



**YAPAY ZEKA TEKNİKLERİ İLE COVID-19  
HASTALIK TAHMİNİ**

**2023  
YÜKSEK LİSANS TEZİ  
BİLGİSAYAR MÜHENDİSLİĞİ**

**Abdullah Türker TOKU**

**Tez Danışmanı  
Dr. Öğr. Üyesi Ferhat ATASOY**

**YAPAY ZEKA TEKNİKLERİ İLE COVID-19 HASTALIK TAHMİNİ**

**Abdullah Türker TOKU**

**Tez Danışmanı  
Dr. Öğr. Üyesi Ferhat ATASOY**

**T.C.  
Karabük Üniversitesi  
Lisansüstü Eğitim Enstitüsü  
Bilgisayar Mühendisliği Anabilim Dalında  
Yüksek Lisans Tezi  
Olarak Hazırlanmıştır**

**KARABÜK  
Mayıs 2023**

Abdullah Türker TOKU tarafından hazırlanan “YAPAY ZEKA TEKNİKLERİ İLE COVID-19 HASTALIK TAHMİNİ” başlıklı bu tezin Yüksek Lisans Tezi olarak uygun olduğunu onaylarım.

Dr. Öğr. Üyesi Ferhat ATASOY

.....

Tez Danışmanı, Bilgisayar Mühendisliği Anabilim Dalı

Bu çalışma, jürimiz tarafından **Oy Birliği** ile Bilgisayar Mühendisliği Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir. 30/05/2023

Ünvanı, Adı SOYADI (Kurumu)

İmzası

Başkan : Doç. Dr. İlker TÜRKER ( KBÜ)

.....

Üye : Doç. Dr. Abdullah ELEN ( BANÜ)

ONLİNE

Üye : Dr. Öğr. Üyesi Ferhat ATASOY ( KBÜ)

.....

KBÜ Lisansüstü Eğitim Enstitüsü Yönetim Kurulu, bu tez ile, **Yüksek Lisans** derecesini onamıştır.

Prof. Dr. Müslüm KUZU

.....

Lisansüstü Eğitim Enstitüsü Müdürü

*“Bu tezdeki tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu; ayrıca bu kuralların ve ilkelerin gerektirdiği şekilde, bu çalışmadan kaynaklanmayan bütün atıfları yaptığımı beyan ederim.”*

Abdullah Türker TOKU

## **ÖZET**

**Yüksek Lisans Tezi**

### **YAPAY ZEKA TEKNİKLERİ İLE COVID-19 HASTALIK TAHMİNİ**

**Abdullah Türker TOKU**

**Karabük Üniversitesi**

**Lisansüstü Eğitim Enstitüsü**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Tez Danışmanı:**

**Dr. Öğr. Üyesi. Ferhat ATASOY**

**Mayıs 2023, 112 sayfa**

Bu tezde, sağlık hizmetleri yönetiminde etkinliği sağlamak üzere, Covid-19 hastalığının doğru olarak teşhisinde makine öğrenmesi tekniklerinin başarısının karşılaştırılması amaçlanmıştır. Yapay zeka ve makine öğrenmesi yöntemlerinin sağlık yönetimi ve hizmetleri alanındaki kullanımı gün geçtikçe artmaktadır.

Covid-19 virüsü tüm dünyayı etkisi altına alan, birçok hastada nefes darlığı, öksürük, ateş gibi benzer belirtiler ile ortaya çıkan bir Sars grubuna ait bir virus türüdür. Hastalığın teşhisinde PCR testleri kullanılmaktadır. Covid-19 pandemi koşullarında dünyanın gelişmiş ülkelerinde dahi test kitlerine erişimde sıkıntılar yaşanmıştır. Ayrıca test kitlerinin doğruluk oranının yüksek olmaması da ayrı bir sorun teşkil etmiştir. Bundan dolayı zaman zaman hastalık belirtilerine göre tedavi aşamasına geçilmiştir. Problemin çözümüne katkı sağlamak amacıyla Covid-19 virüsünün teşhisi için rutin kan testlerinin kullanılması çalışmanın motivasyonunu oluşturmaktadır.

Covid-19 risk faktörlerinin yaygınlığını anlamak için gerçekleştirilen ve kamuya açık, 1677 hastada PCR testleri esnasında alınan kan tahlil değerleri ile birlikte 35 değişkenden oluşan veri seti üzerinde makine öğrenmesi ve veri madenciliği yöntemleri kullanılarak regresyon ve sınıflandırma işlemleri gerçekleştirilmiştir. Modelin eğitilmesi ve testi amacıyla, veri setinin %70'i eğitim, %30'u test kümelerine bölünmüştür. Çalışmada hastalıkların doğru olarak teşhisi amacıyla kullanılan algoritmalarda %80'in üzerinde sınıflandırma başarısı elde edilmiştir. Bu durum, teşhis ve tedavi süreçlerinde insan kaynaklı hataları azaltmak ve medikal karar süreçlerine destek amacıyla, makine öğrenmesi tekniklerine başvurulabileceğini göstermektedir.

**Anahtar Sözcükler :** Sınıflandırma Regresyon, Covid-19, Makine Öğrenmesi

**Bilim Kodu** : 701.3.019

## **ABSTRACT**

**M. Sc. Thesis**

# **COVID-19 DISEASE PREDICTION WITH ARTIFICIAL INTELLIGENCE TECHNIQUES**

**Abdullah Türker TOKU**

**Karabük University  
Institute of Graduate Programs  
Department of Computer Engineering**

**Thesis Advisor:**

**Assist. Prof. Dr. Ferhat ATASOY**

**May 2023, 112 pages**

In this thesis, it is aimed to compare the success of machine learning techniques in the correct diagnosis of Covid-19 disease to ensure effectiveness in health management. The use of artificial intelligence and machine learning methods in the field of health management and services is increasing day by day. The Covid-19 virus is a type of virus belonging to a Sars group, which affects the whole world and presents with similar symptoms such as shortness of breath, cough, and fever in many patients. PCR tests are used in the diagnosis of the disease. In the conditions of the Covid-19 pandemic, there have been difficulties in accessing test kits even in the developed countries of the world. In addition, the fact that the accuracy rate of the test kits is not high has also constituted a separate issue. Therefore, from time to time, the treatment phase was started according to the symptoms of the disease. In order to contribute to the solution of the problem, the use of routine blood tests for the diagnosis of the Covid-19 virus constitutes the motivation of the study. As a result of the notch stroke

regression and classification processes were carried out using machine learning and data mining methods on the data set consisting of 35 variables, together with the blood analysis values taken during PCR tests in 1677 publicly available patients, to understand the prevalence of Covid-19 risk factors. For the purpose of training and testing the model, 70% of the dataset is divided into training and 30% test sets. In the study, over 80% classification success was achieved in the algorithms used for the correct diagnosis of diseases. This situation shows that machine learning techniques can be applied to reduce human-induced errors in diagnosis and treatment processes and to support medical decision processes.

**Key Word** : Classification Regression, Covid-19, Machine Learning

**Science Code** : 701.3.019



## TEŞEKKÜR

Bu tez çalışmasının planlanmasında emeđi geen, bilgisini ve desteđini eksik etmeyen sayın hocam Dr. Öğr. Üyesi Ferhat ATASOY' a sonsuz teşekkürlerimi sunarım.

Sayın jüri üyelerim ve ok deđerli akademisyenler Do. Dr. Abdullah ELEN ve Do. Dr. İlker TÜRKER' e, alışmalarım konusunda bilgi ve deneyimlerini hi bir suretle esirgmeden paylařmaları ve manevi desteklerini hi bir zaman eksik etmemelerinden dolayı sonsuz teşekkürler.

Manevi desteđini esirgemeyen, yeniliki yaklařımları, arařtırmacı kiřiliđi ve farklı bakıř aısıyla tüm alışmalarıma ıřık tutan deđerli akademisyen Do. Dr. Serhat Orkun TAN'a sonsuz teşekkürler.

Deđerli büyüklerim, ablalarım, öğretmenlerim ve en büyük destekilerim Öğretmen Fatma Tülay SOYSAL ve Dr. Teslime ÖZBAY' a hayatımın her ařamasında yanımda oldukları için sonsuz teşekkürler.

Sevgili aileme manevi hibir yardımı esirgmeden yanımda oldukları için tüm kalbimle teşekkür ederim.

## İÇİNDEKİLER

	<b><u>Sayfa</u></b>
KABUL.....	ii
ÖZET.....	iv
ABSTRACT.....	vi
TEŞEKKÜR.....	viii
İÇİNDEKİLER .....	ix
ŞEKİLLER DİZİNİ.....	xiv
ÇİZELGELER DİZİNİ .....	xv
SİMGELER VE KISALTMALAR.....	xvi
BÖLÜM 1 .....	1
GİRİŞ .....	1
1.1. TEZİN AMACI VE KAPSAMI.....	3
BÖLÜM 2 .....	6
COVID-19 HASTALIĞI .....	6
2.1. COVID-19 TEŞHİSİ .....	8
2.2. ÖNCEKİ ÇALIŞMALAR VE MOTİVASYON.....	10
2.3. COVID-19 RİSK FAKTÖRLERİ, ETKİNLİĞİ VE YAYGINLIĞI.....	14
2.3.1. Popülasyonlar ve Demografik Özellikler Arasındaki Farklılıklar.....	17
2.3.2. Epidemiyolojik Veriler ve Risk Faktörlerinin Dağılımı.....	17
2.4. COVID-19 TEŞHİSİ .....	21
2.4.1. COVID-19 Pandemisinde Risk Faktörlerinin Önemi.....	21
2.4.2. Yapay Zeka ve Veri Madenciliği Yöntemlerinin Teşhis Sürecine Katkısı .....	22
BÖLÜM 3 .....	25
VERİ SETİ VE KULLANILAN YÖNTEMLER.....	25
3.1. VERİ SETİ .....	25

## Sayfa

3.1.1. Veri Setinin Tanımı ve Kaynağı .....	25
3.1.2. Değişkenler ve Demografik Bilgiler.....	28
3.2. MAKİNE ÖĞRENMESİ YÖNTEMLERİ.....	30
3.2.1. Sınıflandırma .....	33
3.2.1.1. Çok Katmanlı Algılayıcı - RProp MLP Learner (ÇKA).....	33
3.2.1.2. Karar Ağaçları - Decision Trees (KA).....	33
3.2.1.3. Destek Vektör Makineleri (DVM).....	34
3.2.1.4. k-En Yakın Komşu (k-EYK) .....	34
3.2.1.5. Naïve Bayes (NB).....	35
3.2.1.6. Lojistik Regresyon (LR) .....	35
3.2.1.7. Rastgele Ormanlar - Random Forests (RO).....	35
3.2.1.8. Olasılık Sinir Ağı (OSA) .....	36
3.2.1.9. Yapay Sinir Ağları - Artificial Neural Networks (YSA).....	36
3.2.1.10. Uyarlanabilir Güçlendirme- Adaptive Boosting (AdaBoost) .....	37
3.2.1.11. Gradyan Yükseltme Makineleri - Gradient Boosting Machines (GYM).....	37
3.2.1.12. Aşırı Gradyan Artırma - eXtreme Gradient Boosting (XGBoost).....	38
3.2.1.13. Hafif Gradyan Artırma Makineleri - Light Gradient Boosting Machines (LightGBM) .....	38
3.2.1.14. CatBoost.....	39
3.2.2. Regresyon .....	39
3.2.2.1. Lineer Regresyon .....	39
3.2.2.2. Çoklu Regresyon.....	40
3.2.2.3. Polinom Regresyon.....	40
3.2.2.4. Destek Vektör Regresyonu (DVR).....	40
3.2.3. Kümeleme.....	41
3.2.3.1. k-Ortalamalar .....	41
3.2.3.2. k-Medoids .....	41
3.2.3.3. DBSCAN (Density Based Spatial Clustering of Applications With Noise).....	42
3.2.3.4. OPTICS (Ordering Points To Identify the Clustering Structure) ...	42
3.2.3.5. Yığılmalı Hiyerarşik Kümeleme (YHK) .....	43

	<b><u>Sayfa</u></b>
3.2.3.6. Bölücü Hiyerarşik Kümeleme (BHK) .....	43
3.2.3.7. Bulanık c-Ortalamlar (BcO).....	43
3.2.3.8. Spektral Kümeleme (SK).....	44
3.2.3.9. Balanced Iterative Reducing and Clustering Using Hierarchies Hiyerarşileri Kullanarak Dengeli Yinelemeli Azaltma ve Kümeleme (HKDYAK) .....	45
3.2.3.10. Ortalama Kaydırma (OK) .....	45
3.2.3.11. Affinity Propagation – Afinite Yayılımı (AY).....	46
3.2.3.12. Gaussian Mixture Models – Gauss Karışım Modeli (GKM).....	46
3.2.3.13. COBWEB .....	47
3.2.3.14. Self - Organizing Maps – Kendini Düzenleyen Algoritmalar (KDA) .....	47
 BÖLÜM 4 .....	 49
ANALİZ ORTAMI .....	49
4.1. KNIME YAZILIMI.....	49
4.1.1. Kronoloji.....	51
4.1.2. Özellikler .....	52
4.1.3. Çalışma Prensipleri .....	52
4.1.4. Kullanım Alanları .....	52
4.2. KNIME İLE VERİ ÖN İŞLEME, MODELLEME VE DEĞERLENDİRME .....	54
4.3. KNIME’ DA KULLANILAN ARAÇLAR.....	56
4.3.1. Veri Ön İşleme Araçları .....	56
4.3.1.1. Veri Seti ve Ön İşleme .....	56
4.3.1.2. Veri Temizleme, Eksik Veri Doldurma ve Dönüşümler .....	58
4.3.1.3. Model Geliştirme ve Optimizasyon .....	62
4.3.2. Kullanılan Algoritmalar ve Yöntemlerin Seçimi .....	64
4.3.2.1. Model Eğitimi ve Hiperparametre Optimizasyonu.....	66
4.3.2.2. Çapraz Doğrulama ve Model Seçimi .....	67
4.3.2.3. Model Değerlendirmesi ve Performans Metrikleri.....	68
4.3.2.4. ROC Eğrisi ve AUC Değerleri .....	70

	<b><u>Sayfa</u></b>
BÖLÜM 5 .....	72
UYGULAMA VE DENEYSEL ÇALIŞMALAR .....	72
5.1. VERİ ÖNİŞLEME .....	73
5.2. TEZİN ANA BULGULARI.....	76
5.3. KNIME VE MAKİNE ÖĞRENİMİ YÖNTEMLERİNİN COVID-19 RİSK FAKTÖRLERİNİ ANLAMADA SAĞLADIĞI AVANTAJLAR .....	78
5.4. YAPAY ZEKA UYGULAMALARININ PANDEMİ YÖNETİMİ VE TEŞHİS SÜREÇLERİNE ETKİSİ .....	79
BÖLÜM 6 .....	81
DENEYSEL SONUÇLAR, BULGULAR VE TARTIŞMA .....	81
6.1. YAPAY ZEKA VE VERİ MADENCİLİĞİ YÖNTEMLERİNİN COVID-19 TEŞHİSİNE KATKISI.....	81
6.2. TARTIŞMA VE GELECEK ÇALIŞMALAR VE ÖNERİLER .....	83
6.2.1. Yüksek Lisans Tezi Kapsamından Elde Edilen Başarılar ve Sınırlılıklar .....	83
6.2.2. Risk Faktörlerinin Yaygınlığını Anlamaya Yönelik Yapılan Çalışmaların Önemi.....	84
6.2.3. Gelecek Çalışmalar ve Potansiye İyileştirmeler İçin Öneriler .....	86
BÖLÜM 7 .....	89
SONUÇLAR .....	89
KAYNAKLAR .....	93
ÖZGEÇMİŞ .....	112

## ŞEKİLLER DİZİNİ

	<b><u>Sayfa</u></b>
Şekil 2.1. Covid-19 tanılı bir hastanın BT sonucu.....	12
Şekil 2.2. Günlük vaka oranları.....	18
Şekil 2.3. Günlük ölüm oranları.....	19
Şekil 2.4. Vaka ve Test oranlarının şekilsel gösterimi (İlk 20 Ülke) .....	21
Şekil 3.1. PCR test sonuçlarına göre covid olan hastaların dağılımı.....	26
Şekil 3.2. PCR, Covid ve Cinsiyet değerlerinin grafiksel dağılımı .....	27
Şekil 4.1. Knime yazılımı algoritma düğümleri.....	55
Şekil 4.2. Column Filter düğümü Dialog penceresi.....	58
Şekil 4.3. Missing Value düğümü Dialog penceresi .....	59
Şekil 4.4. Normalizer düğümü Dialog penceresi .....	60
Şekil 4.5. Parameter Oprimization düğümlerinin kullanım şekli.....	61
Şekil 4.6. Parameter Optimization Loop Start Düğümü Dialog Penceresi .....	61
Şekil 4.7. Parameter Optimization Loop End Düğümü Dialog Penceresi .....	62
Şekil 5.1. Knime Analytics Platformunda oluşturulan MÖ yapısı.....	74
Şekil 7.1. 10 özellikten oluşan veri setine ait Parallel Coordinates Plot eğrisi.....	89
Şekil 7.2. Yaş, Covid-19 ve Kan değerleri arasındaki ilişki .....	90

## ÇİZELGELER DİZİNİ

	<b><u>Sayfa</u></b>
Çizelge 2.1. Soğuk algınlığı, Grip ve Covid-19 belirtileri karşılaştırması.....	7
Çizelge 2.2. İlk 20 ülke ve nüfus sayılarına göre vaka oranları .....	18
Çizelge 2.3. İlk 20 ülke, nüfusa göre test sayıları ve vaka oranları. ....	20
Çizelge 3.1. Veri seti ve özellikleri.....	27
Çizelge 7.1. OSR Veri kümesi ve algorilma sonuçları. ....	91
Çizelge 7.2. CBC Veri kümesi ve algorilma sonuçları. ....	91
Çizelge 7.3. Doğrulama performansı için modellere ait en iyi algoritma sonuçları. .	92

## SİMGELER VE KISALTMALAR DİZİNİ

### KISALTMALAR

AIDS	: Acquired Immune Deficiency Syndrome (Kazanılmış Bağışıklık Yetersizliği Sendromu)
AdaBoost	: Adaptive Boosting (Uyarlanabilir Güçlendirme)
ANN	: Area Under the Curve (Eğrinin Altındaki Alan)
AUC	: Balanced Iterative Reducing and Clustering Using Hierarchies (Hiyerarşileri Kullanarak Dengeli Yinelemeli Azaltma ve Kümeleme)
BIRCH	: Balanced Iterative Reducing and Clustering Using Hierarchies (Hiyerarşileri Kullanarak Dengeli Yinelemeli Azaltma ve Kümeleme)
BT	: Bilgisayarlı Tomografi
cDNA	: Complementary Deoksiribo Nükleik Asit
CF	: Clustering Feature (Kümelenme Özelliği)
COBWEB	: Conceptual Clustering System CoV-2 (Kavramsal Kümeleme Sistemi CoV-2)
COVID-19	: Coronavirus Disease 2019
CT	: Cycle Threshold (Döngü Eşiği)
ÇKA	: Çok Katmanlı Algılayıcı
DBSCAN	: Density-Based Spatial Clustering of Applications with Noise (Gürültülü Uygulamaların Yoğunluk Tabanlı Mekansal Kümelenmesi)
DDA	: Dynamic Decay Adjustment (Dinamik Bozunma Ayarı)
DNA	: Deoksiribo Nükleik Asit
EM	: Expectation-Maximization (Beklenti Maksimizasyon)
F1	: Sınıflandırma Modelinin Performansını Ölçmek İçin Kullanılan Bir Metrik
GBM	: Gradient Boosting Machines (Gradyan Yükseltme Makinaları)
GMM	: Gaussian Mixture Models (Gauss Karışım Modelleri)
GYM	: Gradyan Yükseltme Makineleri
HIV	: Human İmmunodeficiency Virus (İnsan Bağışıklık Yetmezliği Virüsü)
IL-6	: İnterlökin-6



KA	: Karar Ağaçları
k-EYK	: k-En Yakın Komşu
k-Means	: K-Means Clustering
KNIME	: Konstanz Information Miner (Konstanz Bilgi Madenciliği)
LightGBM	: Light Gradient Boosting Machines (Hafif Gradyan Artırma Makineleri)
LR	: Lojistik Regresyon
MERS	: Middle East Respiratory Syndrome (Orta Doğu Solunum Sendromu)
MLP	: Multi-Layer Perceptron (Çok Katmanlı Algılayıcı)
MR	: Magnetic Resonance (Manyetik Rezonans)
NB	: Naïve Bayes
OPTICS	: Ordering Points To Identify the Clustering Structure (Kümelene Yapısını Belirlemek İçin Sıralama Noktaları)
OSA	: Olasılık Sinir Ağı
PCR	: Polymerase Chain Reaction (Polimeraz Zincir Reaksiyonu)
PNN	: Probabilistic Neural Network (Olasılıklı Sinir Ağı)
RNA	: Ribonükleik Asid
RO	: Rastgele Ormanlar
ROC	: Receiver Operating Characteristic Curve (Alıcı Çalışma Karakteristik Eğrisi)
RProp	: Rezidüel Propagasyon
RT	: Reverse Transcription (Ters Transkripsiyon)
SARS	: Şiddetli Akut Solunum Sendromu
SOM	: Self-Organizing Maps (Kendi Kendini Düzenleyen Haritalar)
SVM	: Support Vector Machine (Destek Vektör Makineleri)
SVR	: Support Vector Regression (Destek Vektör Regresyonu)
XGBoost	: eXtreme Gradient Boosting (Aşırı Degrade Artırma)
YHK	: Yığılmalı Hiyerarşik Kümeleme
YSA	: Yapay Sinir Ağları

## BÖLÜM 1

### GİRİŞ

Covid-19 pandemisi, dünya genelinde milyonlarca insanı etkileyen ve büyük bir sağlık krizine yol açan önemli bir küresel sorun olarak karşımıza çıkmaktadır. Bu nedenle, pandemi yönetimi ve erken teşhis süreçlerini hızlandırmak ve etkinleştirmek için bilimsel araştırmaların ve teknolojik gelişmelerin kullanılması büyük önem taşımaktadır. Covid-19 pandemisi süresince, sağlık sistemleri üzerindeki aşırı yük, test kapasitesinin sınırlı olması ve hastalığın yayılma hızı gibi zorluklar, teşhis ve tedavi süreçlerini zorlaştırmıştır. Bu nedenle, erken teşhis ve risk faktörlerini belirlemek, pandemi yönetimi ve hastaların tedavisi için büyük önem taşımaktadır. Bu bağlamda, yapay zeka ve veri madenciliği yöntemlerinin Covid-19 risk faktörlerini ve etkilerini daha iyi anlamada, teşhis ve tedavi süreçlerini hızlandırmada önemli bir rol oynadığı görülmektedir. Özellikle, makine öğrenimi ve veri madenciliği uygulamaları, hastalığın seyrini tahmin etme, risk gruplarını belirleme ve tedavi stratejilerini optimize etme gibi konularda katkı sağlamaktadır. Bu çalışma, bu alandaki mevcut bilgiyi ve teknolojileri kullanarak, Covid-19 risk faktörlerinin analizini ve teşhis süreçlerine daha fazla katkıda bulunmayı amaçlamaktadır. Covid-19 pandemisi süresince, literatürde birçok çalışma yapılmış ve bu çalışmalar, hastalığın teşhis ve tedavisinde yapay zeka ve veri madenciliği yöntemlerinin kullanılmasına odaklanmıştır. Özellikle, risk faktörlerinin belirlenmesi ve erken teşhisin önemine vurgu yapılan bu çalışmalar, yapay zeka ve veri madenciliği yöntemlerinin etkinliğini göstermektedir. Örneğin, Wang vd.'nin 2020 yılında yaptıkları bir çalışmada, Covid-19 hastalarının klinik verilerini kullanarak, hastalığın ciddiyetini ve sonucunu tahmin etmek için makine öğrenimi modellerini kullanmışlardır.

Bu tezde, makine öğrenimi modellerinin hastaların risk durumunu hızlı bir şekilde değerlendirmeye ve uygun tedavi stratejileri belirlemeye katkıda bulunduğunu göstermiştir. Başka bir çalışmada, Alimadadi vd. 2020 yılında Covid-19 teşhisinde yapay zeka ve veri madenciliği yöntemlerinin kullanılmasını incelemiş ve bu yöntemlerin hızlı ve doğru teşhis için büyük potansiyel taşıdığını belirtmiştir. Ozturk vd. ise 2020 yılında, tıbbi görüntüleme verilerini kullanarak Covid-19 vakalarını teşhis etmek için derin öğrenme modellerini uygulamış ve başarılı sonuçlar elde etmiştir. Bu bağlamda, yapay zeka ve veri madenciliği yöntemlerinin Covid-19 risk faktörlerini ve etkilerini daha iyi anlamada, teşhis ve tedavi süreçlerini hızlandırmada önemli bir rol oynadığı görülmektedir.

Bu tez çalışması, KNIME (Konstanz Information Miner) platformu üzerinde geliştirilen makine öğrenimi modelleri ve veri madenciliği yöntemleri ile Covid-19 risk faktörlerinin analizi ve teşhis süreçlerine katkısını incelemeyi amaçlamaktadır.

Tez çalışmasının başlangıcında, Covid-19 pandemisi konusu hakkında literatür taraması ve risk faktörlerinin önemi, etkileri ve yaygınlığı hakkında genel bilgiler verilmekte ve bu bilgiler ışığında, makine öğrenimi ve veri madenciliği yöntemlerinin pandemi yönetimi ve yapay zeka ve veri madenciliği yöntemlerinin teşhis süreçlerindeki önemi, risk faktörlerinin yaygınlığı ve etkileri konuları vurgulanmaktadır. Ardından, makine öğrenimi ve veri madenciliği yöntemleri, süreçleri, teknikleri konularına değinilmekte, bu çalışma için kullanılan KNIME platformunun temel özellikleri, avantajları ve kullanım alanları detaylı bir şekilde açıklanmaktadır.

Tezin ilerleyen bölümlerinde, veri seti ve ön işleme süreçleri ele alınarak, Covid-19 risk faktörlerini içeren veri setinin tanımı ve kaynağı, değişkenler ve demografik bilgiler, veri temizleme, eksik veri doldurma ve dönüşümler gibi önemli konular üzerinde durulmaktadır. Bu aşamalar, KNIME platformu üzerinde gerçekleştirilen işlemler ve kullanılan araçlarla açıklanarak, süreçlerin nasıl yapıldığına dair bilgi verilmektedir.

Model geliştirme ve optimizasyon süreçlerinde, kullanılan algoritmalar ve yöntemlerin seçimi, model eğitimi ve hiperparametre optimizasyonu, çapraz doğrulama ve model seçimi gibi önemli konular ele alınmaktadır. Bu süreçlerin her birinde KNIME platformu üzerinde gerçekleştirilen işlemler ve kullanılan araçlarla detaylı açıklamalar yapılmaktadır.

Model değerlendirmesi ve performans metrikleri bölümünde, doğruluk, duyarlılık, özgüllük, F1 skoru, ROC eğrisi ve AUC değerleri gibi önemli metrikler kullanılarak geliştirilen modellerin başarıları ve performansları incelenmektedir. Bu metrikler sayesinde, modellerin Covid-19 teşhisinde ne kadar doğru ve tutarlı tahminlerde bulunduğu değerlendirilmektedir. Ayrıca, yapılan analizler ve değerlendirmeler sonucunda, modellerin güçlü ve zayıf yönleri belirlenerek, gelecekte yapılacak çalışmalarda alınabilecek önlemler ve iyileştirme önerileri sunulmaktadır.

Tartışma ve gelecek çalışmalar bölümünde, tez kapsamında elde edilen başarılar ve sınırlılıklar, risk faktörlerinin yaygınlığını anlamaya yönelik yapılan çalışmaların önemi ve potansiyel iyileştirmeler için öneriler ele alınmaktadır. Bu kısımda, tezde kullanılan yöntemlerin ve KNIME platformunun sağladığı avantajlar ve sınırlılıklar üzerinde durulmakta, mevcut çalışmaların ve yöntemlerin değerlendirilmesi yapılmaktadır.

Sonuç olarak, bu tez çalışması, KNIME platformu üzerinde geliştirilen sınıflandırma modellerinin Covid-19 risk faktörlerinin analizi ve teşhis süreçlerine önemli katkılar sağladığını ortaya koymaktadır. Ayrıca, bu yöntemlerin ve araçların kullanılması, pandemi yönetimi ve erken teşhis süreçlerini hızlandırarak, daha etkili ve verimli stratejilerin belirlenmesine katkıda bulunmaktadır. Bu nedenle, bu alandaki gelecek çalışmaların ve uygulamaların bu tür yöntemler ve teknolojiler üzerinde yoğunlaşması büyük önem taşımaktadır.

## **1.1. TEZİN AMACI VE KAPSAMI**

Bu tezin temel amacı, Covid-19 hastalığının teşhisi ve risk faktörlerinin belirlenmesinde yapay zeka ve veri madenciliği yöntemlerinin etkinliğinin

incelenmesi ve bu yöntemlerle elde edilen sonuçların değerlendirilmesini sağlayarak pandemiyle mücadelede daha etkili, hızlı ve doğru sonuçlar alınmasına ve toplum sağlığının korunmasına katkıda bulunulmasıdır. Bu amaç doğrultusunda, tez kapsamında gerçekleştirilen çalışmalar ve analizler aşağıdaki başlıklar altında toplanmıştır:

1. Mevcut literatürün incelenmesi: Covid-19 pandemisi, risk faktörleri, teşhis yöntemleri ve yapay zeka tekniklerinin kullanımıyla ilgili mevcut literatür incelenerek, tezin temelini oluşturan bilgi birikimi sağlanmıştır.
2. Veri setinin temini ve ön işleme: Kamuya açık, 1677 hastada PCR testleri esnasında alınan kan tahlil değerleri ile birlikte 35 değişkenden oluşan veri seti temin edilmiş ve veri madenciliği çalışmaları için ön işleme süreçleri gerçekleştirmiştir.
3. Öğrenme modellerinin seçimi ve uygulanması: Covid-19 teşhisi ve risk faktörlerinin belirlenmesinde kullanılacak makine öğrenmesi yöntemleri belirlenerek, KNIME programında çalışma yapılmıştır.
4. Model performansının değerlendirilmesi: Yapay zeka ve veri madenciliği yöntemleri ile elde edilen sonuçların performansı, doğruluk, hassasiyet, duyarlılık ve F1 skoru gibi metrikler kullanılarak değerlendirilmiştir.
5. Risk faktörleri ve hastalık ilişkilerinin analizi: Elde edilen sonuçlar ve modeller kullanılarak, Covid-19 hastalığı ile risk faktörleri arasındaki ilişkiler analiz edilmiş ve önemli bulgular ortaya konulmuştur.
6. Sonuçların değerlendirilmesi ve öneriler: Yapılan çalışmalar ve analizler sonucunda elde edilen bulgular değerlendirilerek, Covid-19 teşhisi ve risk faktörlerinin belirlenmesinde yapay zeka ve veri madenciliği yöntemlerinin potansiyel katkılarına ilişkin öneriler sunulmuştur.

7. Yapay zeka ve veri madenciliği yöntemlerinin sınırlılıkları: Her ne kadar yapay zeka ve veri madenciliği yöntemleri önemli avantajlar sunsa da, bu tekniklerin sınırlılıklarının da göz önünde bulundurulması önemlidir. Tezde, kullanılan yöntemlerin sınırlılıkları ve potansiyel zorluklarının analizi yapılarak, gelecekte bu alanlarda yapılabilecek çalışmalar için öneriler sunulmuştur.
8. Etik ve güvenlik konuları: Yapay zeka ve veri madenciliği yöntemleri kullanılırken, hastaların gizliliği ve veri güvenliği gibi etik ve güvenlik konularına özen gösterilmiştir. Tezde, bu konulara dikkat çekilerek, yapay zeka ve veri madenciliği uygulamalarının etik ve güvenlik yönünden nasıl yönetilebileceği üzerine düşünceler sunulmuştur.
9. Uygulama ve politika önerileri: Tezin sonunda, elde edilen bulgular ve değerlendirmeler doğrultusunda, Covid-19 teşhisi ve risk faktörlerinin belirlenmesinde yapay zeka ve veri madenciliği yöntemlerinin kullanılmasına yönelik uygulama ve politika önerileri geliştirilmiştir. Bu öneriler, sağlık hizmeti sunan kurumlar ve karar vericiler için yol gösterici olabilir.

