DM METODOLOJİSİ İLE YAPAY ZEKA UYGULAMALARI GERÇEKLEŞTİRME SÜRECİ 18
2.5. MAKİNE ÖĞRENMESİ 20
2.5.1. Makine Öğrenmesi Terimleri ve Sınıflandırması 22
2.5.2. Denetimli Öğrenme 26
2.5.2.1. Karar Ağaçları 26
2.5.1.2. Naive Bayes 28
2.5.2.3. En Yakın Komşu 29
2.5.2.4. Destek Vektör Makineleri 31
2.5.2.5. AdaBoost 32
2.5.2.6. Random Forest 33
2.5.2.7. CatBoost 34
2.5.2.8. XGBoost 35
2.5.2.9. LightGBM 36
2.5.3. Denetimsiz Öğrenme 37
2.5.4. Pekiştirmeli Öğrenme 38
2.5.5. Makine Öğrenmesinde Model Doğrulama Yöntemleri 38
2.5.6. Makine Öğrenmesinde Model Performans Ölçüm Yöntemleri 39
2.5.7. Makine Öğrenmesinde Hiperparametre Ayarlaması 41

ÜÇÜNCÜ BÖLÜM
YÖNTEM

3.1. İŞİ ANLAMA 44
3.2. VERİLERİN TOPLANMASI 46
3.3. VERİ SETİ 48
3.3.1. Eksik Verileri Analizi 52
3.3.2. Aykırı Değer Analizi 54
3.3.3. Özellik Mühendisliği 55
3.3.4. Kategorik Verilerin Numerik Değere Dönüştürülmesi 57
3.3.5. Veri Normalleştirme 57
3.3.6. Özellik Seçimi 58
3.4. MODEL KURMA 58
3.5. MODEL PERFORMANS DEĞERLENDİRMESİ 58

SONUÇ VE ÖNERİLER 63
KAYNAKÇA 65

KISALTMALAR

Adaboost	Adaptive Boosting (Uyarlanabilir Yükseltme)
AI	Artificial Intelligence (Yapay Zeka)
ANN	Artificial Neural Network (Yapay Sinir Ağı)
AutoML	Automated Machine Learning (Otomatik Makine Öğrenmesi)
CAD	Computer-Aided Design (Bilgisayar Destekli Tasarım)
CNC	Computer Numrecal Control (Sayısal Kontrollü Makine)
CRISP-DM	Cross-Industry Standart Process for Data Mining (Veri Madenciliği için Sektörler Arası Standart Süreç)
DA	Dicsriminant Analysis (Diskriminat Analizi)
DBN	Deeğ Belief Network (Derin İnanç Ağları)
DCNC	Deep Convolutional Neural Network (Derin Evrişimsel Sinir Ağı)
DL	Deep Learning (Derin Öğrenme)
FN	False Negative (Yanlış Negatif)
FP	False Positive (Yanlış Pozitif)
HTTP	Hypertext Transfer Protocol (Hipermetin Transfer Protokolü)
IoT	Internet Of Things (Nesnelerin İnterneti)
kNN	k-Nearest Neighbors (k-En Yakın Komşuluk)
LR	Logistic Regression (Lojistik Regresyon)
LSTM	Long Short Term Memory (Uzun-Kısa Süreli Bellek)
MAE	Mean Absolute Error (Ortalama Mutlak Hata)
MAPE	Mean Absolute Percentage Error (Ortalama Mutlak Yüzde Hatası)
MLP	Multi Layer Perceptron (Çok Katmanlı Algılayıcı)
PLC	Programmmable Logic Controller (Programlanabilir Mantık Kontrolleri)
$\boldsymbol{R^2}$	R-squared (R-Kare)
RF	Random Forest (Rastgele Orman)
RMSE	Root Mean Squared Error (Kök Ortalama Kare Hatası)
RNN	Recurrent Neural Network (Yinelenen Sinir Ağı)
RR	Ridge Regression (Ridge Regresyon)
s.	Sayfa No

SVM	Support Vectorm Machines (Destek Vektör Makineleri)
TCP	Transmission Control Protocol (Aktarım Kontrol Protokolü)
TN	True Negatif (Doğru Negatif)
TP	True Possitive (Doğru Pozitif)
Vd.	Ve Diğerleri
XGBoost	eXtreme Gradient Boosting (Aşırı Gradyan Artımı)
XML	eXtensible Markup Language (Genişletilebilir İşaretleme Dili)
OEE	Overall Equipment Efficiency (Toplam Ekipman Etkinliği)

TABLOLAR LİSTESİ

Tablo 1: Yapay Zeka Teknolojileri ve İşletme Çözümleri s. 15
Tablo 2: Makine Öğrenimi Algoritmalarının Kullandığı Sektörler s. 22
Tablo 3: İşletim Sistemine Bağlı Değişkenler s. 48
Tablo 4: Sensörlere Bağlı Değişkenler s. 49
Tablo 5: Ayırma Yöntemi Kullanılarak Oluşturulan Tüm Sınıf Frekanslarının Değerlendirildiği Model Başarı Performansları s. 59
Tablo 6: Ayırma Yöntemi Kullanılarak Oluşturulan İkili Sınıf Frekanslarının Değerlendirildiği Model Başarı Performansları s. 59
Tablo 7: K Katmanlı Çapraz Doğrulama Yöntemi Kullanılarak Oluşturulan İkili Sınıf Frekanslarının Değerlendirildiği Model Başarı Performansları s. 60
Tablo 8: Modellerde Kullanılan Hiperparametreler s. 61
Tablo 9: Hiperparametre Kullanılarak Oluşturulan Modellerin Başarı Performansları s. 62

ŞEKİLLER LİSTESİ

Şekil 1: P-F Eğrisi Bağlamında Kestirimci Bakım s. 9
Şekil *2*: Veri Analitiğine Kavramsal Bakış s. 18
Şekil 3: CRISP-DM Sürecinin Basamakları s. 20
Şekil 4: Yapay Zeka ve Alt Alanları Arasındaki İlişki s. 21
Şekil *5*: Ngram Görüntüleyicide "Machine Learning" İfadesinin Kullanımı s. 21
Şekil 6: Makine Öğrenmesi Algoritmalarında Hiyerarşik Yapı s. 25
Şekil *7*: Karar Ağaç Yapısı s. 28
Şekil 8: Naive Bayes Yapısı s. 29
Şekil 9: kNN Algoritması Komşu Modeli Örneği s. 31
Şekil 10: Destek Vektör Makinesi ile Veri Gösterimi s. 32
Şekil 11: AdaBoost Algoritmasının Gösterimi s. 33
Şekil 12: Rastgele Ormanın Bir Örneği s. 34
Şekil 13: CatBoost algoritmasının açıklaması s. 35
Şekil 14: XGBoost'un temel yapısı s. 36
Şekil 15: Karar Ağaçları Büyüme Yöntemleri s. 36
Şekil 16: LightGBM Algoritmasında Histogram Yöntemi s. 37
Şekil 17: K Katmanlı Çapraz Doğrulama ve Ayırma Tekniği Modeli s. 39
Şekil 18: Karmaşıklık Matrisi s. 40
Şekil 19: Hiperparametre Ayarlama Şeması s. 41
Şekil 20: Uygulama İş Akışı s. 43
Şekil 21: Mazak CNC Makinesi Dış Görünümü s. 44
Şekil 22: GG25 Döküm Parçası İşlem Öncesi ve Sonrası s. 46
Şekil 23: MTConnect Veri İşlem Süreci s. 48
Şekil 24: Veri Seti Değişken Türleri s. 51
Şekil 25: Korelasyon Analizi s. 52
Şekil 26: Eksik Veri Kontrolü s. 53
Şekil 27: Değişkenlerin Histogram Grafik Dağılımları s. 54
Şekil 28: Değişkenlerin Aykırı Değer İşlemleri Sonrası Dağılımları s. 55
Şekil 29: Hedef Değişkenin Frekans Değerleri s. 56
Şekil 30: Hedef Değişkenin Gruplandırılmış Frekans Değerleri s. 57

Şekil 31: Özellik Seçimi Sonrası Seçilen Değişkenler s. 58

GİRİŞ

Problem Tanımı ve Araştırma Sorunu;

Endüstri 4.0, akıllı fabrikalar ve üretim organizasyonlarının oluşturulmasına yönelik bir dönüşüm sürecidir ve birçok yeni teknolojinin birbiriyle etkileşimli bir şekilde kullanımını gerektirmektedir. Teknoloji ile gelen değişim, tüm sektörlerin geleceği için tercih edilebilecek bir seçenek değil, zorunlu bir dönüşüm sürecini içermektedir (Doğan & Baloğlu, 2020). Bu dönüşüm sürecinde otomasyon, robotik, yapay zeka, büyük veri analizi ve nesnelerin interneti (IoT - Internet of Things) yer almaktadır. Endüstri 4.0 teknolojilerinin kullanımı, endüstrinin daha verimli, hızlı, esnek ve ölçeklenebilir olmasını sağlayan birçok fayda sunmasının yanında, üretim süreçlerinin optimize edilmesine yardımcı olmakta, üretim hatlarının ve tesislerin daha az hata yapmasını sağlamakta, enerji ve kaynak tüketimini azaltarak çevresel sürdürülebilirliği artırmaktadır (Taş, 2018). Tüm bu faydalar endüstrinin rekabetçi yapısını artırmakta ve daha etkili bir şekilde organizasyonun ve müşterinin ihtiyaçlarının karşılanmasına olanak tanımaktadır.

Endüstri 4.0 teknolojileri farklı alanlarda kullanılabilir. Bu alanlardan bir tanesi arıza tespit ve bakım sürelerinin yönetimidir. Endüstriyel işletmelerde IoT teknolojilerinin ve sensör teknolojilerinin birlikte kullanımı, hataların tespit edilmesinde, arıza ve bakım sürelerinin yönetiminde fayda sağlamaktadır. Bir başka ifadeyle, IoT ile elde edilen sensör verileri yapay zeka teknolojilerini kullanarak daha isabetli arıza tespit ve bakım imkanı sağlamakta ve bu şekilde organizasyonun verimliliğine katkı sunmaktadır.

Endüstride bakım çalışmaları, ekipmanların ve makinelerin verimli bir şekilde çalışmasını sağlamak ve arızaların önceden tespit edilerek önlenmesi için yapılan önemli bir faaliyettir. Bakım çalışmasının, üretimde sürekliliği sağlamak, üretim maliyetlerini minimize etmek, ürün kalitesini artırmak ve kullanılan ekipmanların daha uzun sürelerde kullanımını sağlamak için önemli olduğu belirtilmiştir (Emir vd., 2012). Bakım yapılmayan ekipmanlar, arızalara ve erken hasarlara neden olmaktadır ve bu da tamir ve yenileme maliyetini artırmaktadır. Dolayısıyla, tesislerin bakım faaliyetlerini optimize ederek verimli hale getirmesi işletmeler için büyük bir avantaj sağlayabilir.

Endüstriyel şirketler bakım çalışmaları için çeşitli yöntemler uygulamaktadır. Standart bakım kavramları düzeltici bakım, önleyici bakım, durum bazlı bakım, kestirimci bakım ve öngörülü bakım olmak üzere 5 başlıkta toplanmaktadır (Merkt, 2019, s. 695).

1. Düzeltici Bakım: Bir hatanın tanımlanmasından sonra gerçekleştirilen ve bir öğenin gereken bir işlevi yerine getirebilecek duruma getirilmesini amaçlayan bakım çalışmasıdır.

2. Önleyici Bakım: Belirlenmiş aralıklarla veya belirlenmiş kriterlere göre yapılan ve bir öğenin başarısızlık olasılığını veya işlevinin bozulmasını azaltmayı amaçlayan bakım olarak tanımlanmaktadır.

3. Durum Bazlı Bakım: Durum izleme ya da test-analiz yapma yöntemleri sonucunda bakım işlemlerinin bir kombinasyonunu içeren bakım çalışması olarak tanımlanmaktadır.

4. Tahminsel (Kestirimci) Bakım: Makinenin bozulmasına ilişkin önemli parametrelerin analizleri ya da bilinen özelliklerinin değerlendirilmesi sonucunda elde edilen bir tahmin ardından yapılan durum bazlı bakım çalışması olarak tanımlanmaktadır.

5. Öngörülü Bakım: Düzeltici ve kestirimci bakım modellerinin sonuçlarına dayalı olarak uygulanan bakım çalışması olarak tanımlanmaktadır.

Kestirimci bakım yöntemleri, tesislerin bakım faaliyetlerini planlamak, yönetmek ve optimize etmek açısından etkin araçlardan biridir ve günümüzde kestirimci bakım diğer bakım stratejilerine kıyasla daha etkili ve doğru bir bakım planı oluşturuyor olması sebebiyle ön plana çıkmaktadır (Carvalho vd., 2019). Ne var ki, bakım uygulamalarında atılması gereken en önemli adım bakım teknolojilerinin yeniden değerlendirilmesi ve en uygun yöntemin uygulanmasıdır (Ümit Gürsoy vd., 2019). İlgili literatürde çeşitli sektörlerde (enerji, ulaşım, imalat, vb.) kestirimci bakım konusunda yapılan çalışmalar incelenmiştir. Enerji sektöründe, hidrolik santrallerinin bakım işlemlerini optimize etmek için ve ulaşım sektöründe bir hava yolu şirketi uçak motorlarının arızalarını önceden tahmin etmek için makine öğrenmesi yöntemleri kullanılmıştır; bu çalışmalar özellikle rulmanlar, motorlar ve rüzgar türbinleri gibi ekipmanların üzerinde yoğun hesaplamalı deneylerle gerçekleştirilmiştir (Rana Dündar vd., 2021). Türkiye ve uluslararası mevcut çalışmaların genelinde model

oluşturmak için kullanılan veri bölümlendirme oranlarına bakıldığında direkt öğrenme ve test verileri olarak ayrım yapıldığı ve bu süretle algoritma başarı oranlarının hesaplandığı göze çarpmaktadır. Oysa kestirimci bakım uygulamaları literatürü dışındaki birçok makine öğrenimi çalışmasında, hiperparametre ayarı ile birlikte öğrenme ve test verileri üzerindeki oranlar değiştirilerek makine öğrenmesi modellerini analiz eden ve model performansına ilişkin sonuçlar sunulmaktadır. Bir başka ifadeyle, mevcut yapılan araştırmalarda öğrenme ve test verileri üzerindeki oranlar değiştirilerek makine öğrenmesi modellerini karşılaştırmalı analiz eden ve model performansı sonuçlarındaki değişimlere odaklanan kapsamlı bir çalışmaya rastlanmamıştır. Dolayısıyla, bu tez çalışmasının amacı, makine öğrenmesi modellerinin performansı üzerinde öğrenme ve test oranlarının değiştirilmesinin etkisini incelediği gibi, hiperparametre ayarlarının (örneğin, öğrenme oranı, epoch sayısı ve gizli katman sayısı vb.) değiştirilmesinde model performansının sonuçlarına nasıl yansıdığını incelemektedir. Yapılan araştırmalar sonucunda "Sensör verilerine dayalı farklı makine öğrenmesi algoritmaları kullanılarak kestirimci bakım uygulamasının verimliliği arttırılabilir mi?" sorusu araştırma sorusu olarak belirlenmiştir.

Kavramsal Çerçeve ve Yöntem

Kestirimci bakım uygulamasında öncelikle, bir makinenin işlevselliğini sağlamak için gerekli olan veriler toplanır. Veriler genellikle, makinanın işlemesi sırasında oluşan sıcaklık, basınç ve akış hızı gibi değişkenlerden oluşmaktadır. Verilerin toplanması için de sensörler kullanılmaktadır. Toplanan veriler veri ön işleme işlemi yapıldıktan sonra makine öğrenimi modeline aktarılmaktadır. Makine öğrenimi modeli, verileri analiz ederek makinenin sağlığı hakkında tahmin sonucunu oluşturmaktadır. Bu tahminler, belirli bir makinenin bakıma ihtiyacı olup olmadığı konusuna yardımcı olmaktadır.

Bu çerçeve içinde yapılanan mevcut çalışma, fren parçası üretim tesislerinde kestirimci bakımın verimliliğini ve etkinliğini arttırmak için optimal modeller sunmaktadır. Fren parçası üretimi, otomotiv endüstrisi için önemli bir sektördür ve üretim sürecinde kullanılan ekipmanların verimli bir şekilde çalışması kritik öneme sahiptir. Çalışma, soğutucu yağ seviyesi, soğutucu yağ sıcaklığı, trafo sıcaklığı, motor sıcaklığı, sistem durumu ve tarih değişkenleri gibi toplam 22 adet değişken, 30.622

gerçek kayıttan oluşmaktadır. Çalışmanın ilk aşaması, tesisin ekipmanlarının durumunu izlemek için uygun verilerin toplanmasıdır. Toplanan veriler, makine öğrenimi algoritmalarında kullanılarak, tesis içinde kullanılan makinenin bakım durumunun tahmin edilmesinde kullanılmaktadır. Ayrıca model oluşturmak amacıyla tüm veri seti belirlenen oranlarda öğrenme ve test veri kümeleri şeklinde bölümlendirilir. Daha önceki çalışma uygulamalarından farklı olarak modeller üzerinde hiperparametrelerin varsayılan değerleri ve model kurulumunda kullanılan öğrenme ve test verilerinin oranları değiştirilerek model performans değerleri ve kurulan modelin uygulanabilirliği incelenmektedir.

Çalışmanın Katkısı;

Çalışma, Ege Fren Sanayi ve Ticaret Anonim Şirketinin kestirimci bakım performansı ile ilgilenmektedir. İlgili işletme otomotiv sanayisine fren parçaları bileşeni üretimi yapmaktadır. Müşteriler tarafından gönderilen parça modelleri, işletme tarafından sayısal kontrollü makine (CNC – Computer Numerical Control) için modellenmektedir. Üretimi yapılacak fren kaldırma bileşeni Avrupa standartlarına göre tanımlanan bir dökme demir sınıfı olan GG25 materyalinden oluşmaktadır. Bahsi geçen materyal makineye yerleştirildikten sonra modeli oluşturacak kesici takım cihaza takılmakta ve cihazın programlaması yapılmaktır. Cihaz çalışmasını başlattığında modele uygun bir şekilde döküm parçasını nihai sonuca ulaştırmaktadır. Kalite kontrol birimine ulaştırılan nihai ürün onay aldıktan sonra müşterilere iletilmektedir. 7/24 çalışan CNC cihazları günlük, aylık, 3 aylık ve 6 aylık periyotlarda bakıma tabii tutulmaktadır. Şirketin bakım süreçlerindeki temel sorunu ise yapılan bakım işlemlerine ciddi bir bütçe ayırmak zorunda olmalarıdır. Bu çalışma, ilgili şirketin fren parçası üretim tesisinde kestirimci bakım yöntemlerinin kullanılması ve bu yöntemlerde kullanılan makine öğrenmesi model performans değerleri üzerine odaklanmaktadır. Çalışmanın amacı tesiste kullanılan makinenin bakım durumunu tespit etmek amacıyla makine öğrenmesi modelleri kullanılarak bir kestirimci bakım sistemi geliştirmektir.

Makine öğrenmesinde, modellerin doğru tahmin ve sınıflandırmalar yapması veya daha etkili değişken seçimlerinin yapılabilmesi için, model performansının artırılması gerekmektedir. Bu durum ise kullanılan algoritmalara ait parametrelerin uygun bir şekilde ayarlanmasını gerektirmektedir (Yılmaz & Kasapbaşı, 2021).

Hiperparametreler bir modelin eğitilmesi ve yapılandırılması sırasında belirlenen parametrelerdir. Hiperparametre optimizasyonu, genellikle deneme yanılma yöntemi ile yapılmaktadır. Farklı hiperparametre kombinasyonları denenerek en iyi performansı veren kombinasyon seçilmektedir. Bu nedenle, doğru hiperparametrelerin seçimi, modelin en iyi şekilde çalışmasını sağlamak için önem arz etmektedir. Bu çalışma farklı hiperparametre kombinasyonları kullanılarak ve model kurulumunda kullanılan öğrenme ve test veri bölümleme oranları değiştirilerek oluşturulan modellerin performansı üzerindeki etkisini tespit ederek, optimal bir model seçimi sağlamaktadır.

Hiperparametrelerin ayarlanması, makine öğrenmesi algoritmalarının sağladığı parametreleri optimize etmek için kullanılan bir tekniktir ve kullanılan bu teknik modellerin verimli bir şekilde oluşturulmasına ve doğru sınıflandırmanın belirlenmesine olanak tanımaktadır (Jamaleddyn vd., 2023).

Model performansının artırılmasının sektöre çeşitli katkıları bulunmaktadır. Her bir makine öğrenme modeli için hiperparametreler farklı veri kümeleri veya problemler için en iyi performansı elde etmek amacıyla çeşitli optimum değerlere ihtiyaç duyar. Bulunan ihtiyaçlarla daha doğru ve daha iyi sonuçlar elde edilmektedir. Bulunan sonuçlar işletmelerin daha verimli kararlar almasına, daha iyi müşteri hizmeti sunmasına ve rekabet avantajı elde etmesine yardımcı olmaktadır (Feurer & Hutter, 2019, ss. 3-4). Model performansının daha doğru ve yüksek performansta sonuç vermesi sistem güvenliğine katkı sağlamanın yanı sıra, süreçlerdeki kesintileri azaltarak üretimdeki verimliliği artırmaktadır. Aynı zamanda, seri üretim ürünleri söz konusu olduğunda, manuel tespit süreçleri personelleri zorlamaktadır. Ancak modelin yüksek performanslı doğru tespiti, daha hızlı bir şekilde kalitenin kontrolünü yaparak kaliteli ürünlerin üretilmesini sağlamaktadır (Shankarrao Patange & Bharatkumar Pandya, 2023, ss. 623-624). Özetle, makine öğrenmesinin model performansının yüksek olması, daha kolay karar verme süreçleri sunmaktadır. Bu yüksek performanslı sistemler, mühendislik tasarım süreçlerinde önemli bir rol oynamaktadır. Makine öğrenmesinin model performansının yüksek olması, tasarım seçiminde doğru ve optimize edilmiş çözümler sunarak sonuçların kalitesini artırmaktadır. Ayrıca, hesaplama hızı ve doğruluğu, personellerin zaman tasarrufu yapmasını sağlar ve tasarım süreçlerinin verimliliğini artırmaktadır (Romeo vd., 2020, s. 1).

Üretim sektörü, tarih boyunca ekonomik gelişimi ve refahı en fazla destekleyen sektörlerden biri olmuştur ve günümüzde de aynı konumunu korumaktadır (Sadıç, 2022). Dolayısıyla, üretim endüstrisinin ihtiyaçlarını belirlemek, veriye dayalı teknolojik çözümleri tasarlamak ve çözümleri hayata geçirmek, üretimde verim için önem arz etmektedir.