## BÖLÜM 2

### COVID-19 HASTALIĞI

Covid-19, 2019 yılında Çin'in Wuhan şehrinde ortaya çıkan ve kısa süre içerisinde dünya genelinde pandemiye yol açan bir viral enfeksiyon hastalığı olarak tanımlanmaktadır [1,5]. Covid-19'a neden olan virüs, SARS-CoV-2 olarak adlandırılmakta ve insanlar arasında solunum yoluyla kolayca yayılabilmektedir [6,9]. Sürecin ilk aşamasında, Aralık 2019'da Wuhan'daki deniz ürünleri ve canlı hayvan pazarında ortaya çıkan solunum yolu hastalığı vakaları, Çin hükümeti ve sağlık yetkililerinin dikkatini çekmiştir [10,13]. İlk etapta, bu vakaların nedeni bilinmemekteydi ve SARS ve MERS gibi daha önce bilinen koronavirüslerin neden olup olmadığı araştırılmıştır [14].

Ocak 2020'de, Çinli bilim insanları SARS-CoV-2'nin genetik dizisini yayınlanmış ve bu yeni koronavirüsün diğer koronavirüslerle ilişkili olduğunu, ancak daha önce hiçbir insanda görülmemiş bir tür olduğunu açıklamıştır [15,18]. Dünya Sağlık Örgütü (World Health Organization WHO) ve diğer uluslararası sağlık kuruluşları, SARS-CoV-2'nin yayılmasını ve Covid-19 hastalığının etkilerini izlemeye ve değerlendirmeye başlamışlardır [19,20].

Covid-19'un başlangıç belirtileri genellikle ateş, öksürük (kuru veya balgamlı), halsizlik, nefes darlığı, kas ve eklem ağrıları, boğaz ağrısı, baş ağrısı olarak sıralanabilmektedir [21,24]. Bu belirtiler soğuk algınlığı ve grip ile benzerlik göstermektedir ve belirtilerin soğuk algınlığı, grip ve Covid-19 için karşılaştırması Çizelge 2.1'de verilmiştir.

Çizelge 2.1. Soğuk algınlığı, Grip ve Covid-19 belirtileri karşılaştırması

<b>Belirti</b>	<b>Soğuk Algınlığı</b>	<b>Grip</b>	<b>Covid-19</b>
Ateş	Nadir	Yaygın	Yaygın
Baş Ağrısı	Nadir	Yaygın	Az Yaygın
Bitkinlik	Az Yaygın	Yaygın	Yaygın
Boğaz Ağrısı	Yaygın	Az Yaygın	Az Yaygın
Bulantı	Nadir	Az Yaygın	Nadir
Burun Akıntısı	Nadir	Az Yaygın	Nadir
Burun Tıkanıklığı	Yaygın	Az Yaygın	Nadir
Gözlerde Kaşıntı	-	-	Nadir
Gözlerde Sulanma	-	-	Nadir
Hapşırma	Yaygın	Az Yaygın	Nadir
İshal	Nadir	Az Yaygın	Az Yaygın
Koku Kaybı	Nadir	Nadir	Az Yaygın
Kusma	Nadir	Az Yaygın	Nadir
Nefes Darlığı	Nadir	Nadir	Yaygın
Tat Kaybı	Nadir	Nadir	Az Yaygın

Bu belirtiler grip ve soğuk algınlığı gibi diğer solunum yolu enfeksiyonlarına benzerlik göstermesinden dolayı, Covid-19'un erken dönemde tespit edilmesi konusunda zorluklar yaşanmıştır [25,26]. Hastalığın ilerleyen dönemlerinde, hastaların bazılarında zatürre, akut solunum sıkıntısı sendromu [27] ve organ yetmezliği gibi daha ciddi komplikasyonlar gelişebileceği gözlemlenmiştir. Ayrıca, bazı hastaların tat ve koku alma yetisinde geçici bir kayıp yaşaması gibi daha az yaygın semptomlar da bildirilmiştir [28,31]. Covid-19 pandemisi süresince, bilim insanları ve sağlık yetkilileri, SARS-CoV-2'nin yayılmasını önlemek için sosyal mesafe, maske kullanımı ve sıkı hijyen önlemleri gibi çeşitli tedbirler önermiştir [32,36]. Ayrıca, Covid-19 aşılı hızla geliştirilmiş ve dünya çapında dağıtılmıştır [37,38], bu sayede hastalığın yayılmasının önlenmesi ve ciddi hastalık vakalarının azaltılması amaçlanmıştır [39,41]. Pandemi boyunca, hükümetler ve sağlık yetkilileri dünya genelinde toplumları bilgilendirmek ve korumak için hızlı ve etkili bir şekilde hareket etmeye



çalışmışlardır. Bu süre zarfında, Covid-19'un yayılmasını önlemeye yönelik önlemler, kısıtlamalar ve rehberlikler sürekli güncellenmiştir [42,47].

Bilim insanları ayrıca, SARS-CoV-2'nin değişikliklere uğrayarak yeni varyantlar oluşturabileceğini keşfetmişlerdir [48,51]. Bu varyantlar, virüsün bulaşıcılığı ve hastalık şiddeti üzerinde etkili olabilmektedir. Bazı durumlarda, varyantlar mevcut aşular ve tedaviler üzerinde etkili olabilmektedir ve pandemi yönetimi açısından yeni zorluklar ortaya çıkarabilmektedir [52,54]. Bu nedenle, bilim insanları ve sağlık yetkilileri, SARS-CoV-2 varyantlarını yakından izlemişler ve gerektiğinde önlemleri ve tedavi yöntemlerini güncellemişlerdir [55,58]. Covid-19 pandemisi sırasında, halk sağlığı ve tıbbi araştırmalar önemli ölçüde hızlandırılmış ve küresel işbirliği artmıştır [59,62]. Bu süreçte elde edilen bilgiler ve deneyimler, gelecekteki pandemilere ve sağlık krizlerine karşı daha hazırlıklı olmamıza yardımcı olabilecektir [63,67].

## **2.1. COVID-19 TEŞHİSİ**

Covid-19'un ilk çıktığı dönemlerde, hastalığın teşhisi genellikle şüpheli vakaların klinik semptomları ve seyahat geçmişi gibi faktörler dikkate alınarak konulmuştur [68,72]. Bu erken dönemde, kesin teşhis için özellikle RT-PCR (Reverse Transcription-Polymerase Chain Reaction) testleri kullanılmıştır [73,74]. Bu test, virüsün genetik materyalini (RNA) tespit etmek için kullanılmaktadır [75,76] ve hala Covid-19 teşhisinde altın standart olarak kabul edilmektedir [77,78]. İlerleyen zamanlarda, daha hızlı ve uygun maliyetli testler geliştirilerek hastalığın teşhis süreci geliştirilmiştir.

Bu testlerden bazıları şunlardır:

1. Antijen testleri: Bu testler, virüsün belirli proteinlerini tespit ederek ve genellikle hızlı sonuçlar vermektedir. Ancak antijen testlerinin duyarlılığı, RT-PCR testine göre daha düşüktür ve bazen yanlış negatif sonuçlar verebilmektedir [79,80].
2. Serolojik testler (antikor testleri): Bu testler, Covid-19'a karşı vücudun ürettiği antikorları tespit etmektedir. Serolojik testler, geçirilmiş bir enfeksiyonun ya da aşı sonrası bağışıklığın belirlenmesinde kullanılmaktadır. Bu testler, aktif

enfeksiyon teşhisi için kullanılmaması önerilmektedir, çünkü antikorlar enfeksiyonun başlamasından bir süre sonra ortaya çıkmaktadır [81,82].

3. CT (Cycle Threshold) değeri: RT-PCR testleri ile birlikte kullanılan bu değer, virüsün yüksek miktarda RNA'sının ne kadar hızlı tespit edildiğine dair bilgi sağlamaktadır. Düşük Ct değeri, yüksek viral yük ve daha bulaşıcı bir enfeksiyonu işaret edebilir [83,84].

Covid-19 teşhisi için kullanılan yöntemler zaman içinde geliştirilmiştir ve yeni yöntemler geliştirilmeye devam etmektedir. Bu, hastalığın hızlı ve doğru bir şekilde teşhis edilmesine ve etkili önlemlerin alınmasına yardımcı olmaktadır.

Covid-19 hastalığının teşhisi için altın standardı olarak kullanılan PCR testi, genetik materyalin belirli bir bölümünü hızlı ve doğru bir şekilde çoğaltan moleküler bir tekniktir. 1983 yılında Kary Mullis tarafından icat edilmiştir ve çeşitli uygulamalarda örneğin; genetik araştırmalar, tıbbi teşhis, suç mahalli analizi ve v.b. kullanılır [85]. COVID-19 pandemisinde, PCR testi, SARS-CoV-2 virüsünün genetik materyalini tespit etmek için yaygın olarak hastaların enfekte olup olmadığını belirlemek için kullanılır ve bu nedenle virüsün yayılmasını kontrol etmek ve önlemek için önemlidir [86].

Covid-PCR ilişkisi şu şekilde ilerlemektedir:

Örnek toplama: Test, burun ve/veya boğaz sürüntüsü ile gerçekleştirilir [87]. Bu örnekler, virüsün üst solunum yollarında var olabileceği için seçilir [88].

1. RNA izolasyonu: Örnekler laboratuvara gönderilir ve burada virüsün RNA'sı diğer hücrel bileşenlerden izole edilir [89].
2. RNA dönüşümü: SARS-CoV-2 RNA'sı, DNA'ya (cDNA) dönüştürülür [90]. Bu, bir ters transkriptaz enzimi kullanılarak gerçekleştirilir [91].
3. PCR amplifikasyonu: cDNA, özel başlangıçlar ve bir DNA polimeraz enzimi kullanılarak çoğaltılır [92]. Başlangıçlar, SARS-CoV-2 virüsüne özgü genetik

bölgelere hedeflenir. PCR işlemi, genetik materyalin milyonlarca kopyasını üretir [93].

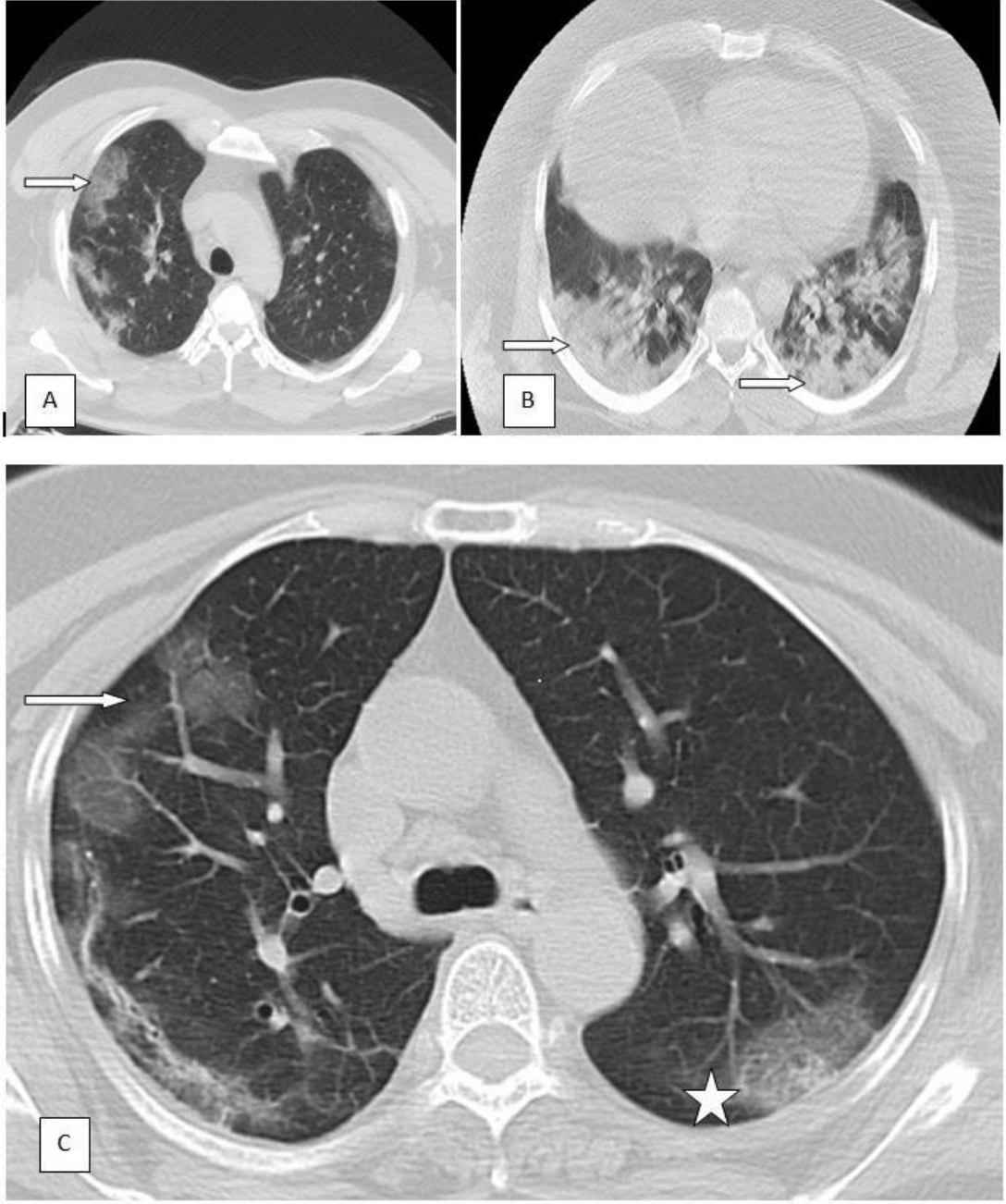
4. Tespit ve sonuç: PCR testi, belirli bir eşik değerine ulaşıldığında sinyal üreten floresan işaretli prob kullanır [94]. Bu sinyal, testin pozitif olduğunu ve virüsün genetik materyalinin örnekte bulunduğunu gösterir. Eğer sinyal tespit edilmezse, test negatif olarak değerlendirilir. PCR testi, COVID-19'un teşhisinde yüksek hassasiyet ve spesifisiteye sahip olduğu için güvenilir bir yöntem olarak kabul edilmektedir [95]. Ancak, örnek toplama ve işleme sırasındaki hatalar ve testin zamanlaması gibi faktörler, test sonuçlarının doğruluğunu etkileyebilmektedir [96,98]. Lippi vd. gerçek zamanlı PCR testlerinin potansiyel pre-analitik ve analitik zafiyetlerini incelemişlerdir. Kimlik eksikliği, numune alma ve işleme için yetersiz prosedürler, kalite ve hacim için toplanan malzemenin yetersizliği ve uygunsuzluğu, girişim yapan maddelerin olması, pipetleme hataları vb. pre-analitik zafiyetler olduğunu ortaya koymuşlardır. Ayrıca bu pre-analitik zafiyetler için, gerçek zamanlı PCR testinin doğruluğu ve özellikle duyarlılığı ideal olmaktan uzaktır. Bu durum ise pandemi sürecinde ihtiyaç duyulan doktor, itfaiyeci, polis vb. kalifiye meslek gruplarında çalışanların yükünü arttırdığı gibi virüsü bulaştırma potansiyeli olan hastaların toplum içinde serbestçe dolaşmasına sebep olmaktadır [99]. Woloshin vd. yaptıkları çalışmada, PCR testinin duyarlılığı için makul bir tahminin %70 olduğu belirtilmektedir [100].

## 2.2. ÖNCEKİ ÇALIŞMALAR VE MOTİVASYON

Covid-19 virüsü havadan bulaşabildiği gibi yakın temas ile de çok hızlı yayılabilmektedir. Test sonucu pozitif olan hastalar için izolasyon süreci başlamaktadır. Karantina süreci de hastaların sosyal, ekonomik ve mental durumlarını olumsuz yönde etkilemektedir. Bundan dolayı hastalığın doğru teşhisi ciddi önem kazanmaktadır. Wynants vd. 2020 yılında yayınladıkları çalışmada Covid-19 hastalığı üzerine yapılan tahmin modeli çalışmalarını incelemişlerdir [101]. Çalışmaya göre literatürdeki tahminleme yapan modellerin kontrol grubu olmaması ve güvenilirlik riskinin yüksek olmasından dolayı tıbbi teşhiste doğrudan kullanılamayacaklarını rapor etmişlerdir. Ayrıca literatürde sunulan modellerin doğrulanması için yeni

verilerle desteklenmesi gerektiğini belirtmişlerdir [102]. Önerilen modellerin çoğunluğu Bilgisayarlı Tomografi (BT) taramalarına veya göğüs röntgenlerine [102,106] dayanmaktadır. Bildirilen umut verici sonuçlara rağmen, özellikle yüksek oranda yanlış negatif sonuçlarla ilişkilendirilen göğüs röntgenine dayalı çözümlerle ilgili olarak, bu ve diğer çalışmalarla ilgili yanlış sonuçlar ile ilgili endişeler dile getirilmiştir [107]. BT görüntülemeye dayalı çözümler, doğru olmasına rağmen, uygulanan modelin özelliklerinden etkilenmektedir. BT'lerin maliyetli olması, hızlı sonuç verememesi, özel ekipman ve donanımlara ihtiyaç duymasından dolayı, görüntüleme tekniğine dayalı çözümler tarama muayeneleri için uygun bir çözüm şekli değildir.

Virüsün ciğerlerdeki etkilerinin oluşmasından önce, BT görüntüleme ile erken teşhis ihtimalinin düşük olması çalışmanın odağını kan tahlili verileri üzerine yönlendirmiştir. Çeşitli klinik çalışmalar [108,110] kan testi tabanlı tanılamamın COVID-19 vakalarının erken tespiti için etkili ve düşük maliyetli bir alternatif sağlayabileceğini vurgulasa da literatürde makine öğrenmesi yöntemlerinin uygulandığı az sayıda çalışmaya rastlanmıştır [111,114]. Şekil 2.1'de Covid-19 tanılı bir hastanın BT sonucu verilmiştir.



Şekil 2.1. Covid-19 tanılı bir hastanın BT sonucu.

Şekil 2.1’de BT sonucuna göre, a) Sağ akciğerde periferde buzlu cam sahası, b) Bilateral akciğer parankiminde posterio bazalde konsolidasyon sahası, c) Sol akciğerde posterior da santrali konsolide etrafı buzlu cam sahası olan iki bulgunun birlikte olduğu görünüm(yıldız), ayrıca sağda buzlu cam sahası (ok) [115].

Yapay zeka uygulamalarının temelinde istatistik, makine öğrenimi, veri tabanı mimarileri çok önemli bir yere sahiptir. İstatistik, verilerden anlamlı sonuçlar

çıkarılması konusunda aktif şekilde kullanılan bir disiplindir. Yazılım mimarilerinde geliştirilen algoritmalar ve makine öğrenme yöntemleri doğrudan istatistik alanına yön vermiştir. İstatistiksel yöntemlerin belirlenmesinde, makine öğrenme algoritmalarının kullanılması daha önce gerçekleştirilmesi çok karmaşık olan istatistiksel çıkarımları yapılabilir duruma getirmiştir [116].

MÖ yöntemleri, sağlık alanında yapılan çalışmalarda büyük öneme sahiptir çünkü aşağıdaki nedenlerle tanı, teşhis ve tedavi süreçlerine önemli katkılar sağlamaktadır [117]

1. Büyük veri analizi: Sağlık alanında toplanan veriler büyük boyutlu ve karmaşıktır. MÖ algoritmaları, bu büyük veri kümesini analiz ederek anlamlı bilgiler elde etmeye yardımcı olmaktadır [118].
2. Kesinlik ve doğruluk: MÖ modelleri, eğitim verileri üzerinde öğrenerek ve genelleme yaparak doğru tahminlerde bulunabilir. Bu, tanı ve teşhis süreçlerinde doğru kararlar almayı sağlar ve tedavi başarısını arttırmaktadır [119].
3. Hızlı ve etkin karar verme: MÖ algoritmaları, büyük miktarda veriyi hızlı bir şekilde analiz ederek anlamlı sonuçlar üretir. Bu, sağlık profesyonellerinin zamanında ve etkin kararlar almasına olanak tanımaktadır [120].
4. Kişiselleştirilmiş tedavi: MÖ, hasta verilerini analiz ederek bireysel hastaların özelliklerine göre en uygun tedavi seçeneklerini belirlemeye yardımcı olmaktadır. Bu, daha etkili ve kişiselleştirilmiş tedavi süreçlerine yol açmaktadır [121].
5. Erken teşhis: MÖ, hastalıkların erken teşhisine yardımcı olabilir. Erken teşhis, tedavinin başarı şansını artırır ve hastaların yaşam kalitesini yükseltir [122].
6. Yeni tedavi yöntemlerinin keşfi: MÖ algoritmaları, tıbbi araştırmalarda yeni tedavi yöntemlerinin keşfi ve geliştirilmesine katkıda bulunabilir [123].

7. Maliyet azaltma: MÖ, tanı ve tedavi süreçlerini optimize ederek sağlık hizmetlerinin maliyetini azaltabilir. Bu, sağlık hizmetlerine daha fazla insanın erişebilmesine olanak tanıyabilmektedir [124].
8. Epidemiyoloji ve halk sağlığı: MÖ, salgın hastalıkların yayılımını ve etkisini tahmin etmeye ve yönetmeye yardımcı olabilir. Bu, halk sağlığı politikalarının ve müdahalelerinin etkinliğini artırır [125].

Bu nedenlerle, MÖ yöntemlerinin sağlık alanında yapılan çalışmalara katkısı büyük öneme sahiptir. Bu teknikler, verilerden anlamlı sonuçlar elde etmeye, tanı ve teşhis süreçlerini iyileştirmeye ve tedavi başarısını artırmaya yardımcı olabilir.

Bu motivasyonla sunulan çalışmada Covid-19 hastalığının teşhisinde kan tahlili verileri ile hastalık teşhisini iyileştirmeye odaklanılmıştır. Elde edilen sonuçlar umut vericidir. Veri seti gerçek verilerle zenginleştirildiğinde sınıflandırma performansının artacağı düşünülmekle birlikte pandeminin etkilerini azaltmak ve hastalığın yayılmasını kontrol altına almak amacıyla risk faktörlerinin belirlenmesi ve anlaşılması büyük önem taşımaktadır.

Sonuç olarak, Covid-19 pandemisinin başlangıcından bugüne kadar süreç, bilim insanları, sağlık yetkilileri ve toplumlar için sürekli öğrenme ve uyum sağlama süreci olmuştur. SARS-CoV-2 virüsünün yayılmasını önlemek ve hastalığın etkilerini azaltmak için önlemler, stratejiler ve tedaviler sürekli güncellenmektedir. Gelecekteki sağlık krizleri ve pandemilere karşı daha dirençli olmak için, bu süre zarfında edinilen bilgiler ve deneyimler büyük önem taşımaktadır.

### **2.3. COVID-19 RİSK FAKTÖRLERİ, ETKİNLİĞİ VE YAYGINLIĞI**

Covid-19 pandemisi, dünya genelinde milyonlarca insanı etkileyen ve önemli sağlık sorunlarına yol açan bir durumdur. Hastalığın yaygınlığı ve ciddiyeti, bazı risk faktörleri ile yakından ilişkilidir. Bu risk faktörleri, Covid-19'un şiddetli seyretme ihtimalini ve hastalığın yayılma hızını etkileyen faktörlerdir. Covid-19 risk faktörleri şu şekilde sıralanabilir:

1. Yaş: Yaşlı bireyler, Covid-19 enfeksiyonundan daha fazla etkilenme riski taşır ve hastalığın daha şiddetli seyretme ihtimali daha yüksektir. 65 yaş ve üzeri bireyler, gençler ve yetişkinlere göre daha yüksek risk altındadır.
2. Cinsiyet: Erkeklerin, kadınlara göre Covid-19 enfeksiyonundan daha fazla etkilendiği ve daha şiddetli semptomlar yaşama ihtimalinin daha yüksek olduğu gözlemlenmiştir.
3. Altta yatan sağlık sorunları: Diyabet, hipertansiyon, kalp hastalıkları, kronik böbrek hastalığı, karaciğer hastalıkları ve kronik akciğer hastalıkları gibi altta yatan sağlık sorunları, Covid-19 enfeksiyonundan daha fazla etkilenme ve hastalığın daha şiddetli seyretme riskini artırır.
4. Obezite: Vücut kitle indeksi (BMI) 30 ve üzeri olan bireylerin, Covid-19 enfeksiyonundan daha fazla etkilendiği ve hastalığın daha şiddetli seyretme ihtimalinin daha yüksek olduğu bilinmektedir. Obezite, bağışıklık sistemi üzerinde olumsuz etkiler yaparak, vücudun enfeksiyonla mücadele kapasitesini azaltabilir.
5. Bağışıklık sistemi durumu: Kanser tedavisi görenler, organ nakli geçirenler, HIV (Human İmmunodeficiency Virus - İnsan Bağışıklık Yetmezliği Virüsü)/AIDS (Acquired Immune Deficiency Syndrome Kazanılmış Bağışıklık Yetersizliği Sendromu) hastaları gibi bağışıklık sistemi zayıf olan bireyler, Covid-19 enfeksiyonundan daha fazla etkilenme ve hastalığın daha şiddetli seyretme riski taşır. Zayıf bağışıklık sistemi, enfeksiyonla mücadelede yetersiz kalarak, hastalığın daha kötü sonuçlar doğurmasına yol açabilir.
6. Sosyoekonomik durum ve yaşam koşulları: Düşük sosyoekonomik düzey, kalabalık yaşam koşulları ve sağlık hizmetlerine erişimde sınırlılıklar, Covid-19 yaygınlığı ve şiddetini etkileyen faktörlerdir. Kalabalık yaşam alanlarında sosyal mesafe kurallarını uygulamak zorlaşabilir ve bu durum, enfeksiyonun yayılmasını kolaylaştırabilir. Ayrıca, düşük sosyoekonomik durum, sağlık



hizmetlerine erişimi zorlaştırarak hastalığın erken teşhis ve tedavisini engelleyebilir.

7. Kronik hastalıklar: Diyabet, kalp hastalıkları, hipertansiyon, kronik böbrek hastalığı, karaciğer hastalıkları ve kronik akciğer hastalıkları gibi kronik sağlık sorunları olan bireylerde, Covid-19'un daha şiddetli seyretme ihtimali yüksektir. Bu durum, bağışıklık sisteminin zayıflamasına ve vücudun enfeksiyona karşı savunmasının azalmasına bağlıdır.
8. Sigara kullanımı: Sigara kullanımı, solunum yolu hastalıkları için önemli bir risk faktörüdür ve Covid-19 enfeksiyonundan daha fazla etkilenme ve hastalığın daha şiddetli seyretme riskini artırabilir. Sigara, solunum yolu hücrelerine zarar vererek virüsün hücrelere girişini kolaylaştırır ve bağışıklık sisteminin işlevini azaltabilir.
9. Etnik köken: Bazı etnik kökenlere sahip bireyler, genetik ve sosyoekonomik faktörler nedeniyle Covid-19 enfeksiyonundan daha fazla etkilenme riski taşıyabilir. Özellikle, bazı etnik gruplarda kronik sağlık sorunlarının daha yüksek prevalansı, enfeksiyonun daha şiddetli seyretme ihtimalini artırabilir.

Bu faktörlerin bilinmesi ve anlaşılması, Covid-19 pandemisiyle mücadelede önemli bir rol oynar. Risk faktörlerini ve etkilerini belirleyerek, öncelikli risk gruplarına yönelik önlemler alınabilir, farkındalık artırılabilir ve halk sağlığı politikaları daha etkili bir şekilde tasarlanabilir. Yapay zeka ve veri madenciliği yöntemleri, bu risk faktörlerini analiz etmek ve daha doğru sonuçlar elde etmek için kullanılabilir. Pandemi sürecinde hızlı ve doğru kararlar almak için önemli bir araç olarak kullanılabilir ve sağlık hizmetlerinin yönetimine büyük katkılar sunabilir.

Yapay zeka ve veri madenciliği yöntemleri kullanılarak, risk faktörlerinin belirlenmesi ve analizi, Covid-19'un yaygınlığı ve etkileri üzerinde daha iyi bir anlayış sağlar. Bu sayede, öncelikli risk gruplarına yönelik önlemler alınabilir, farkındalık artırılabilir ve halk sağlığı politikaları daha etkili bir şekilde tasarlanabilir. Bu tür analizler, pandemi

sürecinde hızlı ve doğru kararlar almak için önemli bir araç olarak kullanılabilir ve sağlık hizmetlerinin yönetimine büyük katkılar sunabilir.

### **2.3.1. Popülasyonlar ve Demografik Özellikler Arasındaki Farklılıklar**

Covid-19 pandemisinin etkileri, farklı popülasyonlar ve demografik özellikler arasında değişiklik göstermektedir. Bu farklılıklar, hastalığın yayılma hızını, şiddetini ve toplumların pandemiye karşı aldığı önlemleri etkileyebilir.

Yapay zeka ve veri madenciliği yöntemleri kullanılarak, bu demografik özellikler ve popülasyonlar arasındaki farklılıkları belirlemek ve analiz etmek mümkündür. Bu analizler, hükümetler ve sağlık hizmeti sağlayıcıları için önemli bilgiler sunarak, hedef odaklı önlemler alınmasına ve kaynakların etkin bir şekilde kullanılmasına yardımcı olabilir. Farklı demografik özelliklere sahip popülasyonlar arasındaki farklılıkları anlamak, salgınla mücadelede etkin stratejiler geliştirilmesine ve pandemiye karşı daha etkili politikalar üretilmesine katkı sağlar. Yapay zeka ve veri madenciliği yöntemleri, bu konuda hızlı ve doğru analizler yaparak, pandemi sürecinde bilinçli ve etkili kararlar almak için önemli bir araç olarak kullanılabilir.

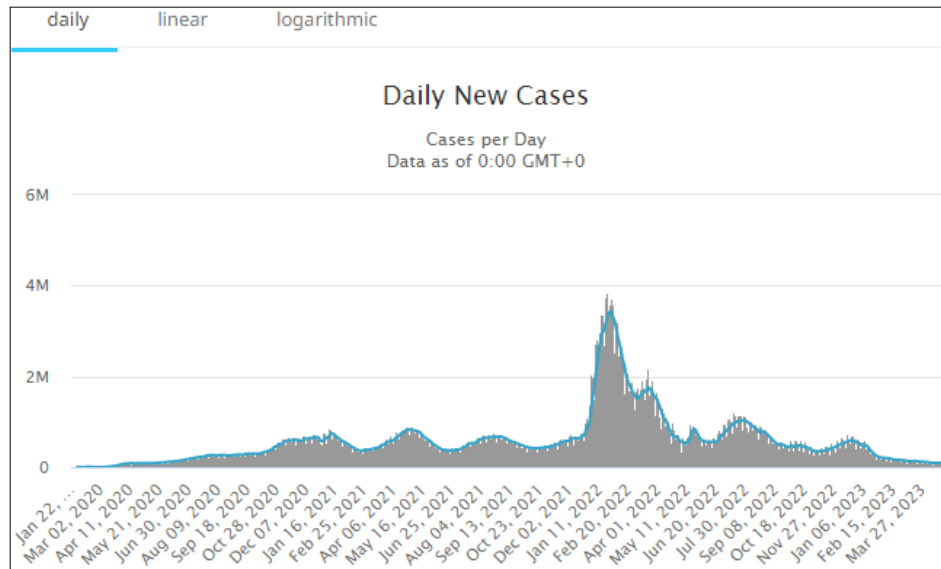
### **2.3.2. Epidemiyolojik Veriler ve Risk Faktörlerinin Dağılımı**

Covid-19 pandemisinin etkilerini ve risk faktörlerinin dağılımını anlamak için, epidemiyolojik verilere dayalı analizler yapılmaktadır. Epidemiyolojik veriler, hastalığın yayılımı, risk faktörlerinin prevalansı ve demografik özelliklerle ilişkisi üzerine bilgi sağlar. Bu verilerin analizi, pandemiye karşı alınacak önlemlerin etkinliğini artırmaya ve hedef odaklı politikalar geliştirmeye katkıda bulunur.