BİRİNCİ BÖLÜM
ENDÜSTRİDE MAKİNE ÖĞRENMESİ İLE KESTİRİMCİ BAKIM: UYGULAMA VE AKADEMİK ÇALIŞMALAR

1.1. ENDÜSTRİ SEKTÖRÜNÜN YENİ İHTİYAÇLARI VE OTOMASYONUN ARTAN ÖNEMİ

Sektörel çalışmalarda ve hayatta sürekli bir değişim söz konusudur. Özellikle endüstride teknolojinin sunduğu hizmetler geliştikçe, endüstrinin de bu değişime uyum sağlama ihtiyacı açığa çıkmaktadır. İnsan aklından yararlanmakla birlikte makinelerin hızlı üretim becerisini de kullanan yenilikçi çalışmalar bu anlamda dikkat çekmektedir. Bunlardan biri de büyük veri yönetimi teknolojilerinin kullanılmasıdır. Verilerin hızlı ve etkili analizi, karar verme süreçlerine anlık etki ederek doğru kararların hemen alınabilmesini sağlamakta, dolayısıyla işletme ve organizasyon için önemli bir avantaj oluşturmaktadır (Aktan vd., 2018). İşletmeler, dağıtım ve lojistik optimizasyonu, işlem sürelerinin azaltılması, müşteri kayıplarının önlenmesi, müşteri takibi, şeffaflık, müşteriye uygun hizmet ve mal sunumu, performans geliştirme ve sürdürülebilir ekonomik yaşam sağlama amacıyla kullandıkları büyük veri analizleri sonucunda elde ettikleri bilgileri kullanarak zaman ve maliyet kısıtlarını kendi lehlerine çevirmekte ve rekabet avantajı elde etmektedir (Akdoğan & Akdoğan, 2018).

Günümüzde artan endüstrileşme, çeşitli ve karmaşık hale gelen sistemler, zamanla çoğalan alternatif yöntemler ve önem kazanan karar verme süreçleri, teknolojinin ilerlemesiyle birlikte otomasyon ağırlıklı uygulamaların artmasına ve bakım giderlerini azaltma ihtiyacının zorunlu hale gelmesine yol açarak, kestirimci bakım uygulamalarını ilgi çeken bir disiplin haline getirmektedir (Bektaş, 2020).

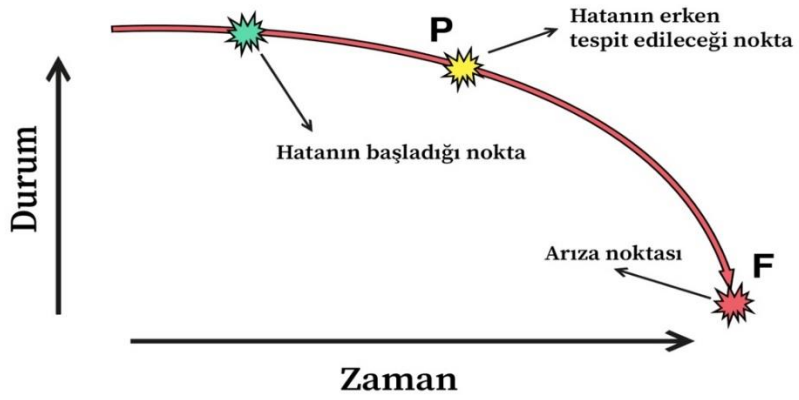
1.2. KESTİRİMCİ BAKIM

Kestirimci bakım, tarihsel verilere, mekanik modellere ve alan bilgisine dayalı sağlık yöntemi ve analiz metodu olarak tanımlanmaktadır (Zhong vd., 2023). Aynı zamanda öngörücü bakım olarak da adlandırılan kestirimci bakım, veri odaklı tekniklerle arızalarını önceden tahmin ederek varlıkların bakım planlarını etkili bir

şekilde yönetmek için uygun bir kavramdır (Kane vd., 2022). Kestirimci bakım uygulamaları için makine öğrenmesi algoritmaları bakım planlamasında arızanın önceden tahmin edilmesi ve önlem alınması amacıyla 1970'lerden beri kullanılmaktadır (Nur Calayır vd., 2021). Kestirimci bakım çalışmasında kullanılan bu algoritmalar, sıcaklık, basınç, yağ kalitesi, gürültü ve titreşim sensörü gibi cihazlardan aldığı verileri kullanarak, makinenin kalan faydalı ömrünün tahmin edilmesini ve bakım ihtiyacının belirlenmesini amaçlamaktadır (Özkat, 2021). Aynı zamanda kestirimci bakım ile karar verme, sensörlerle gerçek zamanlı olarak elde edilen verilerden hareketle gelecekte oluşabilecek arıza modelleri hakkında öngörülerin üretilmesi ve bunlara dayanarak bakım önerileri ve planların geliştirilmesi aşamasını ifade etmektedir (Bousdekis vd., 2019).

Kestirimci bakımın potansiyeli, Şekil 1'de gösterildiği gibi P-F eğrisi kavramıyla açıklanabilir. Varlıkların zamanla kötüleşmesi ve sonuç olarak işlevsel arıza meydana gelmesi durumunun ele alındığı ifade edilmektedir. P-F eğrisi ile, P ile gösterilen olası arıza noktasının tespit edilebileceği zamandan F ile ifade edilen işlevsel arıza noktasına kadar olan sürenin bir 'fırsat penceresi' sunduğu belirtilmektedir. Bir diğer ifadeyle P-F eğrisi, varlıkların zamanla (x ekseni) kötüleşmesini ve sonuç olarak işlevsel arıza meydana gelmesi durumunu (y ekseni) göstermektedir. Arızanın olası zamanları, tarihsel veri analizinden elde edilen bir dağılımdır. Arıza bakımı genellikle F noktasından sonra uygulanırken, kestirimci bakım F'den kaçınmak için belirlenen P-F zaman aralıklarından yararlanır. Bir başka ifadeyle kestirimci bakım, büyük verilerden yararlanarak insanlar tarafından görülebilir semptomlar ortaya çıkmadan önce bile P-F aralığını en üst düzeye çıkararak zamanında ve maliyet-verim kararlarında destekleyici rol almaktadır (Bousdekis & Mentzas, 2021, s. 8).

Şekil 1: P-F Eğrisi Bağlamında Kestirimci Bakım



Kaynak: Brkovic vd., 2015'ten yazar tarafından uyarlanmıştır.

Endüstride kullanılan bakım türleri arasında kıyaslama yapıldığında, kestirimci bakımın bazı avantajları bulunmaktadır. Bu avantajlar şu şekildedir:

- Ekipman yalnızca yakın bir arızadan önce kapatıldığından, durmalar ve süreleri azalır.
- Ekipmanı korumak için harcanan toplam süreyi azaltarak, bakım maliyetlerini ve işletme maliyetlerini azaltır.
- Yıkıcı zararlardan kaçınarak, bakım maliyetlerini azaltır ve bakımları kolaylaştırır.
- Makinelerin kullanılabilirliğini ve güvenliğini artırarak, iş bitim sürelerini optimum hale getirir.
- Ekipmanın ve işlem ömrünün uzatılmasını sağlar (Ümit Gürsoy vd.,, s. 6)

Kestirimci bakım stratejisinin avantajları olduğu gibi, bu yöntemin tercih edilmesinin önünde engel olarak nitelendirilebilecek bazı dezavantajları da bulunmaktadır. Bu stratejinin uygulanması aşamasında, gerekli cihaz ve eğitim maliyetinin yüksek olması nedeniyle önemli engellerle karşılaşmaktadır ve sadece gerekli donanımın satın alınmasının ötesinde, fabrika yönetiminin desteği ve bakım ekibinin kestirimci bakım felsefesini benimsemesi gerekmektedir (Tabak & Özkaymak, 2020).

Kestirimci bakım yönteminde, makine öğrenmesi modelleri oluşturulmasıyla doğru bakım öngörüleri elde edilmesi amaçlanmaktadır. Modellerin daha verimli

sonuç vermesi amacıyla hiperparametreler kullanılmaktadır. Hiperparametreler, makine öğrenmesinde en uygun model parametrelerinin belirlenmesi ve en iyi hiperparametre değerlerinin tespit edilmesi için kullanılan bir süreç olarak tanımlanmaktadır (Çınaroğlu, 2017). Dolayısıyla, hiperparametre kullanılmayan modellerin model performansı düşük olabilmektedir. Ancak bu husus mevcut akademik çalışmalarda yeteri kadar vurgulanmamaktadır. Bir sonraki bölümde bu alanda yapılan çalışmalar için hangi veri analizi yöntemlerinin ve yöntemlere bağlı hangi algoritmaların kullanıldığı ve modellerin doğruluk oranları ile ilgili hangi çalışmaların yapıldığı ele alınmaktadır.

1.3. LİTERATÜR TARAMASI

Mevcut çalışma ile ilgili literatür taraması, konunun farklı alanlarla kesişmesi sebebiyle, birkaç farklı alanda gerçekleştirilmiş ve endüstri dünyasındaki bakım uygulama yöntemleri, kestirimci bakım uygulamaları yöntemleri, makine öğrenmesi algoritmaları ve model performans değerleri odağında çalışmalar incelenmiştir. Kestirimci bakım alanında, farklı veri toplama yöntemleri ile çeşitli makine öğrenmesi ve derin öğrenme yöntemleri kullanılarak benzer çalışmalar yapıldığı gibi, farklı çalışmalarda yapılarak çeşitli sonuçlar elde edilmiştir.

1.3.1. Veri Toplama Yöntemleri Açısından Kestirimci Bakım Çalışmaları

Yapılan çalışmalar incelendiğinde modellerde kullanılacak veriler için farklı veri toplama süreçleri kullanıldığı gözlemlenmiştir. Çalışmaların temelinde sensör verileri kullanılmıştır. Verilerin toplanması için IoT'nin diğer toplama yöntemlerinden daha fazla kullanıldığı gözlemlenmiştir. Güven ve Şahin (2022), toplu taşıma araçlarında plansız arızalardan kaynaklanan sorunların en aza indirilmesi için IoT teknolojisini kullanarak sensör verileri ile kestirimci bakım çalışması yapmıştır. Bu çalışmada IoT teknolojine bağlı sensör verileri test ve eğitim veri kümelerine ayrılarak kesitirimci bakım uygulaması gerçekleştirmişlerdir. Aynı zamanda Nikfar ve diğerleri (2022), endüstriyel motorları izlemek için IoT ve sensör teknolojilerini kullanmışlardır. Çalışmalarında ivmeölçer ve jireskop sensörlerinin titreşim verilerini toplayarak motorların durumunu incelemişlerdir. Çalışma sonucunda IoT ve sensör

teknolojilerinin endüstriyel motor izleme ve bakımında etkin bir şekilde kullanılabileceğin göstermişlerdir. Soylu ve diğerleri (2021), ses, sıcaklık ve nem sensörü verilerini kaydetmek amacıyla Raspberry Pi 3 Model B adlı IoT tabanlı mini bir bilgisayar kullanmıştır. Bu cihazın düşük maliyeti, küçük boyutu, enerji tasarrufu desteği ve çeşitli bağlantı seçenekleri özelliklerinin olması nedeniyle tercih edildiği belirtilmiştir.

Literatürde IoT teknolojisi kullanımı dışında farklı veri toplama süreçlerinden yararlanan çalışmalar da bulunmaktadır. Yeardley ve diğerleri (2022), bir endüstriyel tesiste sensör verilerinin elde edilmesinin zorluğundan bahsederek simülasyon yöntemi ile veri toplama sürecini gerçekleştirmiştir. Simülasyon yöntemi ile 61 farklı sensör oluşturulmuş ve model verisi olarak 27.073 adet okuma gerçekleştirilmiştir. Ceyhan ve diğerleri(2022), verilerin döküm fabrikasındaki makinelerden toplanması için OPC UA haberleşme standardı kullanmıştır. Programlanabilir Mantık Kontrolleri (PLC – Programmable Logic Controller) bir diğer tabiriyle dijital bilgisayarlar kullanılarak veri setleri hazırlanmış ve OPC UA haberleşme standardı aracılığıyla farklı marka ve modellerdeki makineler ile veri iletişimi sağlanarak veri üzerinde standartlaştırma işlemi yapılmıştır. OPC UA yöntemi veri iletişimini marka ve modelden bağımsız hale getiren bir teknoloji olduğu için tercih edilmiştir. Makinelerden veri toplamak yerine hazır veri kullanan çalışmalar da olmuştur. Herrero ve diğerleri (2022), Kaggle sitesinde bulunan, gerçek zamanlı bir vaka çalışmasını barındıran hazır veri seti kullanmışlardır. Çalışmadaki veri setinde küçük bir köye su temininden sorumlu su pompasına takılan 52 sensörden alınan değerler bulunmaktadır. Gerçek makineye erişim olmadığı için, sistem davranışını simüle etmek için veri setini örnek bir çalışma davranışı olarak kullanmışlardır.

1.3.2. Veri Analizi Yöntemleri açısından Bakım Çalışmaları

Literatürdeki kestirimci bakım uygulama araştırmaları analiz teknikleri açısından değerlendirildiğinde, makine öğrenmesi, derin öğrenme ve istatistiksel yöntemlerin ağırlıklı olarak kullanıldığı görülmektedir. Bahsi geçen veri analizi yöntemleri farklı algoritmalardan yararlanılarak uygulanmıştır.

Yapılan çalışmalar incelendiğinde veri analizi için makine öğrenmesi yöntemlerinin sıklıkla kullanıldığı gözlemlenmiştir. Nur Calayır ve diğerleri (2021), çalışmalarında makine öğrenmesi yöntemlerinden faydalanmıştır. Çalışmalarında doğrusal yöntemler ve sınıflandırma yöntemi gibi birçok algoritma kullanmışlardır. Doğrusal yöntemlerde Lineer Regresyon ve Polinomial Regresyon yöntemlerini tercih ederek sonuçlarını kök ortalama kare hatası hesaplama (RMSE – Root Meaan Squared Error) yöntemi ile değerlendirmiş ve yüksek oranda doğruluk sonucuna ulaşmışlarıdır. Çalışmalarında doğrusal yöntemden yararlanan diğer bir çalışma da Kasap ve diğerleri (2021), makine öğrenmesi yönteminden yararlanarak çalışmalarını gerçekleştirmiştir. Çalışmalarında Ridge Regresyon (RR – Ridge Regression) ve Lojistik Regresyon (LR – Logistic Regression) yöntemlerinden yararlanarak Accuracy, Precision, Recall ve f1 skorları açısından sonuçlarını test etmişledir. Accuracy değerlerinin diğer yöntemlere göre daha yüksek oranda doğru sonuç verdiğini tespit etmişlerdir. Ayvaz ve diğerleri (2021), çalışmalarında makine öğrenmesi yöntemlerinden yararlanarak, yapay sinir ağı (ANN – Artificial Neural Network), destek vektör makineleri (SVM – Support Vector Machines) ve rastgele orman (RF- Random Forest) algoritmalarından yararlanmıştır. Makine öğrenmesi veri analiz yöntemlerinde klasik algoritmalar kullanımlarının yanı sıra çalışmalarına regresyon analizine ait çok katmanlı algılayı (MLP – Multi-Layer Perceptron), Gradient Boosting tabanlı olan aşırı gradyan artımı (XGBoost – eXtreme Gradient Boosting) ve uyarlanabilir yükseltme (Adaboost – Adaptive Boosting) gibi yöntemler de kullanarak daha kapsamlı çalışmalar da yapmışlardır. Kullandıkları algoritmaların doğruluğunu test etmek için r-kare (R^2- R-squared), ortalama mutlak hata (MAE – Mean Absolute Error), ortalama mutlak yüzde hatası (MAPE – Mean Absolute Percentage Error) ve RSME yöntemlerine başvurmuşlardır. Değerlendirmeler ve karşılaştırmalar sonucunda R^2 değeri ile RF kullanılarak oluşturulan test veri setinden elde edilmiştir. Kimera ve diğerleri (2020), ANN, SVM, RF algoritmaları dışında k-en yakın komşuluk (KNN – k-Nearest Neighbors) algoritmalarını da çalışmalarına eklemiştir. Tahmin sonuçları ile ilgili bilgiler paylaşılmış fakat doğruluk yöntemleri ve oranları paylaşılmamıştır. Yapılan çalışmalar incelendiğinde veri analizi için derin öğrenme (DL – Deep Learning) yöntemlerinin de kullanıldığı gözlemlenmiştir. Sanzana ve diğerleri (2022), DL yöntemleri ile kestirimci bakım çalışması yapmıştır. Yapılan çalışmada ANN,

yinelenen sinir ağı (RNN – Recurrent Neural Network) ve uzun-kısa süreli bellek (LSTM – Long Short Term Memory) algoritmalarını kullanmıştır. Wang ve diğerleri (2023) de DL yöntemini kullanarak kestirimci bakım uygulaması üzerine bir çalışma gerçekleştirmiştir. Çalışmalarında derin inanç ağları (DBN – Deep Belief Network), RNN ve derin evrişimsel sinir ağı (DCNN – Deep Convolutional Neural Network) gibi yöntemlerle zaman serileri odağında kestirimci bakım çalışması gerçekleştirmişlerdir. RMSE ve score yöntemleri ile doğruluk oranlarını test etmişlerdir. Sonucunda ise yüksek doğruluk oranlarına ulaştıklarını belirtmiştir. Tüm bu çalışmalardan farklı olarak istatistiksel yöntem kullanan çalışmalar da olmuştur. Xayyasith ve diğerleri (2018), kesitirmci bakım çalışması için ANN, SVM ve KNN dışında istatistiksel yöntem olan diskriminant analizi (DA – Discriminant Analysis) ile çalışmışlardır. Yapılan çalışma sonucunda da accuracy doğrulama yöntemi kullanılarak %100 doğruluğa ulaşmışlardır.

Literatürde kestirimci bakım ile ilgili çalışmalar birçok sektörde yapılmıştır. Bunlar arasında havacılık, otomotiv, üretim, enerji, tıbbi cihazlar, savunma ve endüstriyel ekipmanlar gibi sektörler bulunmaktadır. Ayrıca bina yönetimi ve bakımı gibi tesis yönetimi sektörü de kestirimci bakım tekniklerini kullanmaktadır.

Yayınlanmış birçok çalışma, kestirimci bakımda kullanılan verilerin öğrenme ve test oranlarının ve hiperparametre ayarlarının model performansına etkisini incelemektedir. Ancak bu çalışmaların birçoğu, verilerin öğrenme ve test oranlarını ve hiperparametre değerlerini modellere tanımlı varsayılan değerlerle kullanmış ve hem oranlar hem de hiperparametre değerleri üzerinde herhangi bir değişiklik yapmamışlardır. Bu nedenle, verilerin farklı oran dağılımlarının model performansı üzerindeki etkisi ya da hiperparametrelerin varsayılan değerler dışındaki değerleri ile model performansı üzerindeki etkisi hakkında eksiklikler tespit edilmiştir. Bu çalışma, farklı makine öğrenmesi algoritmaları kullanılarak, kestirimci bakım çalışmalarında hiperparametre ve veri oranı kombinasyonlarının her bir algoritmaya olan etkisini araştırarak, bu konuda literatüre daha kapsamlı bir katkı sunmayı amaçlamaktadır.

İKİNCİ BÖLÜM
ENDÜSTRİ 4.0 VE MAKİNE ÖĞRENMESİ

2.1. YAPAY ZEKA

Günümüzde yapay zeka kavramı birçok iş alanında kullanılmaktadır. Özellikle endüstride yapay zeka uygulamaları ile işletmeler çalışma verimi artırmayı hedeflerken aynı zamanda iş yönetiminde karar mekanizmalarında da yapay zeka çalışmalarından yararlanmaktadır.

Yapay zeka bilgisayarların akıl yürüterek, bir probleme çözüm üretmesi, bu çıkarımları anlamlandırması ve herkes tarafından anlaşılabilir bir yapıda insan gibi davranış göstermesi, diğer bir ifadeyle en yüksek seviyede bilişsel davranış özelliği göstermesidir (Arslan, 2020). Özetle yapay zeka akıl odaklı programlanmış ve insan davranışı sergileyen yapılar olarak tanımlanabilmektedir.

Yapay zeka kavramı ilk olarak İkinci Dünya Savaşı sırasında Alan Mathison Tureng tarafından kodları deşifre eden "Bombe" adlı makinenin icat edilmesi ile ortaya atılmıştır ve bu süreçten sonra birçok bilim insanı yapay zeka alanında çeşitli çalışmalar yapmaya başlamıştır (Coşkun & Gülerlioğlu, 2021). Günümüzde yapay zeka çalışmaları birçok sektörde kullanılmaktadır. Özellikle üretimde makinelerin kullanılması ve bu makinelerden üretilen verilerin sektörde verimliliği arttıracak yönde analiz edilmesi endüstri yeteneklerinin artışında önemli rol oynamaktadır. Yapay zeka sadece endüstride değil telefonlarda, oyun konsollarında, asistan olarak kullanılan yazılımlarda, görüntü işleme modellerinde ve dil çeviri programlarında da sıklıkla kullanılmaktadır (Kaya, 2021, s. 68).

Bugün tüm dünyada organizasyonlar yapay zeka uygulamalarından sıklıkla yararlanmaktadır. Yapay zeka çalışmaları Tablo 1'de gösterildiği gibi sınıflandırmış ve işletmelerin nasıl yararlanacağı konusunda öneri sunulmuştur.

Tablo 1: Yapay Zeka Teknolojileri ve İşletme Çözümleri

Teknoloji	Tanım	Çözüm Örneği
Algı		
Bilgisayarlı Görü	Görüntüyü elde etme, işleme, analiz etme ve anlamak	• Video analizleri aracılığıyla ticari faaliyetlerin ilgili risk ve güvenlik açıklarında farkındalık sağlamak. • Perakende satışlarda video analizleri müşteri davranışlarını etkili ve sistematik bir şekilde değerlendirmek.
Ses İşleme	Sesleri tanımlama, tanıma, analiz etme ve konuşma	• Çağrı merkezlerine entegre konuşma tanıma teknolojileri ile arayanların kimliklerini tanımlamak.
Sensör İşleme	Kamera ve mikrofon dışında alınan bilgilerin işlenmesi ve analiz edilmesi	• Ekinlerin daha hassas bakımı için sıcaklık ve nem durumlarının takibini sağlamak.
Anlam		
Doğal Dil İşleme	Sözlü ve yazılı formlardan dili anlama ve üretme	• Akıllı telefonlarındaki kişisel asistanlar aracılığıyla doğal dili teknolojisi kullanarak hizmet ve rehberlik hizmeti sağlamak
Davranış		
Çıkarsama Motoru	İş kuralları gibi sabit bir bilgi tabanından cevap türetme	• Otomatik olarak borç ya da kredi onayı vermek. • Vize verilmesini onaylamak.
Uzman Sistemler	Bir insanın karar verme yeteneğini taklit etme ve mevcut bilgilerle akıl yürüterek karmaşık sorunları çözme	• Tıbbi teşhis ve yasal araştırma uzman sistemlerinin milyonlarca veriyi eleme yeteneği yardımıyla veri kaynaklarını sentezlemek ve kullanıcıya bilgi sunmak.
Makine Öğrenmesi	Deneyime dayanarak karar süreci değiştirme	• Yazılım araçları ile verimliliği arttırmak

Kaynak: Bataller ve Harris, 2016'dan yazar tarafından derlenmiştir.