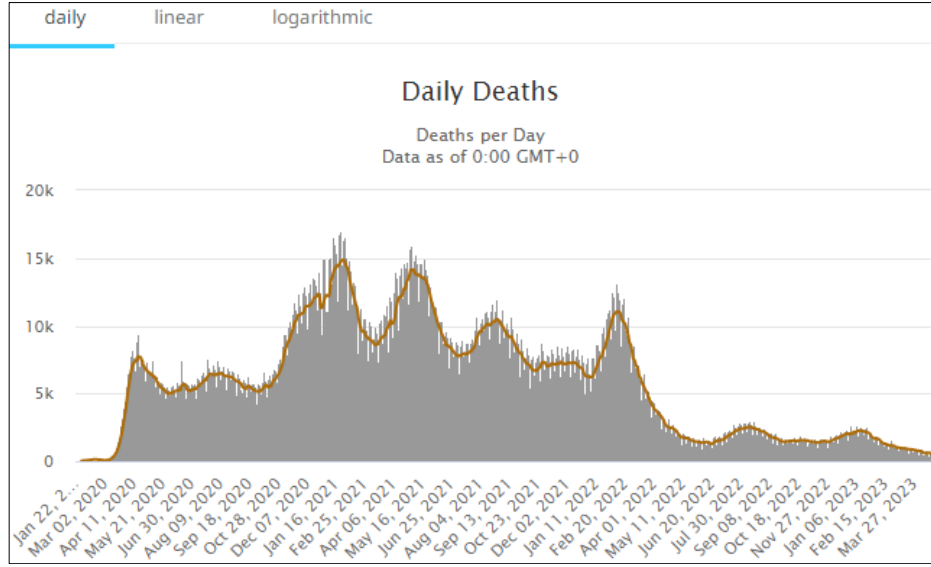
1. Vaka ve ölüm sayıları: Covid-19 pandemisinde, vaka ve ölüm sayıları önemli epidemiyolojik verilerdir. Bu veriler, farklı bölgelerde ve popülasyonlarda hastalığın yaygınlığı ve şiddeti hakkında bilgi sunmaktadır. Çizelge 2.2.'de vaka değişim oranını gösteren ilk 20 ülke listelenmiş, Şekil 2.2 ve Şekil 2.3'de ise günlük vaka ve ölüm oranlarının grafiğine yer verilmiştir.

Çizelge 2.2. İlk 20 ülke ve nüfus sayılarına göre vaka oranları [126].

#	Ülke, Diğer	Haftalık Vaka % Değişim	Nüfus
1	Ecuador	46,40%	18,113,361
2	North Macedonia	29,30%	2,081,304
3	Guadeloupe	20,10%	399,794
4	Sudan	7,10%	45,992,020
5	Cameroon	6,10%	27,911,548
6	Honduras	4,80%	10,221,247
7	Dominican Republic	2,40%	11,056,370
8	Mayotte	1,90%	286,259
9	Cook Islands	900%	17,571
10	Saint Martin	700%	39,73
11	Uganda	640%	48,432,863
12	St. Barth	600%	9,945
13	Cambodia	500%	17,168,639
14	Malawi	500%	20,180,839
15	Myanmar	330%	55,227,143
16	Bhutan	300%	787,941
17	Timor-Leste	300%	1,369,429
18	Cabo Verde	259%	567,678
19	Togo	200%	8,680,837
20	Vietnam	179%	98,953,541



Şekil 2.2. Günlük vaka oranları [126]



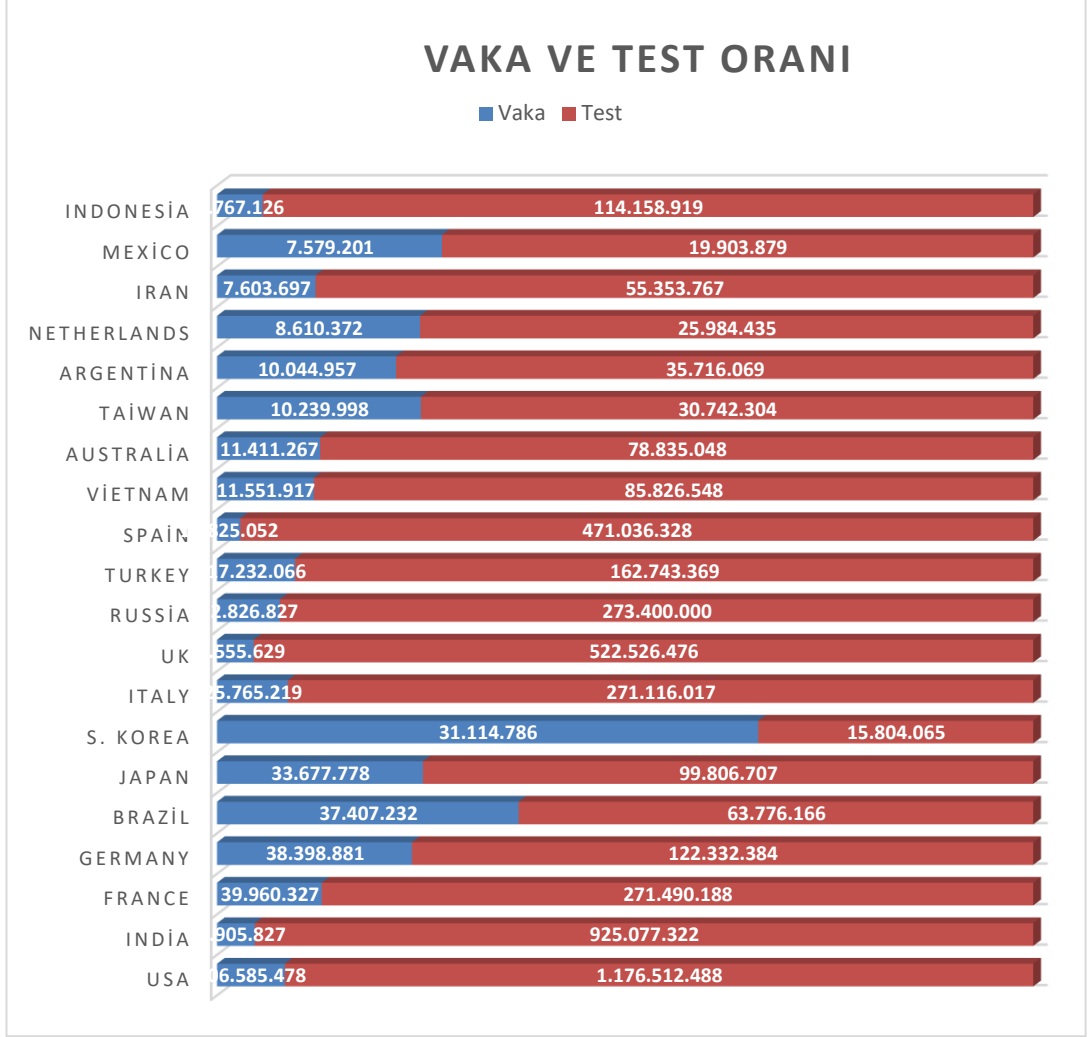
Şekil 2.3. Günlük ölüm oranları [126]

2. Hastalığın yayılım hızı ve yoğunluğu: Pandeminin seyri boyunca, hastalığın yayılım hızı ve yoğunluğu önemli epidemiyolojik verilerdir. Bu veriler, farklı bölgelerde ve demografik gruplarda hastalığın nasıl yayıldığı ve hangi faktörlerin hastalığın seyrini etkilediği hakkında bilgi sağlar.
3. Risk faktörlerinin prevalansı: Epidemiyolojik veriler, farklı popülasyonlarda ve demografik gruplarda risk faktörlerinin prevalansı hakkında bilgi sunar. Bu veriler, hangi faktörlerin hastalığın yayılımını ve şiddetini etkilediğini anlamaya ve bu faktörleri azaltmaya yönelik stratejiler geliştirmeye yardımcı olur.
4. Test sayıları ve pozitiflik oranları: Test sayıları ve pozitiflik oranları, hastalığın yaygınlığı ve bulaşıcılığı hakkında önemli bilgiler sunmaktadır. Bu veriler, pandemi sürecinde halk sağlığı politikalarının etkinliğini değerlendirmekte ve virüsün kontrol altına alınmasında kritik bir rol oynamaktadır. Çizelge 2.3'de toplam vaka sayısı en yüksek olan ilk 20 ülkenin nüfus oranı ve test sayılarına göre dağılımı belirtilmiştir.

Çizelge 2.3. İlk 20 ülke, nüfusa göre test sayıları ve vaka oranları[126].

#	Ülke, Diğer	Toplam Vaka	Toplam Test	Nüfus
1	USA	106,585,478	1,176,512,488	334,805,269
2	India	44,905,827	925,077,322	1,406,631,776
3	France	39,960,327	271,490,188	65,584,518
4	Germany	38,398,881	122,332,384	83,883,596
5	Brazil	37,407,232	63,776,166	215,353,593
6	Japan	33,677,778	99,806,707	125,584,838
7	S. Korea	31,114,786	15,804,065	51,329,899
8	Italy	25,765,219	271,116,017	60,262,770
9	UK	24,555,629	522,526,476	68,497,907
10	Russia	22,826,827	273,400,000	145,805,947
11	Turkey	17,232,066	162,743,369	85,561,976
12	Spain	13,825,052	471,036,328	46,719,142
13	Vietnam	11,551,917	85,826,548	98,953,541
14	Australia	11,411,267	78,835,048	26,068,792
15	Taiwan	10,239,998	30,742,304	23,888,595
16	Argentina	10,044,957	35,716,069	46,010,234
17	Netherlands	8,610,372	25,984,435	17,211,447
18	Iran	7,603,697	55,353,767	86,022,837
19	Mexico	7,579,201	19,903,879	131,562,772
20	Indonesia	6,767,126	114,158,919	279,134,505

Yapay zeka ve veri madenciliği yöntemleri, epidemiyolojik verilerin analizinde kullanılabilir. Bu yöntemler, risk faktörlerinin dağılımını belirlemeye ve farklı popülasyonlar ve demografik gruplar arasındaki ilişkileri anlamaya yardımcı olur. Bu analizler, pandemiye karşı etkili stratejiler ve politikalar geliştirmede önemli bir rol oynar. Ayrıca, toplumların pandemiye daha hızlı ve doğru bir şekilde tepki vermesini sağlar ve kaynakların en etkin şekilde kullanılmasına yardımcı olur. Yapay zeka ve veri madenciliği yöntemlerinin kullanılması, epidemiyolojik verilerin analizinde büyük veri setlerinin hızlı ve doğru bir şekilde işlenmesini sağlayarak, Covid-19 pandemisiyle mücadelede önemli bir katkı sunar. Vaka ve Test oranlarının dağılımını içeren ilk 20 ülkeye ait şekilsel gösterim Şekil 2.4'de belirtilmiştir.



Şekil 2.4. Vaka ve Test oranlarının şekilsel gösterimi (İlk 20 Ülke) [126]

## 2.4. COVID-19 TEŞHİSİ

### 2.4.1. COVID-19 Pandemisinde Risk Faktörlerinin Önemi

Covid-19 risk faktörleri, hastalığa yakalanma, hastalığın şiddetini ve hastalığın seyrini etkileyen unsurları ifade etmektedir. Covid-19 risk faktörlerinin yaygınlığını anlamak ve bu faktörler arasındaki ilişkileri ortaya çıkarmak, pandemi yönetimi açısından önemlidir. Risk faktörlerini belirleyerek, öncelikli müdahale alanlarını belirlemek ve kaynakları etkili bir şekilde kullanmak mümkün hale gelir. Bu sayede, risk altındaki bireylerin korunması ve erken tedavi süreçlerine dahil edilmesi sağlanabilir. Ayrıca, risk faktörleriyle ilgili farkındalık yaratılması ve önleyici önlemlerin alınması, salgının kontrolü ve yayılımının sınırlandırılması açısından da kritik öneme sahiptir. Bu

bağlamda, MÖ yöntemleri, Covid-19 pandemisi ve risk faktörleri üzerine yapılan çalışmalarda, anlamlı ve değerli bilgiler sunarak pandemi yönetimine katkıda bulunmaktadır [127]. Diğer bir açıdan, Covid-19 risk faktörlerinin anlaşılması ve değerlendirilmesi, pandemi ile mücadelede önemli bir rol oynamaktadır. Bu risk faktörleri, hastalığın yayılma hızını, şiddetini ve toplumlar üzerindeki etkisini etkilemektedir. Ayrıca, risk faktörlerinin belirlenmesi ve yönetilmesi, salgınla mücadelede kullanılan stratejilerin etkinliğini artırarak, toplum sağlığını ve ekonomik istikrarı korumaya katkıda bulunmaktadır [128,130]. Risk faktörlerinin analizi ve değerlendirmesi, makine öğrenimi ve veri madenciliği yöntemlerinin kullanılması ile daha hızlı ve doğru bir şekilde gerçekleştirilebilir. Bu yöntemler, büyük ve karmaşık veri setlerinin analizinde etkili olduğu için, Covid-19 risk faktörleri üzerinde derinlemesine çalışmalar yaparak, daha etkili önlemler ve politikalar geliştirmeye yardımcı olmaktadır.

#### **2.4.2. Yapay Zeka ve Veri Madenciliği Yöntemlerinin Teşhis Sürecine Katkısı**

Yapay zeka - Artificial Intelligence (AI) ve Veri Madenciliği (VM) yöntemleri, son yıllarda sağlık sektöründe önemli gelişmeler kaydetmiştir. Özellikle teşhis süreçlerinde, makine öğrenimi ve veri analizi yöntemleri sayesinde, daha hızlı ve doğru sonuçlar elde etmek mümkün hale gelmiştir. Covid-19 pandemisi sırasında da bu teknolojilerin kullanılması, hastalığın teşhisi ve yönetimi açısından büyük önem taşımaktadır [131].

Yapay zeka ve veri madenciliği yöntemlerinin teşhis sürecine katkıları şu şekildedir:

1. Hızlı ve doğru teşhis: AI ve VM yöntemleri, büyük miktarda veriyi hızlı bir şekilde analiz edebilir ve bu sayede hastalığın teşhisine katkı sağlayabilir. Bu durum, özellikle Covid-19 pandemisi gibi yoğun ve hızlı müdahale gerektiren durumlarda büyük avantaj sağlayabilir.
2. Risk faktörlerinin belirlenmesi: AI algoritmaları, hastaların demografik bilgileri, tıbbi geçmişi ve diğer verileri kullanarak hastalığa yakalanma ve hastalığın seyri açısından önemli olan risk faktörlerini belirleyebilir.

3. Erken uyarı ve önleyici tedbirler: VM yöntemleri ile elde edilen bilgiler, risk faktörleri ve hastalık belirtileri açısından erken uyarı önlemleri belirleyebilir. Bu sayede, hastaların durumlarının takibi ve gerekli önleyici tedbirlerin alınması kolaylaşır.
4. Tahmin ve projeksiyonlar: AI ve VM yöntemleri, mevcut verilerden yola çıkarak gelecekteki hastalık seyrini ve salgın durumunu tahmin etmeye yardımcı olabilir. Bu sayede, sağlık hizmetleri ve kaynaklarının daha etkin bir şekilde planlanması ve yönlendirilmesi sağlanabilir.
5. Kişiselleştirilmiş tedavi ve bakım: AI ve VM yöntemleri ile elde edilen bilgiler, hastaların bireysel ihtiyaçları ve risk faktörleri dikkate alınarak kişiselleştirilmiş tedavi ve bakım hizmetleri sunulmasına imkan tanıyabilir.
6. Tıbbi görüntüleme ve analiz: AI algoritmaları, tıbbi görüntüleme teknikleri (Örneğin Röntgen, BT, Manyetik Rezonans (MR)) ile elde edilen görüntülerin analizinde kullanılabilir. Bu sayede, Covid-19 enfeksiyonunun erken teşhisinde ve hastalığın seyrinin takibinde önemli bir rol oynayabilir. Özellikle akciğer tutulumunu değerlendiren görüntülerde, yapay zeka destekli analizlerle daha hızlı ve doğru sonuçlara ulaşılabilir.
7. Epidemiyolojik modelleme ve salgın yönetimi: AI ve VM yöntemleri, salgınların yayılımını ve etkilerini anlamak için kullanılabilir. Özellikle pandemi dönemlerinde, bu teknolojilerin kullanımıyla salgının seyrini tahmin etmek, risk bölgelerini belirlemek ve salgınla mücadelede etkili stratejiler geliştirmek mümkün hale gelebilir.
8. İlaç ve aşı geliştirme süreçlerine katkı: AI ve VM yöntemleri, ilaç ve aşı geliştirme süreçlerinde de önemli katkılar sağlayabilir. Bu yöntemlerle, potansiyel ilaç ve aşı adaylarının belirlenmesi, etkinliklerinin test edilmesi ve yan etkilerinin değerlendirilmesi süreçleri hızlandırılabilir.



9. Hasta takibi ve izlem: AI destekli teşhis süreçleri ve veri analizi, hasta takibi ve izlem süreçlerinde de önemli avantajlar sunabilir. Bu sayede, hastaların durumlarının düzenli olarak değerlendirilmesi, hastalık seyrinde önemli değişikliklerin erken tespit edilmesi ve gerekli müdahalelerin zamanında yapılması sağlanabilir.

10. Sağlık hizmetlerinin iyileştirilmesi ve optimizasyonu: AI ve VM yöntemlerinin kullanılmasıyla, sağlık hizmetlerinin kalitesi ve etkinliği artırılabilir. Bu teknolojiler sayesinde, hasta ve risk gruplarına yönelik öncelikler belirlenebilir, kaynakların en etkili şekilde kullanılması sağlanabilir ve sağlık sistemlerinin genel performansı iyileştirilebilir.

AI ve VM yöntemleri, Covid-19 pandemisi gibi küresel sağlık krizlerinde teşhis sürecine önemli katkılar sağlayabilir. Bu teknolojilerin kullanımıyla, hastalığın erken teşhisi, hastaların durumlarının düzenli izlenmesi ve tedavi süreçlerinin kişiselleştirilmesi mümkün hale gelir. Bu sayede, pandemiyle mücadelede daha etkili ve hızlı sonuçlar alınabilir, sağlık sistemlerinin yükü hafifletilebilir ve toplum sağlığı açısından önemli kazanımlar elde edilebilir.

## BÖLÜM 3

### VERİ SETİ VE KULLANILAN YÖNTEMLER

Makine öğrenimi ve veri madenciliği yöntemleri, büyük veri setlerinden bilgi elde etmek ve karmaşık problemleri çözmek için kullanılan güçlü araçlardır. Bu bölümde, bu yöntemlerin temel prensipleri ve Covid-19 risk faktörlerinin analizinde nasıl kullanıldığı üzerinde durulacaktır.

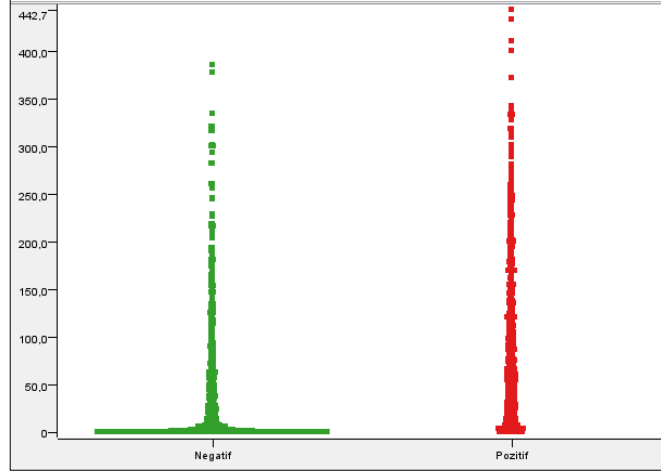
#### 3.1. VERİ SETİ

##### 3.1.1. Veri Setinin Tanımı ve Kaynağı

Veri seti, Zenodo adlı açık erişimli dijital veri deposundan alınmıştır ve aşağıdaki bağlantıdan erişilebilir: <https://zenodo.org/record/4081318#.X4RWqdD7TIU>. Bu veri seti, Covid-19 ile ilgili bir araştırma projesi için toplanmıştır ve 1677 hastada PCR testleri sırasında alınan kan tahlil değerleri ile birlikte 35 değişkenden oluşmaktadır. Bu veri seti, Covid-19 hastalarının ve hastalığı taşıyan bireylerin teşhisi ve tedavisi için kullanılacak potansiyel biyobelirteçleri ve risk faktörlerini belirlemeye yönelik çalışmalar için oldukça yararlıdır.

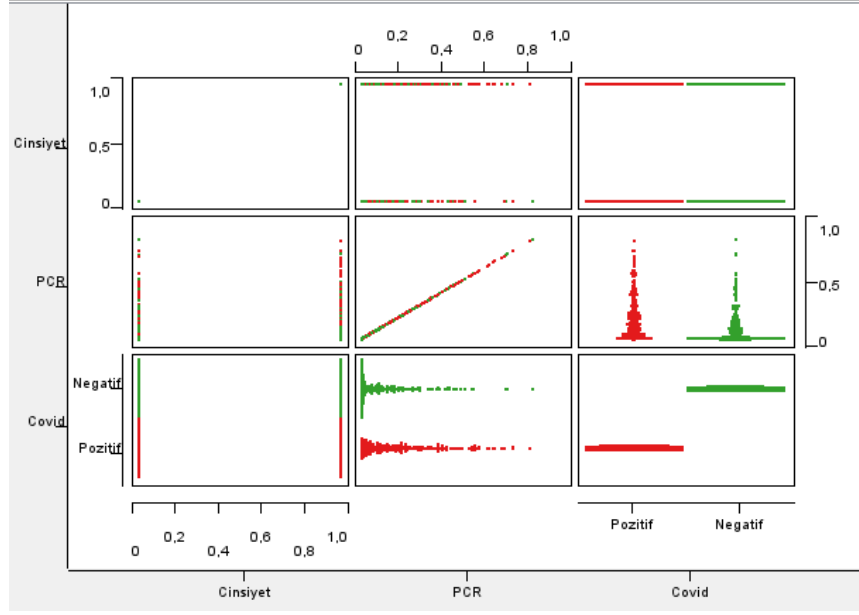
Veri setinin kaynağı, Porto Üniversitesi ve Portekiz Küresel Sağlık ve Tropikal Tıp Enstitüsü tarafından gerçekleştirilen bir çalışmadır. Bu çalışma, Covid-19 pandemisinin başlangıcından bu yana yapılan bir dizi araştırmayı temel almaktadır ve farklı yaş grupları, cinsiyet ve etnik kökenlerden katılımcıların kan örneklerini içermektedir. Veri seti, Covid-19 hastalarının klinik özelliklerinin, laboratuvar sonuçlarının ve demografik bilgilerinin yanı sıra hastalığın seyri ve prognozuyla ilgili bilgileri de içermektedir.

Veri setinde Covid-19 sonucu pozitif 814, negatif 863 kayıt yer almaktadır. Yaş aralığı 0-100 arası değişen 1677 hastanın 712'si kadın ve 965'i erkektir. Cinsiyet, Covid ve PCR değerleri arasındaki ilişki Şekil 3.1'de gösterilmiştir. Düşey kısımda Pcr sonuçlarını, yatay kısımda ise hastaların covid durumu belirtilmiştir.



Şekil 3.1. PCR test sonuçlarına göre covid olan hastaların dağılımı.

Veri seti içinde başvuru tarihi, cinsiyet, yaş, kalsiyum, kreatin kinaz, kreatin, alkalın fosfat, gamma glutamil transferaz, glikoz, aspartat amiotransferaz, alanin aminotransferaz, laktat dehidrojenaz, PCR, KAL, NAT, üre, beyaz kan hücresi sayısı, kırmızı kan hücresi sayısı, hemoglobin, hematokrit, ortalama eritrosit hacmi, ortalama eritrosit hemoglobin konsantrasyonu, trombosit, nötrofil sayısı (%), lenfosit sayısı (%), monosit sayısı (%), eozinofil sayısı (%), bazofil sayısı (%), nötrofil sayısı, lenfosit sayısı, monosit sayısı, eozinofil sayısı, bazofil sayısı, şüphe özellikleri ve hedef durumu bulunmaktadır. Covid durumu ve sayısal olarak kodlanmış cinsiyet verileri hariç diğer verilerin tamamı sayısal değerlerden oluşmaktadır. Şekil 3.2'de Scatter Matrix düğümüne ait PCR, Covid ve Cinsiyet değerlerinin grafiksel dağılımına yer verilmiştir.



Şekil 3.2. PCR, Covid ve Cinsiyet değerlerinin grafiksel dağılımı

Veri setinin kullanılması, yapay zeka ve veri madenciliği yöntemlerinin uygulanmasıyla, Covid-19 hastalığının teşhisi ve risk faktörlerinin belirlenmesinde önemli katkılar sağlamıştır. Bu veri seti, araştırmacıların ve sağlık profesyonellerinin, hastalığın yayılmasını kontrol etmek, hastaların durumunu değerlendirmek ve uygun tedavi stratejileri geliştirmek için daha iyi bilgi ve araçlara sahip olmasına yardımcı olmaktadır. Veri setine ait özellikler Çizelge 3.1’de belirtilmiştir.

Çizelge 3.1. Veri seti ve özellikleri.

Sıra	Özellik	CBC Veri Seti	En Küçük Değer	En Büyük Değer
1	Cinsiyet	X	0	1
2	PCR	X	0	442,7
3	Covid	X	Negatif	Pozitif
4	Yas	X	1	100
5	Kalsiyum	X	0	20,25
6	Kreatinin	X	0	15,7
7	Glikoz	X	0	855
8	Beyaz Kan Hücreleri	X	0	77,67
9	Kırmızı Kan Hücreleri	X	0	7,06
10	Hemoglobin	X	0	18,55
11	Kreatin Kinaz		0	7004,33

12	Alkalin Fosfataz	0	914
13	Gama Glutamiltransferaz	0	3041
14	Aspartat Aminotransferaz	0	1018,5
15	Alanin Aminotransferaz	0	559,5
16	Laktat Dehidrogenaz	0	5457
17	KAL	0	6,84
18	NAT	0	181,9
19	UREA	0	464,5
20	Hematokrit	0	56,7
21	Ortalama Korpüsküler Hacim	0	131,5
22	MCH	0	49,7
23	Ortalama Korpüskolar Hemog. Kon.	0	46,75
24	PLT1	0	801,5
25	Nötrofil	0	98,5
26	Lenfosit	0	67
27	Monosit	0	42,7
28	Eozinofil	0	12,7
29	Bazofil	0	3,1
30	NET	0	73
31	Lenfosit	0	11,4
32	Monosit	0	14,5
33	Eozinofil	0	1,45
34	Bazofil	0	0,7
35	Suphe	0	1

### 3.1.2. Değişkenler ve Demografik Bilgiler

Veri setinde toplamda 35 değişken bulunmaktadır. Bu değişkenlerin bir kısmı demografik bilgileri, diğerleri ise hastaların klinik özelliklerini ve laboratuvar sonuçlarını temsil etmektedir. Değişkenlerin ve demografik bilgilerin detaylı bir açıklaması madeler halinde belirtilmiştir:

#### 1. Demografik bilgiler:

- Yaş: Hastaların yaşı
- Cinsiyet: Hastaların cinsiyeti (kadın veya erkek)

#### 2. Klinik özellikler:

- Sistolik kan basıncı: Hastaların ölçülen sistolik kan basıncı değeri

- Diyastolik kan basıncı: Hastaların ölçülen diyastolik kan basıncı değeri
- Kalp atış hızı: Hastaların kalp atış hızı
- Ateş: Hastaların ateş durumu
- Oksijen saturasyonu: Hastaların oksijen saturasyon seviyesi
- Solunum hızı: Hastaların solunum hızı

### 3. Laboratuvar sonuçları:

- Hemoglobin: Hastaların hemoglobin seviyeleri
- Hematokrit: Hastaların hematokrit değeri
- Trombosit sayısı: Hastaların trombosit sayısı
- Lökosit sayısı: Hastaların lökosit sayısı
- Lenfosit sayısı: Hastaların lenfosit sayısı
- Nötrofil sayısı: Hastaların nötrofil sayısı
- Eozinofil sayısı: Hastaların eozinofil sayısı
- Bazofil sayısı: Hastaların bazofil sayısı
- Monosit sayısı: Hastaların monosit sayısı
- C-reaktif protein (CRP): Hastaların C-reaktif protein değeri
- Laktat dehidrogenaz (LDH): Hastaların laktat dehidrogenaz seviyeleri
- Aspartat aminotransferaz (AST): Hastaların aspartat aminotransferaz değeri
- Alanin aminotransferaz (ALT): Hastaların alanin aminotransferaz değeri
- Gama-glutamil transferaz (GGT): Hastaların gama-glutamil transferaz değeri
- Kreatinin: Hastaların kreatinin seviyeleri
- Üre: Hastaların üre değeri
- Prokalsitonin: Hastaların prokalsitonin seviyeleri
- D-dimer: Hastaların D-dimer değeri
- Ferritin: Hastaların ferritin seviyeleri
- Fibrinojen: Hastaların fibrinojen değeri
- Troponin: Hastaların troponin seviyeleri
- ProBNP: Hastaların proBNP değeri
- Interlökin-6 (IL-6): Hastaların interlökin-6 seviyeleri
- Tümör nekroz faktörü-alfa (TNF-alfa): Hastaların tümör nekroz faktörü-alfa değeri

Veri setinde yer alan bu deęişkenler ve demografik bilgiler, Covid-19'un risk faktörlerinin anlaşılması ve hastalığın teşhisinde kullanılacak yapay zeka ve veri madencilięi yöntemlerinin başarısını ölçmek için önemlidir. Bu deęişkenler sayesinde, farklı popülasyonlardaki risk faktörlerinin yaygınlığı ve Covid-19 ile ilişkisi incelenmiştir. Ayrıca, bu deęişkenlerin her biri, hastaların Covid-19 durumunu tahmin etmek için kullanılacak modelin girdisi olarak deęerlendirilmiştir. Örneğin, yaş ve cinsiyet gibi demografik bilgiler, laboratuvar sonuçları ve klinik özelliklerle birlikte modele dahil edilerek, her hastanın Covid-19 durumunu tahmin etmeye yönelik daha doğru ve kapsamlı bir yaklaşım sağlamıştır. Bu, hem teşhis sürecinde doğruluęu artırabilir hem de belirli risk faktörleriyle ilişkili olan grupları tespit etmeye yardımcı olabilir.

Sonuç olarak, Covid-19 risk faktörleri ve yaygınlığına dair bu detaylı deęişkenler ve demografik bilgiler, tez kapsamında KNIME programı kullanılarak gerçekleştirilerek makine öğrenimi ve veri madencilięi yöntemlerinin başarılı uygulanması ve deęerlendirilmesi için büyük önem taşımaktadır. Bu sayede, hastalığın teşhisi ve risk faktörlerinin anlaşılması konularında önemli katkılar sağlanabilecektir.