2.2. YAPAY ZEKANIN KARAR VERME ÜZERİNE ETKİSİ

Yapay zeka alanı birçok disiplinin bir arada kombine çalıştığı bir sistem olarak görülmektedir. Bu disiplinler istatistik, mantık, bilgisayar bilimleri, sinirbilim gibi alanlardan oluşmaktadır. Örgütler başarıya ulaşabilmek için gerçekleşmesini

istedikleri amaçlarını ölçümlemek için bu disiplinlerden yararlanmaktadır (İnce vd., 2021, s. 54).

Teknolojik ilerlemelerin yaşanması, bilgilerin erişilebilir hale gelerek dijital ortamda herhangi bir bilgisayar tarafından okunabilir hale getirilmesi olarak adlandırılan dijitalleşmeyi ve buna bağlı olarak dünya çapında üretim yapısının hızlanmasını beraberinde getirmektedir. Üretim yapısındaki değişimler ise doğal olarak tüketimi ve tüketim kararlarını etkilemektedir. Dijitalleşme alanındaki küresel rekabetle birlikte, bu alanda yapılacak çalışmalar ülkelerin ekonomik büyümesi için en önemli faktörlerden biri haline gelmiştir. Aynı zamanda organizasyonlar sürdürülebilir çalışmalar gerçekleştirmek için de çeşitli kararlar almak durumundadır. Bu karar mekanizmasında insana destek olma ve gerçekçi sonuçlar ortaya çıkaran yapay zeka uygulamaları belirsiz olabilecek süreçleri daha net hale getirerek karar mekanizmasını güçlendirmektedir (Gür & Yıldız, 2022, s. 203).

Karar süreçlerinde yapay zeka uygulamaları verilerden anlamlı sonuçlar çıkarabilmektedir; fakat sezgisel ve duygusal olarak karar süreçlerine herhangi bir etki koyamamaktadır. Bu yüzden bu süreçlerde insan becerilerinden yararlanılmaktadır. Burada yapay zeka ve insanın iş birliği içinde en uygun karar sonucunu ortaya çıkaracağı düşünülmektedir (Bogehoj, 2016, ss. 9-10).

2.3.YAPAY ZEKA UYGULAMALARINDA VERİ ANALİTİĞİ

Veriden faydalı bilgi çıkarımına ulaşmak için çeşitli yöntemler uygulanmaktadır. Burada yapay zeka alanında veri analitiği verileri anlamak ve uygun sonuçlara ulaşmak için çalışmalarda önemli rol üstlenmektedir. Houtmeyers ve diğerleri (2021), veriden anlamlı bilgiye ulaşabilmek için veri üzerindeki çalışmaları 4 başlıkta toplamıştır.

2.3.1. Betimleyici Analitik

Betimleyici analitik çalışması yapılmadan önce tutarlı, güvenilir ve geçerli verilerin toplanması işlemlerinin gerçekleştirilmesi gerekmektedir. Veri kaynağında toplanan veriler betimleyici analitik aşamasında veriler hakkında bilgi sunmaktadır. Veriler arasındaki ilişkiler ve çeşitlik istatistiksel değerlerin raporlandığı alandır. Bu

raporlamada veriler üzerinde ortalama hesaplama, medyan çıkarımı ve veri görselleştirme gibi değerler bulunmaktadır. Özetle betimleyici analitik alanında "Nedir?" sorusunun cevabı aranmaktadır.

2.3.2. Teşhis/Tanı Analitiği

Betimleyici analitikte verilerin rapor sonuçları kullanılarak veri sonuçlarında ortaya çıkan sonuçların nedenlerine Teşhis/Tanı analitiği alanı hizmet etmektedir. Bu alanda meydana gelen sonuç için "Neden?" sorusu sorulur ve cevaplar üzerinden sebepler ortaya çıkartılmaktadır.

2.3.3. Tahminsellik Analitiği

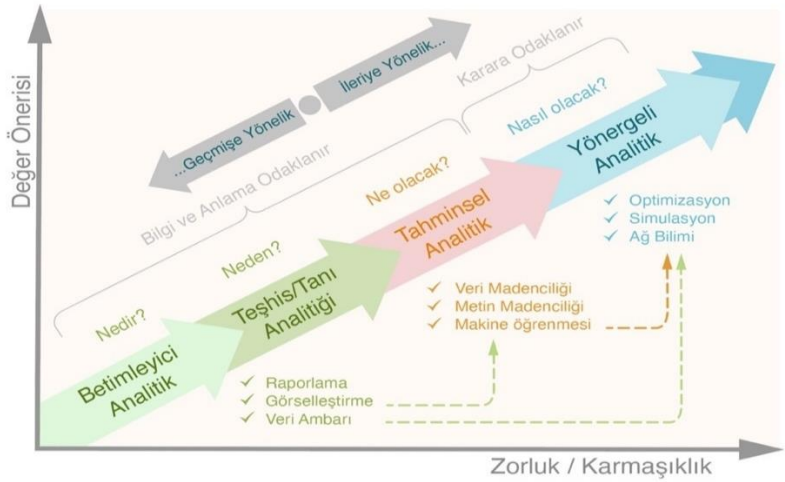
Bu alanda veriler kullanılarak geleceğe yönelik tahminler oluşturulmaktadır. Burada "Ne olacak?" sorusuna cevap aranmaktadır.

2.3.4. Yönergeli Analitik

Tahminsellik analitiğinde ortaya çıkan tahminler üzerinden yeni yönergelerin oluşturulduğu alan yönergeli analitik alanıdır. Bu alanda yeni oluşturulan yönergelerle verimliliğin arttırılması hedeflenmektedir. Yönergeli Analitik çalışmaları "Nasıl olacak?" sorusuna cevap aramaktadır.

Şekil 2'de veri analitiği zorluk/karmaşıklık ve buna bağlı değer önerisi bağlamında görselleştirilmiştir. Şekilde betimleyici analitiğe doğru geçmişe yönelik çıkarımlar gösterilirken yönergeli analitiğe doğru ileriye yönelik kararların olduğu gösterilmektedir. Süreçlerde de hangi çalışmaların yapıldığı detayları ile gösterilmektedir.

Şekil 2: Veri Analitiğine Kavramsal Bakış



Kaynak: Delen & Ram, 2018'den yazar tarafından uyarlanmıştır.

2.4. CRISP-DM METODOLOJİSİ İLE YAPAY ZEKA UYGULAMALARI GERÇEKLEŞTİRME SÜRECİ

İş dünyasında yapay zeka uygulamalarını gerçekleştirmek için iş süreçlerini tasarlamak gerekmektedir. Bu akışı tasarlamak için çeşitli metodolojilerden yararlanılmaktadır. Bunların başında da bir veri madenciliği uygulamasını inşa etmek için kapsamlı bir plan ile iş sürecini tasarlayan veri madenciliği için sektörler arasında standart süreç (CRISP-DM – Cross-Industry Standard Process for Data Mining) metodolojisi gelmektedir (Shearer C, 2000). İlgili metodoloji ile bir projenin iş süreç döngüsü basamakları:

- İşi Anlama

o Hem organizasyon bazlı hem de endüstriyel anlamda işletmenin geçmişi hakkında bilgi sahibi olunur.

o İş problemleri ve fırsatları hakkında detaylar öğrenilir.

o Verilerden hangi hedeflere ulaşılması gerektiği belirlenir.

o Başarı kriterleri tanımlanır.

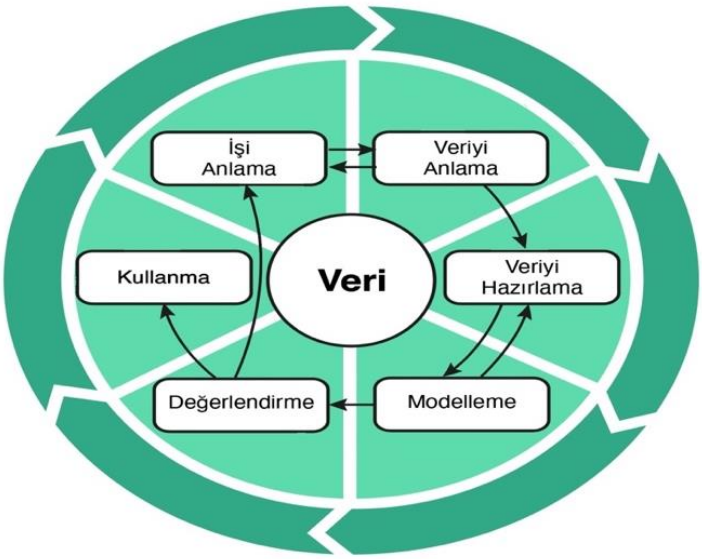
o İş ihtiyaçları, olasılıklar ve kısıtlar tanımlanır.

o Mevcut iş kaynaklarının ve ticari terimlerin ne olduğu belirlenir.

- Veriyi Anlama ve Hazırlama
  - Veri toplama süreçleri gerçekleştirilir.
  - Grafiklerle veriyi anlama ve yorumlama işlemleri yapılır.
  - Ortalama, standart sapma, minimum, maksimum gibi özet istatistiksel değerler hesaplanır.
  - Aykırı değerler bulunur.
  - Veri temizleme işlemleri gerçekleştirilir.
  - Var olan veriler kullanılarak yeni özelliklere sahip veriler türetilir.
  - Veriler arasındaki ilişkiler gözlemlenir ve korelasyon değerleri hesaplanır.
- Modelleme
  - Öğrenme verileri ve test veri tanımlanır.
  - Veriler için algoritma modeli oluşturulur.
  - Modelin parametreleri belirlenir.
  - Modele uygun değişkenler ayarlanır.
- Değerlendirme
  - Model performansları hesaplanır.
  - Model sonuçları değerlendirilir.
- Modeli Kullanma ve Yayma
  - Kurulan model sisteme entegre edilir (Karahan Adalı & Balaban, 2019, ss. 95-96).

Şekil 3'te CRISP-DM sürecinin basamakları ve birbirleriyle olan bağlantıları gösterilmektedir.

Şekil 3: CRISP-DM Sürecinin Basamakları



Kaynak: Tounsi vd., 2020'den yazar tarafından uyarlanmıştır.

2.5. MAKİNE ÖĞRENMESİ

Yapay zeka çeşitli alt birimlere ayrılarak farklı alanlarda sektöre hizmet etmektedir. Bu alanlar başlıca görüntü ve ses işleme, sensör teknolojileri, uzman sistemler gibi alanlardan oluşmaktadır. Makine öğrenmesi de yapay zeka çalışmalarının bir alt dalı olarak sektördeki yerini almaktadır.

Yapay zeka, makine öğrenmesinden daha geniş bir araştırma alanıdır. Yapay zeka, makinelerin beyin fonksiyonlarını taklit etmesini amaçlamaktadır. Algoritmalara dayanarak, bilgisayarlar ve cihazlar görevleri akıllı bir şekilde yerine getirebilmektedir. Makine öğrenimi ise yapay zekanın bir parçası olup, makinelerin veriyi soyutlama ve genelleme yeteneklerini geliştirerek algoritmalarını veriye göre değiştirebilme becerisini hedeflemektedir. Makine öğrenimi algoritmaları, makinelerin insan gibi düşünmeye yaklaşmasını sağlamak için kullanılmaktadır (Velu Sindhu, 2020, s. 88). Makine öğrenmesini statik programlama yönergelerini izleyen modellerin aksine, veriye dayalı kararlar alabilen dinamik algoritmalar geliştiren bir bilgisayar disiplini olarak da tanımlamıştır (Thessen, 2016).

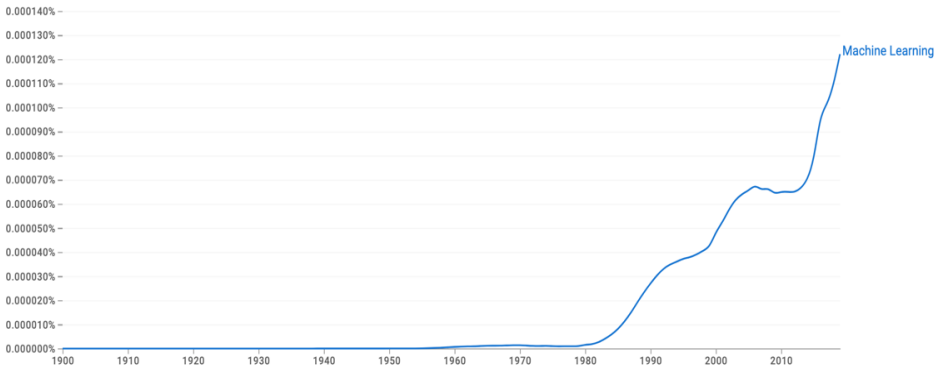
Şekil 4: Yapay Zeka ve Alt Alanları Arasındaki İlişki



Kaynak: Rattan vd., 2022'den yazar tarafından uyarlanmıştır.

Yapay zeka kavramı ilk olarak İkinci Dünya Savaşı sonrası ilk kez karşımıza çıksa da 1980 yılı itibariyle kullanımını arttırmıştır. Google Ngram görüntüleyicide makine öğrenimi ifadesinin yıllara göre kullanımı Şekil 5'te gösterilmektedir.

Şekil 5: Ngram Görüntüleyicide "Machine Learning" İfadesinin Kullanımı



Kaynak: Google, 2022

Makine öğreniminin her geçen gün daha da fazla yaygınlaşması nedeniyle, organizasyonlarda ve iş hayatında bu kavramın karşımıza çıkması olası hal almaktadır. Özellikle büyük ölçekli verilerle çalışan iş alanlarında bu veriler karar süreçlerinde önemli bir rol almaktadır. Bu aşamada makine öğrenimi veri ile sektör ilişkisini ve

aynı zamanda sektör davranışlarını anlamlandırabilmesi veya karar verme aşamalarında otonom bir hal alması bu algoritmaların kullanımı değerli kılmaktadır. Makine öğrenimi çeşitli sektörlere uyarlanarak daha hızlı ve verimli iş süreçlerinin oluşmasına olanak sağlar (Payal vd., 2022, ss. 5831-5832).

Makine öğrenimi algoritmalarını kullanarak karar sistemlerini geliştiren ya da verimliliğini arttıran birçok iş alanı bulunmaktadır. Tablo 2'de makine öğreniminin kullanıldığı sektörler ve kullanım alanları yer almaktadır.

Tablo 2: Makine Öğrenimi Algoritmalarının Kullandığı Sektörler

Sektör	**Kullanım Alanları**
Sağlık	• Hastalık verileri incelenerek hastalığın ne olduğuna ait sonuçlara ulaşmak. • Hastalık durumları incelenerek tedavi yöntemlerinde iyileşme yöntemleri geliştirmek.
Otomasyon	• Makinenin hatalarının tespitini yapıp gerekli düzenlemeleri yapmak. • İnsandan bağımsız sorunların çözümünün bulunmasına katkı sağlamak.
Finans ve Bankacılık	• Finans için uygun değerlerin tahminini hesaplamak. • Dolandırıcılık işlemlerinde risk tespit etmek. • Kredi puanı kapsamında yönetimsel karar sağlamak.
Ulaşım	• Güvenli rotaların oluşturulması ile birlikte uygun varış zamanını ve en hızlı rotayı keşfetmek. • Aracın bozulma durumunu tahmin etmek. • Trafik doluluk oranlarını belirlemek.
Görüntü Tanıma	• Görüntünün tahmin edilmesi ve gönderi etiklerinde tahminin kullanıcı ile paylaşılmasını sağlamak.
Konuşma Tanıma	• Seslerin metne dönüştürülmesini sağlamak. • Sesin kime ait olduğunun belirlenmesine katkı sağlamak.
E-Ticaret	• Farklı firmaların önerilerini kullanıcıya çeşitli platformlarda sunmak.
İnsan Kaynakları	• İşe alım süreçlerinde başvuru sahiplerinin iş için uygunluğunu tahmin etmek. • Şirkette çalışan bireylerin iş performanslarını belirlemek.

Kaynak: Nozari vd., 2021'den yazar tarafından derlenmiştir.

2.5.1. Makine Öğrenmesi Terimleri ve Sınıflandırması

Makine öğrenimi çalışmalarını daha etkili ilerletmek için bu alanda kullanılan terimlerin ne anlama geldiğinin açıklanması gerekmektedir.

Yapılandırılmış Veri (Structured Data): Bir veri tabanı gibi tanımlanmış, indekslenmiş, referans gösterilebilir ya da kolayca ve doğru bir şekilde aranabilir verilerdir. Bu veriler çoğunlukla sayısal değerler, tarihler ya da kısa metin dizilerinden oluşmaktadır.

Yapılandırılmamış Veri (Unstructured Data): Veri tabanında tutulmayan, belli bir referans değeri olmayan düzensiz verilerden oluşmaktadır. Video ve resim gibi veriler yapılandırılmamış verilere örnek verilmektedir.

Özellikler (Features): Bir veri setindeki girdi değişkenleri, değişken ya da bağımsız değişken olarak da adlandırılmaktadır. Özellikler basit bir şekilde sayısal aralıklı değerlerden oluşabildiği gibi kategorik değer değişkenlerinden de oluşabilmektedir.

Kesin Referans (Ground Truth): Bir makine öğrenmesi modelinin tahmin edilmesi ya da sınıflandırılmasını eğitmek ya da sonuçlarını test etmek için kullanılan sonuç ya da çıktı değişkeninden oluşmaktadır.

Sınıf/Etiket (Class/Label): Bir sınıflandırma modelin modelin çıktısı olarak tanımlanmaktadır. Çıktılar birbirinin zıttı ise sınıf, çoklu çıktı ise etiket kavramı kullanılmaktadır.

Maliyet/Kayıp Fonksiyonu (Loss/Cost Funciton): Bir modelde eğitim sırasındaki tahmin ile gerçek değerler arasındaki farkı hesaplayan fonksiyondur. Bu fonksiyon modelin optimize edilmesi için minimize edilmektedir.

Parametreler (Parameters): Bir modelin eğitim sırasında performansını optimize etmek için kullanılan temel faktörlerdir. Bu faktörler regresyon modelindeki katsayılara benzemektedir

Hiperparametreler (Hyperparameters): Eğitim parametreleri olarak da bilinen hiperparametreler modelin veriden nasıl öğreneceğini belirleyen ve ileri seviye modellerin tasarımının nasıl yapıldığını belirleyen yapılardır. Bu faktörler sadece eğitimden önce ayarlanmaktadır ve eğitim seti üzerindeki performansa göre özellikler belirlemektedir.

Eğitim Seti (Training Set): Bir modelin eğitimi için kullanılan veri setinin alt kümesidir.

Test Seti (Test Set): Bir modelin performansını değerlendirmek için kullanılan verilerdir. Bu veriler üzerinde herhangi bir ayarlama ya da eğitim uygulaması

yapılmamış olmalıdır. Bu nedenle, tüm veri setinden alınmış küçük bir alt küme veri seti olabildiği gibi değerlendirme için kullanılabilecek tamamen harici benzer veri seti de olabilir.

Aşırı Öğrenme (Overfitting): Bir modelin, eğitim setinde yer alan verilere özel bir şekilde eğitildiği ve bundan dolayı genel bir modelin oluşturulamadığı durumdur. Bu durum, eğitim verilerinde mükemmel performans sunarken, test setlerinde düşük performans yaratmaktadır.

Yetersiz Öğrenme (Underfitting): Bir model, hiperparametrelerin tümünü optimize etmeye rağmen eğitim verilerinde hala kötü performans gösteriyorsa yetersiz öğrenme durumuna maruz kalmıştır. Bu durum, modeldeki verilerin ilişkilerini göstermek için uygun olmadığı anlamına gelmektedir.

Duyarlılık (Recall): Model tahminlerinin duyarlılığı ya da doğru pozitif oranı olarak tanımlanmaktadır.

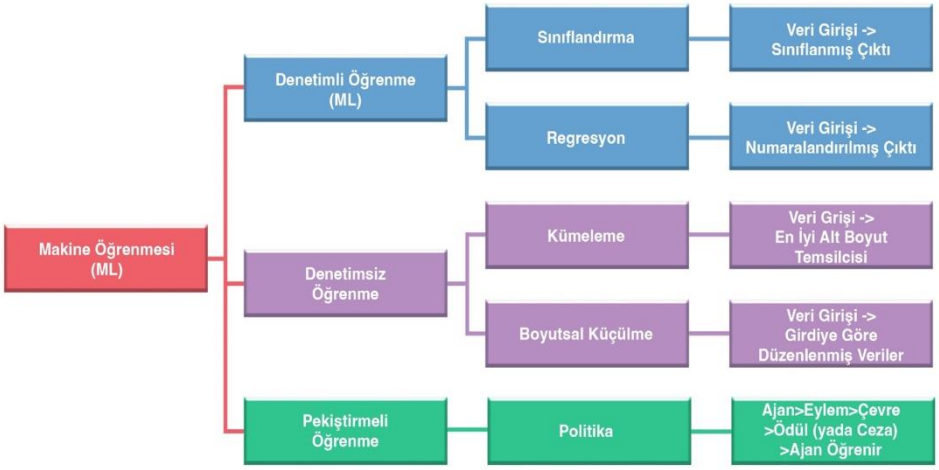
Kesinlik (Precision): Model tahminlerinin pozitif öngörü değerini yansıtmaktadır.

Doğruluk (Accuracy): Performans değerlendirmek için kullanılmaktadır. Doğru tahminlerin toplam tahmin sayısına oranını vermektedir.

F-Skoru (F-Score): Doğruluk ölçütü olarak kullanılan diğer bir yöntemdir. Precision ve Recall değerlerinin harmonik ortalamasını temsil etmektedir (Rattan vd., 2022, s. 72).

Makine öğrenimi algoritmaları verilerden öğrenme ve bu doğrultuda sektörde gelişme yeteneği sağlayan algoritmalarının incelenmesidir. Bu algoritmaların kullanılması sonucunda çeşitli sonuçlara ya da kararlara ulaşılmaktadır. Kullanılan karmaşık verilerin oluşturduğu çeşitli kalıpların bulunması verilerden sonuç üretilmesini sağlamaktadır. Öğrenme yaklaşımı, girdi ve çıktı değerlerinin türleri ve problemin türü makine öğrenmesi algoritmalarının kullanımını etkilemektedir. Bu parametrelere bakıldığında makine öğrenmesi algoritmaları denetimli, denetimsiz ve takviyeli öğrenme olmak üzere 3 ana kategoriden oluşmaktadır(Sah, 2020, s. 1). Şekil 6'da makine öğrenmesi algoritmalarının hiyerarşik yapısı bulunmaktadır.

Şekil 6: Makine Öğrenmesi Algoritmalarında Hiyerarşik Yapı



Kaynak: Rashidi vd., 2019'dan yazar tarafından derlenmiştir.

Veriler çeşitli türlerde kaydedilmektedir. Bu verilerin çeşitlerine göre uygulanacak algoritmalar da çeşitlilik göstermektedir.