### **3.2. MAKİNE ÖĞRENMESİ YÖNTEMLERİ**

Makine Öğrenimi (MÖ), AI altında bir alan olarak bilgisayarların veri ile öğrenme becerisi kazandığı bir tekniktir. Temelde, MÖ algoritmaları, belirli bir görevi yerine getirmek için verilerden bağımsız özellikler ve modeller oluşturur. Öğrenme süreçleri sırasında, algoritma veri setindeki ilişkileri keşfeder ve performansını görev için deęerlendirmek üzere metrikler kullanır. MÖ , sınıflandırma, regresyon, kümeleme ve boyut indirgeme gibi çeşitli problemlere uygulanabilir [132,133]. VM, büyük ve karmaşık veri setlerinin analizinde kullanılan otomatik veya yarı otomatik yöntemlerin uygulanmasıdır. Bu yöntemler, veri setlerindeki gizli kalıpları, ilişkileri ve trendleri keşfetmeye çalışır. VM süreçleri, veri ön işleme, modelleme, deęerlendirme ve yorumlama gibi adımları içerir. VM, farklı endüstrilerde ve disiplinlerde, stratejik karar verme, müşteri ilişkileri yönetimi, tıbbi teşhis ve pazarlama gibi amaçlarla kullanılmaktadır [134,135].

VM, büyük ve karmaşık veri setlerindeki gizli bilgi ve örüntüleri keşfetmek için kullanılan bir dizi teknik ve süreçten oluşmaktadır. VM süreci, genellikle aşağıdaki adımları içerir:

1. Veri toplama: VM sürecinin ilk aşaması, analiz edilecek verilerin toplanmasıdır. Bu süreç, veri kaynaklarının belirlenmesi, veri entegrasyonu ve veri temizleme gibi alt süreçleri içerir.
2. Veri ön işleme: Veri ön işleme aşaması, veri setinin analize hazır hale getirilmesini amaçlar. Bu süreç, eksik verilerin doldurulması, gürültülü verilerin düzeltilmesi, veri dönüşümü ve özellik seçimi gibi işlemleri içerir.
3. Veri analizi: Veri analizi süreci, veri setindeki örüntülerin ve yapıların keşfedilmesini amaçlar. Bu aşamada, istatistiksel analiz, görselleştirme ve makine öğrenimi yöntemleri gibi teknikler kullanılır.
4. Modelleme ve tahmin: Modelleme süreci, öğrenilen bilgilerin yeni veriler üzerinde tahminlerde bulunmak için kullanılmasını amaçlar. Bu süreç, makine öğrenimi algoritmalarının eğitilmesi ve doğrulanması, performans ölçütlerinin değerlendirilmesi ve model optimizasyonu gibi işlemleri içerir.
5. Değerlendirme ve raporlama: Değerlendirme süreci, analiz sonuçlarının ve model performansının değerlendirilmesini amaçlar. Bu aşamada, analiz sonuçları, hedeflenen amaçlara ve sorunlara göre yorumlanır ve raporlanır.[136]

VM yöntemleri, genellikle aşağıdaki kategorilere ayrılır:

1. Sınıflandırma: Sınıflandırma yöntemleri, veri nesnelere önceden belirlenmiş sınıflara ayırmayı amaçlar. Örnek yöntemler: Karar Ağaçları, Destek Vektör Makineleri, k-En Yakın Komşu (İng. K-Nearest Neighbour (k-EYK)) ve Naïve Bayes.



2. Kümeleme: Kümeleme yöntemleri, veri nesnelerini benzerliklere göre gruplara ayırmayı amaçlar. Örnek yöntemler: k-Ortalamalar (k-Means), DBSCAN, hiyerarşik kümeleme ve yoğunluk tabanlı kümeleme.
3. Birliktelik kuralları: Birliktelik kuralları yöntemleri, veri setindeki öğeler veya olaylar arasındaki ilişkileri ve bağımlılıkları keşfetmeyi amaçlar. Örnek yöntemler: Apriori algoritması, Eclat ve FP-Growth.
4. Anomali tespiti: Anomali tespiti yöntemleri, veri setindeki beklenmeyen veya anormal öğeleri ve olayları tespit etmeyi amaçlar. Örnek yöntemler: İstatistiksel testler, izolasyon ormanları (Isolation Forest) ve Otomatik Kodlayıcılar (Autoencoders).
5. Regresyon: Regresyon yöntemleri, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi modellemeyi ve tahmin etmeyi amaçlar. Örnek yöntemler: Lineer regresyon, çoklu regresyon, polinom regresyon ve destek vektör regresyonu.
6. Metin madenciliği: Metin madenciliği yöntemleri, doğal dil işleme tekniklerini kullanarak metin verilerinden bilgi ve örüntüler çıkarmayı amaçlar. Örnek yöntemler: Kelime frekans analizi, Latent Dirichlet Allocation (LDA) ve Word2Vec.
7. Zaman serisi analizi: Zaman serisi analizi yöntemleri, zaman içinde gözlenen veri noktalarının özelliklerini ve yapılarını analiz etmeyi amaçlar. Örnek yöntemler: Otoregresif Entegre Kayan Ortalama (ARIMA), Exponential Smoothing State Space Model (ETS) ve Long Short-Term Memory (LSTM) ağları.

MÖ ve VM süreçleri ve yöntemlerinin anlaşılması, Covid-19 risk faktörlerinin yaygınlığını ve etkilerini analiz etme sürecinde önemlidir. Bu bilgi, tez çalışması kapsamında kullanılacak yöntemlerin seçimi ve uygulanması için temel bir yapı sağlamaktadır. [134,135].

### **3.2.1. Sınıflandırma**

Sınıflandırma, veri nesnelerini önceden tanımlanmış sınıflara (etiketlere) göre atama sürecidir. Sınıflandırma algoritmaları, öğretici (supervised) öğrenme kategorisine girer, çünkü bu tür algoritmalar eğitim verilerine dayalı olarak öğrenmekte ve tahminlerde bulunmaktadır. Sınıflandırma algoritmaları, özellikle tıbbi teşhis, spam e-posta filtreleme ve el yazısı tanıma gibi alanlarda yaygın olarak kullanılır. [134,135].

#### **3.2.1.1. Çok Katmanlı Algılayıcı - RProp MLP Learner (ÇKA)**

RProp MLP Learner (Multi-Layer Perceptron) algoritması, bir Yapay Sinir Ağı (ANN) eğitim yöntemi olan Rezidüel Propagasyon (RProp) algoritmasını kullanır. ÇKA sınıflandırma ve regresyon problemlerini çözmek için kullanılabilen, girdi, çıktı ve bir veya birden fazla gizli katmandan oluşan süpervize edilmiş bir yapay sinir ağıdır. RProp, hızlı ve ölçekten bağımsız olarak sinir ağı ağırlıklarını güncellemek için bir yöntemdir. RProp algoritması, gradyanın işaretini kullanarak ağırlıkları günceller ve gradyanın büyüklüğünü dikkate almaz. Bu, ÇKA'nin daha hızlı ve daha verimli bir şekilde eğitilmesine olanak tanır. [134,135].

#### **3.2.1.2. Karar Ağaçları - Decision Trees (KA)**

Karar Ağaçları algoritması, hem sınıflandırma hem de regresyon problemlerinde kullanılabilen bir MÖ algoritmasıdır. KA, veriyi hiyerarşik bir yapı içinde ayrıştırarak öğrenir ve tahminlerde bulunur. Bu yapıda, her düğümde bir öznitelik üzerinde bir koşul bulunur ve her dal kararın bir sonucunu temsil eder. Yaprak düğümler ise tahmin edilen sınıf etiketleri veya sürekli hedef değişken değerlerini içerir. KA, sınıflandırma problemlerinde veri noktalarını belirli sınıflara atamak için kullanılabilir ve regresyon problemlerinde ise sürekli hedef değişkenlerin değerlerini tahmin etmek için kullanılabilir. Kümeleme algoritmalarından farklı olarak, karar ağaçları önceden belirlenmiş sınıf etiketleri veya hedef değişken değerleri kullanarak öğrenir ve tahminlerde bulunur [134,135].

### **3.2.1.3. Destek Vektör Makineleri (DVM)**

Destek Vektör Makineleri (İng. Support Vector Machine - SVM) algoritması, bir MÖ algoritmasıdır ve başlıca sınıflandırma problemlerinde kullanılır. DVM, veri noktalarını iki veya daha fazla sınıfa ayırmak için kullanılabilir. DVM algoritması, sınıflar arasındaki marjı en büyük hale getiren bir hiperdüzlem (ayırıştırıcı düzlem) bulmaya çalışır. Bu hiperdüzlem, sınıflar arasındaki en uygun ayrımı sağlar ve aynı zamanda sınıflar arasındaki boşluğu maksimize eder. DVM, doğrusal ve doğrusal olmayan sınıflandırma problemleri için kullanılabilir. Doğrusal olmayan problemlerde, DVM, özgün veri kümesini yüksek boyutlu bir uzaya dönüştürerek sınıflandırma yapar. Bu dönüşüm, çekirdek fonksiyonları (kernel functions) olarak bilinen özel fonksiyonlar kullanılarak gerçekleştirilir [134,135].

### **3.2.1.4. k-En Yakın Komşu (k-EYK)**

k-En Yakın Komşu (k-EYK) algoritması, makine öğreniminde kullanılan temel bir algoritmadır ve genellikle sınıflandırma problemleri için kullanılır. k-EYK, veri kümesindeki mevcut örneklerin etiketlerine dayalı olarak yeni örneklerin sınıflandırılmasını sağlar. k-EYK algoritmasının çalışma prensibi şu şekildedir: Yeni bir veri noktası sınıflandırılması gerektiğinde, algoritma bu veri noktasına en yakın olan k örnek arasındaki mesafeyi hesaplar. En yakın komşuların çoğunluk sınıfı, yeni veri noktasının sınıf etiketi olarak atanır. k değeri, algoritmanın dikkate alacağı komşu sayısını belirler ve bu değer kullanıcı tarafından seçilir.

k-EYK algoritması esas olarak bir sınıflandırma algoritmasıdır; ancak, kümeleme problemlerinde de kullanılabilir. k-EYK, kümeleme için kullanıldığında, benzer veri noktalarını bir araya getiren ve böylece veri kümesini doğal gruplara bölen bir yaklaşım sağlar. Yine de, k-EYK daha çok sınıflandırma problemleri için kullanılır ve yaygın kümeleme algoritmalarıyla (ör. k-ortalamlar, DBSCAN) karşılaştırıldığında daha az etkilidir. [134,135].

### **3.2.1.5. Naïve Bayes (NB)**

Naïve Bayes algoritması, temel olarak bir sınıflandırma algoritmasıdır ve olasılık teorisi temelinde çalışır. İsmi, Bayes teoreminin kullanımından ve özelliklerin bağımsız olduğu varsayımından (naive: saf, basit) alır. Algoritma, veri kümesindeki özelliklerin birbirinden bağımsız olduğunu varsayarak, Bayes teoremini kullanarak her bir özellik için olasılıkları hesaplar ve bu olasılıkların çarpımını alarak her bir sınıf için koşullu olasılığı elde eder. Yeni bir veri noktasının sınıfını belirlemek için, algoritma bu koşullu olasılıkları karşılaştırır ve en yüksek olasılığa sahip sınıfı seçer.

NB algoritması, sınıflandırma problemleri için kullanılmasına rağmen, kümeleme problemleri için de kullanılabilir. Bununla birlikte, NB'in başarısı, özelliklerin bağımsızlık varsayımına bağlı olduğundan, kümeleme için daha yaygın ve etkili olan diğer yöntemler (ör. k-ortalamlar, DBSCAN) genellikle tercih edilir. [134,135].

### **3.2.1.6. Lojistik Regresyon (LR)**

Lojistik Regresyon algoritması, temel olarak bir sınıflandırma algoritmasıdır. İkili sınıflandırma problemlerini çözmek için kullanılır, ancak çok sınıflı sınıflandırma problemleri için de uyarlanabilir. LR, bağımsız değişkenlerin (özelliklerin) ve bağımlı değişkenin (sınıf etiketi) arasındaki ilişkiyi modellemek için kullanılır. LR, adından da anlaşılacağı gibi, regresyon analizine dayanır, ancak sınıflandırma problemlerine uygulanır. Lineer regresyonun ters lojik fonksiyonu ile modifiye edilmiş bir versiyonudur. Bu nedenle, lojistik regresyon, veri kümesindeki özellikler ve sınıflar arasındaki ilişkiyi temsil eden bir model oluşturur. LR, doğrusal olarak ayrılabilir problemler için özellikle etkili olup, sınıflandırma problemlerinde yaygın olarak kullanılır. [134,135].

### **3.2.1.7. Rastgele Ormanlar - Random Forests (RO)**

Rastgele Ormanlar (Random Forests) algoritması, hem sınıflandırma hem de regresyon problemlerinde kullanılabilen bir MÖ algoritmasıdır. Temel olarak, birden çok karar ağacını bir araya getirerek daha güçlü ve daha istikrarlı bir model oluşturur.

Bu yöntem, "ensemble learning" veya "topluluk öğrenme" olarak adlandırılan bir yaklaşımın parçasıdır. RO, sınıflandırma problemlerinde veri noktalarını belirli sınıflara atamak için kullanılabilir ve regresyon problemlerinde ise sürekli hedef değişkenlerin değerlerini tahmin etmek için kullanılabilir. Kümeleme algoritmalarından farklı olarak, RO, önceden belirlenmiş sınıf etiketleri veya hedef değişken değerleri kullanarak öğrenir ve tahminlerde bulunur. [134,135].

### **3.2.1.8. Olasılık Sinir Ağı (OSA)**

Olasılık Sinir Ağı (OSA) (İng. Probabilistic Neural Network (PNN)), bir MÖ algoritmasıdır ve özellikle sınıflandırma problemlerinde kullanılır. OSA, temel olarak Parzen penceresi ve Bayes teoremi ile örneklerin sınıflandırılmasını sağlayan istatistiksel bir yaklaşıma dayanır. OSA, Radyal Temelli Fonksiyonlar (İng. Radial Basis Function) ağlarına benzer bir yapıya sahiptir ve tipik olarak hızlı eğitim süreçleri ve güçlü tahmin kabiliyetleri ile bilinir. DDA (Dynamic Decay Adjustment) ise, PNN Learner'a uygulanan bir optimizasyon yöntemidir. DDA, ağırlıklı ortalamalar kullanarak her sınıf için en iyi dağılım parametrelerini (ör. varyanslar) otomatik olarak belirleyebilir ve böylece daha iyi sınıflandırma performansı sağlar. [134,135].

### **3.2.1.9. Yapay Sinir Ağları - Artificial Neural Networks (YSA)**

Yapay Sinir Ağları (İng. Artificial Neural Networks - ANN), insan beyninin bilgi işleme şeklinden ilham alarak oluşturulan ve öğrenme yeteneğine sahip olan karmaşık matematiksel modellerdir. YSA, katmanlara ayrılmış düğümlerden (nöronlar) oluşur ve bu düğümler arasında ağırlıklı bağlantılar bulunur. Ağlar, verileri işlemek ve öğrenmek için düğümler ve bağlantılar arasındaki ağırlıkları güncelleyerek eğitilir. YSA, sınıflandırma, regresyon ve kümeleme gibi çeşitli görevlerde kullanılabilen çok yönlü algoritmalarıdır. Sınıflandırma ve regresyon problemlerinde, ağ, veri kümesindeki özellikler ve hedef değişken (sınıf etiketi veya sayısal değer) arasındaki ilişkiyi temsil eden bir model oluşturur. Kümeleme problemlerinde ise, ağ, veri kümesindeki örneklerin doğal gruplarını belirleyerek kullanılabilir.

Bu nedenle, YSA hem sınıflandırma hem de regresyon algoritması olarak kullanılabilir ve aynı zamanda kümeleme algoritması olarak da kullanılabilir. YSA'nın uygulanabilirliği ve performansı, ağın yapılandırılması, eğitim algoritması ve hiperparametrelerin seçimi gibi faktörlere bağlıdır. [134,135].

### **3.2.1.10. Uyarlanabilir Güçlendirme- Adaptive Boosting (AdaBoost)**

AdaBoost (Adaptive Boosting) algoritması, sınıflandırma ve regresyon problemleri için kullanılabilen bir topluluk öğrenme yöntemidir. AdaBoost, zayıf öğrencileri (sınıflandırıcılar veya regresyon modelleri) bir araya getirerek daha güçlü ve doğru bir model oluşturmayı amaçlar. Bu algoritma, özellikle sınıflandırma problemleri için oldukça popülerdir. AdaBoost, öğrencileri iteratif olarak eğitir ve her iterasyonda örneklerin ağırlıklarını günceller. Yanlış sınıflandırılan veya tahmin edilen örneklerin ağırlıkları artırılırken, doğru sınıflandırılan veya tahmin edilen örneklerin ağırlıkları azaltılır. Bu, modelin zor örnekler üzerinde daha fazla odaklanmasını sağlar. Öğrenciler, son modelde ağırlıklı bir oylama veya toplama ile birleştirilir. AdaBoost, başlıca sınıflandırma ve regresyon algoritması olarak kullanılır, ancak kümeleme algoritması olarak kullanılmaz. Başarılı ve verimli bir algoritma olarak kabul edilen AdaBoost, genellikle karar ağaçları gibi basit sınıflandırıcılarla birlikte kullanılır [137].

### **3.2.1.11. Gradyan Yükseltme Makineleri - Gradient Boosting Machines (GYM)**

Gradient Boosting Machines (GBM) algoritması, hem sınıflandırma hem de regresyon problemleri için kullanılabilen güçlü ve esnek bir topluluk öğrenme yöntemidir. GYM, zayıf öğrencileri (sınıflandırıcılar veya regresyon modelleri) bir araya getirerek daha güçlü ve doğru bir model oluşturmayı amaçlar. Bu algoritma, özellikle sınıflandırma ve regresyon problemleri için oldukça popüler ve etkilidir. GYM, öğrencileri iteratif olarak eğitir ve her iterasyonda, mevcut modelin hatalarını düzeltmeye çalışarak bir sonraki öğrenciyi ekler. Bu süreç, hataların (kalanlar) negatif gradientleri üzerinde öğrencinin eğitilmesiyle gerçekleştirilir. GYM, genellikle karar ağaçları gibi basit sınıflandırıcılar veya regresyon modelleri kullanarak toplulukları oluşturur.

GYM algoritması, başlıca sınıflandırma ve regresyon algoritması olarak kullanılır, ancak kümeleme algoritması olarak kullanılmaz. GYM, genellikle daha büyük ve karmaşık veri kümeleri üzerinde etkili sonuçlar üretir ve diğer topluluk öğrenme yöntemlerine göre daha hızlı ve doğru tahminler yapabilir [138].

### **3.2.1.12. Aşırı Gradyan Artırma - eXtreme Gradient Boosting (XGBoost)**

XGBoost (eXtreme Gradient Boosting) algoritması, Gradient Boosting Machines (GBM) algoritmasına dayanan gelişmiş ve optimize edilmiş bir topluluk öğrenme yöntemidir. XGBoost, hem sınıflandırma hem de regresyon problemleri için kullanılabilen güçlü ve etkili bir modeldir. XGBoost, GBM'ye göre daha hızlı eğitim süreleri ve daha iyi performans sunar. XGBoost, zayıf öğrencileri (genellikle karar ağaçları) bir araya getirerek daha güçlü ve doğru bir model oluşturmayı amaçlar. Ayrıca, XGBoost, düzenleme terimleri ekleyerek modelin aşırı uyumunu önlemeye yardımcı olur ve daha iyi genelleme sağlar. Bu, modelin daha kararlı ve daha iyi tahminler yapmasına olanak tanır. XGBoost algoritması, başlıca sınıflandırma ve regresyon algoritması olarak kullanılır, ancak kümeleme algoritması olarak kullanılmaz. XGBoost, genellikle daha büyük ve karmaşık veri kümeleri üzerinde etkili sonuçlar üretir ve diğer topluluk öğrenme yöntemlerine göre daha hızlı ve doğru tahminler yapabilir [139].

### **3.2.1.13. Hafif Gradyan Artırma Makineleri - Light Gradient Boosting Machines (LightGBM)**

LightGBM, Microsoft tarafından geliştirilen ve Gradient Boosting Machines (GBM) algoritması temelli hafif bir topluluk öğrenme yöntemidir. Hem sınıflandırma hem de regresyon problemleri için kullanılabilen etkili ve verimli bir modeldir. LightGBM, diğer GBM türevlerine göre daha hızlı eğitim süreleri ve daha iyi performans sunar. LightGBM, zayıf öğrencileri (genellikle karar ağaçları) bir araya getirerek daha güçlü ve doğru bir model oluşturmayı amaçlar. LightGBM'in önemli özelliklerinden biri, daha büyük ve karmaşık veri kümeleri üzerinde hızlı eğitim süreleri ve ölçeklenebilirlik sunmasıdır. Ayrıca, LightGBM, düşük bellek kullanımı ve daha hızlı tahminler sağlar. LightGBM algoritması, başlıca sınıflandırma ve regresyon

algoritması olarak kullanılır, ancak kümeleme algoritması olarak kullanılmaz. LightGBM, özellikle büyük ve karmaşık veri kümeleri üzerinde etkili sonuçlar üretir ve diğer topluluk öğrenme yöntemlerine göre daha hızlı ve doğru tahminler yapabilir [140].

#### **3.2.1.14. CatBoost**

CatBoost, Yandex tarafından geliştirilmiş Gradient Boosting Machines (GBM) tabanlı bir topluluk öğrenme algoritmasıdır. Hem sınıflandırma hem de regresyon problemleri için kullanılabilen etkili ve verimli bir modeldir. CatBoost, özellikle kategorik değişkenlerin işlenmesinde diğer GBM türevlerine göre daha iyi performans sunar. CatBoost, zayıf öğrencileri (genellikle karar ağaçları) bir araya getirerek daha güçlü ve doğru bir model oluşturmayı amaçlar. CatBoost'un önemli özelliklerinden biri, kategorik verilerin doğrudan işlenmesi için özel olarak tasarlanmış olmasıdır. Bu, kategorik özelliklerin verimli ve doğru bir şekilde kodlanmasını sağlar ve daha doğru tahminler üretir. CatBoost algoritması, başlıca sınıflandırma ve regresyon algoritması olarak kullanılır, ancak kümeleme algoritması olarak kullanılmaz. CatBoost, özellikle kategorik değişkenlerin önemli olduğu veri kümeleri üzerinde etkili sonuçlar üretir ve diğer topluluk öğrenme yöntemlerine göre daha doğru tahminler yapabilir [141].

#### **3.2.2. Regresyon**

Regresyon, bağımlı değişkenin değerini, bir veya birden fazla bağımsız değişkenin değerine göre tahmin etmeyi amaçlar.

##### **3.2.2.1. Lineer Regresyon**

Lineer Regresyon, istatistiksel öğrenme ve MÖ alanlarında kullanılan temel bir regresyon algoritmasıdır. Sürekli bağımlı değişkenlerin ve bir veya birden fazla bağımsız değişken arasındaki ilişkiyi modellemeye yardımcı olur. Lineer regresyon, regresyon problemlerine uygulanır ve sürekli sonuçlar elde etmek için kullanılır. Lineer regresyon algoritması, bağımlı ve bağımsız değişkenler arasında doğrusal bir ilişki varsayarak, en iyi uyumlu doğruyu (veya düzlemi) bulmaya çalışır. Bu, veri



noktaları arasındaki toplam kare hatalarını en aza indiren bir doğru bulmakla eşdeğerdir [142].

### **3.2.2.2. Çoklu Regresyon**

Çoklu Regresyon, istatistiksel öğrenme ve makine öğrenimi alanlarında kullanılan bir regresyon algoritmasıdır. Tek bir bağımlı değişkenin değerini birden fazla bağımsız değişkene dayalı olarak tahmin etmek için kullanılır. Çoklu regresyon, regresyon problemlerine uygulanır ve sürekli sonuçlar elde etmek için kullanılır. Çoklu regresyon, bağımlı ve bağımsız değişkenler arasında doğrusal bir ilişki varsayarak, en iyi uyumlu doğruyu (veya düzlemi) bulmaya çalışır. Bu, veri noktaları arasındaki toplam kare hatalarını en aza indiren bir doğru bulmakla eşdeğerdir. Lineer regresyonun çok değişkenli bir genişlemesidir [142].

### **3.2.2.3. Polinom Regresyon**

Polinom Regresyon algoritması, istatistiksel öğrenme ve MÖ alanlarında kullanılan bir regresyon algoritmasıdır. Polinom regresyon, bağımlı ve bağımsız değişkenler arasında doğrusal olmayan ilişkileri modellemek için kullanılır. Bu, veriye daha iyi uyum sağlayan karmaşık eğriler oluşturarak doğrusal regresyonun genişletilmesidir. Polinom Regresyon, bağımlı değişkenin değerini, bağımsız değişkenlerin kuvvetlerine göre tahmin etmeye çalışır. Örneğin, ikinci dereceden bir polinom regresyon modeli, bağımsız değişkenin karesini içerebilir. Bu, bağımlı değişkenin bağımsız değişkenin farklı kuvvetlerine göre nasıl değiştiğini modellemeye yardımcı olur [142].

### **3.2.2.4. Destek Vektör Regresyonu (DVR)**

Destek Vektör Regresyonu (Support Vector Regression - DVR) algoritması, MÖ alanında kullanılan bir regresyon algoritmasıdır. DVR, DVM algoritmasının regresyon problemleri için uyarlanmış bir versiyonudur. DVM temel olarak sınıflandırma algoritması olarak bilinirken, DVR regresyon problemleri için kullanılır. DVR, veri noktalarını temsil eden örnekleri kullanarak, bağımsız değişkenler ve bağımlı değişken arasındaki ilişkiyi modellemeye çalışır. Amaç, hedef değişkenin gerçek değerlerine en

yakın olan ve buna göre en düşük hata payını sağlayacak şekilde tahminler üreten bir regresyon modeli oluşturmaktır [142].

### **3.2.3. Kümeleme**

Kümeleme, veri noktalarını benzer özelliklere göre gruplara ayırmayı amaçlar. Kümeleme algoritmaları, öğretici olmayan (unsupervised) öğrenme kategorisine girer, çünkü bu tür algoritmalar, etiketli eğitim verilerine dayalı olmaksızın çalışır.

#### **3.2.3.1. k-Ortalamalar**

k-Ortalamalar (k-Means) algoritması, MÖ ve VM alanında kullanılan popüler bir kümeleme (clustering) algoritmasıdır. Algoritma, veri noktalarını k adet küme (cluster) olacak şekilde gruplandırmaya çalışır. Kümeleme, benzer özelliklere sahip veri noktalarını bir araya getirme işlemidir. k-ortalamar algoritması, rastgele başlangıç noktaları (küme merkezleri) seçerek işe başlar ve ardından bu merkezleri iteratif olarak güncelleyerek veri noktalarını en yakın küme merkezine atar. Bu süreç, küme merkezleri hareket etmeyi durdurana veya belirlenen iterasyon sayısına ulaşana kadar devam eder [143].

#### **3.2.3.2. k-Medoids**

k-Medoids (PAM, Partitioning Around Medoids) algoritması, bir kümeleme algoritmasıdır. k-Means algoritmasına benzer bir şekilde çalışır, ancak kümelerin merkezleri olarak veri noktalarının kendilerini (medoid adı verilen) kullanır. Bu özellik, k-Medoids algoritmasının aykırı değerlere karşı daha dayanıklı olmasını sağlar. k-Medoids algoritması, k adet küme oluşturacak şekilde veri kümesini böler. İlk olarak rastgele k adet medoid seçilir ve daha sonra her veri noktası en yakın medoid'e atanır. Bu adımlar tekrar edilerek, her veri noktasının atandığı medoid'e olan uzaklıkların toplamı minimize edilir [144].

### **3.2.3.3. DBSCAN (Density Based Spatial Clustering of Applications With Noise)**

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algoritması, yoğunluk tabanlı bir kümeleme algoritmasıdır. Veri noktalarının yoğun bölgelerini bulmak ve bu bölgelerdeki yoğunluk farklılıklarını kullanarak kümeleri ayırt etmek için tasarlanmıştır. Ayrıca, DBSCAN, veri kümesindeki gürültüyü (aykırı değerleri) de dikkate alarak, veri noktalarını kümeler halinde gruplar. DBSCAN algoritması, belirli bir yoğunlukta ve belirli bir mesafede bulunan komşu veri noktalarını kullanarak kümeleri tanımlar. Algoritma, her veri noktasının etrafında belirli bir yarıçapa sahip bir daire içinde yer alan komşu veri noktalarının sayısını hesaplar ve bu sayı belirli bir eşik değerinin üzerindeyse, bu veri noktası bir yoğun bölge olarak kabul edilir. Bu yoğun bölgeler genişletilir ve birbirine bağlı olanlar birleştirilerek kümeler oluşturulur. Yoğunlukla bağlantılı olmayan veri noktaları, gürültü olarak kabul edilir [145].

### **3.2.3.4. OPTICS (Ordering Points To Identify the Clustering Structure)**

OPTICS (Ordering Points To Identify the Clustering Structure) algoritması, yoğunluk tabanlı bir kümeleme algoritmasıdır ve DBSCAN algoritmasının bir genelleştirmesidir. OPTICS, veri noktalarının yoğunluklarına göre sıralanmış bir veri kümesi oluşturarak kümelerin yapısal özelliklerini analiz eder. Bu algoritma, farklı yoğunluklardaki küme yapılarını algılayabilme yeteneğiyle DBSCAN'dan ayrılır. OPTICS, veri noktalarını bir yoğunluk-tabanlı mesafe ve ulaşılabilirlik mesafesi kullanarak sıralar. Bu sıralama, veri kümesindeki küme yapılarına ilişkin bilgileri korur ve daha sonra bu yapıları çıkartmak için kullanılabilir. OPTICS algoritması, belirli bir yoğunluk ve ulaşılabilirlik mesafesindeki komşu veri noktalarını kullanarak kümeleri tanımlar ve genişletir. Bu sayede, farklı yoğunluk düzeylerindeki kümeleri ayırt edebilir ve bu özelliği, daha esnek ve uyumlu bir kümeleme algoritması yapar [146].

### **3.2.3.5. Yığılmalı Hiyerarşik Kümeleme (YHK)**

Yığılmalı Hiyerarşik Kümeleme (İng. Agglomerative Hierarchical Clustering) algoritması, bir kümeleme algoritmasıdır ve veri noktalarını hiyerarşik bir şekilde gruplara ayırmak için kullanılır. Sınıflandırma veya regresyon algoritması değildir. YHK, alttan yukarı (bottom-up) bir yaklaşımla çalışır. İlk olarak, her veri noktasını kendi başına bir küme olarak kabul eder. Daha sonra, algoritma her adımda en yakın iki kümenin birleşimini gerçekleştirerek daha büyük kümeler oluşturur. Bu süreç, tüm veri noktaları tek bir küme halinde toplanana kadar devam eder. Kümeleme sürecinde kullanılan uzaklık ölçütüne ve kümeleme adımlarında nasıl birleştirme yapılacağına karar vermek önemlidir. Öklid uzaklığı, Manhattan uzaklığı gibi farklı uzaklık ölçümleri kullanılabilir. Ayrıca, küme birleşiminde tek-bağlantı, tam-bağlantı, ortalama-bağlantı gibi farklı bağlantı yöntemleri uygulanabilir. YHK algoritması, hiyerarşik küme yapısı ve veri noktaları arasındaki benzerlik veya uzaklık ölçütüne dayalı kümeler oluşturur. Bu nedenle, kümeleme algoritması olarak kabul edilir [147].

### **3.2.3.6. Bölücü Hiyerarşik Kümeleme (BHK)**

Bölücü Hiyerarşik Kümeleme algoritması, bir kümeleme algoritmasıdır ve veri noktalarını hiyerarşik bir şekilde gruplara ayırmak için kullanılır. Sınıflandırma veya regresyon algoritması değildir. BHK, yukarıdan aşağıya (top-down) bir yaklaşımla çalışır. İlk olarak, tüm veri noktalarını içeren tek bir küme ile başlar. Daha sonra, algoritma her adımda mevcut kümeleri alt kümelerine böler. Bu süreç, her veri noktası kendi başına bir küme haline gelene kadar devam eder. BHK algoritması, önceden belirlenmiş bir küme sayısına ulaşıldığında veya belirli bir durma kriteri karşılandığında durabilir. Bu algoritma için uygun bölme yöntemi ve küme sayısı seçimi önemlidir [147].