Makine öğrenimi algoritmaları temel başlıkta 2 değişken türünden oluşmaktadır. Bu değişkenler sayısal değişkenler ve kategorik değişkenlerden oluşmaktadır. Sayısal değişkenler; nicel (kantitatif) değerlerden oluşan, gözlemlenebilir ve ölçülebilir değişkenlerdir. Bu değişkenler aralık ve oran ölçek türlerini barındırmaktadır. Aralık değerler başlangıç noktası sıfır olmayan negatif ve pozitif değer aralıklarından oluşmaktadır. Başlangıç noktası sıfır olan ölçek türleri ise oran olarak tanımlanmaktadır. Kategorik değişkenler ise, nitel (kalitatif) değerlerden oluşan, kelimeler vasıtasıyla değerlendirilebilen değişkenlerdir. Bu değişkenler nominal ve ordinal ölçüm türlerini barındırmaktadır. Nominal değerler kategorik sınıflar arasında herhangi bir derece olmayan ölçüm türü olarak tanımlanmaktadır. Ordinal ölçüm türleri ise kategorik değişkenler arasında fark ve derece içeren ölçüm türü olarak tanımlanmaktadır.

Veri türlerine göre makine öğrenimi algoritmalarında kullanılan yöntemler de değişmektedir. Bu yöntemler denetimli, denetimsiz ve pekiştirmeli olmak üzere 3 modelden oluşmaktadır (Aldahiri vd., 2021, s. 184).

2.5.2. Denetimli Öğrenme

Denetimli öğrenme, belirli bir girdi-çıktı eğitim örneği kümesine bağlı olarak, bir sistemin girdi-çıktı ilişkisi bilgilerini elde etmek için kullanılan makine öğrenimi yöntemlerinden biridir. Bu algoritmada kullanılan çıktılar, girdi verilerinin etiketi olarak düşünüldüğünden, girdi-çıktı eğitim verilerine etiketli eğitim verileri de denmektedir.

Denetimli öğrenme yöntemleri 2 başlıkta toplanmaktadır. Bu başlıklardan biri regresyon, diğeri ise sınıflandırma olarak tanımlanmaktadır. Sınıflandırma yönteminde veri girişlerinin sonucu kategorik olarak kategoriler halinde sonuçlanmaktadır. Regresyon yönteminde ise veri girişlerinin sonucu gerçek değerler halinde sonuçlanmaktadır (Q. Liu & Wu, 2012, s. 3244)

2.5.2.1. Karar Ağaçları

Karar ağacı, kök düğüm, dallar ve iç düğümleri olan bir ağaç gibi sınıflandırma ve regresyon modelleri oluşturmak için kullanılmaktadır (Kocakır Gacar & Deveci Kocakoç, 2020). Karar ağaçları, makine öğrenmesinde veri kümesindeki verilerin sınıflandırılması veya tahmini için görselleştirilen bir yol haritalamaktadır. Her düğüm, verilerin belirli bir özelliğine dayalı bir karar veya filtreleme mekanizması içerir ve yaprak düğümlerinde tahmin sonucunu bulundurmaktadır. Karar ağaçları, en iyi sonucu bulabilmek için veriler arasındaki ilişkileri keşfetmeye yardımcı olmaktadır ve sonuçları anlamak için kolay bir şekilde okunabilir ve anlaşılabilir bir yapı sunabilmektedir (Rokach & Maimon, 2015, s. 166).

Verilerin sınıflandırılması veya tahmin edilmesi amacıyla otomatik olarak karar ağacı yapısı oluşturmak için geliştirilmiş birçok farklı karar ağacı algoritması bulunmaktadır. Karar ağacı algoritmalarının amaçları arasında genelleştirme hatasını minimize etmek ve en uygun karar ağacı yapısını oluşturmaktır. Karar ağacının oluşturulması sırasında, dallanmalar için belirlenecek kriterler en önemli adım oluşturmaktadır. Belirlenecek kriterler için farklı yöntemler bulunmaktadır, bunlar arasında bilgi kazancı, bilgi kazancı oranı, gini indeksi, Tosting kuralı, Ki-Kare olasılık tablosu istatistiği sayılabilmektedir (Onan, 2015, ss. 10-11).

Eğitim veri setindeki niteliklerin hangisinin en fazla bilgi kazanç sağladığı belirlenerek, karar ağaçlarının oluşturulması sırasında düğümlerin seçimi yapılır ve burada Entropi ölçümü kullanılır (Çınar, 2019). Entropi, verilerin hangi sınıf içinde olduğunun belirsizliğini ölçen bir ölçüt olarak tanımlanmaktadır. Her düğümde entropi hesaplanarak, belirsizliğin en düşük olduğu düğüm seçilmektedir ve ağacın oluşmasına bu düğümden devam edilmektedir. Bu yöntemle, ağaç en uygun şekilde verileri ayırarak, en yüksek doğruluk oranını sağlamaktadır. Entropi formülü aşağıdaki gibi hesaplanmaktadır:

$$Entropi(S) = -\sum_{i=1}^{n} p_i x log_2 p_i$$

Formülde:

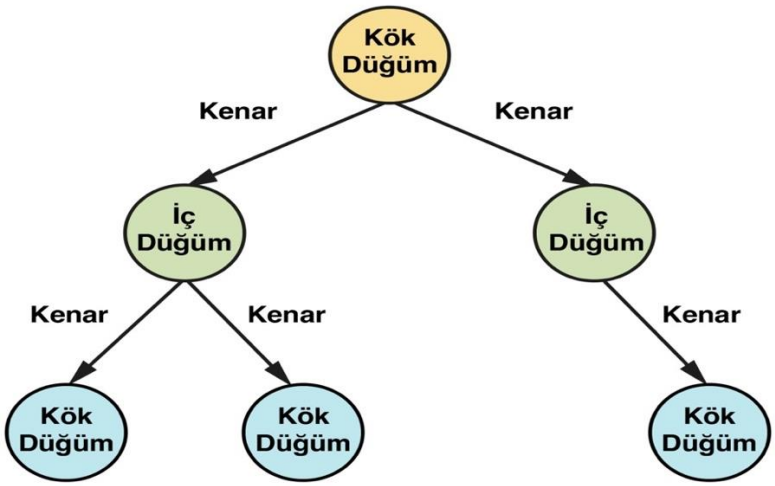
p_i: Verinin i. Sınıfına ait olasılığı

S: Entropi değerini ifade etmektedir.

Entropi değeri 0 ile 1 arasında değişmektedir. Belirsizliğin en yüksek olduğu durumlarda bu değer 1'e yaklaşmaktadır, belirsizliğin en düşük olduğu durumlarda da bu değer 0'a yaklaşmaktadır.

Şekil 7'de ağaç yapısının kök yapısı, dalları ve iç düğümlerinin genel yapısı gösterilmektedir. Burada kök düğüm en üst düğüm ve veri kümesinin tüm özelliklerine ilişkin karar noktasını belirtmektedir. İç düğümler, kök düğümün dışında kalan, her özelliğe göre bir karar yapısını temsil etmektedir. Dallar ise iç düğümlerden çıkan ve sonuçları temsil eden bağlantıları ifade etmektedir.

Şekil 7: Karar Ağaç Yapısı



Kaynak: Javed Mehedi Shamrat vd., 2022'den yazar tarafından derlenmiştir.

2.5.1.2. Naive Bayes

Naive Bayes sınıflandırması, istatistiksel yöntemler kullanarak verileri etiketlemek için kullanılan bir sınıflandırma türüdür ve kolay uygulanabilir olması nedeniyle sınıflandırma problemlerinde tercih edilmektedir (Çelik & Altunaydın, 2018). Bu algoritma, sınıflandırma yapmak için olasılık teorisini kullanmaktadır ve sınıflandırma yapılacak veri özelliklerine dayalı olarak olasılık hesaplamaktadır. Öncelikle, sınıflandırılacak veri örneğinin her bir özelliği için olasılık değeri hesaplanmaktadır. Bu olasılık değerleri, daha önce eğitim için kullanılan veri seti üzerinden oluşturulmaktadır. Oluşturulan bu olasılık değerleri kullanılarak, veri örneklerinin her bir sınıfa ait olma olasılığı hesaplanmaktadır. Sınıflar arasında oluşturulan bu değerler karşılaştırılarak, en yüksek olasılığa sahip sınıf, veri örneğinin tahmin sınıfı olarak tanımlanmaktadır.

Naive Bayes algoritması, Bayes teoreminin temel prensiplerine dayanmaktadır (Androutsopoulos vd., 2000). Bu prensibe dayanarak oluşturulan formül aşağıdaki gibidir.

$$P(c|x) = \frac{P(x|c)xP(c)}{P(x)}$$

$$P(c|x) = P(x_1|c)xP(x_2|c)x \dots xP(x_n|c)xP(c)$$

Formülde:

c: Tahmin edilmeye çalışılan sınıf

x: Tahmin eden sınıf

P(c|x) = x olayının gerçekleşmesi durumunda c olayının gerçekleşmesi olasılığı

P(c) = c olayının gerçekleşme olasılığı

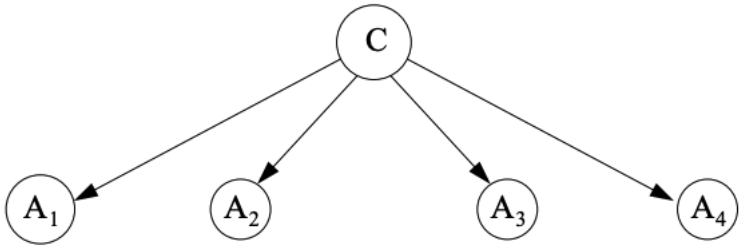
P(x|c) = c olayının gerçekleşmesi durumunda x olayının gerçeklemesi olasılığı

P(x) = x olayının gerçeklemesi olasılığı

Naive Bayes algoritması, bu formülü kullanarak verilen özelliklere sahip verinin her bir sınıfa ait olma olasılığını hesaplamaktadır. En yüksek olasılıktaki sınıf, veri için uygun sınıf olarak seçilmektedir.

Şekil 8'de Naive Bayes algoritmasının örnek yapısı gösterilmektedir. Bu yapıda her bir özellik düğümü, sınıf düğümü hariç hiçbir üst düğüme sahip değildir. Şekilde C özellik düğümünü, A ise sınıf düğümü değerlerini göstermektedir.

Şekil 8: Naive Bayes Yapısı



Kaynak: H. Zhang, t.y.den yazar tarafından derlenmiştir.

2.5.2.3. En Yakın Komşu

kNN algoritması, benzer özelliklere sahip verilerin aynı sınıfa ait olma eğiliminde olduğu prensibine dayanır ve sorgunun k en yakın komşunun ana sınıfıyla eşleştirilerek sınıflandırma yapmasını sağlamaktadır (S. Zhang vd., 2018). Veri kaydının sınıflandırılması için k en yakın komşularının çoğunluğuna dayalı olarak sınıflandırma yapılır ancak bu değerin seçimi sınıflandırmanın başarısını önemli ölçüde etkilemektedir (Guo vd., 2004).

kNN algoritması, sınıflandırma ve regresyon için parametrik olmayan bir yöntemdir (Pandey & Jain, 2017). Dolayısıyla, bu algoritma veri setinin yapısına göre

değişiklik göstermektedir ve verilerin özelliklerine göre uyum sağlayabilmektedir. Bu nedenle kNN algoritmasının avantajları ve dezavantajları bulunmaktadır.

Avantajları:

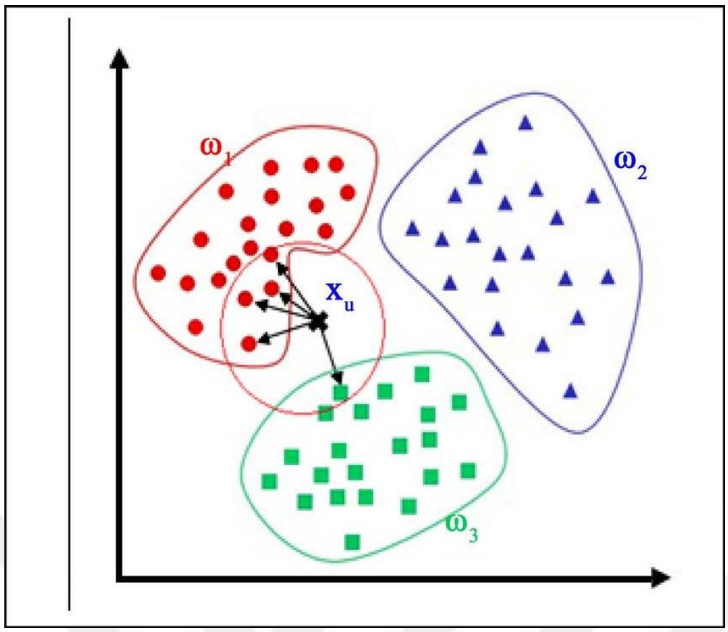
- Basit, anlaşılır, kolay uygulanabilirdir. Parametre tahmini yapmak veya model eğitimi yapmak gerekmez.
- Nadir olayların sınıflandırması için uygundur.
- Özellikle çok sınıflı sınıflandırma problemleri için uygundur.

Dezavantajları:

- Tembel bir algoritma olduğunda, test örneğinin sınıflandırılması için hesaplama miktarı büyük, bellek gereksinimi yüksek ve işlem süresi yavaş olabilmektedir.
- Örnek dengesi sağlanmadığında, bir sınıfın örnek sayısı çok büyükken, diğer sınıfların örnek sayısı çok az olduğunda, yeni bir örneğin girdisi, k komşularındaki örneklerin çoğunluğu büyük kapasiteli sınıfın örneklerinden oluşabilmektedir.
- Yorumlanması düşük olabildiği gibi karar ağacı gibi kurallar da oluşturamamaktadır (Sun vd., 2018, ss. 2-3).

Şekil 9'da, k en yakın komşu sınıflandırıcısı bu veri kümesini üç farklı renk ile temsil eden gruplara ayırmaktadır. Oklar X'in en yakın komşularını temsil etmektedir. Bu durumda, x k değeri kümesinde (şekilde k = 5) en yakın komşuya sahip olduğu için kırmızı olarak sınıflandırılmaktadır.

Şekil 9: kNN Algoritması Komşu Modeli Örneği



Kaynak: W. Zhang, 2017'den yazar tarafından derlenmiştir.

2.5.2.4. Destek Vektör Makineleri

SVM, sınıflandırma ve regresyon analizi için kullanılan verileri analiz ederek, doğrusal olmayan sınıflandırmayı mümkün kılmaktadır ve sınıflar arasındaki maksimum marj mesafesi ile sınıflandırma hatasını en aza indiren bir denetimli öğrenme modeli olarak tanımlanmaktadır (Mahesh, 2018). Destek vektör makineleri, makine öğrenmesi alanındaki başarılı uygulamalar sonrasında uzaktan algılama alanında kullanılan bir sınıflandırma yöntemidir ve veriyi yüksek boyuta dönüştürerek oluşturulan bir hiperdüzlem ile sınıfları birbirinden ayırmayı amaçlar; bu aşamada, kernel fonksiyonları olarak adlandırılan fonksiyonlar kullanılmaktadır ve matematiksel ifadesindeki parametreler kullanıcı tarafından belirlenmektedir (Kavzoğlu & Çölkesen, 2010).

SVM, hiperdüzlemi hesaplamak için aşağıda formülü kullanmaktadır.

$$w.x + b = 0$$

Formülde:

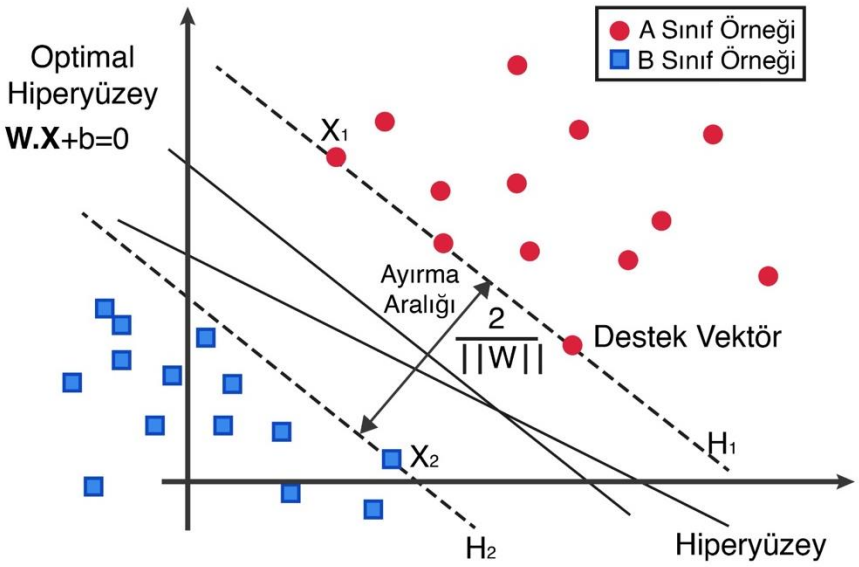
w: özelliklerin ağırlığını içeren vektör

x: özellik vektörü

b: hiperdüzlem sapması değerini göstermektedir.

Şekil 11'de, A sınıfından ve B sınıfından oluşan iki farklı sınıf bulunmaktadır. H1 ve H2 doğruları oluşturularak bulunan ayırma aralığı değeri belirlenmiştir. Bu doğrular üzerinde gösterilen veri noktaları ise destek vektörleri olarak belirtilmektedir.

Şekil 10: Destek Vektör Makinesi ile Veri Gösterimi



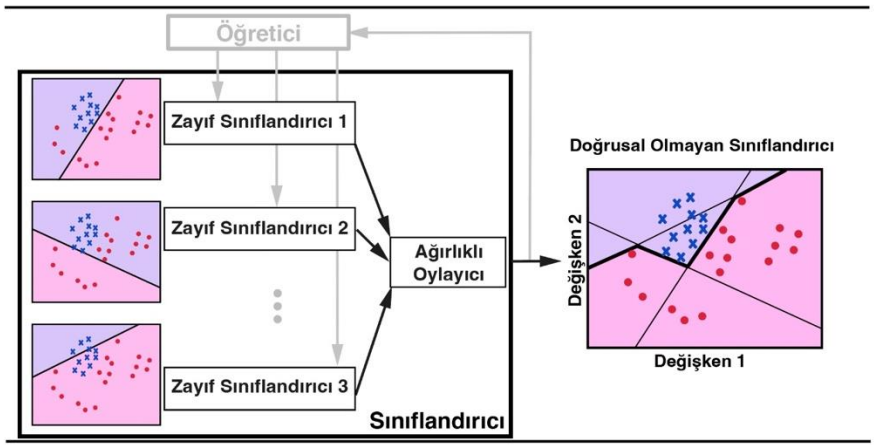
Kaynak: García-Gonzalo vd., 2016'dan yazar tarafından derlenmiştir.

2.5.2.5. AdaBoost

AdaBoost algoritması, öğrenme sistemlerinin tahmin yeteneğini geliştirmek için etkili bir araç olarak kullanılan ve öğrenme sürecini koordine etmek için en tipik yöntemlerden biridir. Adaboost algoritması, öğrenenin önceden bilgiye ihtiyaç duymadan değişkenlerin ağırlıklarını ayarlama yöntemiyle çalışmaktadır. Böylelikle zayıf sınıflandırıcıları güçlü bir sınıflandırıcıya dönüştürme hedefi gerçekleştirilmektedir. Kullanılan yöntem, öğrenme sürecinin daha verimli ve etkili olmasını sağlamaktadır ve birçok uygulamada başarılı sonuçlar elde etmektedir (Chengsheng vd., 2017, s. 1). Algoritmanın çalışma mantığı Şekil 11'de gösterilmektedir. Burada, zayıf sınıflandırıcılar veri örneklerinin yanlış sınıflandırıldığı bir iterasyonda eğitilmekte ve sonraki iterasyonun eğitimi sırasında bu örneklerin ağırlıkları artırılmaktadır. Zayıf sınıflandırıcı kararları elde edildikten sonra, bu kararlar ağırlıklı bir oylayıcıya aktarılmakta ve nihai sınıflandırma kararı

elde edilmektedir. Oylayıcı ağırlıkları da AdaBoost eğitimi sırasında belirlenmekte ve oylayıcı donanımı, ilgili zayıf sınıflandırıcının kararına bağlı olarak bir ağırlık ekleyen veya çıkaran bir imzalı toplayıcıdan oluşmaktadır. Şekil 11'de gösterildiği gibi, lineer bir model üzerinde zayıf sınıflandırıcılar durumunda, AdaBoost algoritması birden çok zayıf sınıflandırıcıdan oluşan bir karar sınırı oluşturarak güçlü bir sınıflandırıcı seviyesinde uyum sağlamaktadır (Z. Wang vd., 2015, s. 3).

Şekil 11: AdaBoost Algoritmasının Gösterimi



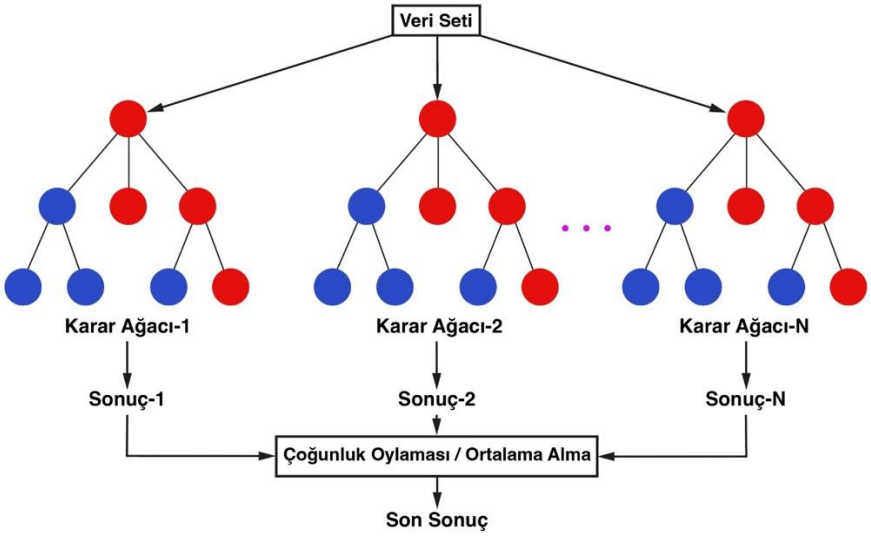
Kaynak: Z. Wang vd., 2015'ten yazar tarafından derlenmiştir.