### **3.2.3.7. Bulanık c-Ortalamlar (BcO)**

Bulanık c-Ortalamlar (İng. Fuzzy c-Means) algoritması, bir kümeleme algoritmasıdır ve veri noktalarını belirli sayıda kümeye ayırmak için kullanılır. Sınıflandırma veya regresyon algoritması değildir. BcO algoritması, k-Ortalamlar algoritmasının

"bulanık mantık" tabanlı bir varyasyonudur. Bu algoritma, veri noktalarını kesin bir şekilde belirli kümelerle eşleştirmek yerine, her veri noktasının farklı kümelerle olan üyelik derecelerini hesaplar. Bu, her veri noktasının birden fazla kümeye ait olabileceği anlamına gelir, ancak üyelik dereceleri toplamı 1'e eşit olmalıdır.

Bulanık c-Ortalamalar algoritması, aşağıdaki adımları içerir:

1. Başlangıçta rastgele küme merkezleri seçilmelidir.
2. Her veri noktasının her küme merkezine göre üyelik derecelerini hesaplanmalıdır.
3. Yeni küme merkezlerini, üyelik derecelerine göre güncellenmelidir.
4. Algoritma, küme merkezlerindeki değişim belirli bir eşik değerin altına düştüğünde veya belirli bir iterasyon sayısına ulaşıldığında durur.

BcO algoritması, veri noktalarını kümeler halinde gruplandırır ve bu nedenle kümeleme algoritması olarak kabul edilir [148].

### **3.2.3.8. Spektral Kümeleme (SK)**

Spektral Kümeleme algoritması, bir kümeleme algoritmasıdır ve veri noktalarını belirli sayıda kümeye ayırmak için kullanılır. Sınıflandırma veya regresyon algoritması değildir. SK, grafiğe dayalı bir kümeleme yöntemi olarak kabul edilir. Bu algoritma, verilerin benzerlik matrisine dayalı olarak grafik teorisi ve özdeğerlerin (spektral) özelliklerini kullanarak kümeler oluşturur. SK, karmaşık ve düşük yoğunluklu veri yapılarında bile iyi sonuçlar verebilir.

SK algoritması, aşağıdaki adımları içerir:

1. Veri noktaları arasındaki benzerlikleri hesaplayarak benzerlik matrisi oluşturulmalıdır.
2. Benzerlik matrisine dayalı olarak grafiğin Laplace matrisini hesaplanmalıdır.
3. Laplace matrisinin en düşük özdeğerlerine karşılık gelen özvektörleri bulunmalıdır.

4. Bulunan özvektörlerden oluşan matrisi kullanarak, veri noktalarını k-Ortalamlar veya diğer kümeleme algoritmalarıyla kümeler halinde gruplandırılmalıdır.

SK algoritması, veri noktalarını kümeler halinde gruplandırır ve bu nedenle kümeleme algoritması olarak kabul edilir [149].

### **3.2.3.9. Balanced Iterative Reducing and Clustering Using Hierarchies - Hiyerarşileri Kullanarak Dengeli Yinelemeli Azaltma ve Kümeleme (HKDYAK)**

Hiyerarşileri Kullanarak Dengeli Yinelemeli Azaltma ve Kümeleme (HKDYAK) algoritması, büyük veri kümeleri için hızlı ve ölçeklenebilir bir hiyerarşik kümeleme algoritmasıdır. HKDYAK, veriyi özetlemek ve sıkıştırmak için CF (Kümelenme Özelliği) ağacı adı verilen bir ağaç yapısı kullanır. Bu ağaç, veri noktalarının yoğun bölgelerini temsil eden kümeleri (CF) içerir. HKDYAK algoritması, kümeleme sürecini hızlandırmak ve bellek kullanımını azaltmak için veri noktalarının bir özetini kullanır. Ağaç yapısı, veri noktalarının kümeler halinde saklandığı düğümleri ve bu düğümleri bağlayan dalları içerir. Bu sayede HKDYAK, büyük veri kümeleri için daha hızlı ve verimli bir kümeleme algoritması sağlar [150].

### **3.2.3.10. Ortalama Kaydırma (OK)**

Ortalama Kaydırma (İng. Mean Shift) algoritması, yoğunluk tabanlı bir kümeleme algoritmasıdır ve veri noktalarının yoğun bölgelerini belirlemeye çalışır. Algoritma, veri noktalarının lokal yoğunluğunu maksimize etmeye çalışarak küme merkezlerini belirler. Bu süreç, her veri noktası için bir pencere (genellikle bir çekirdek fonksiyonu) kullanarak yoğunluk tahmini yaparak gerçekleştirilir. Ortalama Kaydırma algoritması, veri kümesindeki küme sayısını belirlemek için yoğunluğa dayalı bir yaklaşım kullanır ve bu sayede küme sayısını önceden belirlemeye gerek kalmaz. Bu özelliği, algoritmanın farklı şekil ve yoğunlukta kümeleri tespit etmesini sağlar [151].

### **3.2.3.11. Affinity Propagation – Afinite Yayılılımı (AY)**

AY algoritması, bir kümeleme algoritmasıdır ve veri noktalarını temsilci örnekler temelinde kümeler halinde gruplamayı amaçlar. Algoritma, mesaj geçiş yöntemi kullanarak, veri noktaları arasındaki benzerlikler üzerinden en uygun temsilci örnekleri belirlemeye çalışır. Bu süreçte, iki tür mesaj kullanılır: "uygunluk" ve "sorumluluk". "Uygunluk" mesajları, bir veri noktasının diğer veri noktalarının temsilcisi olma uygunluğunu değerlendirirken, "sorumluluk" mesajları ise bir veri noktasının diğer noktaların temsilcisi olmaya ne kadar istekli olduğunu değerlendirmek için kullanılır. Algoritma, uygunluk ve sorumluluk mesajlarını iteratif olarak güncelleyerek nihai temsilci örnekleri ve kümeleme çözümüne ulaşır. Affinity Propagation, veri kümesindeki küme sayısını önceden belirlemeye gerek kalmadan çalışabilir. Bu özelliği, algoritmanın farklı şekil ve yoğunlukta kümeleri tespit etmesini sağlar [152].

### **3.2.3.12. Gaussian Mixture Models – Gauss Karışım Modeli (GKM)**

GKM algoritması, bir kümeleme algoritmasıdır. Veri setindeki noktaların, birden fazla Gauss dağılımı (normal dağılım) karışımı olarak modellenebileceği varsayımına dayanır. GKM, her Gauss dağılımının parametrelerini (ortalama ve kovaryans) tahmin etmeye çalışır ve veri noktalarını bu dağılımların her birine atama olasılığına göre kümeler halinde gruplar. GKM algoritması, "Expectation-Maximization (EM)" adı verilen iteratif bir optimizasyon süreci kullanarak dağılım parametrelerini ve veri noktalarının dağılımlara ait olma olasılıklarını öğrenir. Algoritma, EM süreci belirli bir kriterle karşılanana veya maksimum iterasyon sayısına ulaşına kadar devam eder. GMM, veri kümesindeki temel yapıyı ve dağılımları yakalamak için kullanılır, bu nedenle kümeleme algoritması olarak kabul edilir. Sınıflandırma veya regresyon problemleri için doğrudan kullanılamaz, ancak veri ön işleme veya öznetelik çıkarımı gibi adımlarda yardımcı olabilir [153].

### 3.2.3.13. COBWEB

COBWEB algoritması, hiyerarşik bir kümeleme algoritmasıdır. COBWEB, 1980'lerde Fisher tarafından önerilen ve kavramsal gruplamayı (conceptual clustering) destekleyen bir algoritmadır. COBWEB, sadece veri noktalarını gruplandırmakla kalmaz, aynı zamanda her kümenin öznitelikleri üzerinde yapılandırılmış bir kavramsal tanım sağlar.

COBWEB, aşağıdaki adımları içeren bir süreçle çalışır:

1. Rastgele bir öznitelik ve değeri seçerek ağacın kök düğümünden başlar.
2. Yeni bir veri noktası alır ve ağaç boyunca uygun dallara yerleştirir.
3. Düğümleri bölebilir veya birleştirebilir, böylece daha iyi yapılandırılmış bir kavramsal hiyerarşi oluşturur.
4. Algoritma, veri noktaları tükenene kadar bu süreci yineler.

COBWEB, kavramsal gruplama algoritması olduğu için, veri noktalarını kümeler halinde gruplandırır ve kümelerin özniteliklerine göre yapılandırılmış bir tanım sağlar. Bu nedenle, COBWEB bir kümeleme algoritması olarak kabul edilir ve doğrudan sınıflandırma veya regresyon problemleri için kullanılamaz. Ancak, veri analizinde veya özellik seçimi gibi uygulamalarda kullanılabilir [154].

### 3.2.3.14. Self - Organizing Maps – Kendini Düzenleyen Algoritmalar (KDA)

Self-Organizing Maps (KDA) algoritması, 1980'lerde Teuvo Kohonen tarafından geliştirilen unsupervised (öğretmensiz) bir yapay sinir ağı modelidir. KDA, özellikle yüksek boyutlu verilerin düşük boyutlu uzaylara haritalanması (boyut indirgeme) ve verilerin kümeleme işlemlerinde kullanılır. KDA, öğrenme sürecinde veri noktalarını düşük boyutlu uzayda (genellikle 2 boyutlu) organize ederek ve benzer veri noktalarını birbirine yakın hale getirerek çalışır. KDA, kümeleme algoritması olarak kabul edilir çünkü veri noktalarını benzerliklerine göre gruplandırır. Ayrıca, boyut indirgeme ve veri görselleştirme amacıyla da kullanılabilir. KDA doğrudan sınıflandırma veya regresyon problemleri için kullanılmaz, ancak bu tür problemler için ön işleme ve veri analizi aşamasında kullanılabilir [155].



Covid-19 risk faktörlerinin yaygınlığını ve etkilerini analiz etme sürecinde Sınıflandırma, regresyon ve kümeleme tekniklerinin kullanılması önemlidir. Bu tekniklerin doğru uygulanması ve yorumlanması, tez çalışması kapsamında hastalığın teşhisi ve risk faktörlerinin anlaşılmasında değerli sonuçlar elde etmeye yardımcı olacaktır. Bu tekniklerin kronolojisi, tarih boyunca istatistiksel modellere ve daha sonra makine öğrenimi ve yapay zeka alanındaki gelişmelere dayanmaktadır. İlk regresyon ve sınıflandırma yöntemleri 19. ve 20. yüzyıllarda ortaya çıkmıştır, kümeleme yöntemleri ise daha yakın dönemlerde, özellikle 1960'lardan itibaren yaygınlaşmıştır.

## BÖLÜM 4

### ANALİZ ORTAMI

#### 4.1. KNIME YAZILIMI

KNIME (Konstanz Information Miner) Yazılımı ve Özellikleri:

Genel Bakış: KNIME, 2004 yılında Konstanz Üniversitesi'nde geliştirilmeye başlanan açık kaynaklı, platformlar arası çalışabilen ve veri analizi, işleme ve entegrasyonu için güçlü bir araç olan bir veri madenciliği ve makine öğrenimi yazılımıdır. KNIME, büyük veri setlerini analiz etmek, model oluşturmak ve raporlamak için kullanılmaktadır [156]. KNIME, veri analizi, veri işleme, veri entegrasyonu ve MÖ için kullanılan açık kaynaklı ve platformlar arası çalışabilen bir yazılım aracıdır. KNIME, düğüm tabanlı bir yapı sunarak kullanıcıların sürükle ve bırak yöntemiyle iş akışları oluşturmaya ve veri analizi süreçlerini gerçekleştirmesine olanak tanır.

Tez çalışmamızda KNIME platformunu kullanmamızın temel nedenleri şu şekilde listelenmiştir.

1. Kullanıcı dostu ve görsel arayüz: KNIME, sezgisel bir sürükle ve bırak arayüzü sunarak, karmaşık veri işleme, analiz ve modelleme süreçlerini kolaylaştırır. Bu özellik, hem başlangıç hem de ileri düzey kullanıcılar için uygun hale getirir.
2. Geniş algoritma yelpazesi: KNIME, veri madenciliği ve makine öğrenimi algoritmalarının geniş bir yelpazesine sahip olduğu için, tez çalışmamızda kullanılacak en uygun yöntemleri seçmeyi ve test etmeyi kolaylaştırır.

3. Esneklik ve özelleştirme: KNIME, kullanıcıların veri işleme ve analiz süreçlerini özelleştirmelerine ve geliştirmelerine olanak tanır. Bu, tez çalışmamızda özgün ve özel ihtiyaçlara uygun çözümler üretebilme avantajı sağlar.
4. Açık kaynak ve ücretsiz: KNIME, açık kaynak bir platform olduğu için, tez çalışmamız için ekstra maliyet gerektirmez ve süreçlerin şeffaflığını artırır.
5. Topluluk desteği ve eklentiler: KNIME, geniş ve aktif bir kullanıcı topluluğuna sahip olduğu için, sorunlarla karşılaştığımızda veya özel ihtiyaçlarımız olduğunda destek ve kaynaklara kolayca erişebiliriz. Ayrıca, topluluk tarafından geliştirilen eklentiler ve özellikler sayesinde platformun işlevselliğini genişletebiliriz.
6. Entegrasyon yeteneği: KNIME, diğer popüler veri analizi ve makine öğrenimi araçlarıyla (Python, R, SQL, Hadoop, Spark vb.) entegrasyon sağlayarak, çalışmalarımızda kullanabileceğimiz yöntemlerin ve kütüphanelerin kapsamını genişletir.
7. Kodlama bilgisi gerektirmez: KNIME, kullanıcıların herhangi bir programlama dili bilmeden veri analizi ve makine öğrenimi süreçlerini gerçekleştirebilmelerine olanak sağlar. Bu sayede, kodlama bilgisi olmayan kullanıcılar bile veri işleme ve modelleme süreçlerinde rahatlıkla çalışabilir.
8. Görsel akış tabanlı programlama: KNIME, görsel akış tabanlı programlama yaklaşımını benimser, bu da kullanıcıların algoritmaları ve veri işleme adımlarını görsel bloklar (düğümler) olarak sürükleyip bırakarak birleştirmelerine olanak tanır. Bu, özellikle de kodlama deneyimi olmayan kullanıcılar için işlemlerin ve analizlerin daha kolay anlaşılmasını ve yönetilmesini sağlar.
9. Önceden hazırlanmış düğüm ve örnekler: KNIME, önceden hazırlanmış düğüm ve örneklerle birlikte gelir, böylece kullanıcılar başlamak için bir temele sahip olur ve karmaşık kodlama gerektirmeyen analizler gerçekleştirebilir. Bu,

kodlama bilgisi olmayan kullanıcıların hızla başlamasına ve verimli bir şekilde çalışmasına olanak tanır.

10. Kademeli öğrenme eğrisi: KNIME, kullanıcıların ihtiyaçlarına ve bilgi düzeyine uyum sağlayan kademeli bir öğrenme eğrisi sunar. Platform, başlangıç seviyesinden ileri düzeye kadar farklı kullanıcılar için uygun öğrenme materyalleri ve belgeler sunarak, kodlama bilgisi olmayanlar için bile öğrenmeyi ve başlamayı kolaylaştırır.

Bu nedenlerle, tez çalışmamızda veri madenciliği ve makine öğrenimi yöntemlerinin uygulanması için KNIME platformu tercih edilmiştir.

#### **4.1.1. Kronoloji**

KNIME, 2004 yılında Almanya'nın Konstanz şehrindeki Konstanz Üniversitesi'nde geliştirilmeye başlanmıştır. Başlangıçta akademik bir projeyken, zamanla büyük popülarite kazanmış ve endüstriyel ve akademik kullanıcılar tarafından benimsenmiştir. 2008 yılında KNIME AG şirketi kurulmuştur ve KNIME yazılımının gelişimi ve desteği bu şirket tarafından sağlanmaktadır [156].

KNIME programında, Covid-19 risk faktörlerinin yaygınlığını anlamak için gerçekleştirilen ve kamuya açık, 1677 hastada PCR testleri esnasında alınan kan tahlil değerleri ile birlikte 35 değişkenden oluşan veri seti üzerinde MÖ ve VM yöntemleri kullanarak yapay zeka tekniklerinin Covid-19 hastalık teşhisi için kullanılmasına yönelik bir çalışması yapılmıştır. KNIME, bu çalışmada veri ön işleme, özellik seçimi, model eğitimi ve performans değerlendirmesi gibi adımlarda kullanılmış ve elde edilen sonuçlar, farklı algoritmalar ve yöntemlerle karşılaştırılmıştır. KNIME programının kullanılması, tez çalışması boyunca analiz süreçlerini kolaylaştırmış ve deneylerin hızlı bir şekilde gerçekleştirilmesine olanak tanımıştır.

### 4.1.2. Özellikler

KNIME, kullanıcı dostu bir arayüz sunar ve sürükle-bırak yöntemiyle veri işleme ve analizi için modüller (düğüm) eklemeye olanak tanımaktadır. Bu düğümler birbirine bağlanarak karmaşık analiz ve işleme iş akışları oluşturabilmektedir. KNIME, R, Python, Java ve diğer dillerde yazılan özel düğümlerle entegrasyon sağlar, böylece kullanıcılar kendi algoritmalarını veya fonksiyonlarını ekleyebilirler. Kullanıcı dostu ve sezgisel bir arayüz sunarak, kullanıcıların teknik olmayan bir geçmişe sahip olsalar bile analizleri kolayca gerçekleştirmelerine olanak tanımaktadır. Bu, tez çalışması sürecinde analizlere daha hızlı ve etkili bir şekilde başlamayı sağlamıştır. KNIME, kullanıcıların özel düğümler ve iş akışları oluşturarak analizlerini özelleştirmelerine olanak tanımaktadır. Açık kaynaklı bir platform olduğu için, kullanıcılar tarafından geliştirilmiş yeni düğümler ve eklentiler ile sürekli olarak genişletilebilir. Python ve diğer dillerle entegrasyon sağlayarak, kullanıcıların mevcut programlama becerilerini ve dış kaynaklı analiz araçlarını kullanmalarına olanak tanımaktadır. KNIME, ücretsiz ve açık kaynaklı bir platform olduğu için, tez çalışması sırasında ek maliyetler olmadan güçlü analitik araçlara erişim sağlamaktadır. Bu, tez çalışmasının bütçesini ve kaynaklarını optimize etmeye yardımcı olmuştur.

### 4.1.3. Çalışma Prensipleri

KNIME, düğüm tabanlı bir yapıda çalışır. Kullanıcılar, veri okuma, veri ön işleme, MÖ algoritmaları, istatistiksel analiz ve görselleştirme gibi işlemlere sahip düğümleri iş akışına ekleyerek ve düğümleri birbirine bağlayarak analiz süreçlerini gerçekleştirebilirler. Her düğüm, belirli bir görevi gerçekleştirmek için tasarlanmıştır ve düğümleri birleştirerek karmaşık işlemler gerçekleştirilebilir.

### 4.1.4. Kullanım Alanları

KNIME, çeşitli endüstrilerde ve araştırma alanlarında kullanılmaktadır. Başlıca kullanım alanları şu şekilde sıralanabilir:

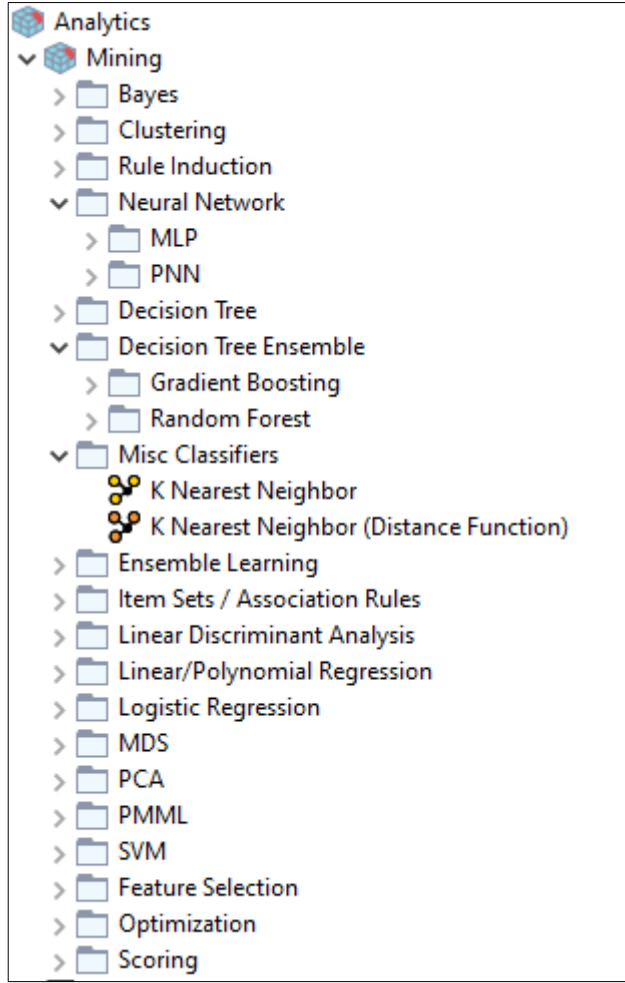
1. Veri Madenciliği: Büyük veri setlerini analiz etmek, veri dönüşümü ve ön işleme yapmak, özellik seçimi ve görselleştirme işlemlerinde kullanılmaktadır.
2. Makine Öğrenimi: KNIME, sınıflandırma, regresyon, kümeleme ve boyut indirgeme gibi MÖ algoritmalarının uygulanması için popüler bir platformdur.
3. Veri Entegrasyonu: KNIME, çeşitli veri kaynaklarından gelen verilerin entegrasyonunu ve dönüşümünü sağlar, böylece tek bir ortamda analiz yapılabilmektedir.
4. Kimya ve Biyoenformatik: KNIME, moleküler modelleme, biyoenformatik ve kimyasal bilgi işlem alanlarında da yaygın olarak kullanılmaktadır.
5. Müşteri İlişkileri Yönetimi (CRM) ve Pazarlama: KNIME, müşteri segmentasyonu, hedefleme ve pazarlama kampanyalarının optimizasyonu gibi alanlarda kullanılmaktadır.
6. Finans ve Risk Yönetimi: KNIME, finansal analiz, kredi riski tahmini ve portföy yönetimi gibi finansal uygulamalar için kullanılmaktadır.

KNIME, iş akışlarının kolayca paylaşılmasına ve yeniden kullanılmasına olanak tanır. Bu sayede, başkalarının oluşturduğu iş akışlarını inceleyebilir, üzerinde değişiklikler yaparak bireysel analizler için kullanılabilir. Temel işlevselliğine ek olarak, geniş bir yelpazede eklenti ve özel düğümler sunmaktadır. Bu sayede, daha spesifik görevler için özelleştirilmiş iş akışları oluşturmak mümkündür. Ayrıca, Python, R ve Java gibi dillerle kod yazarak özel düğümlerinizi oluşturulabilmektedir. Veri analizi süreçlerinde kullanıcıya interaktif ve özelleştirilebilir görselleştirmeler sunmaktadır. İş akışlarının içinde yer alan düğümlerle, grafikler ve tablolar oluşturularak analiz süreçlerini anlaşılır hale getirebilmektedir. Veri işleme ve analiz süreçlerini paralel olarak gerçekleştirerek performansı arttırmaktadır. Ayrıca, Apache Hadoop ve Apache Spark gibi büyük veri teknolojileri ile entegre çalışarak büyük veri setlerini işlemek için uygun bir ortam sağlamaktadır.

Bu bölümde, KNIME programının tanımı, kullanım alanları, çalışma prensipleri ve özellikleri hakkında detaylı bilgi sunulmaktadır. Bu bilgiler, tez çalışmasında KNIME programının nasıl kullanıldığına ve analiz süreçlerinin gerçekleştirilmesine yönelik temel bilgileri sağlamaktadır.

#### **4.2. KNIME İLE VERİ ÖN İŞLEME, MODELLEME VE DEĞERLENDİRME**

KNIME, çeşitli makine öğrenimi algoritmalarını uygulamak için düğümler sunmaktadır. Bu algoritmalar sınıflandırma, regresyon ve kümeleme gibi görevler için kullanılabilir. KNIME'de yerleşik olarak gelen algoritmaların yanı sıra, özel düğümler ve eklentiler ile ek algoritmalar da kullanılabilir. KNIME, sınıflandırma, regresyon ve kümeleme gibi temel MÖ görevlerini gerçekleştirmek için çeşitli algoritma düğümlerinin kullanımına olanak sağlamaktadır. Destek Vektör Makineleri (SVM), PNN Learner (DDA), Karar Ağaçları, Random Forest, Naive Bayes, K-En Yakın Komşu (K-EYK), Logistic Regression ve Yapay Sinir Ağları (ANN) gibi yaygın algoritmaları içermektedir. Bu algoritmalar, Covid-19 teşhis sürecinde kullanılarak hastaların durumlarını tahmin etmeye yardımcı olmuştur. Şekil 4.1'de KNIME yazılımının Düğüm Deposu penceresine ait Madencilik yapısı belirtilmiştir.



Şekil 4.1. Knime yazılımı algoritma düğümleri

KNIME, eğitilmiş modellerin performansını değerlendirmek için çeşitli metrikler ve görselleştirme araçları sunmaktadır. Bu araçlar, modelin doğruluğunu, hassasiyetini, hatırlama oranını, F1 puanını ve ROC eğrisini hesaplamak ve görselleştirmek için kullanılmıştır. Ayrıca, karışıklık matrisi ve sınıflandırma raporu gibi detaylı analizler de gerçekleştirilmiştir.

KNIME, modellerin hiperparametrelerini otomatik olarak optimize etmek için Grid Search, Random Search ve Bayesian Optimization gibi yöntemler sunmaktadır. Bu yöntemler, en iyi model performansını sağlamak için hiperparametrelerin ideal değerlerini bulmaya yardımcı olmuştur.



## 4.3. KNIME' DA KULLANILAN ARAÇLAR

### 4.3.1. Veri Ön İşleme Araçları

KNIME, eksik veri doldurma, özellik seçimi, ölçeklendirme, normalizasyon ve kategorik değişken dönüşümü gibi önemli veri ön işleme görevlerini gerçekleştirebilmek için geniş bir yelpazede düğüm ve araç sunmaktadır. Bu araçlar, veri setinin temizlenmesi ve modelleme süreci için hazırlanmasında büyük kolaylık sağlamıştır.

#### 4.3.1.1. Veri Seti ve Ön İşleme

1. Veri setinin anlaşılması ve analizi: İlk adım, kullanılacak veri setini detaylı bir şekilde incelenmiştir. Bu aşamada, veri setinde bulunan değişkenlerin ve gözlemlerin sayısı, eksik veriler, veri türleri ve veri setinin yapısı gibi temel bilgileri anlamak önemli olmuştur.
2. Eksik verilerin ele alınması: Veri setinde eksik değerlerin uygun bir şekilde ele alınması gerekiyordu. Eksik değerler, gözlem veya özellik ortalaması ile doldurulmuş ya da eksik değerlerin bulunduğu anlamsız gözlemler veri setinden çıkartılmıştır. Eksik veri işleme yöntemleri seçilirken, veri setinin yapısı ve eksik veri mekanizması göz önünde bulundurulmuştur.
3. Özellik seçimi: Veri setindeki özelliklerin (değişkenlerin) önemini ve etkisini değerlendirmek önem arz etmiştir. İlgili ve etkili özellikleri seçmek, modelin doğruluğunu ve performansını arttırmıştır. Özellik seçimi için filtre, gömülü ve sarıcı yöntemler gibi çeşitli teknikler kullanılmıştır.
4. Veri dönüşümü: Veri setindeki kategorik değişkenlerin, makine öğrenimi algoritmalarına uygun bir biçimde sayısal değerlere dönüştürülmüştür.
5. Veri ölçeklendirme ve normalizasyon: Veri setindeki değişkenlerin ölçeklerinde bazı farklılıkların olduğu gözlemlenmiştir. Bu durum, bazı MÖ algoritmalarının

performansını olumsuz etkilemektedir. Veri ölçeklendirme ve normalizasyon yöntemleri (min-max ölçeklendirme, z-skor normalizasyonu vb.) ile değişkenlerin ölçekleri uyumlu hale getirilmiştir.

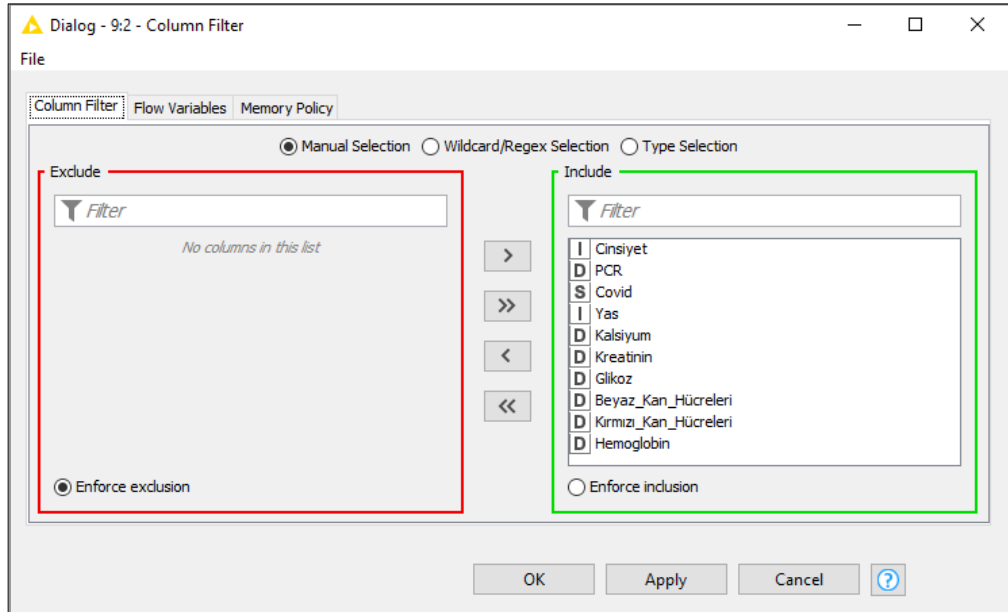
6. Veri bölme: Veri setini eğitim ve test setlerine ayırmak, modelin performansını değerlendirmekte ve genelleştirme yeteneğini ölçmekte önemlidir. Tezde kullanılan veri setinin %70'i eğitim seti, geri kalan %30 ise test seti olarak kullanılmıştır.
7. Aykırı değerlerin işlenmesi: Veri setindeki aykırı değerler, MÖ modelinin performansını olumsuz etkilemektedir. Aykırı değerlerin tespiti ve işlenmesi için çeşitli teknikler kullanılmıştır. Öncelikle, veri setinin görselleştirilmesi ile aykırı değerlerin varlığı incelenmiştir. Aykırı değerleri işlemek için kullanılacak yöntemler arasında aykırı değerleri kaldırmak, aykırı değerlerin üzerinde dönüşüm gerçekleştirmek (logaritmik, kök alma vb.), ya da aykırı değerleri medyan veya ortalama değerlerle değiştirmek gibi yöntemler kullanılmıştır.
8. Özellik mühendisliği: Mevcut özelliklerin dönüşümü, yeni özelliklerin oluşturulması veya mevcut özelliklerin birleştirilmesi ile veri seti üzerinde daha derinlemesine analizler yapılmış ve model performansları artırılmıştır. Özellik mühendisliği, özelliklerin anlamını ve önemini anlamak için veri kümesine ve iş probleminin özelliklerine dayalı olarak gerçekleştirilmiştir.
9. Model seçimi ve hiperparametre optimizasyonu: MÖ algoritmalarının ve hiperparametrelerinin doğru bir şekilde seçilmesi ve ayarlanması, modelin performansını ve doğruluğunu önemli ölçüde etkilemiştir. Model seçimi ve hiperparametre optimizasyonu için kapsamlı bir araştırma ve deneme yanılma süreci izlenmiştir. Hiperparametre optimizasyonu için kullanılan yöntemler arasında ızgara arama (grid search), rastgele arama (random search) ve bayesyen optimizasyon bulunmaktadır.

Bu süreçler, veri ön işleme, modelleme ve değerlendirme adımlarında başarıyla uygulanmış, Covid-19 hastalık teşhisi için kullanılacak yapay zeka tekniklerinin etkinliği ve doğruluğu arttırabilme yeteneğine sahip olduğu gözlemlenmiştir.