2.5.2.6. Random Forest

Rastgele ormanlar, her bir ağacın rastgele bir vektörün değerlerine bağlı olduğu ve tüm ormandaki ağaçlar için bağımsız olarak ve aynı dağılıma sahip bir şekilde örneklenen ağaç tahmincilerinin bir kombinasyonudur; bu yöntem, bireysel ağaçların gücüne ve aralarındaki korelasyona dayalı olarak genelleme hatasını azaltırken, rastgele özellik seçimiyle daha iyi performans gösterir ve gürültüye karşı daha dayanıklıdır (Cutler vd., 2012). Özellikle, orijinal eğitim veri seti, örnekleme aracılığıyla yerine koymalı olarak farklı bagged örnekler oluşturmak amacıyla öncelikle örneklenmektedir. Ardından, her bir örnekleme için yüksek varyansa ve düşük yanlılık hatasına sahip bir derin ağaç eğitilmektedir. Bagged örneklerin birbirinden farklı olması sebebiyle her bir ağacın tahmini farklı bir sonuç değeri üretmektedir. Son olarak, tüm karar ağaçlarının çoğunluk oylaması kuralı kullanılarak

final tahmini oluşturulmaktadır (Sapountzoglou vd., 2020, s. 1). Şekil 12'de, bir rastgele orman algoritmasının bir örneğini göstermektedir.

Şekil 12: Rastgele Ormanın Bir Örneği

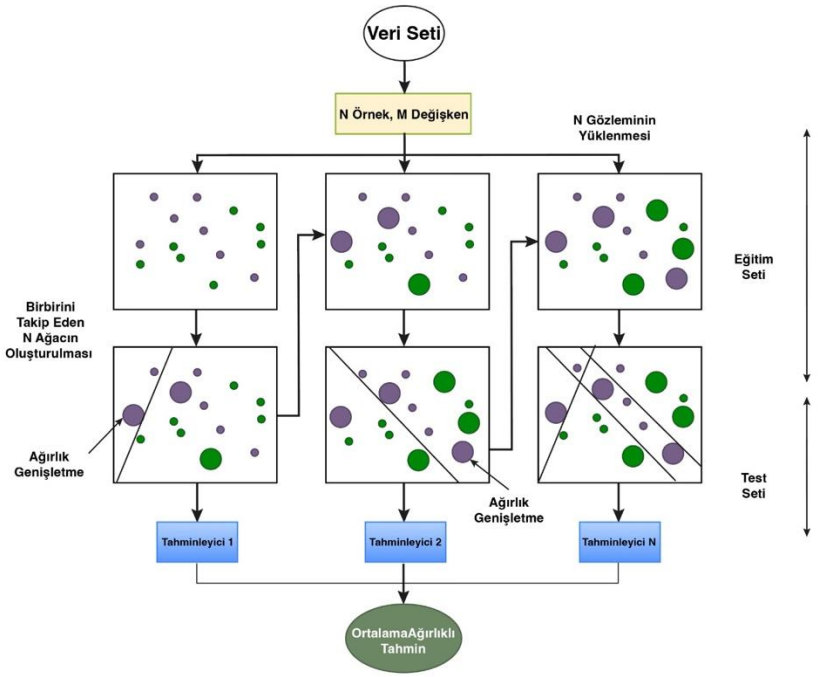


Kaynak: Khan vd., 2021'den yazar tarafından derlenmiştir.

2.5.2.7. CatBoost

CatBoost, gradyan artırma yöntemini kullanarak bir dizi zayıf tahminciyi birleştirerek güçlü bir tahminci oluştururken, özel bir düğüm bölünme stratejisi kullanarak kategorik özellikleri doğrudan kullanabilen ve tahmincilerin ağaçların derinliklerini ve karmaşıklıklarını otomatik olarak ayarlayarak, öğrenme planlaması sağlayan bir algoritma olarak tanımlanmaktadır (Prokhorenkova vd., 2018). CatBoost, ağaçtaki ilk bölünme dışında, sonraki bölünmeler için mevcut ağaçtaki kategorik özellikleri ve kombinasyonları veri setindeki tüm kategorik özelliklerle birleştirerek kullanır; tüm kombinasyon değerlerini hızlı bir şekilde sayılara dönüştürebilir ve seçilen ağaç bölünmelerini kategorik bir özellik olarak ele alarak sayısal ve kategorik özellik kombinasyonları oluşturabilir (Alshari vd., 2021). Şekil 12, CatBoost algoritmasının açıklamasını göstermektedir.

Şekil 13: CatBoost algoritmasının açıklaması

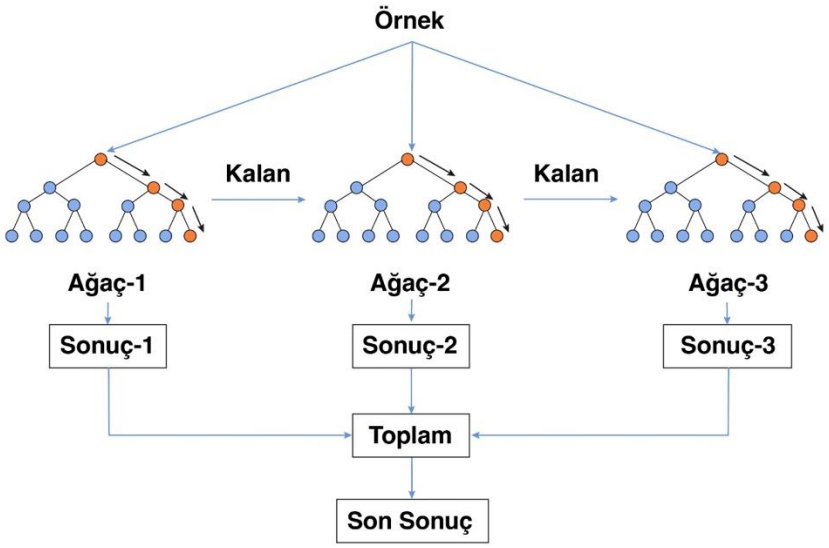


Kaynak: Shahani vd., 2022'den yazar tarafından derlenmiştir.

2.5.2.8. XGBoost

XGBoost modeli, ikinci dereceden optimizasyon kullanımı ve k regresyon ağacı inşası ile her bir yaprak düğümünün en iyi tahmin değerini oluşturmaktadır. En iyi parametrelerin seçimi için çapraz doğrulama kullanılmaktadır ve aşırı uyumun önlenmesi için düzenlileştirme yapılmaktadır. Yüksek verimlilik, iyi sonuçlar, büyük ölçekli verilerin işlenmesi yeteneği ve özel kayıp fonksiyonlarının desteklenmesi gibi avantajları bulunmaktadır. Ayrıca XGBoost'un amaç fonksiyonu ve kayıp fonksiyonu bulunmaktadır. Bu fonksiyonlar modelin veriye uyumunu ve tahmin yeteneğini belirlerken, düzenlileştirme terimi modelin karmaşıklığını kontrol etmektedir. Böylelikle daha etkili bir model elde edilmektedir (Y. Wang vd., 2023, s. 72). Şekil 14'te XGBoost'un temel yapısı gösterilmektedir. Burada, ağaç-1'in artık değeri, artığı azaltmak için ağaç-2'ye iletilmektedir ve bu son sürece kadar devam etmektedir.

Şekil 14: XGBoost'un Temel Yapısı

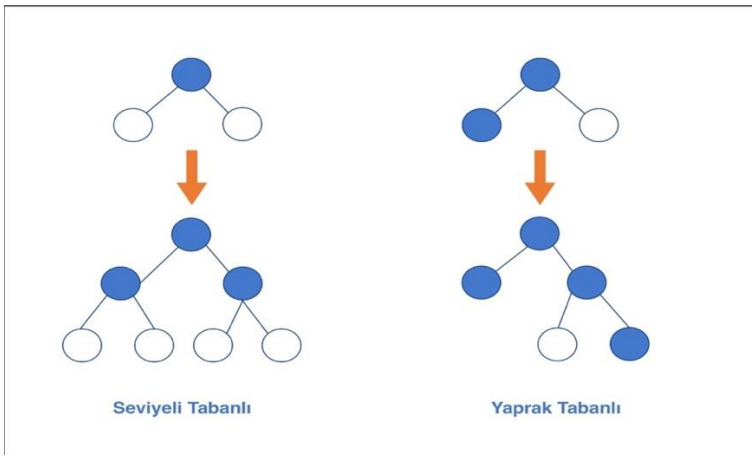


Kaynak:W. Wang vd., 2020'den yazar tarafından derlenmiştir.

2.5.2.9. LightGBM

Karar ağaçlar 2 farklı yöntemle büyüme gerçekleştirmektedir. Bunlardan biri seviye temelli stratejiyle ağacın dengeli bir şekilde büyütülmesiyken bir diğeri yaprak odaklı strateji ile kaybı en aza indiren yapraklardan başlayarak bölünmelerin dikine gerçekleştirilmesidir. Şekil 15'te karar ağaçlarının büyüme yöntemleri gösterilmektedir.

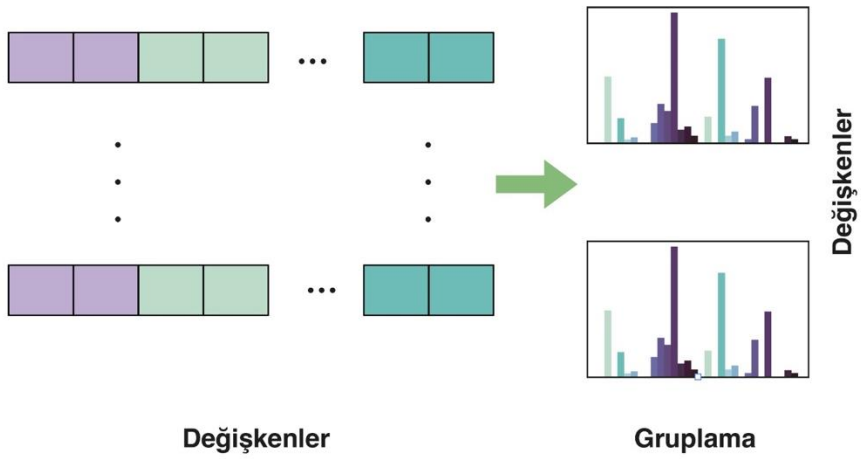
Şekil 15: Karar Ağaçları Büyüme Yöntemleri



Kaynak: Yin vd., 2021'den yazar tarafından derlenmiştir.

LightGBM, histogram kullanarak sürekli değerlere sahip değişkenleri ayrık değerlere dönüştürerek, yaprak odaklı seviye büyütme yaklaşımı ile, değişkenleri kategori gruplarına ayırmaktadır. Her bir kategori için de histogram oluşturmaktadır. Histogramlar, değişkenlerin daha hızlı ve verimli şekilde işlenmesini sağlamaktadır. Ancak, veri sayısı düşük olduğunda bu işlem aşırı öğrenmeye neden olmaktadır. Dolayısıyla, LightGBM, büyük veri setlerini daha hızlı ve verimli bir şekilde işlemektedir (Farzinpour vd., 2023, s. 5). Şekil 16'da algoritmanın kullandığı histogram yapısı gösterilmektedir. Şekil X'te histogram yöntemi aracılığıyla sürekli değişkenlerin aralıklara ayrılarak kategoriler halinde dönüştürülmesi gösterilmektedir.

Şekil 16: LightGBM Algoritmasında Histogram Yöntemi



Kaynak:K. Wang vd., 2023'ten yazar tarafından derlenmiştir.

2.5.3. Denetimsiz Öğrenme

Denetimsiz öğrenme, sınıflandırılmamış veya etiketlenmemiş veri noktaları içeren veri kümelerinde desenlerin tanınması için harici rehberlik olmadan veri noktalarını sınıflandıran, etiketleyen ve gruplandıran algoritmaları ifade etmektedir(Dridi, 2021). Bu stratejide, sisteme sadece girdi değerleri sunulmakta ve öğrenme işlemi sadece bu girdiler üzerinde gerçekleştirilmektedir (Pulat & Deveci Kocakoç, 2021). Diğer bir deyişle sınıflandırma veya gruplama yapmak için önceden belirlenmiş bir etiket veya kategori bulunmamaktadır.

2.5.4. Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme, doğal hayattaki hayvan davranışlarından ilham almaktadır. Organizmalar, eylemleri gerçekleştirerek çevreleriyle etkileşime girmekte, mevcut durumlarından bir sonraki duruma geçerek ödül şeklinde geri bildirim almaktadır. Organizmalar, ödüllerin kalitesini değerlendirmekte ve olumlu ödüller benzer durumlarda aynı eyleme yol açarken, negatif ödüller alternatif eylemleri keşfetme konusunda teşvik etmektedir. Çevreleriyle etkileşim aracılığıyla daha iyi politikalar öğrenme sebebiyle pekiştirmeli öğrenme, doğal öğrenme davranışı sergileyerek teorik temel yapısını oluşturmaktadır (Tong vd., 2023, s. 3).

Denetimsiz ve Pekiştirmeli öğrenme tez kapsamındaki çalışmalara dahil olmadığı için her iki öğrenme yönteminin de detaylarına yer verilmemiştir.

2.5.5. Makine Öğrenmesinde Model Doğrulama Yöntemleri

Değerlendirme yöntemleri, bir modelin genelleme yeteneğini tahmin etmek için kullanılmaktadır. Bu yetenek, modelin eğitim sırasında kullanılan verilerle karşılaştırılabilir sonuçlar üretme becerisini ifade etmektedir. Eğer mevcut tüm veriler modelin eğitimi için kullanılırsa, modelin performansı yeni veriler üzerinde değerlendirilemez ve dolayısıyla modelin genelleme yeteneği belirsiz kalmaktadır. Bu nedenle, değerlendirme yöntemleri, modelin performansını yeni veriler üzerinde değerlendirmek için ayrı bir veri kümesi kullanmaktadır. Dolayısıyla veriler belli oranda bölünerek veri kümelerine ayrılmaktadır. Değerlendirme için kullanılan veri kümesine "test seti" denirken, modelin eğitiminde kullanılan veri kümesine ise "eğitim seti" denmektedir. Veri bölümlendirme için de ayırma (hold out) ve k katmanlı çarpraz doğrulama (k-fold cross validation) yöntemi bulunmaktadır (Chlis, 2013, s. 26).

Şekil 17'de k katmanlı çapraz doğrulama ve ayırma yöntemleri bulunmaktadır. Ayırma yöntemi, veri setini eğitim ve test seti olarak iki parçaya ayırarak modelin performansını değerlendirmektedir. Ayırma yöntemi aracılığıyla veri setinin belirli bir oranda bölünmesiyle eğitim seti üzerinde model eğitimi gerçekleştirilmekte ve test seti üzerinde modelin performansı değerlendirilmektedir. K katmanlı çapraz doğrulama yönteminde ise veri setini k parçaya bölerek daha kapsamlı bir doğrulama sağlar. Şekil 17'de k değeri 5 olarak alınarak her bir bölünen parça sırayla test seti olarak

kullanılırken, diğer parçalar eğitim seti olarak kullanılır ve modelin performansı her bir parça üzerinde değerlendirilmektedir.

Şekil 17: K Katmanlı Çapraz Doğrulama ve Ayırma Tekniği Modeli



Kaynak: Dierking vd., 2023'ten yazar tarafından derlenmiştir.

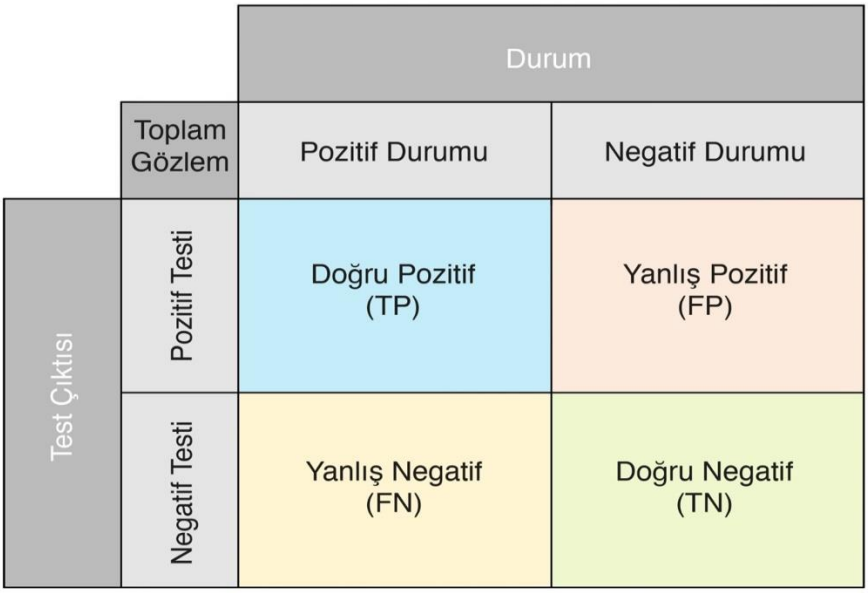
2.5.6. Makine Öğrenmesinde Model Performans Ölçüm Yöntemleri

Algoritmaların sonuç performanslarını değerlendirmek için çeşitli yöntemler geliştirilmiştir. Makine öğrenimi ve özellikle istatistiksel sınıflandırma problemlerinde, algoritmanın performansını görselleştirmek için kullanılan bir tablo düzeni olan hata matrisi veya karmaşıklık matrisi olarak da bilinen bir matris bulunur. Bu matriste aşağıdaki ifadeler kullanılmaktadır:

- Doğru Pozitif (TP – True Positive): Gerçek pozitifleri doğru bir şekilde tahmin eden modelin sayısını vermektedir.
- Yanlış Pozitif (FP – False Positive): Gerçek negatif değerini yanlış bir şekilde pozitif olarak tahmin eden modelin sayısını vermektedir.
- Yanlış Negatif (FN – False Negative): Gerçek pozitif değerini yanlış bir şekilde negatif olarak tahmin eden modelin sayısını vermektedir.
- Doğru Negatif (TN – True Negative): Gerçek negatifleri doğru bir şekilde tahmin eden modelin sayısını vermektedir (Ali vd., 2019, s. 3).

Şekil 18'de bir karmaşıklık matrisi ve ondan hesaplanabilen çeşitli yaygın metriklerin denklemlerini göstermektedir.

Şekil 18: Karmaşıklık Matrisi



Kaynak: Y. Liu & Yang, 2022'den yazar tarafından derlenmiştir.

Karmaşıklık matrisi kullanılarak çeşitli algoritma performans değerlendirme formülü çıkarılmaktadır. Bunlar:

o Doğruluk (Accuracy): Karmaşıklık matrisinden doğru tahmin edilen örneklerin toplam örnek sayısına oranıyla hesaplanmaktadır.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

o Hassasiyet (Precision): Karmaşıklık matrisinden doğru pozitif tahmin edilen örneklerin toplam pozitif tahmin edilen örnekler sayısına oranıyla hesaplanmaktadır.

$$Precision = \frac{TP}{(TP + FP)}$$

o Duyarlılık (Recall): Karmaşıklık matrisinden doğru pozitif tahmin edilen örneklerin toplam gerçek pozitif örnekler sayısına oranıyla hesaplanmaktadır.

$$Recall = \frac{TP}{(TP + FN)}$$

o F1 Puanı (F1 Score): Hassasiyet ve duyarlılık değerlerinin harmonik ortalaması alınarak hesaplanmaktadır (Figueroa & Pope, 2017, s. 13).

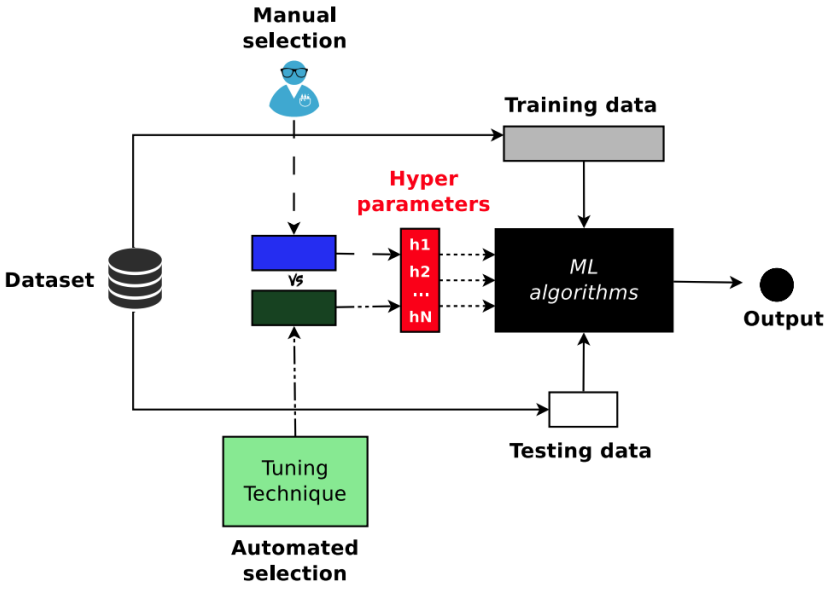
$$F1\ Score = \frac{2 * (Precision * Recall)}{Precision + Recall}$$

2.5.7. Makine Öğrenmesinde Hiperparametre Ayarlaması

Kurulan modellerin ve kullanılan algoritmaların performansını karşılaştırmak için genellikle doğruluk değeri kullanılmaktadır, ancak doğruluk tek başına yeterli bir performans ölçütü değildir. Bunun sebebi veri setindeki sınıflar arasındaki dengesizlik etkisini göz ardı edebilir olmasından kaynaklanmaktadır. Dolayısıyla, sadece doğruluk değerlerine bakmak dışında hassasiyet, duyarlılık ve F1 puanı değerlerine bakılarak model performansı değerlendirilmektedir (Kartal & Ozen, 2017, s. 111).

Her makine öğrenimi sisteminin hiperparametreleri vardır ve otomatik makine öğrenimi (AutoML – Automated Machine Learning) alanındaki en temel görev, bu hiperparametreleri otomatik olarak ayarlayarak performansı optimize etmektir (Feurer & Hutter, 2019).

Şekil 19: Hiperparametre Ayarlama Şeması



Kaynak: Gomes Mantovani, 2018'den yazar tarafından derlenmiştir.

Denetimli sınıflandırma problemlerinde, makine öğrenmesi algoritmaları yeni bir gözlemin hangi etikete ait olduğunu belirlemek için etiketli verilerle eğitilmektedir. Bu algoritmaların tahmin edici performansını doğrudan etkileyen hiperparametreler bulunmaktadır. Uygun bir hiperparametre ayarı bulmak belirli bir bilgiye, sezgiye ve çoğunlukla deneme yanılma deneylerine dayanmaktadır. Aynı zamanda otomatik

olarak yapılan hiperparametre çözümleri bulunmaktadır (Gomes Mantovani, 2018, s. 31).

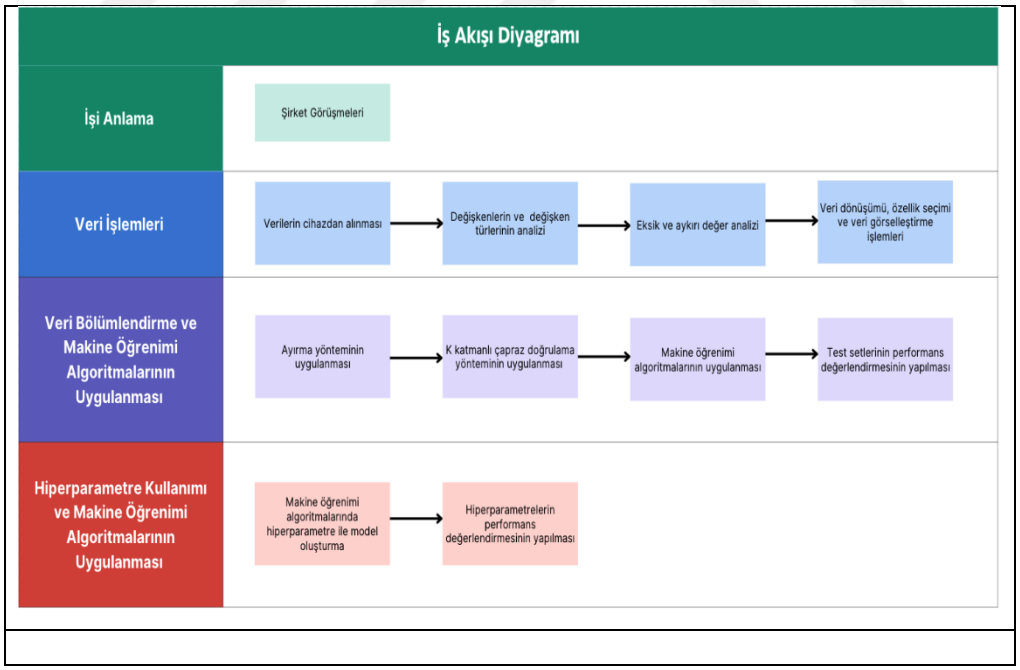
Şekil 19, genel hiperparametre ayarlama şemasını göstermektedir.

ÜÇÜNCÜ BÖLÜM
YÖNTEM

Çalışmada, Ege Fren Sanayi ve Ticaret Anonim Şirketi'nin CNC makineleri tarafından üretilen fren kaldırma bileşenlerinin üretiminde kullanılan makinelerin bakım durumlarının tahmin edilmesi amacıyla, makine öğrenmesi algoritmaları kullanılarak bir kestirimci bakım uygulaması gerçekleştirilmiştir. Bu uygulama, makinenin arıza ihtiyacını tespit etmek ve bakım faaliyetlerini optimize etmek amacıyla veri odaklı ve analitik yaklaşımları benimsemektedir. Çalışmanın odak noktası, CNC makinelerinin bakım durumlarını tahmin etmek için geliştirilen bir kestirimci modelin doğruluk değerini artırmaktır. Bu doğruluk değeri, modelin ekipman arızalarını tespit etme yeteneğini ve doğru tahminler yapma kabiliyetini ifade etmektedir.

Çalışmada Şekil 20'de belirtilen iş akışı gerçekleştirilmiştir.

Şekil 20: Uygulama İş Akışı



Kaynak: Yazar tarafından derlenmiştir.

Süreç içindeki çalışmaların gerçekleştirilmesinde CRISP-DM metodolojisinden yararlanılmıştır.

3.1. İŞİ ANLAMA

Uygulamada Mazak marka CNC makinesinin verileri kullanılmıştır. Mazak CNC makinesi, fren kaldırma parçalarının yüksek hassasiyetle üretimini sağlamak için kullanılan otomatik kontrol sistemidir. Mazak CNC makinesi, bir bilgisayar kontrollü işleme birimi, hassas hareket sistemleri ve kesme araçlarından oluşmaktadır. İşleme birimi, iş parçasının üretiminde kullanılacak olan kesme yolunu ve hareketleri belirlemek için programlanmaktadır. Şekil 21'de Mazak CNC makinesinin üretim tesisinden alınan dış görüntüsü bulunmaktadır.

Şekil 21: Mazak CNC Makinesi Dış Görünümü



Kaynak: Yazar tarafından derlenmiştir.

Mazak CNC makinesinin çalışma mantığı aşağıdaki adımları içermektedir.

1. Tasarım ve Programlama: İlk adımda, üretilecek parçanın tasarımı ve üretim süreci programlaması yapılmaktadır. Bilgisayar destekli tasarım (CAD – Computer Aided Design) yazılımları kullanılarak, parça tasarımı oluşturulmakta ve üretim sürecine yönlendirecek olan G-kodu ile işlem programı oluşturulmaktadır.

2. Hammaddenin Yerleştirilmesi: Mazak CNC makinesinin çalışma alanına, GG25 döküm parçası takılmakta ve sabitlenmektedir. Bu döküm parçası, kesme işlemlerinin gerçekleştirileceği kesici araçlara doğru konumlandırılmaktadır.

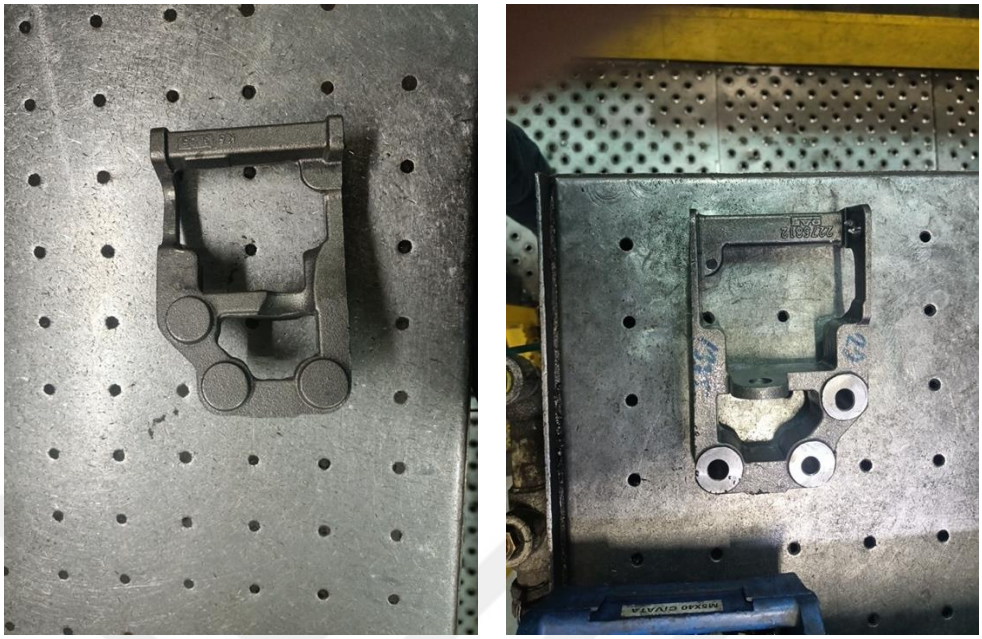
3. Kesme Araçlarının Seçimi: Hammaddenin işlenmesi için uygun kesici araçlar seçilmektedir. Farklı kesici uçlar ve kesme stratejileri, hammaddenin şekline, malzemesine ve işleme gereksinimlerine bağlı olarak belirlenmektedir.

4. Hareket Kontrolü: Mazak CNC makinesi, X, Y ve Z eksenlerindeki lineer hareketler ile kesme işleminin istenen konumda gerçekleştirilmesini sağlamaktadır. Ayrıca, makineye entegre edilmiş döner eksende dönme hareketleri de kullanılabilmektedir.

5. Kesme İşlemi: GG25 döküm parçasının üzerinde belirtilen kesme yolları takip edilir ve kesme araçları doğru kesme hızı, kesme derinliği ve besleme hızıyla parçayı işlemektedir. Mazak CNC makinesi, çeşitli kesme operasyonları gerçekleştirebilir, bunlar arasında delme, frezeleme, tornalama, delik açma, yüzey işleme gibi işlemler bulunmaktadır. Şekil 22'de GG25 döküm parçasının Mazak CNC kesme işlemi öncesi ve sonrası bulunmaktadır.

6. İşlem Sonrası Kontrol: Kesme işlemi tamamlandıktan sonra, iş parçasının kalitesini kontrol etmek için denetimler yapılmaktadır. Ölçüm cihazları kullanılarak parçanın boyutları, toleransları ve yüzey kalitesi kontrol edilmektedir. Kontrol işlemi tamamlandıktan sonra müşteri sevkiyatı yapılmaktadır.

Şekil 22: GG25 Döküm Parçası İşlem Öncesi ve Sonrası



Kaynak: Yazar tarafından derlenmiştir.

Şirkette bulunan Mazak CNC makinesinin içerisinde servo motor ve M51 motor bulunmaktadır. Aynı zamanda yağ sıcaklık ve seviye, soğutucu sıcaklık ve seviye, hidrolik yağ sıcaklık ve seviye, elektrik pano sıcaklık, motor sıcaklık ve trafo sıcaklığı ölçümü yapan sensörler bulunmaktadır. İlgili sensör verileri ve işletim sistemi tarafından hesaplamaları yapılan X, Y, Z konumları ve motor devir sayıları gibi veriler saniye bazında veri tabanına kaydedilmektedir.

3.2. VERİLERİN TOPLANMASI

Cihazdaki verileri kaydetmek amacıyla üretim teknoloji bağlantısı (MTConnect – Manufacturing Technology Connect) adı verilen açık bir iletişim protokolü kullanılmaktadır. İlgili protokol, farklı makine türleri ve üretim ekipmanları arasındaki veri paylaşımını standartlaştırmaktadır. Veriler genişletilebilir işaretleme dili (XML – eXtensible Markup Language) formatındadır.

MTConnect veri toplama sürecinde farklı roller üstlenen bileşenler bulunmaktadır. Bu bileşenler:

1. Mazatrol: Mazak CNC makinelerinin kontrol sistemidir. MAZATROL, Mazak'ın özel olarak geliştirdiği bir kontrol yazılımıdır ve CNC makinesinin iş

parçasını işlemek ve hareket etmek için kullanılan araçtır. MAZATROL, işleme programlarının oluşturulması, makinenin hareket kontrolünü sağlama ve diğer CNC işlevlerini gerçekleştirme yeteneğine sahiptir.

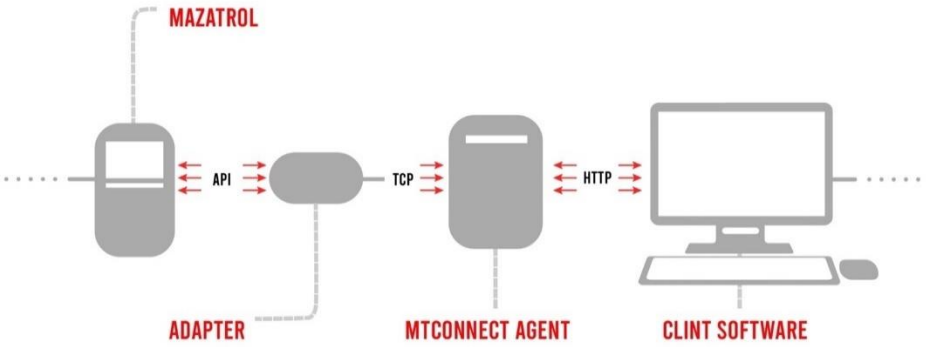
2. Adapter: Mazak CNC makinesinin MAZATROL kontrol sistemini MTConnect uyumlu bir veri toplama sistemine bağlamak için kullanılan bir ara bileşendir.

3. MTConnect Agent: CNC makinesi veya diğer üretim ekipmanlarının MTConnect protokolünü desteklemesini sağlayan bir yazılım bileşenidir. MTConnect Agent, CNC makinesinden gelen verileri alır, işler ve MTConnect uyumlu bir veri formatında sunar. Bu bileşen, veri toplama ve iletim sürecini sağlar ve verilerin diğer sistemlere aktarılmasını kolaylaştırır.

4. Client Software: MTConnect verilerini alan, işleyen ve kullanıcılar tarafından görüntülenebilen bir yazılım uygulamasıdır. Bu uygulama, operatörlerin ve yöneticilerin CNC makinesi verilerine erişmesini ve gerçek zamanlı izlemesini sağlayan bir bileşendir.

Mazatrol tarafından gerçekleştirilen işlemler uygulama programlama arayüzü (API – Application Programming Interface) aracılığıyla adaptöre iletilir. Aktarım kontrol protokolü (TCP – Transmission Control Protocol) aracılığıyla adaptörde toplanan veriler MTConnect Agent'ta toplanır. Burada veriler MTConnect ile uyumlu hale getirilir. Uyumlu veriler hipermetin transfer protokolü (HTTP – Hypertext Transfer Protocol) ile kullanıcıların görebileceği bir bilgisayar ortamına aktarılır. MTConnect veri transfer işlemi Şekil 23'te gösterilmektedir.

Şekil 23: MTConnect Veri İşlem Süreci



Kaynak: Yazar tarafından derlenmiştir.

3.3. VERİ SETİ

Mazak CNC makineleri üzerinden 2 farklı alandan veri kaydı gerçekleşmektedir. Bu yöntemlerden biri işletim sistemi üzerinden otomatik gelen verilerden oluşmaktadır. Diğer veri kayıt yöntemi ise cihaza bağlanan sensörlerden oluşmaktadır. Tablo 3'te işletim sistemine bağlı değişkenler ve açıklamaları ifade edilmektedir.

Tablo 3: İşletim Sistemine Bağlı Değişkenler

Değişken Adı	**Açıklama**
TarihSaat	Verinin kaydedilme zamanıdır.
X Eksen Yüklenmesi	Makinenin yatay eksen üzerindeki mekanik yükün ölçüm değeridir.
Y Eksen Yüklenmesi	Makinenin dikey eksen üzerindeki mekanik yükün ölçüm değeridir.
Z Eksen Yüklenmesi	Makinenin ileri-geri eksen üzerindeki mekanik yükün ölçüm değeridir.
Spindle Yüklenmesi	Makinenin döner milinin üzerindeki mekanik yük ölçüm değeridir.

Spindle Devri	Makinenin döner milinin üzerindeki dönüş hızını ifade etmektedir.
Spindle Sıcaklığı	Makinenin döner milinin üzerindeki sıcaklık değerini ifade etmektedir.
Spindle Sıcaklığı Durumu	Makinenin döner milinin üzerindeki aşırı ısınma durumunu ifade etmektedir.
Servo Durumu	Servo motorunun mevcut durumunun işlevselliğini ifade etmektedir.
Spindle Durumu	Spindle motor bileşenin mevcut durumunun işlevselliğini ifade etmektedir.
Emergency Durumu	Makine üzerindeki acil durum stop tuşuna basılma değerini ifade etmektedir.
Sistem Durumu	Anlık verilerle sistemin çalışma durumunu belirtmektedir.

Kaynak: Yazar tarafından derlenmiştir.

Tablo 4'te sensörler tarafından kaydedilen değişkenler ve açıklamaları bulunmaktadır. İlgili sensör parçaları seviye ölçme ve sıcaklık sensörlerinden oluşmaktadır. Bu sensörler cihaza şirket tarafından yerleştirilmiştir.

Tablo 4: Sensörlere Bağlı Değişkenler

Değişken Adı	**Açıklama**
Anamil Soğutucu Yağ Seviyesi	Motorun çalışmasını sağlayan ana milin soğutma sistemine bağlı yağ seviyesini ifade etmektedir.
Anamil Soğutucu Yağ Sıcaklığı	Motorun çalışmasını sağlayan ana milin soğutma sistemine bağlı yağ sıcaklığını ifade etmektedir.
Aparat Hidrolik Yağ Seviyesi	Makine aparatlarının hidrolik çalışma sisteminde kullanılan yağ seviyesini ifade etmektedir.
Aparat Hidrolik Yağ Sıcaklığı	Makine aparatlarının hidrolik çalışma sisteminde kullanılan yağ sıcaklığını ifade etmektedir.
Elektrik Pano Sıcaklığı	Makineye bağlı trafo birimin iç sıcaklığını ifade etmektedir.

Hidrolik Yağ Seviyesi	Makinen genel hidrolik sistemindeki yağ seviyesini ifade etmektedir.
Hidrolik Yağ Sıcaklığı	Makinen genel hidrolik sistemindeki yağ sıcaklığını ifade etmektedir.
M51 Motor Sıcaklığı	Makine motorunun sıcaklığını ifade etmektedir.
Tezgah Trafo Sıcaklığı	Tezgaha bağlı trafonun sıcaklığını ifade etmektedir.

Kaynak: Yazar tarafından derlenmiştir.

MTConnect aracılığıyla toplanan veriler virgülle ayrılmış değerler (CSV – Comma-Seperated Values) formatında kullanılmaktadır. Oluşturulan CSV dokümanında toplamda 22 değişken ve 30.622 kayıt bulunmaktadır. İlgili kayıtlar 27.02.2023 tarihi ile 05.05.2023 arasındaki verilerden oluşmaktadır. Şekil 24'te görüldüğü gibi veri setinde kategorik ve numerik olmak üzere iki tür veri yapısı bulunmaktadır. Şekil 24'te aynı zamanda her bir veri setindeki dolu değer sayısı da gösterilmektedir.

Şekil 24: Veri Seti Değişken Türleri

```
 0   Gres Yağ Seviyesi              25317 non-null  float64
 1   TarihSaat                      30622 non-null  object
 2   X Eksen Yüklenmesi             21959 non-null  float64
 3   Y Eksen Yüklenmesi             21958 non-null  float64
 4   Z Eksen Yüklenmesi             21958 non-null  float64
 5   Spindle Yüklenmesi             21956 non-null  float64
 6   Spindle Devri                  21956 non-null  float64
 7   Spindle Sıcaklığı              21885 non-null  float64
 8   Spindle Sıcaklığı Durumu       21365 non-null  object
 9   Servo Durumu                   21365 non-null  object
 10  Spindle Durumu                 21365 non-null  object
 11  Emergency Durum                21357 non-null  object
 12  Sistem Durumu                  21956 non-null  object
 13  Anamil Soğutucu Yağ Seviyesi   23459 non-null  float64
 14  Anamil Soğutucu Yağ Sıcaklığı  23460 non-null  float64
 15  Aparat Hidrolik Yağ Seviyesi   25317 non-null  float64
 16  Aparat Hidrolik Yağ Sıcaklığı  25317 non-null  float64
 17  Elektrik Pano Sıcaklığı        23459 non-null  float64
 18  Hidrolik Yağ Seviyesi          25317 non-null  float64
 19  Hidrolik Yağ Sıcaklığı         25317 non-null  float64
 20  M51 Motor Sıcaklığı            25317 non-null  float64
 21  Tezgah Trafo Sıcaklığı         25317 non-null  float64
dtypes: float64(16), object(6)
```

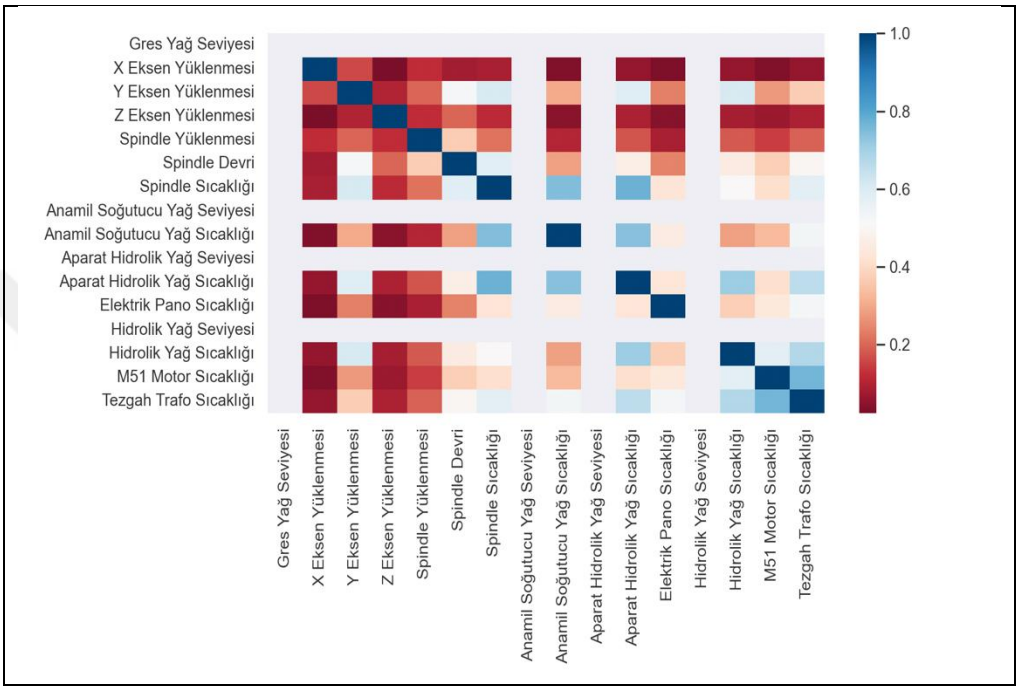
Kaynak: Yazar tarafından derlenmiştir.

Değişkenlerden sistem durumu hedef değişken olarak kullanılmaktadır. Diğer değişkenler ise bağımsız değişken olarak tanıtılmaktadır.

Değişkenler arasındaki ilişkiyi gözlemlemek için korelasyon analizi görseli oluşturulmuştur. Şekil 25'te koyu kırmızı negatif, koyu mavi ise pozitif korelasyonu göstermektedir. Renklerin koyulaşması korelasyon değerinin yüksek olduğunu göstermektedir. Bu da değişkenlerin birbirine benzer olduklarını göstermektedir.

Şekilde M51 Motor Sıcaklığı ile Tezgah Trafo Sıcaklığı arasında benzer bir ilişki olduğu koyu maviye yakın bir renkle gösterilerek korelasyon analizine örnek verilmektedir.

Şekil 25: Korelasyon Analizi



Kaynak: Yazar tarafından derlenmiştir.

3.4. VERİ ÖNİŞLEME

Makine öğrenmesi modellerinin daha sağlıklı ve güvenilir sonuçlar üretmesi için veri setinde bazı düzenlemeler yapılmıştır. İlgili ön işleme süreçleri, veri setindeki eksik, aykırı değerleri düzeltmek ve veriyi modele uygun hale getirmek için bir dizi adımlar uygulanmamıştır.

3.3.1. Eksik Verileri Analizi

Şekil 26'da değişkenlerdeki eksiklikler gösterilmektedir. Veriler internet aracılığıyla kaydedildiğinden dolayı internet kesintilerinde veri kaydı eksik değer olarak işlenmiştir. Bundan dolayı veri setindeki eksik değerlerin sayısı fazla

görünmektedir. En çok veri eksikliği ise Emergency Durum değişkeninde bulunmaktadır.

Şekil 26: Eksik Veri Kontrolü

Gres Yağ Seviyesi	5305
TarihSaat	0
X Eksen Yüklenmesi	8663
Y Eksen Yüklenmesi	8664
Z Eksen Yüklenmesi	8664
Spindle Yüklenmesi	8666
Spindle Devri	8666
Spindle Sıcaklığı	8737
Spindle Sıcaklığı Durumu	9257
Servo Durumu	9257
Spindle Durumu	9257
Emergency Durum	9265
Sistem Durumu	8666
Anamil Soğutucu Yağ Seviyesi	7163
Anamil Soğutucu Yağ Sıcaklığı	7162
Aparat Hidrolik Yağ Seviyesi	5305
Aparat Hidrolik Yağ Sıcaklığı	5305
Elektrik Pano Sıcaklığı	7163
Hidrolik Yağ Seviyesi	5305
Hidrolik Yağ Sıcaklığı	5305
M51 Motor Sıcaklığı	5305
Tezgah Trafo Sıcaklığı	5305

Kaynak: Yazar tarafından derlenmiştir.