#### 4.3.1.2. Veri Temizleme, Eksik Veri Doldurma ve Dönüşümler

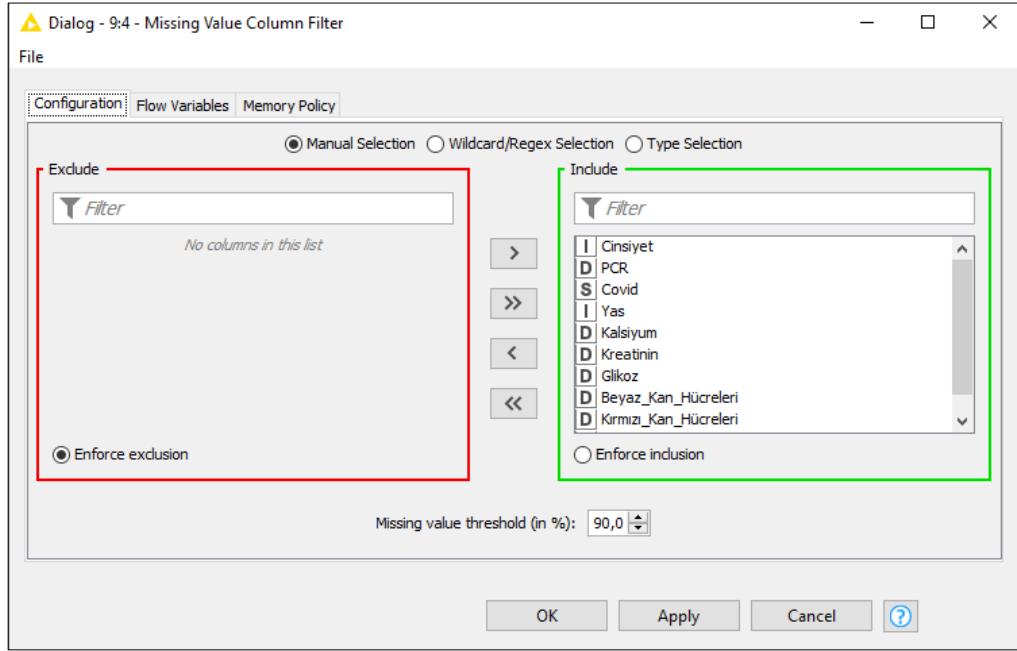
Veri temizleme, eksik veri doldurma ve dönüşümler konusunda yapılacak çalışma, veri setinin analiz ve modelleme aşamalarında doğru ve etkili sonuçlar elde etmek için önemlidir. Bu süreçte, veri setinin öncelikle temizlenmesi, eksik verilerin doldurulması ve gerekli dönüşümlerin gerçekleştirilmesi gerekmektedir. Aşağıda bu süreçlerin detayları ve KNIME araçlarının kullanımı hakkında bilgi verilmektedir.

1. Veri temizleme: İlk aşama olarak, veri setinde yer alan hatalı, tutarsız veya anlamsız verilerin temizlenmesi gerekmektedir. Bu aşamada, KNIME'da bulunan "Column Filter" ve "Missing Value Column Filter" araçları kullanılmıştır. Ayrıca, veri setindeki aykırı değerlerin tespiti ve çözülmesi için "Duplicate Row Filter", "Row Filter", "Tukey Outlier Detection" ve "Z-Score Normalization" gibi araçlar da kullanılabilir. Column Filter düğümüne ait dialog penceresi Şekil 4.2'de yer almaktadır.



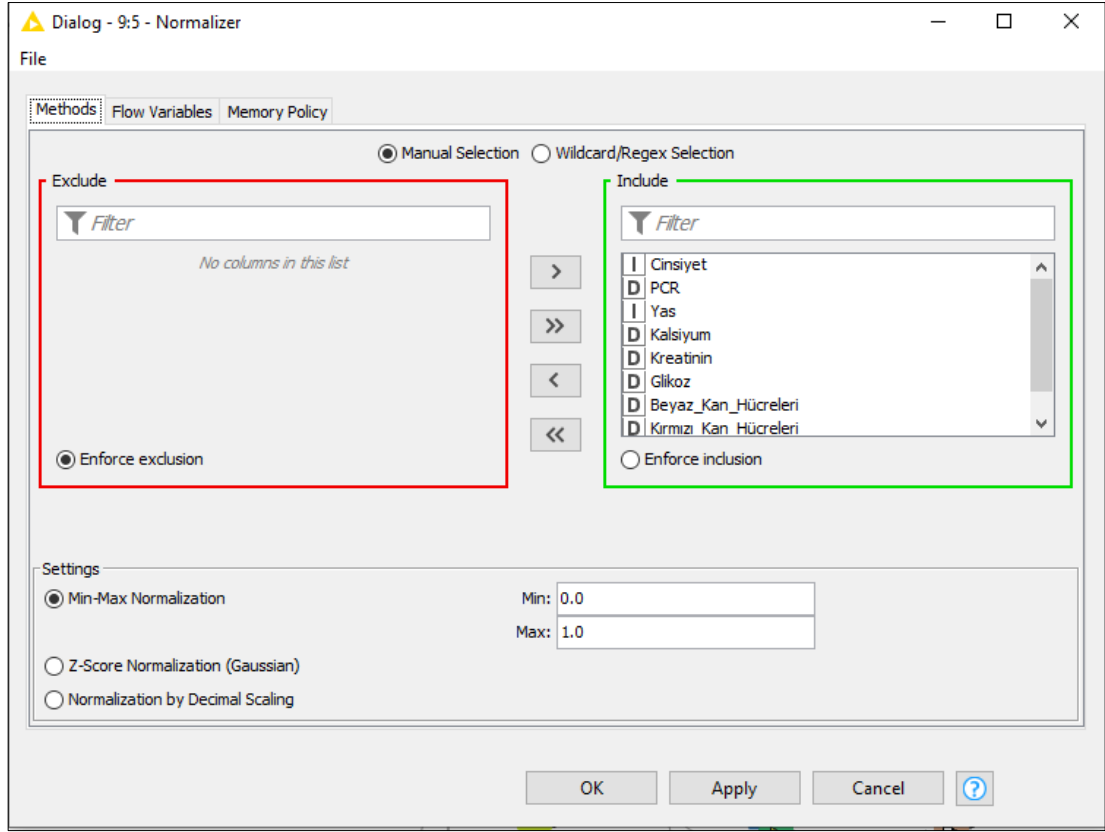
Şekil 4.2. Column Filter düğümü Dialog penceresi

2. Eksik veri doldurma: Veri setinde eksik olan değerlerin uygun yöntemlerle doldurulması, analizlerin doğruluğunu ve güvenilirliğini artırmaktadır. KNIME'da eksik verileri doldurmak için "Missing Value Column Filter" düğümü kullanılmıştır. Bu düğüm ile eksik değerlerin ortalama, medyan, mod gibi istatistiksel yöntemlerle ya da sabit değerlerle doldurulması sağlanabilir. Missing Value düğümüne ait dialog penceresi Şekil 4.3'de yer almaktadır.



Şekil 4.3. Missing Value düğümü Dialog penceresi

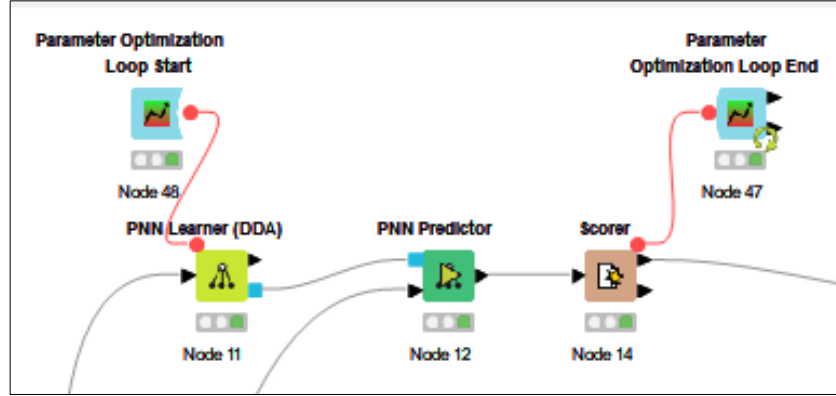
3. Veri dönüşümleri: Veri setindeki değişkenlerin uygun ölçek ve formatta olması, analizlerin ve modellemenin başarısını etkilemektedir. Veri dönüşümleri için KNIME'da "Normalizer" düğümü kullanılmıştır. Ayrıca "Log Transform", "Min-Max Scaling" ve "Z-Score Scaling" gibi düğümler de "Normalizer" düğümü içerisinde ayrı ayrı test edilerek algoritmanın performansı ölçülmüştür. Kategorik değişkenlerin sayısal değerlere dönüştürülmesi için "One-Hot Encoding" veya "Ordinal Encoding" gibi düğümler de kullanılabilir. Normalizer düğümü Dialog penceresine ait görsel Şekil 4.4'de yer almaktadır.



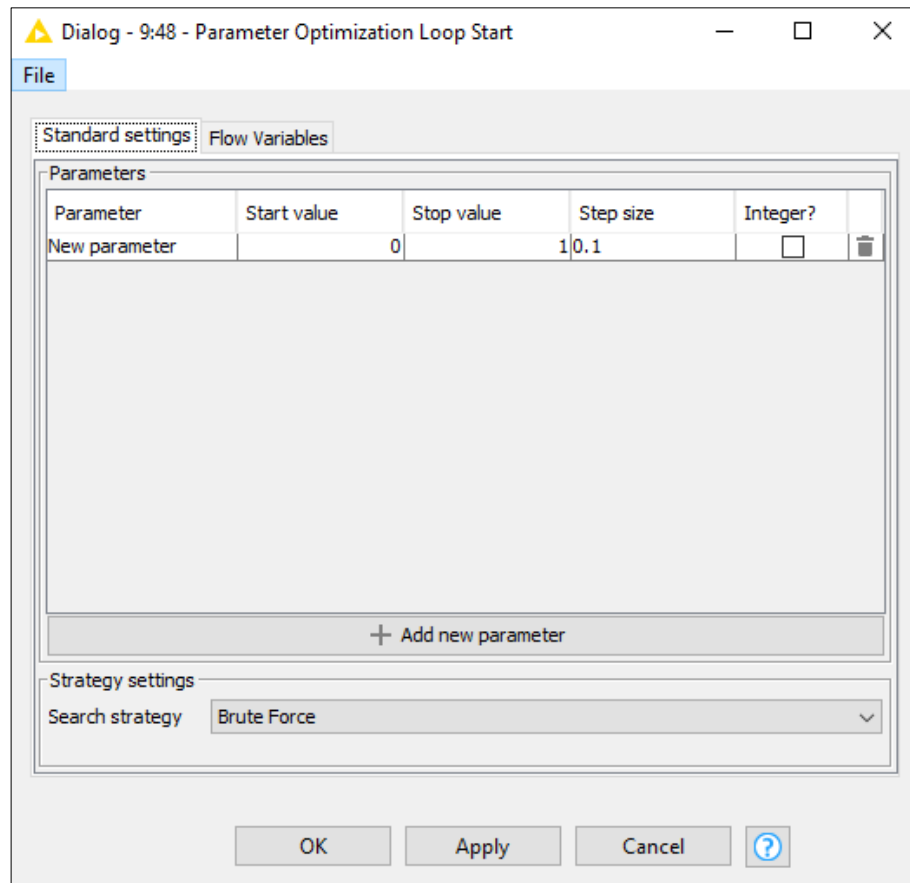
Şekil 4.4. Normalizer düğümü Dialog penceresi

4. "Parameter Optimization Loop Start" ve "Parameter Optimization Loop End: Bu bileşenleri, hiperparametre optimizasyonu yapmak için kullanılan döngü bileşenleridir. Bu bileşenler, farklı hiperparametre değerlerini denemek ve en iyi performansı sağlayan hiperparametre kombinasyonunu bulmak için kullanılmıştır. *Parameter Optimization Loop Start (Parametre Optimizasyonu Döngüsü Başlangıcı)*: Bu bileşen, hiperparametreleri tanımlamak ve döngüyü başlatmak için seçilmiştir. İçerisine yerleştirilen bileşenler, hiperparametrelerin farklı değerlerini alacak şekilde yapılandırılmıştır. Bu bileşenin çıkışı, döngüdeki her bir iterasyon için kullanılacak hiperparametre değerlerini içeren bir tablo üretmeyi sağlamıştır. *Parameter Optimization Loop End (Parametre Optimizasyonu Döngüsü Sonu)*: Bu bileşen, döngüyü sonlandırmak ve en iyi hiperparametre kombinasyonunu seçmek için eklenmiştir. İçerisine yerleştirilen bileşenler, modelin eğitimini gerçekleştirecek ve performansını değerlendirecek olan bileşenlerdir. Bu bileşenin girişi, her bir iterasyonda kullanılan hiperparametre değerlerini içeren bir tablodur. Çıkışı ise en iyi performansı

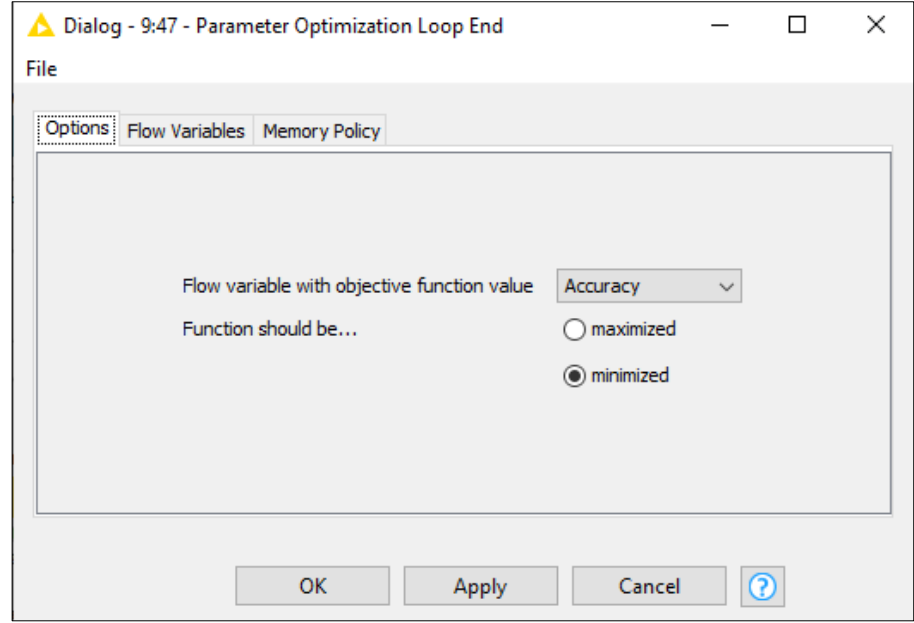
gösteren hiperparametre kombinasyonunu ve modelin sonuçlarını içeren bir tablo üretir. Şekil 4.5’de Parameter Oprimization düğümünün kullanım şekline ait görsel yer almaktadır. Parameter Oprimization Loop Start ve End düğümünün özellik pencereleri ve kullanımını Şekil 4.6 ve Şekil 4.7’de belirtilmiştir.



Şekil 4.5. Parameter Oprimization düğümünün kullanım şekli



Şekil 4.6. Parameter Oprization Loop Start Düğümü Dialog Penceresi



Şekil 4.7. Parameter Optization Loop End Dügümü Dialog Penceresi

5. Nelere dikkat edilmeli? Veri temizleme, eksik veri doldurma ve dönüşümler sürecinde, veri setinin yapısını ve değişkenlerin özelliklerini dikkate almak önemlidir. Ayrıca, yapılan işlemlerin ve kullanılan yöntemlerin, analizlerin ve modellemenin doğruluğunu ve güvenilirliğini olumsuz etkilemeyecek şekilde seçilmesi gerekmektedir.

Sonuç olarak, veri temizleme, eksik veri doldurma ve dönüşümler sürecinde KNIME araçlarının doğru ve etkili kullanımı, tez kapsamında gerçekleştirilecek analizlerin ve modellemenin başarısını arttırmıştır. Bu süreçlerin titizlikle ve dikkatle gerçekleştirilmesi, Covid-19 hastalığına ilişkin risk faktörlerinin daha doğru ve güvenilir bir şekilde belirlenmesine ve hastalık teşhisinde AI tekniklerinin etkinliğinin değerlendirilmesine katkı sağlayacağı düşünülmektedir.

#### 4.3.1.3. Model Geliştirme ve Optimizasyon

Model geliştirme ve optimizasyon konusunda yapılacak çalışma, veri setine uygulanacak MÖ algoritmalarının seçimi, modelin eğitimi ve değerlendirilmesi,

hiperparametre optimizasyonu ve model performansının incelenmesi aşamalarını içermektedir. Bu süreçte şu adımlar izlenmiştir:

1. Algoritma seçimi: Tezde kullanılacak makine öğrenimi algoritmalarının seçimi oldukça önemliydi. Sınıflandırma, Regresyon ve Kümeleme gibi yöntemler arasından, Covid-19 hastalığı teşhisi için en uygun olan Sınıflandırma ve Regresyon yöntemleri belirlenmiştir.
2. Model eğitimi ve değerlendirme: Seçilen algoritmalar kullanılarak modelin eğitimi gerçekleştirilmiştir. Veri seti eğitim ve test verisi olarak ayrılmıştır. Modelin doğruluğunu ve genelleştirilebilirliğini değerlendirmek için çapraz doğrulama gibi yöntemler uygulanmıştır.
3. Hiperparametre optimizasyonu: Modelin performansını artırmak için hiperparametre optimizasyonu yapılmıştır. KNIME, bu amaçla Grid Search ve Random Search gibi yöntemlerin kullanılmasına olanak sağlar. Bu süreçte modelin en iyi parametrelerini belirlemek önemli olmuştur.
4. Model performansının incelenmesi: Geliştirilen modelin performansını değerlendirmek için çeşitli metrikler kullanılmıştır. Başarı oranı, hassasiyet, duyarlılık ve F1 skoru gibi metrikler, modelin başarısını ölçmek için kullanılmıştır. Ayrıca, karmaşıklık matrisi ve ROC eğrisi gibi görselleştirmeler de modelin performansını değerlendirmede yardımcı olmuştur.
5. KNIME araçlarının kullanılması: KNIME platformu, model geliştirme ve optimizasyon sürecinde kullanılacak bir dizi araç sunmaktadır. Veri bölme, çapraz doğrulama, hiperparametre optimizasyonu ve model performans metriklerinin hesaplanması için gerekli düğümler KNIME platformunda mevcuttur.

Bu süreçler dikkatli ve titizlikle uygulanmıştır. Covid-19 hastalığının teşhisinde kullanılan yapay zeka modellerinin doğruluğunu ve güvenilirliğini artırarak, tezin başarısına önemli ölçüde katkı sağladığı düşünülmektedir.



### 4.3.2. Kullanılan Algoritmalar ve Yöntemlerin Seçimi

Kullanılan algoritmalar ve yöntemlerin seçimi için uygun algoritmaların belirlenmesinde dikkate alınan önemli kriterler şunlardır:

1. Problemin türü: Covid-19 teşhisi, sınıflandırma problemi olarak ele alınmıştır. Bu nedenle, sınıflandırma ve regresyon algoritmalarının kullanılması uygun olacağı düşünülmüştür.
2. Model karmaşıklığı: Basit ve anlaşılır modeller tercih edilmiştir. Karmaşık modellerin eğitimi ve yorumlanması daha zor olduğundan ve aşırı uyuma (overfitting) yol açacağı için tercih edilmemiştir.
3. Model seçimi: Covid-19 teşhisi için kullanılacak en uygun modelin seçimi önemlidir. Bu seçim, daha önce belirlenen performans ölçütlerine göre yapılmıştır.
4. Hız ve ölçeklenebilirlik: Hızlı eğitim ve tahmin süreleri sağlayan, büyük veri setleriyle çalışabilen algoritmalar tercih edilmiştir.
5. Performans: Yüksek doğruluk, hassasiyet ve duyarlılık sağlayan algoritmalar seçilmiştir.

Covid-19 teşhisi için kullanılacak algoritmaların seçimi, yukarıda maddeler halinde belirtilen kriterlere göre yapılmıştır. Aşağıda, bu kriterlere göre seçilen popüler sınıflandırma algoritmaları listelenmiştir:

1. Lojistik Regresyon (LR): İstatistiksel bir sınıflandırma yöntemi olan lojistik regresyon, basit ve hızlıdır. İyi ölçeklenebilir ve yorumlanabilirlik açısından avantajlıdır.
2. Karar Ağaçları (KA): Karar ağaçları, veriyi dallanarak sınıflandıran yapıları kullanır. Basit ve anlaşılır olması nedeniyle tercih edilmiştir.

3. Rastgele Orman (RO): Karar ağaçlarının birleştirilmesiyle oluşturulan rastgele orman algoritması, yüksek performans sağlar ve aşırı uyum problemini azaltmaktadır.
4. Destek Vektör Makineleri (SVM): SVM, iyi genelleştirme kabiliyeti ve yüksek performans sağlamaktadır ancak ölçeklenebilirlik konusunda zorluklarla karşılaşabilmektedir.
5. k-En Yakın Komşu (k-EYK): Basit ve sezgisel bir yöntem olan k-EYK, küçük veri setleri için uygundur, ancak büyük veri setlerinde ölçeklenebilirlik sorunları yaşanabilmektedir.
6. Probabilistic Neural Network (PNN) – Dynamic Decay Adjustment (DDA): PNN'nin öğrenme ve sınıflandırma süreçlerini hızlandıran ve verimliliğini artıran bir yöntemdir. DDA, aktivasyon fonksiyonlarının yayılım oranını adaptif olarak ayarlayarak, PNN'nin sınıflandırma performansını ve esnekliğini optimize etmektedir.
7. RProp MLP Learner: RProp, hızlı ve ölçekten bağımsız olarak sinir ağı ağırlıklarını güncellemek için uygun bir yöntemdir.
8. Naïve Bayes: Yeni bir veri noktasının sınıfını belirlemek için algoritma bu koşullu olasılıkları karşılaştırarak en yüksek olasılığa sahip sınıfı seçmektedir.

Seçilen algoritmalar, Covid-19 teşhisi için en uygun olanları belirlemek amacıyla karşılaştırmalı bir analiz yaparak değerlendirilmiştir. Bu analiz sonucunda, Covid-19 teşhisi için en uygun algoritmalar belirlenmiştir. Bu süreçte dikkate alınan adımlar maddeler halinde belirtilmiştir:

1. Performans ölçütleri: Algoritmaların karşılaştırılması sırasında doğruluk, hassasiyet, duyarlılık, F1 skoru ve AUC-ROC gibi performans ölçütleri

kullanılmıştır. Bu ölçütler, algoritmaların ne kadar başarılı olduğunu değerlendirmeye yardımcı olmuştur.

2. Çapraz doğrulama: Algoritmaların performansının doğru ve güvenilir bir şekilde ölçülmesi için k katlı çapraz doğrulama (k-fold cross-validation) kullanılmıştır. Bu yöntem, modelin farklı veri alt kümeleri üzerindeki performansını değerlendirerek daha sağlam sonuçlar elde etmeye olanak tanımıştır.
3. Hiperparametre optimizasyonu: Algoritmaların başarısını artırmak için hiperparametre optimizasyonu yapılmıştır. Bu süreçte, en uygun hiperparametre değerlerini belirlemek için Grid Search, Random Search veya Bayesian Optimization gibi yöntemler de kullanılabilir.
4. Özellik önemi: En iyi performansı sağlayan model seçildikten sonra, özellik önemini analiz edilmiştir. Bu analiz, Covid-19 teşhisi için hangi faktörlerin en önemli olduğunu belirlemeye yardımcı olmuştur.
5. Model yorumlanabilirliği: En iyi modelin yorumlanabilirliğini artırmak için, LIME (Local Interpretable Model-agnostic Explanations) veya SHAP (SHapley Additive exPlanations) gibi yöntemler kullanılmıştır. Bu yöntemler, modelin tahminlerini açıklamak için yerel açıklamalar sunarak, modelin neden belirli bir tahminde bulunduğunu anlamaya yardımcı olmuştur.

Sonuç olarak, yukarıdaki adımlar uygulanarak, Covid-19 teşhisi için en uygun algoritmalar ve yöntemler seçilmiştir. Bu süreç, doğru ve güvenilir teşhisler yapmak için en iyi modelin kullanılmasını sağlayarak, pandemiyle mücadelede önemli bir katkı sağlaması hedeflenmiştir.

#### **4.3.2.1. Model Eğitimi ve Hiperparametre Optimizasyonu**

Model eğitimi ve hiperparametre optimizasyonu süreçlerinde dikkate alınan önemli adımlar şu şekilde sıralanmıştır:

1. Veri bölümlenmesi: Veri seti, eğitim ve test kümelerine bölünmüştür. Verinin %70'i eğitim kümesi için ve %30'u test kümesi için kullanılmıştır. Bu aşamada, verinin doğru bir şekilde dağıtılması ve rastgele örnekleme yapılması önemlidir.
2. Ölçeklendirme ve normalizasyon: Model eğitimi sırasında, özelliklerin ölçeklendirilmesi ve normalleştirilmesi gerekmektedir. Bu işlem, algoritmanın daha hızlı ve doğru bir şekilde çalışmasına yardımcı olmuştur. Standartlaştırma ve Min-Max ölçeklendirme gibi yöntemler kullanılmıştır.
3. Model eğitimi: Eğitim kümesi üzerinde modelin eğitilmesi sürecinde, algoritmanın parametrelerinin uygun şekilde ayarlanması önemlidir. Bu süreçte, modelin öğrenme hızı ve yineleme sayısı gibi hiperparametreler optimize edilmiştir.
4. Model iyileştirme: Modelin performansını daha da artırmak için, özellik seçimi, özellik mühendisliği ve düzenleme gibi yöntemler kullanılmıştır.

Sonuç olarak, model eğitimi ve hiperparametre optimizasyonu süreçlerinde dikkate alınması gereken önemli adımlar ve dikkat edilmesi gereken hususlar belirtilmiştir. Bu adımları dikkatli bir şekilde takip ederek, Covid-19 teşhisi için kullanılacak olan modelin performansını optimize etmek ve güvenilir sonuçlar elde etmek mümkündür. Başarılı bir model geliştirme ve optimizasyon süreci, pandemi ile mücadelede önemli bir katkı sağlayarak, hızlı ve doğru teşhislerin gerçekleştirilmesine yardımcı olacaktır.

#### **4.3.2.2.Çapraz Doğrulama ve Model Seçimi**

Çapraz doğrulama ve model seçimi süreçleri, modelin performansını değerlendirmek ve genelleştirme kabiliyetini ölçmek için önemlidir. Bu aşamalara başlarken, aşağıdaki adımları izlenmiştir.

1. Çapraz doğrulama yöntemi seçimi: Çapraz doğrulama yöntemi olarak k-katlı çapraz doğrulama (k-fold cross-validation), stratified k-katlı çapraz doğrulama (stratified k-fold cross-validation) veya zaman serisi çapraz doğrulama (time

series cross-validation) gibi yöntemlerden her biri ayrı ayrı test edilerek en iyi doğrulama veren yöntemler her bir algoritma için ayrı ayrı seçilmiştir. K-katlı çapraz doğrulama, veri kümesini eşit büyüklükte k parçaya böler ve her parçayı sırayla test verisi olarak kullanırken, geri kalan parçaları eğitim verisi olarak kullanmaktadır.

2. Final modelin eğitimi ve değerlendirmesi: Seçilen en iyi model ve hiperparametrelerle, tüm eğitim veri kümesi üzerinde model tekrar eğitilmiştir. Daha sonra test veri kümesi üzerinde modelin performansını değerlendirilmiştir. Bu, modelin gerçek dünya verileri üzerinde nasıl bir performans sergilediğine dair farklı çalışmalarla karşılaştırılmıştır.

Bu adımlar dikkatlice takip edilerek, Covid-19 teşhisi için kullanılan modelin performansını değerlendirilmiş ve en uygun model seçilmiştir. Bu süreç, modelin genelleştirme kabiliyetini artırarak, farklı veri kümeleri üzerinde daha doğru ve güvenilir sonuçlar elde etmemizi sağlamıştır.

#### **4.3.2.3. Model Değerlendirmesi ve Performans Metrikleri**

Model değerlendirme, eğitilen makine öğrenimi modelinin performansını ölçmek ve modelin başarısını değerlendirmek için kullanılan süreçtir. Model değerlendirmesinde, performans metriklerini kullanarak modelin ne kadar doğru ve güvenilir sonuçlar ürettiğini belirlenmiştir. Covid-19 teşhisi için kullanılacak bir modelde dikkate alınan temel performans metrikleri şunlardır:

1. Doğruluk (Accuracy): Modelin doğru tahminlerinin toplam tahminlere oranıdır. Yüksek doğruluk, modelin iyi performans gösterdiğini belirtmektedir. Ancak, dengesiz veri kümelerinde doğruluk yanıltıcı olabilmektedir.
2. Duyarlılık (Sensitivity), Geri Çağırma (Recall) Oranı veya Gerçek Pozitif Oranı (True Positive Rate, TPR): Pozitif vakaların ne kadarının doğru bir şekilde pozitif olarak tahmin edildiğini göstermektedir. Duyarlılık, hastalığı olan ve model tarafından doğru olarak teşhis edilen bireylerin oranını ölçmektedir.

Gerçek pozitif örneklerin ne kadarının doğru olarak tahmin edildiğini gösterir. Yüksek duyarlılık, modelin pozitif vakaları kaçırmama eğiliminde olduğunu belirtmektedir.

3. Özgüllük (Specificity) veya Gerçek Negatif Oranı (True Negative Rate, TNR): Negatif vakaların ne kadarının doğru bir şekilde negatif olarak tahmin edildiğini göstermektedir. Özgüllük, sağlıklı bireylerin model tarafından doğru olarak sağlıklı olarak tanımlanma oranını ölçmektedir.
4. F1 Skoru: Duyarlılık ve özgüllüğün harmonik ortalamasıdır ve dengesiz veri kümesi durumlarında modelin performansını değerlendirmek için daha iyi bir ölçümdür. F1 skoru, yanlış pozitif ve yanlış negatif tahminlerin etkisini dikkate alarak, modelin genel performansını ölçmektedir.

Bu metrikler, Covid-19 teşhisi için kullanılan MÖ modelinin performansını değerlendirmek ve iyileştirmek adına önemli kriterlerdir. Modelin doğru ve güvenilir tahminler yapabilmesi için, yukarıdaki metriklerin değerleri göz önünde bulundurularak modelin sürekli olarak optimize edilmesi ve geliştirilmesi sağlanmıştır. Bu sayede, modelin gerçek dünya durumlarında yüksek başarı ve etkinlik sağlaması mümkün olacaktır.

Model değerlendirmesi sırasında dikkate alınan önemli noktalar şunlardır:

1. Dengesiz veri kümelerinde, doğruluk metriğinin yanıltıcı olabileceği düşünüldüğünden dolayı, F1 skoru gibi daha dengeli metriklerin kullanılması gerektiği göz önüne alınmıştır.
2. Performans metriklerini yorumlarken, Covid-19 teşhisi için yanlış negatif ve yanlış pozitif tahminlerin etkisi göz önünde bulundurulmuştur.
3. Model seçimi ve optimizasyon aşamalarında çapraz doğrulama kullanarak, modelin genelleştirme kabiliyeti değerlendirilmiştir.

4. Bu bilgiler ışığında, model değerlendirmesi ve performans metriklerini kullanarak Covid-19 teşhisi için eğitilen modelin başarısını ve güvenilirliği doğru bir şekilde ölçülebilmektedir. Unutulmamalıdır ki, modelin gerçek dünya verileriyle karşılaştırıldığında yüksek performans göstermesi için, model seçimi ve optimizasyon süreçlerinde doğru yöntemler ve algoritmalar kullanılması çok büyük bir önem arz etmiştir.
5. Bunun yanı sıra, modelin başarısını ölçmek için kullanılan metrikler, tıbbi teşhis ve tedavi bağlamında etik ve mali sonuçları olan yanlış pozitif ve yanlış negatif tahminlerin etkisi göz önünde bulundurulmuştur. Bu nedenle, performans metriklerinin değerlendirilmesi, modelin başarısı ve gerçek dünya durumlarındaki uygulanabilirliği hakkında daha geniş bir perspektif sunmuştur.