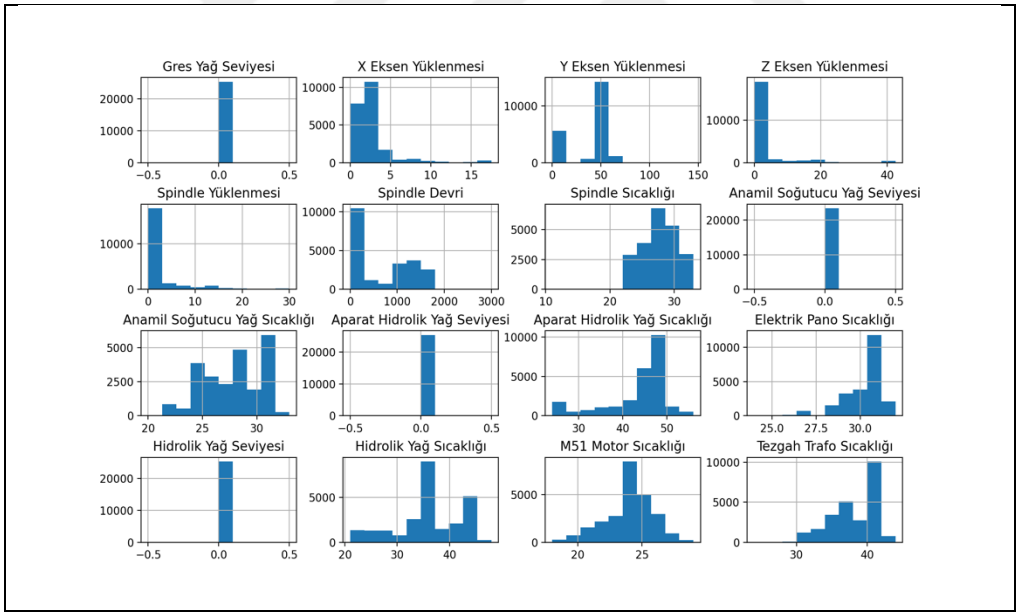
Eksik veri analizi sonucu tüm eksik değerlerin bulunduğu satırlar veri setinden silinmiştir. 30.622 kayıt silme işlemi sonrasında 19472 adete düşürülmüştür. Silme işlemi sonrası hiçbir sütunda eksik değer görülmemektedir.

Veri setinde bulunan Gres Yağ Seviyesi, Anamil Soğutucu Yağ Seviyesi, Aparat Hidrolik Yağ Seviyesi ve Hidrolik Yağ Seviyesi değerleri 0 (sıfır) değer dışında herhangi bir değer almamıştır. İlgili şirkete, ilgili bilgilendirme yapılmıştır.

3.3.2. Aykırı Değer Analizi

Aykırı değerler, genellikle diğer örneklerden büyük ölçüde farklı olan veri noktalarıdır ve modelin performansını olumsuz etkileyebilmektedir. Aykırı değer tespiti ve düzeltilmesi, makine öğrenmesi modelinin yanlış yönlendirilmesini önlemek için kullanılmaktadır. Aykırı değer işlemlerinden önce veri setindeki numerik değerlerin aykırı dağılımları Şekil 27'de gösterilmektedir.

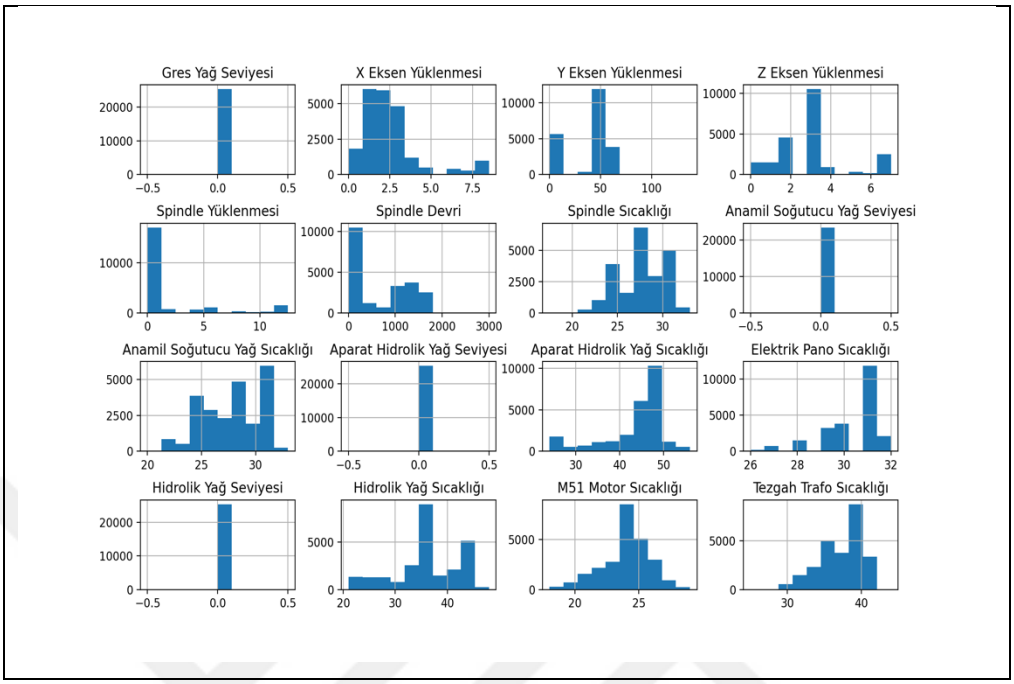
Şekil 27: Değişkenlerin Histogram Grafik Dağılımları



Kaynak: Yazar tarafından derlenmiştir.

Aykırı değer tespiti için her bir değişkenin veri setindeki dağılımlarının 0,05 ve 0,95 aralığı dışındaki değerlerler tespit edilmiş ve bu sınırın dışındaki değerler sınır değerlerine eşitlenmiştir. Bu durumda değişkenlerdeki tüm aykırı değerler ortadan kaldırılmıştır.

Şekil 28: Değişkenlerin Aykırı Değer İşlemleri Sonrası Dağılımları



Kaynak: Yazar tarafından derlenmiştir.

3.3.3. Özellik Mühendisliği

Sistem durumu, kullanılan veri setinde hedef değişkeni oluşturmaktadır. Bu hedef değişkenin sınıf frekansları ve her bir frekansa ait değerler de Şekil 29'da gösterilmektedir.

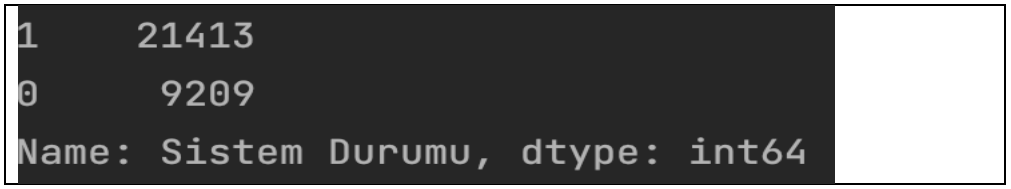
Şekil 29: Hedef Değişkenin Frekans Değerleri

Normal	21413
Warning:HAFIZA KORUMA	193
Warning:GECERSIZ BICIM	176
Warning:HAFIZA KORUMA (OTO OPERASYON)	54
Warning:NO MDI DATA	39
Warning:AYNI PROGRAM VAR	23
Fault:KESME BLOKU BASLANGICI KILITLI	12
Warning:KESME ILERLEMESI ASIRIYÜK '0'DA	9
Warning:GECERSIZ ADRES GIRISI	8
Warning:VERILER YENILENEMEZ	7
Warning:OTOMATIK HESAPLAMA IMKANSIZ	6
Fault:ASIRI YUK	5
Warning:IMLEC POZISYONU YANLIS	5
Warning:MAZATROL PROGRAMLAMA SECILDI	2
Warning:ÇARPISMA	1
Warning:CANNOT CHANGE UNIT	1
Warning:VERILER EKSIK	1
Warning:SOFT LIMIT	1

Kaynak: Yazar tarafından derlenmiştir.

Şirketle yapılan görüşmeler doğrultusunda normal değerler makinenin çalışması için optimum değerleri içermektedir. Diğer durumlarda ise cihazın bakıma girmesi gerektiği bilgisi verilmiştir. Bu durumda cihazın tüm frekanslarının kontrol edildiği durumlar için model kurulduğu gibi özellik mühendisliği yöntemi ile veri setinin sistem durumu normal değerler için 1, diğerler için 0 olarak gruplandırıldığı durumlar için de ayrı bir model kurulmuştur. Her iki durum için de başarı hesaplaması yapılmıştır. 1 ve 0 için gruplama yapıldıktan sonraki frekans dağılımı Şekil 30'da gösterilmektedir.

Şekil 29: Hedef Değişkenin Gruplandırılmış Frekans Değerleri



Kaynak: Yazar tarafından derlenmiştir.

3.3.4. Kategorik Verilerin Numerik Değere Dönüştürülmesi

Birçok makine öğrenmesi algoritması, verileri sayısal formatta işlemek üzerine tasarlanmıştır. Dolayısıyla, kategorik olan verilerin sayısal değerlere dönüştürülmesi ve böylelikle kategorik değişkenlerin makine öğrenmesi algoritmalarında kullanılabilmesi sağlanmaktadır. Verileri numerik değere dönüştürmek için Label Encoder ya da One-Hot Encoder yöntemleri tercih edilmektedir. Spindle Sıcaklık Durumu, Servo Durumu, Spindle Durumu ve Emergency Durum değişkenleri veri setindeki kategorik değişkenlerdir. İlgili veri setinde kategorik değer içeren değişkenler Label Encoder yöntemi ile numerik değerlere dönüştürülmüştür.

3.3.5. Veri Normalleştirme

Veri normalleştirme, farklı ölçeklere sahip olan değerler aynı ölçekte ifade etmek için kullanılan bir veri ön işleme tekniğidir. Makine öğrenmesi veya istatistiksel analiz gibi veri tabanlı işlemlerde, değerlerin farklı ölçeklere sahip olması modelin doğruluğunu olumsuz etkileyebilmektedir. Veri normalleştirme birden çok yöntem bulunmaktadır. Bu çalışma kapsamın min-max normalleştirme yöntemi kullanılmaktadır. Min-max normalleştirme yönteminde değerleri 0 ile 1 arasında normalleştirmektedir. Veri setindeki tüm değerler min-max normalleştirme yöntemi ile normalleştirilmiştir.

3.3.6. Özellik Seçimi

Veri setindeki gereksiz değişkenlerin ya da hedef değişken üzerindeki etkisiz değişkenlerin çıkarılması ve modelin performansını artıracak değişkenlerin seçilmesi işlemidir. SelectKBest yöntemi ile model kurulacak 5 değişken seçilerek özellik seçimi işlemi tamamlanmıştır. Şekil 31'de seçilen 5 değişken bulunmaktadır.

Şekil 30: Özellik Seçimi Sonrası Seçilen Değişkenler

```
Index(['TarihSaat', 'Spindle Sıcaklığı', 'Anamil Soğutucu Yağ Sıcaklığı',
       'M51 Motor Sıcaklığı', 'Tezgah Trafo Sıcaklığı'],
      dtype='object')
```

Kaynak: Yazar tarafından derlenmiştir.

3.4. MODEL KURMA

Veri ön işleme süreçleri makine öğrenmesi algoritmalarına daha uygun veri seti hazırlamak için uygulanmıştır. Bu süreçte Bölüm 2 de bahsi geçen DT, NB, kNN, SVM, AdaBoost, RF, CatBoost, XGBoost ve LightGBM algoritmaları kullanılmıştır. Modelin eğitilmesinde ayırma ve k katmanlı çapraz doğrulama yöntemi kullanılmıştır. Modelin hedef değişkeni için öncelikle tüm sistem durumu sınıflarına göre model kurulmuştur. Daha sonra hedef değişken normal değerler ve hatalı değerler şeklinde 2 frekans haline dönüştürülerek model kurulumu gerçekleşmiştir.

3.5. MODEL PERFORMANS DEĞERLENDİRMESİ

Hedef değişkenin kategorik değerlerden oluşması sebebiyle model performansında sınıflandırma ölçüm yöntemleri kullanılmıştır. Kurulan her bir modelin test sonuçlarının accuracy, precision, recall ve f1-score değerleri hesaplanmıştır.

Bu bölümde makine öğrenmesi algoritmalarının öncelikle ayırma yöntemi ile ilgili başarı sonuçları hesaplanmıştır. Ayırma yöntemi sistem durumu değişkenin öncelikle tüm sınıf frekanslarına uygulanmıştır. Daha sonra hedef değişken normal değerler ve hatalı değerler şeklinde 2 frekans haline dönüştürülerek model performansları değerlendirilmiştir. Ayırma yönteminde veri bölümlendirme %80 eğitim, %20 test veri seti olarak belirlenmiştir. Tablo 5'te tüm sınıf frekansları kullanılan durum için model performans değerleri bulunmaktadır. Tablo 6'da ise hedef değişkenin normal ve hatalı olma durumlarını içeren 2 frekanslı yapısının model performans değerleri bulunmaktadır.

Tablo 5: Ayırma Yöntemi Kullanılarak Oluşturulan Tüm Sınıf Frekanslarının Değerlendirildiği Model Başarı Performansları

Algoritma	Accuracy	Precision	Recall	F1-Score
DT	0.906	0.907	0.909	0.908
NB	0.882	0.882	0.900	0.896
kNN	0.906	0.906	0.909	0.908
SVM	0.885	0.885	0.900	0.890
AdaBoost	0.904	0.903	0.910	0.907
RF	0.906	0.907	0.909	0.908
CatBoost	0.901	0.901	0.910	0.906
XGBoost	0.902	0.901	0.915	0.906
LightGBM	0.904	0.906	0.907	0.907

Kaynak: *Yazar tarafından derlenmiştir.*

Tablo 6: Ayırma Yöntemi Kullanılarak Oluşturulan İkili Sınıf Frekanslarının Değerlendirildiği Model Başarı Performansları

Algoritma	Accuracy	Precision	Recall	F1-Score
DT	0.950	0.950	0.950	0.950
NB	0.922	0.922	0.950	0.936
kNN	0.946	0.946	0.949	0.948
SVM	0.925	0.925	0.950	0.937
AdaBoost	0.944	0.945	0.950	0.947
RF	0.946	0.947	0.949	0.948

CatBoost	0.941	0.941	0.950	0.946
XGBoost	0.947	0.946	0.950	0.948
LightGBM	0.944	0.946	0.947	0.947

Kaynak: Yazar tarafından derlenmiştir.

K katmanlı çapraz doğrulama yönteminde k değeri 10 olarak belirlenerek model kurma işlemi gerçekleştirilmiştir. Hedef değişkenin sınıf değerlerinin tamamının kullanımı k katmanlı çapraz doğrulama yapılmak istendiğinde veri bölümlendirme işlemi dengesiz veri seti probleminden dolayı yapılamamıştır. Dolayısıyla ilgili yöntem sadece hedef değişkenin 2 frekanslı yapısı ile yapılmıştır. Tablo 7'de ise hedef değişkenin normal ve hatalı olma durumlarını içeren 2 frekanslı yapısının model performans değerleri bulunmaktadır.

Tablo 7: K Katmanlı Çapraz Doğrulama Yöntemi Kullanılarak Oluşturulan İkili Sınıf Frekanslarının Değerlendirildiği Model Başarı Performansları

Algoritma	Accuracy	Precision	Recall	F1-Score
DT	0.655	0.993	0.651	0.767
NB	0.968	0.972	0.996	0.984
kNN	0.663	0.990	0.659	0.772
SVM	0.972	0.972	1.0	0.98
AdaBoost	0.912	0.992	0.917	0.937
RF	0.732	0.993	0.722	0.814
CatBoost	0.853	0.993	0.855	0.910
XGBoost	0.740	0.993	0.738	0.825
LightGBM	0.817	0.989	0.821	0.888

Kaynak: Yazar tarafından derlenmiştir.

Model kurma sürecinde modellerin öğrenme ve test etme becerilerinin artırılması için hiperparametre kullanımları gerçekleştirilmiş. Tablo 8'de her bir algoritma için optimum değerlerin oluşturulması için farklı hiperparametreler kullanılmıştır. Tablo 8'de algoritmalar ve manuel olarak belirlenen hiperparametreler gösterilmektedir.

Tablo 8: Modellerde Kullanılan Hiperparametreler

Algoritma	Kullanıla
DT	Criterion: gini,entropy,log_loss
NB	Hiperparametre ihtiyacı olmadığından değişiklik yapılmamıştır.
kNN	N_neighbors: 2,5,10,100, Weights: uniform, distance, Algorithm: ball_tree, kd_tree, brute, Metric: euclidean, manhattan, chebyshev, minkowski, P: 1,2
SVM	C: 1, Kernel: linear, poly, rbf, sigmoid, Degree: 3,5
AdaBoost	Random_state : 1
RF	N_estimators:10,50,100, Critesion: gini,entropy, log_loss, Max_features: sqrt, log2
CatBoost	Learning_rate: 0.03,0.1, Depth: 4,6,8, 12_leaf_reg: 1,3,5,7,9
XGBoost	Min_child_weight: 1,5,10, Gamma: 0.5,1,1.5,2,5, Subsample: 0.6,0.8,1.0, Colsample_bytree: 0.6,0.8,1.0, Max_depth: 3,4,5
LightGBM	Num_leaves: 20,40,60,80,100, Min_child_samples: 5,10,15, Max_depth: -1,5,10,20, Learning_rate: 0.05,0.1,0.1, Reg_alpha: 0,0.01,0.03

Kaynak: Yazar tarafından derlenmiştir.

Tablo 8’de gösterilen hiperparametreler ilgili algoritmalara uygulanmış ve model performans değerleri Tablo 9’da gösterilmiştir.

Tablo 9: Hiperparametre Kullanılarak Oluşturulan Modellerin Başarı Performansları

Algoritma	Accuracy	Precision	Recall	F1-Score
DT	0.997	0.997	0.999	0.999
NB	0.998	0.998	0.999	0.998
kNN	0.995	0.996	0.998	0.997
SVM	0.995	0.996	0.996	0.996
AdaBoost	0.993	0.992	1.000	0.996
RF	0.993	0.995	0.997	0.996
CatBoost	0.991	0.995	0.996	0.996
XGBoost	0.992	0.995	0.991	0.993
LightGBM	0.999	0.999	1.000	0.998

Kaynak: Yazar tarafından derlenmiştir.

SONUÇ VE ÖNERİLER

Gerçekleştirilen uygulamada, kestirimci bakım çalışması ile ilgili amaca uygun modeller kurulmuş ve çözümlere uygun ilham verici sonuçlara ulaşılmıştır.

Çalışmaların sonuçlarına bakıldığında, yapılan uygulama insan etkeni dışında bir makinenin arıza oluşturup, oluşturmayacağına karar vermektedir. Bu işlem hem makinenin parçalarına zarar verme oranını azaltarak ciddi bir giderin önüne geçmekte hem de arızadan dolayı modele verilen zararı önlemektedir.

Çalışma ile ilgili kurulan makine öğrenmesi modellerinde elde edilen sonuçlar farklı makine öğrenmesi algoritmaları ile denenmiş ve sonuçların doğruluk oranları incelenmiştir. Bu incelemelerde veri setinin büyük olmasından ve etkili bir veri ön işleme sürecinin tasarlanmasından kaynaklı genel olarak algoritmalar %89-%91 oranında başarılı sonuca ulaşmaktadır. İlgili sonuçlar yöneticilerin ve bakım personellerinin kararlarını doğrudan etkilediği için daha yüksek performanslı sonuçlara ulaşılabilmesi için veri bölümlendirme ve hiperparametre kullanımı yöntemleri uygulanmıştır. Yapılan işlemler sonrasında lightGBM algoritması ile %100'e yakın performans başarı değerine ulaşılmıştır.

Günümüzde, endüstrinin karar verme süreçlerinde etkili yöntemler arayışı hızla artmaktadır. Bu durum, yapay zeka uygulamaları, makine öğrenmesi ve derin öğrenme gibi teknolojilere olan ihtiyacı da beraberinde getirmektedir. Özellikle üretim süreçlerine entegre edilen robotik sistemler, bakım maliyetlerinin yükselmesine neden olmaktadır. Bu artan maliyetleri kontrol altına almak, bakım maliyetlerini azaltmak ve üretim sürekliliğini sağlamak için kestirimci bakım uygulamalarına olan talebin her geçen gün artması kaçınılmazdır.

Çalışma yönetim bilişim sistemleri odağında değerlendirildiğinde, tahmine dayalı verimlilik süreci ön plana çıkmaktadır. Bu süreçteki katkılar aşağıdaki gibi listelenebilir:

- Kestirimci bakım arızaları önceden tespit etme ve önleme yönüyle iş durmalarını engelleyerek iş sürekliliği artırır ve şirketin verimliliğini yükseltir.
- Düzenli bakımın önüne geçerek gereksiz yere değişen parçaların ve bakım maliyetlerin önüne geçmektedir. Bu da şirketin bakım maliyetlerini azaltmaktadır.

o Kestirimci bakım, arızadan kaynaklı oluşan ekipman yıpranmalarının önüne geçerek ekipmanın kullanım ömrünü uzatır. Dolayısıyla, süreçteki fazla harcamanın önüne geçmektedir.

o Arızaların tahmin edilmesi ve bakım zamanlarının önceden planlanması, bakım ekiplerinin işlerini daha iyi planlamasına ve kaynaklarını etkin bir şekilde kullanmasına yardımcı olmaktadır. Dolayısıyla iş akışında ve zaman yönetiminde verimliliğin artmasını sağlamaktadır.

o Beklenmedik arızaların önlenmesi, müşterilere sürekli ve kesintisiz hizmet sağlama kabiliyetini de artırmaktadır. Bu da müşteri memnuniyetini ve sadakatini sağlamaktadır.

İlgili çalışmada, 3 aylık gerçek verilerden yararlanılmıştır ve prototip aşamasındadır. Tez kapsamında geliştirilen kestirimci bakım uygulamasının daha sürdürülebilir olması için, gelecek çalışmalarda,

- En az 1 yıllık gerçek veri kullanılması,
- Makine öğrenmesi algoritmaları dışında derin öğrenme algoritmalarının da kullanılması,
- Sadece kategorik hedef değişken değil, zaman serilerine bağlı sonuçlar üzerine çalışılması,
- Sistemdeki anlık veri ile karar veren bir sistemin tasarlanması,
- Hiperparametrelerin manuel yerine otomatik optimizasyon yöntemleri ile seçilmesi,
- Mobil cihazlara arıza tahmin sonucunun iletilmesi, sistemin daha verimli olmasını sağlayabilir.
- Çalışma sağlanan şirkete kestirimci bakım uygulandığında yıllık maliyete etkisi üzerine çalışma gerçekleştirilebilir.
- Toplam ekipman etkinliği (OEE – Overall Equipment Efficiency) analizi ile verimlilik değerlendirilmesi yapılabilir.

KAYNAKÇA

Akdoğan, N. ve Akdoğan, M. U. (2018). Büyük Veri-Bilişim Teknolojisindeki Gelişmelerin Muhasebe Uygulamalarına ve Muhasebe Mesleğine Etkisi. *Muhasebe ve Denetime Bakış*, *18*(55), 1-14.

Aktan, E. (2018). Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu. *Bilgi Yönetimi Dergisi*, *1*(1), 1-22.

Aldahiri, A., Alrashed, B. ve Hussain, W. (2021). Trends in Using IoT with Machine Learning in Health Prediction System. *Forecasting*, *3*(1), 181-207.

Ali, N., Neagu, D. ve Trundle, P. (2019). Evaluation of k-Nearest Neighbour Classifier Performance for Heterogeneous Data Sets. *SN Applied Sciences*, *1*(12), 1559.

Alshari, H., Yahya Saleh, A. ve Odabas, A. (2021). Comparison of Gradient Boosting Decision Tree Algorithms for CPU Performance Linguistics Analysis View Project Comparison of Gradient Boosting Decision Tree Algorithms for CPU Performance. *Erciyes Tip Dergisi*, *37*(1), 157-168.

Androutsopoulos, I., Koutsias, J., Chandrinos, K. ve Spyropoulos, C. (2000). *An Experimental Comparison of Naive Bayesian and Keyword-Based Anti-Spam Filtering with Personal E-mail Messages*. 160-168.

Arslan, K. (2020). Eğitimde Yapay Zeka ve Uygulamaları. *Batı Anadolu Eğitim Bilimleri Dergisi*, *11*(1), 71-88.

Ayvaz, S. ve Alpay, K. (2021). Predictive Maintenance System for Production Lines in Manufacturing: A Machine Learning Approach Using IoT Data in Real-Time. *Expert Systems with Applications*, *173*(6).

Bataller, C. ve Harris, J. (2016). Turning Artificial Intelligence into Business Value. Today. *Accenture*, 1-16.

Bektaş, O. (2020). Kestirimci Bakım İçin Döner Mekanizma Bozulma Eğrisinin Tanımlanması. *Avrupa Bilim ve Teknoloji Dergisi, 19*, 420-428.

Bogehoj, L. (2016). Artificial Intelligence vs. Human Intelligence (Man vs. Machine). *Discussion Paper – On Artificial Intelligence*.

Bousdekis, A., Lepenioti, K., Apostolou, D. ve Mentzas, G. (2019). Decision Making in Predictive Maintenance: Literature Review and Research Agenda for Industry 4.0. *IFAC-PapersOnLine, 52*(13), 607-612.

Bousdekis, A. ve Mentzas, G. (2021). Enterprise Integration and Interoperability for Big Data-Driven Processes in the Frame of Industry 4.0. *Frontiers in Big Data, 4*.

Brkovic, A., Gajic, D., Gligorijevic, J., Savic-Gajic, I., Georgieva, O. ve Di Gennaro, S. (2015, Kasım 13). Early Fault Detection and Diagnosis in Bearings Based on Logarithmic Energy Entropy and Statistical Pattern Recognition. *Proceedings of 2nd International Electronic Conference on Entropy and Its Applications*.

Carvalho, T. P., Soares, F. A. A. M. N., Vita, R., Francisco, R. da P., Basto, J. P. ve Alcalá, S. G. S. (2019). A Systematic Literature Review of Machine Learning Methods Applied to Predictive Maintenance. *Computers and Industrial Engineering, 137*, 1-9.

Ceyhan, H. ve Kasapbaşı, M. C. (2022). Üretim Sistemlerinde Makine Öğrenmesi ile Kestirimci Bakım Uygulaması ve Modellemesi. *Avrupa Bilim ve Teknoloji Dergisi, 33*, 167-175.

Chengsheng, T., Huacheng, L. ve Bing, X. (2017). AdaBoost Typical Algorithm and Its Application Research. *MATEC Web of Conferences, 139*.

Chlis, N. (2013). *Comparison of Statistical Methods for Genomic Signature Extraction*.

Coşkun, F. ve Gülerlioğlu, H. D. (2021). Geçmişten Günümüze Yapay Zekanın Gelişimi ve Eğitim Alanında Kullanılması. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 54*(3), 948-966.

Cutler, A., Cutler, D. R. ve Stevens, J. R. (2012). Random Forests. *DBLP*, *45*, 157-176.

Çelik, Ö. ve Altunaydın, S. S. (2018). A Research on Machine Learning Methods and Its Applications. *Journal of Educational Technology & Online Learning*, *1*(3), 25-40.

Çınar, A. (2019). Veri Madenciliğinde Sınıflandırma Algoritmalarının Performans Değerlendirmesi ve R Dili İle Bir Uygulama. *Marmara Üniversitesi Öneri Dergisi* , *14*(51), 90-111.

Çınaroğlu, S. (2017). Sağlık Harcamasının Tahmininde Makine Öğrenmesi Regresyon Yöntemlerinin Karşılaştırılması. *Uludağ University Journal of The Faculty of Engineering*, *22*(2), 179-200.

Delen, D. ve Ram, S. (2018). Research Challenges and Opportunities in Business Analytics. *Journal of Business Analytics, 1*(1), 2-12.

Dierking, I., Dominguez, J., Harbon, J. ve Heaton, J. (2023). Deep Learning Techniques for the Localization and Classification of Liquid Srystal Phase Transitions. *Frontiers in Soft Matter*, *3*, 1-11.

Doğan, O. ve Baloğlu, N. (2020). Endüstri 4.0 Kavramsal Farkındalık Ölçeği. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, *22*(38), 58-81.

Dridi, S. (2021). *Unsupervised Learning-A Systematic Literature Review*. OSF Preprints. https://www.researchgate.net/publication/357380639

Emir, T., Marangoz, M. C. ve Işık, M. Z. (2012). Endüstriyel İşletmelerde Bakım Kültürünün Değişimi. *Journal of Life Sciences*, *1*(2), 365-371.

Farzinpour, A., Mohammadi Dehcheshmeh, E., Broujerdian, V., Nasr Esfahani, S. ve Gandomi, A. H. (2023). Efficient Boosting-Based Algorithms for Shear Strength Prediction of Squat RC Walls. *Case Studies in Construction Materials*, *18*, 1-16.

Feurer, M. ve Hutter, F. (2019). Hyperparameter Optimization. *Springer International Publishing*, 3-33.

Figueroa, M. ve Pope, C. (2017). Root System Water Consumption Pattern Identification on Time Series Data. *Sensors*, *17*(6), 1-21.

García-Gonzalo, E., Fernández-Muñiz, Z., Nieto, P. J. G., Sánchez, A. B. ve Fernández, M. M. (2016). Hard-Rock Stability Analysis for Span Design in Entry-Type Excavations with Learning Slassifiers. *Materials*, *9*(531), 1-19.

Gomes Mantovani, R. (2018). *Use of Meta-Learning for Hyperparameter Tuning of Classification Problems Use of Meta-Learning for Hyperparameter Tuning of Classification Problems View Project Applications of ML Techniques in Real Problems View project*.

Google. (2022). *Google Books Gram Viewer*. Google. https://books.google.com/ngrams/graph?content=machine+learning&year_start=1900&year_end=2019&corpus=26&smoothing=6#

Guo, G., Wang, H., Bell, D. A., Bi, Y., Bell, D. ve Greer, K. (2004). *KNN Model-Based Approach in Classification*.

Gür, Y. ve Yıldız, A. (2022). Dijitalleşme ve Tüketim. Içinde *Türkiye'de Kamu Kurumlarında Dijitalleşme: Hazine ve Maliye Bakanlığı Örneği* (ss. 149-200). Dora Yayınevi.

Güven, Ö. ve Şahin, H. (2022). Predictive Maintenance Based On Machine Learning In Public Transportation Vehicles. *Mühendislik Bilimleri ve Araştırmaları Dergisi*, *4*(1), 89-98.

Herrero, R. D. ve Zorrilla, M. (2022). An I4.0 Data Intensive Platform Suitable for the Deployment of Machine Learning Models: A Predictive Maintenance Service Case Study. *Procedia Computer Science*, *200*, 1014-1023.

Houtmeyers, K. C., Jaspers, A. ve Figueiredo, P. (2021). Managing the Training Process in Elite Sports: From Descriptive to Prescriptive Data Analytics. *International Journal of Sports Physiology and Performance*, *16*(11), 1719-1723.

İnce, H., İmamoğlu, S. E. ve İmamoğlu, S. Z. (2021). Yapay Zeka Uygulamalarının Karar Verme Üzerine Etkileri: Kavramsal Bir Çalışma. *International Review of Economics and Management*, *9*(1), 50-63.

Jamaleddyn, I., El ayachi, R. ve Biniz, M. (2023). An Improved Approach to Arabic News Classification Based on Hyperparameter Tuning of Machine Learning Algorithms. *Journal of Engineering Research*, *11*(2).

Javed Mehedi Shamrat, F. M., Ranjan, R., Hasib, K. M., Yadav, A. ve Siddique, A. H. (2022). Performance Evaluation Among ID3, C4.5, and CART Decision Tree Algorithm. *International Conference on Pervasive Computing and Social Networking*, *317*, 127-142.

Kane, A., Kore, A., Khandale, A., Nigade, S. ve Joshi, P. P. (2022). *Predictive Maintenance Using Machine Learning*.

Karahan Adalı, G. ve Balaban, M. E. (2019). Elektrik Malzemeleri Sektöründe R ve Shiny ile Pazar Sepet Analizine Yönelik Dinamik Bir Uygulama. *Bilişim Teknolojileri Dergisi*, *12*(2), 93-102.

Kartal, E. ve Ozen, Z. (2017). Dengesiz Veri Setlerinde Sınıflandırma. Içinde *Mühendislik ve Yapay Zeka Uygulamaları* (C. 1, ss. 109-131).

Kasap, M., Çınar, E., Yazıcı, A. ve Özkan, K. (2021). An Experimental Evaluation of Intellegent Fault Detection and Classification and Classification for Induction Motors Utilizing Machine Learning Approaches. *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi*, *29*(2), 126-136.

Kavzoğlu, T. ve Çölkesen, İ. (2010). Destek Vektör Makineleri ile Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi. *Harita Dergisi*, *144*, 73-82.

Kaya, M. (2021). Sanayi 4.0'da Yapay Zekâ ve Türkiye. *Fırat Üniversitesi. İİBF Uluslararası İktisadi ve İdari Bilimler Dergis*, *5*(2), 63-94.

Khan, M. Y., Qayoom, A., Nizami, M. S., Siddiqui, M. S., Wasi, S. ve Raazi, S. M. K.-R. (2021). Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques. *Complexity*, *2021*, 1-18.

Kimera, D. ve Nangolo, F. N. (2020). Predictive Maintenance for Ballast Pumps on Ship Repair Yards via Machine Learning. *Transportation Engineering*, *2*.

Kocakır Gacar, B. ve Deveci Kocakoç, İ. (2020). Regresyon Analizleri mi Karar Ağaçları mı? *Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, *18*(4), 251-260.

Liu, Q. ve Wu, Y. (2012a). Supervised Learning. *Encyclopedia of the Sciences of Learning*, 3243-3245.

Liu, Q. ve Wu, Y. (2012b). Supervised Learning. Içinde *Encyclopedia of the Sciences of Learning* (ss. 3243-3245). Springer US.

Liu, Y. ve Yang, S. (2022). Application of Decision Tree-Based Classification Algorithm on Content Marketing. *Journal of Mathematics*, *2022*, 1-10.

Mahesh, B. (2018). Machine Learning Algorithms-A Review Machine Learning Algorithms-A Review View project Self Flowing Generator View project Batta Mahesh Independent Researcher Machine Learning Algorithms-A Review. *International Journal of Science and Research*, *9*(1), 381-386.

Merkt, O. (2019). On the Use of Predictive Models for Improving theQuality of Industrial Maintenance: an AnalyticalLiterature Review of Maintenance Strategies. *Proceedings of the 2019 Federated Conference on Computer Science and Information Systems, FedCSIS 2019*, *18*, 693-704.

Nikfar, M., Bitencourt, J. ve Mykoniatis, K. (2022). A Two-Phase Machine Learning Approach for Predictive Maintenance of Low Voltage Industrial Motors. *Procedia Computer Science*, *200*, 111-120. https://doi.org/10.1016/j.procs.2022.01.210

Nozari, H., Sadeghi, M. E. ve Ghahremani Nahr, J. (2021). Artificial intelligence and Machine Learning for Real-world problems (A survey). *International journal of innovation in Engineering*, *1*(3), 38-47.

Nur Calayır, G. ve Kabak, M. (2021). Bakım için Makine Öğrenme Tekniklerinin Analizi ve Bir Uygulama. *Journal of Turkish Operations Management, 1*(5), 662-675.

Nur Calayır, G., Kabak, M., Mühendisliği Bölümü, E. ve Üniversitesi, G. (2021). *Journal of Turkish Operations Management Bakım için makine öğrenme tekniklerinin analizi ve bir uygulama* (C. 1, Sayı 5). http://orcid.org/0000-0001-9344-701X

Onan, A. (2015). Şirket İflaslarının Tahminlenmesinde Karar Ağacı Algoritmalarının Karşılaştırmalı Başarım Analizi. *Bilişim Teknolojileri Dergisi, 8*(1), 9-19.

Özkat, E. C. (2021). Makine Öğrenmesi Metodolojisi Kullanılarak Yüksek Hızlı Rulmanlarda Sağlık Göstergesinin Belirlenmesi. *European Journal of Science and Technology, 22*, 176-183.

Pandey, A. ve Jain, A. (2017). Comparative Analysis of KNN Algorithm Using Various Normalization Techniques. *International Journal of Computer Network and Information Security*, *9*(11), 36-42. https://doi.org/10.5815/ijcnis.2017.11.04

Payal, M., Ananth Kumar, T. ve Suresh Kumar, K. (2022). Machine Learning in Industry 4.0. *International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), 11*(5), 5831-5840.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. ve Gulin, A. (2018). CatBoost: Unbiased Boosting with Categorical Features. *32nd Conference on Neural Information Processing Systems*, 1-11.

Pulat, M. ve Deveci Kocakoç, İ. (2021). Türkiye'de Makine Öğrenmesi ve Karar Ağaçları Alanında Yayınlanmış Tezlerin Bibliyometrik Analizi. *Yönetim ve Ekonomi*, *28*(2).

Rana Dündar, D., Sariçiçek, İ., Çinar, E., Yazici, A., Osmangazi Üniversitesi, E., Fakültesi, M.-M., Mühendisliği, E. ve Kelimeler Öz Kestirimci Bakım Makine Öğrenmesi Derin Öğrenme, A. (2021). Kestirimci Bakımda Makine Öğrenmesi: Literatür Araştırması. *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi*, *29*(2), 256-276.

Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P. ve Green, R. (2019). Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. *Academic Pathology*, *6*.

Rattan, P., Penrice, D. D. ve Simonetto, D. A. (2022). Artificial Intelligence and Machine Learning: What You Always Wanted to Know but Were Afraid to Ask. *Gastro Hep Advances*, *1*(1), 70-78.

Rokach, L. ve Maimon, O. (2015). Decision Trees. Içinde *Data Mining and Knowledge Discovery Handbook* (C. 6, ss. 165-192). Springer-Verlag.

Romeo, L., Loncarski, J., Paolanti, M., Bocchini, G., Mancini, A. ve Frontoni, E. (2020). Machine Learning-Based Design Support System for the Prediction of Heterogeneous Machine Parameters in Industry 4.0. *Expert Systems with Applications*, *140*.

Sadıç, Ş. (2022). *Endüstri 5.0 ve Sürdürülebilirlik*. https://www.researchgate.net/publication/361506170_Endustri_50_ve_Surdurulebilirlik

Sah, S. (2020). *Machine Learning: A Review of Learning Types*. www.preprints.org

Sanzana, M. R., Maul, T., Wong, J. Y., Abdulrazic, M. O. M. ve Yip, C. C. (2022). Application of Deep Learning in Facility Management and Maintenance for Heating, Ventilation, and Air Conditioning. *Automation in Construction*, *141*.

Sapountzoglou, N., Lago, J. ve Raison, B. (2020). Fault Diagnosis in Low Voltage Smart Distribution Grids Using Gradient Boosting Trees. *Electric Power Systems Research*, *182*.

Schock, C. (2019). *CRISP-DM – ein Ansatz zur Systematisierung von Digitalisierungsprojekten in Produktionsumfeldern*.

Shahani, N. M., Zheng, X., Guo, X. ve Wei, X. (2022). Machine Learning-Based Intelligent Prediction of Elastic Modulus of Rocks at Thar Coalfield. *Sustainability*, *14*(6).

Shankarrao Patange, G. ve Bharatkumar Pandya, A. (2023). How Artificial Intelligence and Machine Learning Assist in Industry 4.0 for Mechanical Engineers. *Materials Today: Proceedings*, *72*, 622-625.

Shearer C. (2000). The CRISP-DM Model: The New Blueprint for DataMining. *Journal Of Data Warehousing*, *5*(4), 13-23.

Soylu, B., Yiğiter, H., Sarıkaya, V., Sandıkçı, Z. ve Utku, A. (2021). Kestirimci Bakım Planlama İçin Makine Öğrenmesi Temelli Bir Karar Destek Sistemi ve Bir Uyuglama. *Verimlilik Dergisi*, 48-66.

Sun, J., Du, W. ve Shi, N. (2018). A Survey of kNN Algorithm. *Information Engineering and Applied Computing*, *1*.

Tabak, A. ve Özkaymak, M. (2020). Elektrik Motorlarında Uygulanan Bakım Yöntemlerinin İncelenmesi, Karşılaştırılması ve Uzaktan Erişimin Kestirimci Bakıma Etkisi. *Konya Journal of Engineering Sciences*, *8*(4), 889-905.

Taş, H. Y. (2018). Dördüncü Sanayi Devrimi'nin (Endüstri 4.0) Çalışma Hayatına ve İstihdama Muhtemel Etkileri. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, *9*(16), 1817-1836.

Thessen, A. E. (2016). Adoption of Machine Learning Techniques in Ecology and Earth Science. *One Ecosystem*, *1*.

Tong, R., Feng, Y., Wang, J., Wu, Z., Tan, M. ve Yu, J. (2023). A Survey on Reinforcement Learning Methods in Bionic Underwater Robots. *Biomimetics*, *8*(2), 168-197.

Tounsi, Y., Anoun, H. ve Hassouni, L. (2020). CSMAS. *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, 1-7.

Ümit Gürsoy, M., Can Çolak, U., Hilmi Gökçe, M., Akkulak, C. ve Ötleş, S. (2019). Endüstri İçin Kestirimci Bakım. *Uluslararası 3B Yazıcı Teknolojileri ve Dijital Endüstri Dergisi*, *3*(1), 56-66.

Ümit Gürsoy, M., Can Çolak, U., Hilmi Gökçe, M., Akkulak, C., Ötleş, S., Üniversitesi, E., Bilimleri Enstitüsü, F., Yaşam Döngüsü Yönetimi Anabilim Dalı, Ü., Ege Üniversitesi, T., Yaşam Döngüsü Yönetimi Mükemmeliyet Merkezi, Ü. ve Yazar, S. (2019a). Endüstri İçin Kestirimci Bakım. Içinde *Internatıonal Journal Of 3d Prıntıng Technologıes And Dıgıtal Industry* (C. 3).

Ümit Gürsoy, M., Can Çolak, U., Hilmi Gökçe, M., Akkulak, C., Ötleş, S., Üniversitesi, E., Bilimleri Enstitüsü, F., Yaşam Döngüsü Yönetimi Anabilim Dalı, Ü., Ege Üniversitesi, T., Yaşam Döngüsü Yönetimi Mükemmeliyet Merkezi, Ü. ve

Yazar, S. (2019b). Endüstri İçin Kestirimci Bakım. Içinde *Internatıonal Journal Of 3d Prıntıng Technologıes And Dıgıtal Industry* (C. 3).

Velu Sindhu. (2020). An Empirical Science Research on Bioinformatics in Machine Learning. *Journal of Mechanics Continua and Mathematical Sciences*, *7*(1).

Wang, K., Gao, J., Li, H. ve Liang, L. (2023). Stacking Strategy-Assisted Random Forest Algorithm and Its Application. *AIP Advances*, *13*(3).

Wang, T., Li, X., Wang, W., Du, J. ve Yang, X. (2023). A Spatiotemporal Feature Learning-Based RUL Estimation Method for Predictive Maintenance. *Measurement: Journal of the International Measurement Confederation*, *214*.

Wang, W., Chakraborty, G. ve Chakraborty, B. (2020). Predicting the Risk of Chronic Kidney Disease (CKD) Using Machine Learning Algorithm. *Applied Sciences*, *11*(1), 202.

Wang, Y., Guo, L., Zhang, Y. ve Ma, X. (2023). Research on CSI 300 Stock Index Price Prediction Based On EMD-XGBoost. *Frontiers in Computing and Intelligent Systems*, *3*(1), 72-77.

Wang, Z., Zhang, J. ve Verma, N. (2015). Realizing Low-Energy Classification Systems by Implementing Matrix Multiplication Directly Within an ADC. *IEEE Transactions on Biomedical Circuits and Systems*, 1-1.

Xayyasith, S., Promwungkwa, A. ve Ngamsanroaj, K. (2018). Application of Machine Learning for Predictive Maintenance Cooling System in Nam Ngum-1 Hydropower Plant. *2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE)*, 1-5.

Yeardley, A. S., Ejeh, J. O., Allen, L., Brown, S. F. ve Cordiner, J. (2022). Integrating Machine Learning Techniques into Optimal Maintenance Scheduling. *Computers and Chemical Engineering*, *166*.

Yılmaz, F. ve Kasapbaşı, M. (2021). EEG Sinyalleri ile Epilepsi Krizinin Tahminlenmesinde Rassal Orman Algoritması ile Hiper Parametre Optimizasyonun Uygulanması. *Journal of Technology and Applied Sciences*, *3*(2), 189-201.

Yin, L., Ma, P. ve Deng, Z. (2021). JLGBMLoc—A Novel High-Precision Indoor Localization Method Based on LightGBM. *Sensors*, *21*(8).

Zhang, H. (2004). The Optimality of Naive Bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference*, *2*.

Zhang, S., Cheng, D., Deng, Z., Zong, M. ve Deng, X. (2018). A Novel kNN Algorithm with Data-Driven k Parameter Computation. *Pattern Recognition Letters*, *109*, 44-54.

Zhang, W. (2017). Machine Learning Approaches to Predicting Company Bankruptcy. *Journal of Financial Risk Management*, *06*(04), 364-374.

Zhong, D., Xia, Z., Zhu, Y. ve Duan, J. (2023). Overview of Predictive Maintenance Based on Digital Twin Technology. *Heliyon*, *9*(4).