Sonuç olarak, model değerlendirmesi ve performans metriklerinin doğru bir şekilde kullanılması, Covid-19 teşhisi için makine öğrenimi modelinin etkinliğini ölçmek ve modelin gerçek dünya uygulamalarında başarı sağlamasına yardımcı olmak için önemli bir adımdır. Bu değerlendirmeler sayesinde, modelin genelleştirme kabiliyeti ve performansı hakkında daha sağlam sonuçlar elde edebilir ve modeli sürekli olarak iyileştirebiliriz.

#### **4.3.2.4. ROC Eğrisi ve AUC Değerleri**

Model değerlendirmesinde önemli bir performans metriği olan ROC eğrisi ve AUC değerlerine değinelim:

1. ROC Eğrisi (Receiver Operating Characteristic Curve): ROC eğrisi, sınıflandırma modelinin performansını görselleştiren bir grafikdir. Eksenlerde Gerçek Pozitif Oranı (Duyarlılık) ve Yanlış Pozitif Oranı (1 - Özgüllük) bulunur. ROC eğrisi, modelin eşik değerleri değiştirilerek farklı duyarlılık ve özgüllük değerleri elde edilir ve bu değerler eğri üzerinde gösterilir. Bu sayede, modelin farklı eşik değerlerindeki performansı değerlendirilmiştir.

2. AUC (Area Under the Curve) Deęeri: ROC eęrisinin altında kalan alanın büyüklüğüdür ve modelin performansını özetleyen tek bir sayısal deęerdir. AUC deęeri, 0 ile 1 arasında bir deęer alır. Deęer 1'e yaklaştıkça, modelin performansı daha iyi olarak kabul edilir. AUC deęeri, modelin sınıflandırma yeteneğini, pozitif ve negatif örnekler arasındaki ayrımı ölçer. AUC deęeri 0.5 olduğunda, modelin sınıflandırma yeteneęi rastgele tahminle eşdeęerdir.

ROC eęrisi ve AUC deęerleri, Covid-19 teşhis modelinin performansını deęerlendirmede önemli bir rol oynamıştır. Bu metrikler, modelin doğru tahminler yapma becerisini ölçerken, farklı eşik deęerlerindeki başarısını da göz önünde bulundurmıştır. Bu sayede, modelin genel başarısı daha doğru bir şekilde ölçülebilir ve iyileştirilebilir duruma gelmiştir. ROC eęrisi ve AUC deęerleri kullanılarak, modelin performansı optimize edilmiş ve daha doğru ve güvenilir tahminler yapma yeteneęi arttırmıştır.



## BÖLÜM 5

### UYGULAMA VE DENEYSEL ÇALIŞMALAR

Bu yüksek lisans tezi çalışması, Covid-19 risk faktörlerinin incelenmesi ve bu faktörlerin hastalığın yayılması ve şiddeti üzerindeki etkilerinin belirlenmesi amacıyla makine öğrenimi ve veri madenciliği yöntemlerinin kullanımına odaklanmaktadır. Çalışmada kullanılan veri seti, Covid-19 pozitif vakaların demografik ve epidemiyolojik verilerini içermekte olup, zenodo.org adresinden temin edilmiştir.

Çalışma sürecinde veri ön işleme, modelleme ve değerlendirme aşamaları gerçekleştirilmiştir. Bu aşamalarda, eksik veri doldurma, dönüşüm ve temizleme yöntemleri kullanılmıştır. KNIME programı, veri madenciliği ve makine öğrenimi yöntemlerinin uygulanması için tercih edilmiştir.

Çeşitli sınıflandırma ve regresyon algoritmaları test edilerek, model eğitimi ve hiperparametre optimizasyonu süreçleri gerçekleştirilmiş ve çapraz doğrulama yöntemi kullanılarak model seçimleri yapılmıştır. Model değerlendirilmesi için doğruluk, duyarlılık, özgüllük, F1 skoru, ROC eğrisi ve AUC değerleri gibi performans metrikleri kullanılmıştır.

Çalışma sonucunda, Covid-19 risk faktörlerinin yaygınlığı ve makine öğrenimi modellerinin etkinliği konularında önemli bilgiler elde edilmiştir. Yapay zeka ve veri madenciliği yöntemleri, Covid-19 teşhisi ve risk değerlendirmesinde önemli katkılar sağlamıştır.

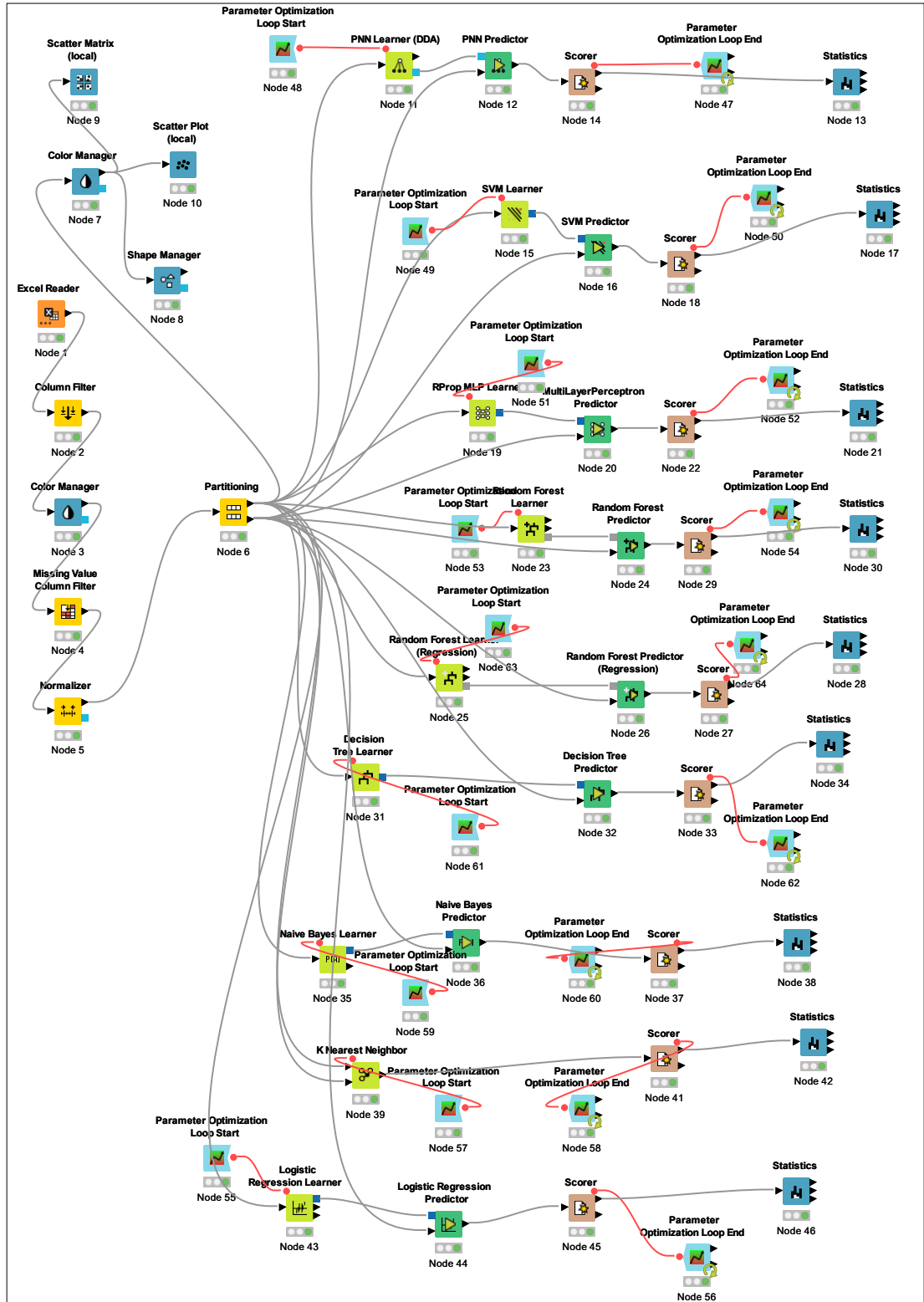
Tartışma ve gelecek çalışmalar bölümünde, yapılan çalışmanın başarıları, sınırlılıkları ve potansiyel iyileştirmeler ele alınmıştır. Öneriler, gelecekteki araştırmalar ve uygulamalar için bir temel oluşturmayı amaçlamaktadır.

Özetle, bu tez çalışması, Covid-19'un anlaşılması ve kontrolü için makine öğrenimi ve veri madenciliği yöntemlerinin etkili bir şekilde kullanılmasını sağlayarak, pandemi yönetimi ve halk sağlığına önemli katkılar sunmaktadır.

## 5.1. VERİ ÖNİŞLEME

Çalışma Knime Analytics 4.6.1 platformunda, Win10 Pro 64-bit, İ5-8250 1.60 GHz, 8 GB ram ve 250GB SSD üzerinde gerçekleştirilmiştir. MÖ modeli için hesaplama, veri normalleştirme, özellik seçimi ve sınıflandırmayı kapsayan dört aşamalı bir işlem hattı tasarlanmıştır. Makine öğrenimi için Knime Analytics Platformunda 8 farklı algoritma kullanılmıştır. Platformun Analytics kütüphanesinden Learner, Predictor, Statistic ve Scorer araçları ile algoritma performansları karşılaştırılmış, grafiksel gösterim için Scatter Plot, Scatter Matrix araçları kullanılmıştır. Son olarak Property Shape Manager ve Color Manager araçları kullanılarak model tamamlanmıştır.

Veri setinin %70'i eğitim, %30'luk kısmı ise test için kullanılmıştır. Şekil 5.1'de Knime Analytics Platformunda oluşturulan MÖ yapısı gösterilmiştir. Oluşturulan MÖ yapısında "Excel Reader" düğümü ile veriler okunduktan sonra "Column Filter" düğümü ile seçilen özellikler üzerinden algoritmaların çalışması sağlanmaktadır. "Missing Value Column Filter" düğümü sayesinde belirlenen oranın altında kalan sütunlar belirlenmiştir. Sonraki adımda "Normalizer" düğümünde her sütun kendi aralığına göre normalize edildikten sonra "Partitioning" düğümü aracılığı ile veri seti belirlenen kriterlere ve orana göre eğitim ve test verisi olacak şekilde bölünmüştür. "Partitioning" düğümüne bağlı her algoritma için 2 düğüm bulunmaktadır. "Learner" düğümü eğitim verileriyle oluşturulan modelin parametreleri "Predictor" düğümüne bağlanarak eğitim sonunda aktarılmaktadır. Bundan sonra "Predictor" düğümüne "Partitioning" düğümünden gelen test verileriyle modelin başarısı ölçülmüş ve "Predictor" düğümüne bağlı "Scorer" düğümü ile karmaşıklık matrisi ve diğer metriklere göre sonuç üretilmiştir. "Statistics" düğümü ise sayısal sütunlar için min, maks, ortalama, standart sapma, varyans, medyan, toplam, eksik satırların sayısı vb. istatistiksel verileri sunarak histogramı oluşturmaktadır.



Şekil 5.1. Knime Analytics Platformunda oluşturulan MÖ yapısı

Deneysel çalışmalar iki farklı veri seti için gerçekleştirilmiştir. Bunlardan ilki; çoğunluğu San Raffaele Hastanesi acil servisinde alınan kan örneklerinden oluşan 35

özelliikli veri seti. Diğeri ise aynı veri seti içinden sadece tam kan sayımı tetkikine dahil olan değerlerden oluşturulmuş bir alt küme olan veri setidir. Her iki veri seti %70 eğitim, %30 test olmak üzere kullanılmıştır. Bu bölümde, tez konusu ve içeriğine göre hazırlanan tüm başlıklar ve alt maddeleri içeren geniş kapsamlı bir tartışma sunulmaktadır.

Covid-19 pandemisi ve risk faktörlerinin önemi üzerine yapılan çalışmalar, pandemi yönetimi ve önleyici tedbirlerin geliştirilmesinde büyük öneme sahiptir. Bu çalışmalarda, KNIME platformu ve makine öğrenimi yöntemlerinin kullanılması, veri analizinin daha hızlı ve etkili gerçekleştirilmesini sağlamıştır. Makine öğrenimi ve veri madenciliği yöntemleri, Covid-19 risk faktörlerinin analizinde ve teşhis süreçlerinde önemli bir rol oynamaktadır. Bu yöntemlerin doğru ve etkili bir şekilde uygulanması, pandemi yönetiminde ve halk sağlığı stratejilerinde büyük fayda sağlayabilir.

KNIME platformu ve MÖ araçları, veri ön işleme, modelleme ve değerlendirme süreçlerinde kullanılmıştır. Bu araçların kullanılması, veri analizinin daha etkili ve verimli bir şekilde gerçekleştirilmesine yardımcı olmuştur. Veri seti ve ön işleme aşamasında dikkat edilmesi gereken önemli kriterler ve aşamalar uygulanmıştır. Model geliştirme ve optimizasyon sürecinde kullanılan algoritmalar ve yöntemler, modelin başarısı ve doğruluğu üzerinde önemli bir etkiye sahiptir. Model değerlendirmesi ve performans metrikleri, modelimizin doğruluğunu ve güvenilirliğini ölçmek için kullanılmıştır.

Doğruluk, duyarlılık, özgüllük, F1 skoru, ROC eğrisi ve AUC değerleri gibi metrikler, modelin başarısını ve etkinliğini değerlendirmek için önemli olmuştur. Covid-19 risk faktörlerinin yaygınlığı ve makine öğrenimi modelleri ile yapılan çalışmalar, pandemi yönetiminde ve halk sağlığı stratejilerinde önemli bilgiler sunabilmektedir. Bu bilgiler, daha etkili ve veri odaklı kararlar alınmasına ve pandemiye karşı mücadelede başarılı olmaya katkıda bulunabilir. AI ve VM yöntemlerinin Covid-19 teşhisine katkısı, pandemi yönetiminde hızlı ve doğru teşhislerin konulması için önemlidir. Bu yöntemler, geleneksel yöntemlere göre daha hızlı ve etkili çözümler sunarak, sağlık sisteminin kapasitesini artırır ve pandemiyle mücadelede başarıya katkı sağlayabilir.

Tartışma ve gelecek çalışmalar bölümünde, bu tez kapsamında elde edilen başarılar ve sınırlılıklar değerlendirilmiştir. Ayrıca, gelecek çalışmalar ve potansiyel iyileştirmeler için öneriler sunulmuştur. Bu öneriler, Covid-19 risk faktörlerinin anlaşılması ve pandemi yönetimi açısından daha etkili ve veri odaklı stratejilerin geliştirilmesine katkıda bulunabilir.

Özet ve genel değerlendirme bölümünde, tezin ana bulguları ve sonuçları sunulmaktadır. KNIME ve makine öğrenimi yöntemlerinin Covid-19 risk faktörlerini anlamada sağladığı avantajlar ve yapay zeka ve veri madenciliği uygulamalarının pandemi yönetimi ve teşhis süreçlerine etkisi değerlendirilmiştir.

Sonuç olarak, bu tez çalışması, Covid-19 risk faktörlerinin anlaşılması ve pandemi yönetimi açısından önemli bir katkı sağlamaktadır. Makine öğrenimi ve veri madenciliği yöntemlerinin etkin kullanımı, pandemiyle mücadelede başarıya ulaşmak için veri odaklı ve bilimsel temelli stratejilerin geliştirilmesine yardımcı olabilir. Gelecek çalışmalar ve potansiyel iyileştirmeler ile bu alandaki bilgi birikimi ve uygulamalar daha da geliştirilebilir, böylece pandemiye karşı daha etkili ve sürdürülebilir çözümler sunulabilir.

## **5.2. TEZİN ANA BULGULARI**

Bu yüksek lisans tezi çalışması kapsamında elde edilen ana bulgular ve sonuçlar şu şekildedir:

1. Covid-19 risk faktörleri üzerinde yapılan analizler, hastalığın yayılması ve şiddeti ile ilgili önemli bilgiler sunmaktadır. Bu faktörlerin tanımlanması, salgının yönetimi ve halk sağlığına yapılacak müdahalelerin etkinliği açısından büyük önem taşımaktadır.
2. MÖ ve VM yöntemleri, Covid-19 teşhisi ve risk değerlendirmesi konularında başarılı sonuçlar elde etmekte ve bu yöntemlerin kullanımı, sağlık alanında karar destek sistemlerine güçlü bir katkı sağlamıştır.

3. KNIME programının kullanımı, MÖ ve VM yöntemlerinin uygulanması sürecini kolaylaştırmış ve hızlandırmıştır. Programın sezgisel ve kullanıcı dostu arayüzü, veri analiz süreçlerini daha erişilebilir hale getirmiştir.
4. Çeşitli sınıflandırma ve regresyon algoritmalarının test edilmesi ve en uygun modelin seçilmesi sürecinde, model eğitimi, hiperparametre optimizasyonu ve çapraz doğrulama yöntemlerinin kullanımı, model performansının artırılmasında önemli rol oynamıştır.
5. Model değerlendirmesinde kullanılan performans metrikleri, doğruluk, duyarlılık, özgüllük, F1 skoru, ROC eğrisi ve AUC değerleri gibi metriklerle, modelin başarısı ve güvenilirliği hakkında önemli bilgiler sunmuştur.

Sonuç olarak, bu tez çalışması, Covid-19 risk faktörlerinin incelenmesi ve hastalığın yönetimi açısından MÖ ve VM yöntemlerinin etkin bir şekilde kullanılmasını sağlayarak, pandemi sürecinde alınacak önlemlerin ve halk sağlığına yapılacak müdahalelerin etkinliğini artırmaya katkı sağlamaktadır. Ayrıca, bu çalışma KNIME gibi veri analizi araçlarının, sağlık alanında karmaşık veri setlerinin işlenmesi ve analiz edilmesi süreçlerini daha erişilebilir ve kullanıcı dostu hale getirdiğini ortaya koymaktadır.

Gelecekteki çalışmalarda, daha büyük ve çeşitli veri setleri kullanarak, modelin başarısını ve uygulanabilirliğini daha da artırmak mümkündür. Bu, daha kapsamlı ve doğru tahminler elde etmeye ve Covid-19 ile mücadelede daha etkili stratejiler geliştirmeye yardımcı olacaktır. Ayrıca, başka sağlık sorunları ve risk faktörleri için de benzer yöntemlerin kullanılması, karar destek sistemlerinin güçlendirilmesi ve sağlık hizmetlerinin kalitesinin artırılması açısından değerli olacaktır. Bu tez çalışması, MÖ ve VM yöntemlerinin sağlık alanında nasıl başarılı bir şekilde uygulanabileceğine dair önemli bir örnek teşkil edebilir.

### **5.3. KNIME VE MAKİNE ÖĞRENİMİ YÖNTEMLERİNİN COVID-19 RİSK FAKTÖRLERİNİ ANLAMADA SAĞLADIĞI AVANTAJLAR**

Bu tez çalışması kapsamında kullanılan KNIME ve makine öğrenimi yöntemleri, Covid-19 risk faktörlerinin anlaşılması ve analizi sürecinde önemli avantajlar sağlamaktadır. Bu avantajlar şu şekilde sıralanabilir:

1. Veri ön işleme ve temizleme: KNIME, veri temizleme, eksik veri doldurma ve dönüşümler gibi ön işleme adımlarını kolaylıkla gerçekleştirmeye olanak tanır. Bu, veri setinin analize uygun hale getirilmesi sürecini hızlandırır ve daha güvenilir sonuçlar elde etmeye yardımcı olur.
2. Otomatik model seçimi ve optimizasyon: KNIME, çeşitli makine öğrenimi algoritmalarını deneyerek en uygun modeli ve hiperparametreleri otomatik olarak seçebilir. Bu, modelin performansını en üst düzeye çıkarmaya ve doğru tahminler elde etmeye katkı sağlar.
3. Yüksek performanslı analiz: Makine öğrenimi yöntemleri, geleneksel istatistiksel yöntemlere kıyasla daha karmaşık ve büyük veri setlerini işlemekte daha başarılıdır. Bu sayede, Covid-19 risk faktörlerini daha kapsamlı ve detaylı bir şekilde incelemek mümkün hale gelir.
4. Çapraz doğrulama ve model değerlendirme: KNIME ve makine öğrenimi yöntemleri, çapraz doğrulama ve model değerlendirme gibi teknikleri kullanarak modelin güvenilirliğini ve genelleştirilebilirliğini test eder. Bu, modelin gerçek dünya verilerinde başarılı bir şekilde uygulanabilmesini sağlar.
5. Kullanıcı dostu arayüz: KNIME, sezgisel ve kullanıcı dostu bir arayüz sunarak, veri analizi ve makine öğrenimi süreçlerini daha erişilebilir hale getirir. Bu, sağlık alanında çalışan profesyonellerin ve araştırmacıların, teknik bilgiye sahip olmaksızın bu yöntemleri kullanarak Covid-19 risk faktörlerini anlamalarına yardımcı olur.

Sonuç olarak, KNIME ve MÖ yöntemlerinin kullanılması, Covid-19 risk faktörlerinin anlaşılması ve analizi sürecini daha etkili, hızlı ve güvenilir hale getirir. Bu da, hastalığın yayılmasını önlemeye yönelik politikaların ve stratejilerin geliştirilmesinde önemli bir rol oynar.

#### **5.4. YAPAY ZEKA UYGULAMALARININ PANDEMİ YÖNETİMİ VE TEŞHİS SÜREÇLERİNE ETKİSİ**

Yapay zeka ve veri madenciliği yöntemleri, Covid-19 pandemisi ile mücadele sürecinde önemli bir rol oynamaktadır. Bu tez çalışması kapsamında yapılan analizler ve uygulamalar, bu yöntemlerin pandemi yönetimi ve teşhis süreçlerine etkisini detaylı bir şekilde ortaya koymaktadır:

1. Risk faktörlerinin anlaşılması: MÖ modelleri, Covid-19 risk faktörlerinin belirlenmesine ve yaygınlıklarının anlaşılmasına katkı sağlar. Bu sayede, risk altındaki popülasyon gruplarına yönelik öncelikli müdahaleler ve koruyucu önlemler geliştirilebilir.
2. Teşhis süreçlerinin iyileştirilmesi: AI ve VM yöntemleri, Covid-19 teşhis süreçlerinde doğruluk, duyarlılık ve özgüllük gibi metrikleri optimize ederek, erken ve doğru teşhis imkanı sunar. Bu, hastalığın yayılmasını kontrol altına almak ve tedavi süreçlerini hızlandırmak için kritik öneme sahiptir.
3. Hızlı ve etkili politika geliştirme: AI ve VM araçları, büyük ve karmaşık veri setlerini hızlı ve etkili bir şekilde analiz edebilir. Bu, pandemi yönetimi ve sağlık politikaları geliştirme sürecinde, doğru ve zamanında kararlar almayı kolaylaştırır.
4. Özelleştirilmiş tedavi yaklaşımları: MÖ modelleri, bireysel hasta profilleri ve risk faktörlerini dikkate alarak, özelleştirilmiş tedavi yaklaşımlarının geliştirilmesine olanak tanımıştır. Bu, hastaların ihtiyaçlarına daha iyi hizmet etmeye ve tedavi süreçlerinin etkinliğini artırmaya yardımcı olur.



5. Pandemi yönetiminde etkili iletişim: AI ve VM yöntemleri, pandemi yönetiminde kamuoyu bilgilendirme ve iletişim stratejilerinin geliştirilmesine katkı sağlamıştır. Bu, risk bilincinin artırılması ve toplumun sağlık önlemlerine uyumunun sağlanması açısından önemlidir.

AI ve MÖ uygulamaları, Covid-19 pandemi yönetimi ve teşhis süreçlerinde önemli avantajlar sunmaktadır. Bu yöntemlerin kullanılması, pandeminin etkilerini azaltmaya ve toplum sağlığını korumaya yardımcı olmaktadır. Bu çalışma kapsamında incelenen KNIME platformu ve makine öğrenimi yöntemleri, bu amaçla kullanılabilir ve etkin sonuçlar elde etme potansiyeline sahiptir. MÖ modellerinin doğru ve hızlı teşhisler sağlaması, tedavi süreçlerinin hızlandırılmasına ve pandeminin yayılımının kontrol altına alınmasına katkıda bulunur. Ayrıca, bu yöntemlerin kullanılmasıyla risk faktörlerinin belirlenmesi ve yaygınlıklarının anlaşılması, risk altındaki gruplara yönelik öncelikli müdahaleler ve koruyucu önlemler geliştirilmesine olanak sağlar.

AI ve MÖ tekniklerinin sağlık politikaları ve pandemi yönetimi alanında hızlı, etkili ve doğru kararlar alınmasına imkan tanınması, bu süreçlerin etkinliğini artırır. Öte yandan, özelleştirilmiş tedavi yaklaşımları ve etkili iletişim stratejileri ile hastaların ihtiyaçlarına daha iyi hizmet edilerek, tedavi süreçlerinin etkinliği ve toplumun sağlık önlemlerine uyumu sağlanır. Bu bağlamda, AI ve MÖ yöntemlerinin Covid-19 pandemi yönetimi ve teşhis süreçlerindeki rolü ve önemi, bu tez çalışması ile daha iyi anlaşılmıştır. Gelecekte, bu yöntemlerin daha geniş ölçekte ve farklı alanlarda kullanılmasıyla, pandemi ve benzeri sağlık sorunlarıyla mücadelede daha etkili sonuçlar elde edilebilir.

## BÖLÜM 6

### DENEYSEL SONUÇLAR, BULGULAR VE TARTIŞMA

#### 6.1. YAPAY ZEKA VE VERİ MADENCİLİĞİ YÖNTEMLERİNİN COVID-19 TEŞHİSİNE KATKISI

Bu yöntemlerin Covid-19 teşhisi ve önleme stratejilerinde nasıl etkili olduğunu inceleyelim:

1. Hızlı ve doğru teşhis: AI ve MÖ yöntemleri, geniş veri setlerinden özellikler ve bilgiler çıkararak Covid-19'un hızlı ve doğru teşhisine katkıda bulunur. Özellikle görüntü işleme ve doğal dil işleme teknikleri, tıbbi görüntüler ve elektronik sağlık kayıtları üzerinde çalışarak teşhis sürecini hızlandırabilir ve doğruluğunu artırabilir.
2. Risk faktörlerinin belirlenmesi: AI ve MÖ yöntemleri, Covid-19 risk faktörlerinin belirlenmesinde önemli rol oynar. Bu yöntemler, hastaların demografik, klinik ve laboratuvar verilerini analiz ederek, hangi faktörlerin hastalığın seyri ve sonuçları üzerinde etkili olduğunu anlamaya yardımcı olur.
3. Epidemiyolojik analiz: AI ve MÖ, Covid-19'un yayılımı ve etkilerini anlamak için epidemiyolojik analizlerde de kullanılabilir. Bu yöntemler, coğrafi ve zamansal faktörlerle ilişkili olarak hastalığın nasıl yayıldığını ve hangi faktörlerin hastalığın seyrini etkilediğini incelemeye yardımcı olur.
4. Önleme ve tedavi stratejileri: AI ve MÖ, hastalığı önleme ve tedavi stratejilerinin geliştirilmesine ve uygulanmasına katkı sağlar. Bu yöntemler, risk faktörlerini ve hastalığın yayılımını analiz ederek, hedefli ve etkili stratejilerin belirlenmesine yardımcı olur.

5. Kişiselleştirilmiş tıp: AI ve MÖ, kişiselleştirilmiş tıp uygulamalarına da katkıda bulunur. Bu yöntemler, hastaların genetik, biyolojik ve yaşam tarzı faktörlerini dikkate alarak, her birey için en uygun tedavi ve önleme stratejilerinin belirlenmesine yardımcı olur.

Yapay zeka ve veri madenciliği yöntemlerinin Covid-19 teşhisine katkısı, hızlı ve doğru teşhis süreçlerinin oluşturulması, risk faktörlerinin belirlenmesi, epidemiyolojik analizlerin yapılması, önleme ve tedavi stratejilerinin geliştirilmesi ve kişiselleştirilmiş tıp uygulamalarının hayata geçirilmesi şeklinde özetlenebilir.

6. Aşı ve ilaç geliştirme: AI ve MÖ yöntemleri, aşı ve ilaç geliştirme süreçlerini hızlandırmada da önemli rol oynamaktadır. Bu yöntemler, potansiyel hedef moleküllerin ve bileşenlerin belirlenmesi, etkinliğin ve güvenliğin değerlendirilmesi ve klinik deneylerin tasarlanması aşamalarında kullanılabilir.
7. Sağlık hizmetlerinin planlanması ve yönetimi: AI ve MÖ, sağlık hizmetlerinin planlanması ve yönetimi açısından da değerli bilgiler sunar. Bu yöntemler, hastane ve yoğun bakım kapasitelerinin, tıbbi malzeme ve personel ihtiyaçlarının ve enfeksiyon kontrol önlemlerinin etkili bir şekilde planlanması ve yönetilmesine yardımcı olur.
8. Halk sağlığı politikaları ve karar verme: AI ve MÖ yöntemlerinin sağladığı bilgiler, halk sağlığı politikalarının ve kararların alınmasında önemli rol oynar. Bu yöntemler, politika yapıcılar ve karar vericilere, hastalığın yayılımı ve etkileri hakkında güncel ve doğru bilgiler sunarak, daha etkili ve hedef odaklı politikaların belirlenmesine katkı sağlar.

Sonuç olarak, yapay zeka ve veri madenciliği yöntemleri Covid-19 teşhisine önemli ölçüde katkı sağlamaktadır. Bu yöntemler, teşhis süreçlerinin hızlandırılması, risk faktörlerinin belirlenmesi, epidemiyolojik analizlerin yapılması, önleme ve tedavi stratejilerinin geliştirilmesi, kişiselleştirilmiş tıp uygulamalarının hayata geçirilmesi, aşı ve ilaç geliştirme süreçlerinin hızlandırılması, sağlık hizmetlerinin planlanması ve yönetimi, ve halk sağlığı politikalarının ve kararların alınmasında etkili olmaktadır.

## **6.2. TARTIŞMA VE GELECEK ÇALIŞMALAR VE ÖNERİLER**

Bu tez kapsamında, Covid-19 risk faktörlerinin belirlenmesi ve epidemiyolojik analizlerin yapılması amacıyla makine öğrenimi ve veri madenciliği yöntemleri kullanılmıştır. Geliştirilen modeller ve analizler, Covid-19 risk faktörlerinin anlaşılması ve yaygınlığının değerlendirilmesine önemli ölçüde katkı sağlamıştır. Ayrıca, bu çalışma KNIME programını kullanarak veri ön işleme, modelleme ve değerlendirme süreçlerini gerçekleştirmiştir. Tez sonuçları, belirli demografik özellikler ve sağlık durumlarına sahip bireylerin Covid-19'a karşı daha yüksek risk altında olduğunu ortaya koymaktadır. Bu bulgular, hastalığın yayılımı ve etkileri hakkında güncel ve doğru bilgiler sunarak, daha etkili ve hedef odaklı politikaların belirlenmesine katkı sağlar. Ayrıca, tezde kullanılan yöntemlerin, Covid-19 teşhisi ve tedavisinde önemli bir rol oynadığı gösterilmiştir.

### **6.2.1. Yüksek Lisans Tezi Kapsamından Elde Edilen Başarılar ve Sınırlılıklar**

Bu yüksek lisans tezinin başarıları ve sınırlılıkları aşağıdaki gibi özetlenebilir:

Başarılar:

1. Covid-19 risk faktörlerinin belirlenmesi ve analizi için başarılı bir şekilde makine öğrenimi ve veri madenciliği yöntemlerinin uygulanması.
2. Farklı sınıflandırma ve regresyon tekniklerinin kullanılması ve performanslarının değerlendirilmesi.
3. KNIME platformunun kullanılması ile etkili ve hızlı model geliştirme süreci sağlanması.
4. Veri ön işleme ve eksik veri doldurma yöntemlerinin başarılı bir şekilde uygulanarak, daha güvenilir sonuçlar elde edilmesi.
5. Model değerlendirmesi ve performans metriklerinin kullanımı ile elde edilen modellerin güvenilirliğinin ve doğruluğunun saptanması.

Sınırlılıklar:

1. Çalışmada kullanılan veri setinin güncelliği ve kapsamı, analizlerin doğruluğunu ve genel geçerliliğini etkileyebilir.
2. Farklı veri kaynaklarının entegrasyonu ile elde edilebilecek daha kapsamlı sonuçlar sınırlıdır. Bu nedenle, farklı veri türlerinin entegrasyonu ile daha zengin analizler gerçekleştirilebilir.
3. Veri ön işleme ve eksik veri doldurma aşamasında yapılan varsayımlar ve seçimler, model sonuçlarını etkiler. Bu nedenle, farklı yöntemlerin ve yaklaşımların değerlendirilmesi önemlidir.
4. Çalışmada kullanılan algoritmalar ve yöntemlerin seçimi, bazı potansiyel tekniklerin ihmal edilmesine yol açabilir. Gelecek çalışmalarda, daha fazla algoritma ve yöntem değerlendirilmeli ve karşılaştırılmalıdır.
5. Çalışmanın sonuçları, belirli bir coğrafi bölge ve popülasyon için geçerli olabilir. Bu nedenle, genellemeler yaparken dikkatli olunmalı ve sonuçlar farklı bölgeler ve popülasyonlar için doğrulanmalıdır.

Bu başarılar ve sınırlılıklar göz önünde bulundurularak, çalışma sonuçları ve gelecek çalışmaların odak noktaları, Covid-19 pandemisinin daha etkili bir şekilde yönetilmesine ve kontrol altına alınmasına katkı sağlayacaktır.

### **6.2.2. Risk Faktörlerinin Yaygınlığını Anlamaya Yönelik Yapılan Çalışmaların Önemi**

Risk faktörlerinin yaygınlığını anlamak, Covid-19 ve benzeri pandemik hastalıkların kontrolü ve yönetimi açısından büyük öneme sahiptir. Bu tür çalışmalar, şu nedenlerle önemlidir:

1. Hedefli önlemler: Risk faktörlerinin yaygınlığını anlamak, hükümetlerin ve sağlık kuruluşlarının hedefli önlemler almasına ve kaynakları en etkili şekilde

kullanmasına yardımcı olur. Öncelikli risk gruplarını belirlemek, koruyucu ve tedavi edici müdahalelerin daha etkili bir şekilde uygulanmasını sağlar.

2. Eğitim ve farkındalık: Risk faktörlerinin yaygınlığı ve önemi hakkında bilgi sahibi olmak, halkın eğitimi ve farkındalığının artırılmasına katkıda bulunur. Bu, bireylerin kendi risklerini anlamalarına ve uygun önlemler alarak hastalığın yayılmasını önlemeye yardımcı olur.
3. Erken teşhis ve tedavi: Risk faktörlerinin yaygınlığını anlamak, sağlık çalışanlarına erken teşhis ve tedavi için daha fazla bilgi sağlar. Bu, özellikle Covid-19 gibi hızla yayılan ve zamanında müdahale gerektiren hastalıklarda önemlidir.
4. Araştırma ve geliştirme: Risk faktörlerinin yaygınlığı hakkında bilgi sahibi olmak, yeni tedaviler, aşılar ve diğer koruyucu önlemler geliştirmek için yapılan araştırmalara yol gösterir. Bu bilgiler, pandemi ile mücadelede en etkili stratejilerin belirlenmesine katkıda bulunur.
5. Sağlık politikaları ve stratejiler: Risk faktörlerinin yaygınlığına dair bilgi, hükümetlerin ve sağlık otoritelerinin sağlık politikaları ve stratejileri oluşturmasına ve uygulamasına yardımcı olur. Bu, salgının yayılmasını kontrol altında tutmak ve halk sağlığını korumak için gereklidir.

Sonuç olarak, risk faktörlerinin yaygınlığını anlamaya yönelik çalışmalar, Covid-19 pandemisi ile mücadelede önemli bir rol oynar. Bu tür çalışmalar, hedefli önlemler almayı, farkındalığı artırmayı, erken teşhis ve tedaviyi sağlamayı, araştırma ve geliştirmeyi desteklemeyi ve sağlık politikaları ve stratejilerinin geliştirilmesine katkıda bulunarak pandeminin kontrol altına alınmasına yardımcı olur.

### 6.2.3. Gelecek Çalışmalar ve Potansiye İyileştirmeler İçin Öneriler

Yüksek lisans tezinin sonuçları ve sınırlılıklarını göz önünde bulundurarak, gelecek çalışmalar ve potansiyel iyileştirmeler için şu önerilerde bulunulabilir:

1. Veri setinin genişletilmesi: Mevcut çalışmada kullanılan veri seti, daha geniş bir coğrafi bölgeyi ve farklı yaş, cinsiyet ve etnik köken gibi demografik faktörleri içerecek şekilde genişletilebilir. Bu, modelin genelleştirilebilirliğini ve farklı popülasyonlarda uygulanabilirliğini artırabilir.
2. Yeni risk faktörlerinin incelenmesi: Gelecek çalışmalarda, mevcut risk faktörlerine ek olarak, Covid-19'un seyri ve sonuçları üzerinde etkili olabilecek diğer potansiyel risk faktörleri de incelenebilir. Bu, daha kapsamlı ve doğru bir risk profili oluşturulmasına yardımcı olabilir.
3. Gelişmiş makine öğrenimi teknikleri: Gelecek çalışmalarda, daha gelişmiş makine öğrenimi teknikleri ve modelleri kullanılabilir. Özellikle, derin öğrenme ve nöral ağlar gibi yöntemler, karmaşık veri yapılarını ve ilişkilerini öğrenmek için daha etkili olabilir.
4. Model optimizasyonu ve hiperparametre ayarı: Gelecek çalışmalarda, model optimizasyonu ve hiperparametre ayarına daha fazla odaklanılabilir. Bu, modelin performansını ve doğruluğunu artırmaya katkıda bulunabilir.
5. Zaman serisi analizi ve tahmin: Gelecek çalışmalarda, Covid-19 vakalarının ve risk faktörlerinin zaman içindeki değişimini analiz etmek için zaman serisi analizi ve tahmin yöntemleri kullanılabilir. Bu, salgının seyrini ve etkilerini daha iyi anlamaya ve yönetmeye yardımcı olabilir.
6. Çoklu seviyeli analiz ve sosyoekonomik faktörler: Gelecek çalışmalarda, sosyoekonomik faktörler ve coğrafi özellikler gibi çoklu seviyeli verilerin analizine odaklanılabilir. Bu, Covid-19'un yayılımı ve etkileri üzerindeki sosyoekonomik ve coğrafi faktörlerin anlaşılmasına yardımcı olabilir.

7. Kullanılabilirliği ve erişilebilirliği artırmak: Geliştirilen modeller ve araçlar, kullanıcı dostu uygulamalar ve platformlar aracılığıyla geniş kitlelere sunulabilir. Bu, sağlık çalışanlarının, hükümetlerin ve genel halkın, Covid-19 ve benzeri pandemik hastalıklarla mücadelede daha etkili ve bilinçli kararlar almalarına yardımcı olabilir.
8. Gerçek zamanlı veri entegrasyonu: Gelecek çalışmalarda, geliştirilen modellerin gerçek zamanlı veri entegrasyonu ve güncelleme yeteneklerine odaklanılabilir. Bu, modellerin güncel verilerle sürekli olarak güncellenmesini ve doğruluğunun korunmasını sağlar.
9. Etki analizi ve politika önerileri: Model sonuçlarına dayalı olarak, Covid-19 ile mücadelede etkili politika önerileri ve stratejiler geliştirilebilir. Bu, salgının yayılmasını kontrol altına almak ve halk sağlığını korumak için önemlidir.
10. Karar destek sistemleri: Geliştirilen modeller ve analizler, sağlık çalışanları ve politika yapıcılar için karar destek sistemlerine entegre edilebilir. Bu, daha hızlı ve bilgi temelli kararlar alınmasını sağlayarak, pandemi yönetiminde etkinliği artırabilir.
11. Bireysel risk tahmini ve kişiselleştirilmiş tıp uygulamaları: MÖ ve VM yöntemleri, bireysel risk faktörlerini ve hastalık seyrini tahmin etmek için kullanılabilir. Bu sayede, hastalar için kişiselleştirilmiş tedavi planları ve önleyici önlemler geliştirilebilir. Gelecek çalışmalarda, kişiselleştirilmiş tıp uygulamalarına yönelik modellerin geliştirilmesi ve uygulanması üzerine yoğunlaşılabilir.
12. Yeni veri kaynaklarının entegrasyonu: Covid-19 risk faktörlerinin belirlenmesi ve analizi için daha fazla veri kaynağının entegrasyonu sağlanabilir. Örneğin, sosyal medya verileri, tıbbi görüntüleme verileri veya genetik bilgiler gibi yeni ve farklı veri türleri kullanılabilir. Bu verilerin entegrasyonu, daha kapsamlı ve doğru analizlerin gerçekleştirilmesine yardımcı olacaktır.

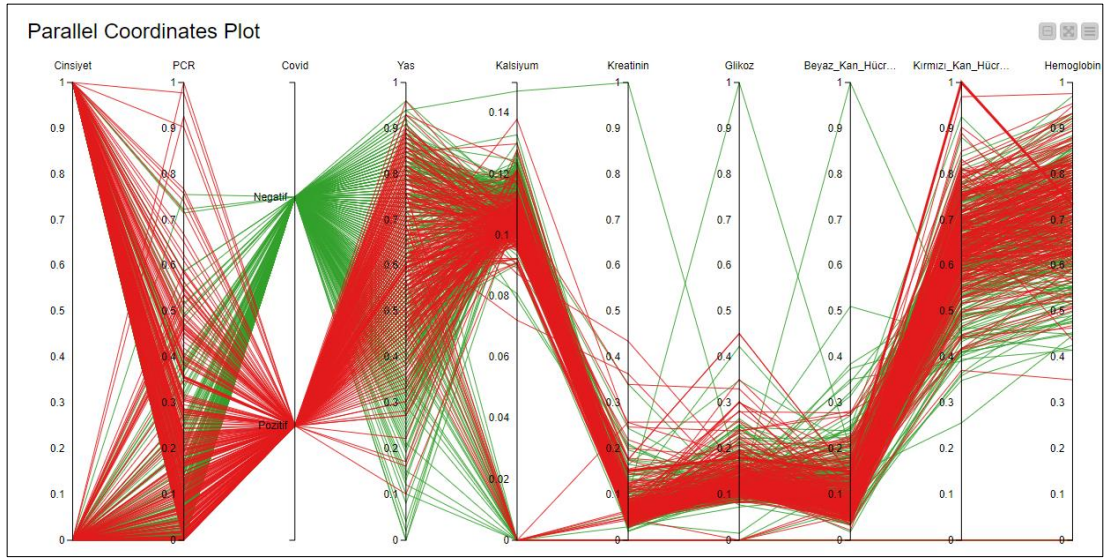


Bu öneriler, mevcut yüksek lisans tezinde yapılan çalışmanın üzerine inşa edilerek, Covid-19 ve benzeri pandemik hastalıklarla mücadelede daha etkili arařtırmalar ve uygulamalar geliřtirilmesine katkıda bulunabilir. Genel olarak, bu tezin temel amacı Covid-19 risk faktörlerinin belirlenmesi ve analizi için makine öğrenimi ve veri madencilięi yöntemlerinin kullanılmasıdır. Bu amaca ulaşmak için yapılan çalışmalar, pandemi ile mücadelede önemli bilgiler sunmuş ve gelecek çalışmalar için yol gösterici olmuştur. Bu doğrultuda, tezin sonuçları ve gelecek çalışmaların odak noktaları, Covid-19 pandemisinin daha etkili bir şekilde yönetilmesine ve kontrol altına alınmasına katkı sağlayacaktır.

## BÖLÜM 7

### SONUÇLAR

Deneysel çalışmalar veri setindeki tüm özellikler ve tam kan sayımı testlerinde olan özelliklerinin bir alt veri kümesinden oluşan verilerle gerçekleştirilmiştir. Şekil 7.1’de CBC verisetine ait Paralel Koordinatlar Grafiğine yer verilmiştir.

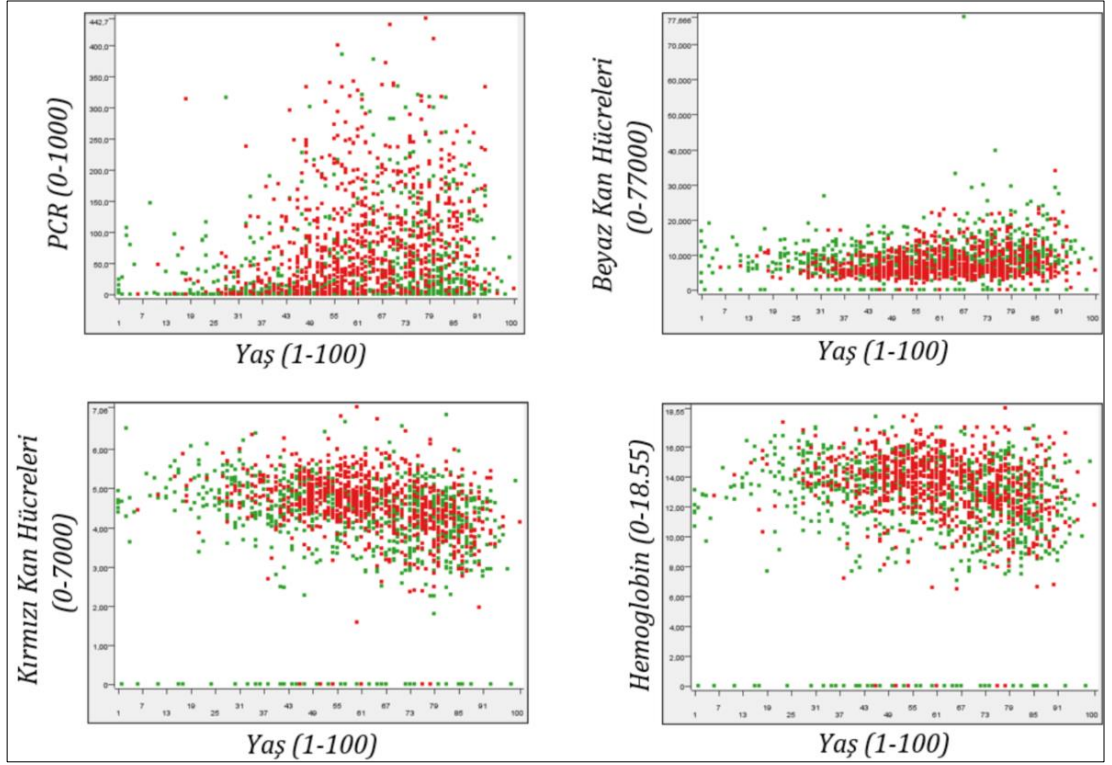


Şekil 7.1. 10 özellikten oluşan veri setine ait Parallel Coordinates Plot eğrisi

Deneysel çalışmalardan önce veri setindeki yaş değişkenine göre Covid19-PCR, Covid19-beyaz kan hücresi sayısı, Covid19-kırmızı kan hücresi sayısı, Covid19-hemoglobin seviyesi ilişkileri incelenmiş ve Şekil 7.2’de bu ilişkiler yeşil negatif ve kırmızı pozitif duruma karşılık gelecek şekilde belirtilmiştir.

Pandemi sürecinde Covid-19 hastalığının çocuklarda görülmediği, çocukların yalnız taşıyıcı olabilecekleri şeklindeki görüşlere ilişkin oluşturulan politikaların doğruluğunun incelenmesi açısından yaş değişkenine göre Covid-19 ve kan değerleri arasındaki ilişkiye bakıldığında hiçbir test sonucu ile doğrudan ilişkili olmadığı

görülmektedir. Elde edilen bu doğrudan ilişkili olmama durumu çoklu parametreye göre karar verilmesi gerektiğini ortaya koymuştur.



Şekil 7.2. Yaş, Covid-19 ve Kan değerleri arasındaki ilişki

Elde edilen sonuçlar MÖ algoritmalarının daha önce görmediği, test verileri ile yapılan tahminlerin sonuçlarını göstermektedir. Yapılan deneysel çalışmalar neticesinde OSR modeli %83' ün üzerinde doğru sınıflandırma başarısı göstermiştir. Çizelge 7.1 ve Çizelge 7.2'de kullanılan veri setine göre algoritmaların sınıflandırma başarıları verilmiştir.

Çizelge 7.1. OSR Veri kümesi ve algorilma sonuçları.

Veri	Model	Doğruluk	Cohen's Kappa	Duyarlılık	Kesinlik	F1 Skor
OSR veri kümesi	Random Forest	<b>83,135</b>	<b>0,659</b>	<b>0,838</b>	0,784	<b>0,810</b>
	Rprop MLP Learner	<b>81,349</b>	0,625	0,794	0,801	0,797
	Logistic Regression	<b>80,754</b>	0,614	0,775	0,818	0,796
	SVM Learner	<b>80,357</b>	0,606	0,770	0,814	0,792
	Decision Tree	76,587	0,526	0,763	0,710	0,735
	K Nearest Neighbor	75,794	0,51	0,751	0,706	0,728
	PNN Learner	73,214	0,464	0,690	0,753	0,720
	Naive Bayes	71,627	0,446	0,634	<b>0,900</b>	0,744

Çizelge 7.2. CBC Veri kümesi ve algorilma sonuçları.

Veri	Model	Doğruluk	Cohen's Kappa	Duyarlılık	Kesinlik	F1 Skor
CBC veri kümesi	Rprop MLP Learner	<b>79,365</b>	<b>0,587</b>	<b>0,788</b>	<b>0,781</b>	<b>0,784</b>
	Random Forest	77,579	0,55	<b>0,784</b>	0,736	0,759
	Logistic Regression	76,984	0,538	<b>0,781</b>	0,723	0,751
	SVM Learner	75,198	0,502	0,751	0,723	0,737
	Decision Tree	73,81	0,474	0,746	0,690	0,717
	PNN Learner	70,04	0,4	0,689	0,686	0,687
	Naive Bayes	61,706	0,247	0,568	<b>0,847</b>	0,680
	K Nearest Neighbor	53,889	0,277	0,623	0,628	0,626

Literatürde aynı veri setiyle yapılan bir diğer çalışmada [157] beş farklı MÖ modeli geliştirilmiş. OSR veri setinin tamamı için, AUC değerinin 0,83 ile 0,90 arasında değiştiği belirtilmiş, CBC veri seti Kesinlik değeri için 0,74 ile 0,86 aralığında değiştiği belirtilmiştir.

Yaptığımız çalışmada Doğruluk değeri için elde ettiğimiz en iyi sonuç %83,135 ile Random Forest algoritması olmuştur. Kesinlik değeri OSR veri setinde ise 0,90 ve CBC veri setinde 0,847 ile Naive Bayes algoritmaları tarafından hesaplanmıştır.

Elde edilen deneysel sonuçlar neticesinde OSA modelinin bizim yaptığımız testlerde başarılı çıktığıdır. Literatürle kıyaslandığında ise OSA modelinin de diğerleri kadar doğru sonuçlar ürettiği sonucuna ulaşılmıştır. Bu sonuç Naive Bayes ve Random Forest algoritmalarının OSR veri kümesindeki sınıflandırma problemini başarılı bir şekilde çözdüğünü göstermektedir. Çizelge 7.3’de doğrulama performansı için modellere ait en iyi algoritma sonuçlarına yer verilmiştir.

Çizelge 7.3. Doğrulama performansı için modellere ait en iyi algoritma sonuçları.

Veri	Model	Doğruluk	Cohen's Kappa	Duyarlılık	Kesinlik	F1 Skor
OSR	Naive Bayes	71,627	0,446	0,634	<b>0,900</b>	0,744
	Random Forest	<b>83,135</b>	<b>0,659</b>	<b>0,838</b>	0,784	<b>0,810</b>
CBC	Naive Bayes	71,627	0,446	0,568	<b>0,847</b>	0,680
	Rprop MLP Learner	81,349	0,625	<b>0,788</b>	0,781	<b>0,784</b>
	Random Forest	<b>83,135</b>	<b>0,659</b>	0,767	0,719	0,742

AI ve MÖ teknikleri ile Covid-19 hastalık tahmini tez sonuçları, Covid-19 teşhisinde kullanılacak modellerin seçiminde ve değerlendirmesinde karar vericilere ve sağlık uzmanlarına yol gösterici olabilir. Bu sayede, erken teşhis ve tedavi süreçlerinin daha etkili bir şekilde yönetilmesine katkı sağlayabilir ve elde edilen bulgular sayesinde daha hassas ve doğru yöntemlerin geliştirilmesi için yeni araştırma alanlarının ortaya çıkmasına da zemin hazırlayabilir. Özellikle, MÖ algoritmalarının farklı veri kümeleri ve parametrelerle nasıl optimize edilebileceğine dair yeni çalışmalar önerilebilir. Covid-19 teşhisi ve tedavisi süreçlerinin geliştirilmesi için politika yapıcılar ve sağlık otoriteleriyle iş birliği yaparak, elde edilen bulguların uygulamaya dökülmesi ve daha geniş kitlelere ulaştırılması önemlidir.

Sonuç olarak, Naive Bayes ve Random Forest algoritmaları, OSR veri kümesinde en iyi performansı gösteren algoritmalar olarak belirlenmiştir. Bu çalışma, farklı makine öğrenimi algoritmalarının başarımını değerlendirmekte ve en uygun modeli seçmede önemli bir kaynak sağlamaktadır. Ayrıca, bu sonuçlar, benzer veri kümeleri ve problemler üzerinde çalışan araştırmacılar için değerli bir rehber niteliğindedir.

## KAYNAKLAR

- [1] K. Roosa *ve diğeri*, “Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020,” *Infect Dis Model*, vol. 5, pp. 256–263, Jan. 2020, doi: 10.1016/j.idm.2020.02.002.
- [2] M. Venkatasubbaiah, P. Dwarakanadha Reddy, and S. V. Satyanarayana, “Literature-based review of the drugs used for the treatment of COVID-19,” *Curr Med Res Pract*, vol. 10, no. 3, pp. 100–109, May 2020, doi: 10.1016/j.cmrp.2020.05.013.
- [3] T. P. B. Thu, P. N. H. Ngoc, N. M. Hai, and L. A. Tuan, “Effect of the social distancing measures on the spread of COVID-19 in 10 highly infected countries,” *Science of the Total Environment*, vol. 742, Nov. 2020, doi: 10.1016/j.scitotenv.2020.140430.
- [4] D. Kang, H. Choi, J. H. Kim, and J. Choi, “Spatial epidemic dynamics of the COVID-19 outbreak in China,” *International Journal of Infectious Diseases*, vol. 94, pp. 96–102, May 2020, doi: 10.1016/j.ijid.2020.03.076.
- [5] M. Ciotti, M. Ciccozzi, A. Terrinoni, W. C. Jiang, C. Bin Wang, and S. Bernardini, “The COVID-19 pandemic,” *Critical Reviews in Clinical Laboratory Sciences*. Taylor and Francis Ltd., pp. 365–388, 2020. doi: 10.1080/10408363.2020.1783198.
- [6] H. T. Xie *ve diğeri*, “SARS-CoV-2 in the ocular surface of COVID-19 patients,” *Eye and Vision*, vol. 7, no. 1, Dec. 2020, doi: 10.1186/s40662-020-00189-0.
- [7] I. Ahammad and S. S. Lira, “Designing a novel mRNA vaccine against SARS-CoV-2: An immunoinformatics approach,” *Int J Biol Macromol*, vol. 162, pp. 820–837, Nov. 2020, doi: 10.1016/j.ijbiomac.2020.06.213.

- [8] J. Segovia-Juarez, J. M. Castagnetto, and G. F. Gonzales, “High altitude reduces infection rate of COVID-19 but not case-fatality rate,” *Respiratory Physiology and Neurobiology*, vol. 281. Elsevier B.V., Oct. 01, 2020. doi: 10.1016/j.resp.2020.103494
- [9] R. Lu *ve diğerleri*, “Epidemiological and clinical characteristics of COVID-19 patients in Nantong, China,” *J Infect Dev Ctries*, vol. 14, no. 5, pp. 440–446, May 2020, doi: 10.3855/jidc.12678.
- [10] V. Mevada, P. Dudhagara, H. Gandhi, N. Vaghamshi, U. Beladiya, and R. Patel, “Drug repurposing of approved drugs Antrafenine and Ergotamine for combating COVID19.” [Online]. Available: <https://www.excelra.com/covid-19-drug-repurposing-database/data>
- [11] J. Zhang *ve diğerleri*, “Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study,” *Lancet Infect Dis*, vol. 20, no. 7, pp. 793–802, Jul. 2020, doi: 10.1016/S1473-3099(20)30230-9.
- [12] K. Roosa *ve diğerleri*, “Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020,” *Infect Dis Model*, vol. 5, pp. 256–263, Jan. 2020, doi: 10.1016/j.idm.2020.02.002.
- [13] X. Yang, J. Zhao, Q. Yan, S. Zhang, Y. Wang, and Y. Li, “A case of COVID-19 patient with the diarrhea as initial symptom and literature review,” *Clin Res Hepatol Gastroenterol*, vol. 44, no. 5, pp. e109–e112, Oct. 2020, doi: 10.1016/j.clinre.2020.03.013.
- [14] M. Cataldi, G. Pignataro, and M. Tagliatela, “Neurobiology of coronaviruses: Potential relevance for COVID-19,” *Neurobiology of Disease*, vol. 143. Academic Press Inc., Sep. 01, 2020. doi: 10.1016/j.nbd.2020.105007.
- [15] D. Delaune *ve diğerleri*, “A novel SARS-CoV-2 related coronavirus in bats from Cambodia,” *Nat Commun*, vol. 12, no. 1, Dec. 2021, doi: 10.1038/s41467-021-26809-4.

- [16] S. Vilcek, “SARS-CoV-2: Zoonotic origin of pandemic coronavirus,” *Acta Virologica*, vol. 64, no. 3. AEPRESS, s.r.o., pp. 281–287, 2020. doi: 10.4149/av\_2020\_302.
- [17] R. Sardar, D. Satish, S. Birla, and D. Gupta, “Integrative analyses of SARS-CoV-2 genomes from different geographical locations reveal unique features potentially consequential to host-virus interaction, pathogenesis and clues for novel therapies,” *Heliyon*, vol. 6, no. 9, Sep. 2020, doi: 10.1016/j.heliyon.2020.e04658.
- [18] S. Dimonte, M. Babakir-Mina, T. Hama-Soor, and S. Ali, “Genetic Variation and Evolution of the 2019 Novel Coronavirus,” *Public Health Genomics*, vol. 24, no. 1–2, pp. 54–66, Feb. 2021, doi: 10.1159/000513530.
- [19] A. Pirzada, A. T. Mokhtar, and A. D. Moeller, “COVID-19 and Myocarditis: What Do We Know So Far?,” *CJC Open*, vol. 2, no. 4. Elsevier Inc, pp. 278–285, Jul. 01, 2020. doi: 10.1016/j.cjco.2020.05.005.
- [20] M. Chahrour *ve diğerleri*, “A Bibliometric Analysis of COVID-19 Research Activity: A Call for Increased Output,” *Cureus*, Mar. 2020, doi: 10.7759/cureus.7357.
- [21] M. Tandan, Y. Acharya, S. Pokharel, and M. Timilsina, “Discovering symptom patterns of COVID-19 patients using association rule mining,” *Comput Biol Med*, vol. 131, Apr. 2021, doi: 10.1016/j.combiomed.2021.104249.
- [22] C. A. Romero-Gameros *ve diğerleri*, “Diagnostic accuracy of symptoms as a diagnostic tool for SARS-CoV 2 infection: a cross-sectional study in a cohort of 2,173 patients,” *BMC Infect Dis*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12879-021-05930-1.
- [23] J. Czubak, K. Stolarczyk, A. Orzel, M. Frączek, and T. Zatoński, “Comparison of the clinical differences between COVID-19, SARS, influenza, and the common cold: A systematic literature review,” *Advances in Clinical and*



- Experimental Medicine*, vol. 30, no. 1. Wroclaw University of Medicine, pp. 109–114, 2021. doi: 10.17219/ACEM/129573.
- [24] S. Brojonegoro, G. Meneng, and K. Bandarlampung, “**Clinical Characteristics of Coronavirus Disease 2019**,” 2020.
- [25] E. Fragaszy *ve diğerleri*, “Symptom profiles and accuracy of clinical definitions for COVID-19 in the community. **Results of the Virus Watch community cohort. on behalf of Virus Watch Collaborative\***”, doi: 10.1101/2021.05.14.21257229.
- [26] S. Nunthavichitra, S. Prapaso, V. Luvira, S. Muangnoicharoen, P. Leungwutiwong, and W. Piyaphanee, “Case report: COVID-19 presenting as acute undifferentiated febrile illness—A tropical world threat,” *American Journal of Tropical Medicine and Hygiene*, vol. 103, no. 1, pp. 83–85, Jul. 2020, doi: 10.4269/ajtmh.20-0440.
- [27] P. Samidoust and A. Samidoust, “10-Since December 2019, **Severe Acute Respiratory Syndrome Coronavirus-2-SARS-CoV-2**,” 2020.
- [28] C. Mietto *ve diğerleri*, “Acute Lower Limb Ischemia as Clinical Presentation of COVID-19 Infection,” *Ann Vasc Surg*, vol. 69, pp. 80–84, Nov. 2020, doi: 10.1016/j.avsg.2020.08.004.
- [29] L. Cipolloni *ve diğerleri*, “Preliminary post-mortem COVID-19 evidence of endothelial injury and factor VIII hyperexpression,” *Diagnostics*, vol. 10, no. 8, Aug. 2020, doi: 10.3390/diagnostics10080575.
- [30] S. Kabashneh, H. Ali, and S. Alkassis, “Multi-Organ Failure in a Patient With Diabetes due to COVID-19 With Clear Lungs,” *Cureus*, May 2020, doi: 10.7759/cureus.8147.
- [31] G. Cavalli *ve diğerleri*, “Interleukin-1 blockade with high-dose anakinra in patients with COVID-19, acute respiratory distress syndrome, and hyperinflammation: a retrospective cohort study,” *Lancet Rheumatol*, vol. 2, no. 6, pp. e325–e331, Jun. 2020, doi: 10.1016/S2665-9913(20)30127-2.

- [32] W. Lyu and G. L. Wehby, “Community use of face masks and COVID-19: Evidence from a natural experiment of state mandates in the US,” *Health Aff*, vol. 39, no. 8, pp. 1419–1425, Aug. 2020, doi: 10.1377/hlthaff.2020.00818.
- [33] D. Loconsole *ve diğerleri*, “Changing features of COVID-19: Characteristics of infections with the SARS-CoV-2 delta (B.1.617.2) and alpha (B.1.1.7) variants in Southern Italy,” *Vaccines (Basel)*, vol. 9, no. 11, Nov. 2021, doi: 10.3390/vaccines9111354.
- [34] M. W. Tenforde *ve diğerleri*, “Morbidity and Mortality Weekly Report Symptom Duration and Risk Factors for Delayed Return to Usual Health Among Outpatients with COVID-19 in a Multistate Health Care Systems Network-United States, March-June 2020,” 2020. [Online]. Available: <https://www.cdc.gov/mmwr>
- [35] E. R. Thompson *ve diğerleri*, “Universal masking to control healthcare-associated transmission of severe acute respiratory coronavirus virus 2 (SARS-CoV-2),” *Infect Control Hosp Epidemiol*, vol. 43, no. 3, pp. 344–350, Mar. 2022, doi: 10.1017/ice.2021.127.
- [36] K. L. Andrejko *ve diğerleri*, “Morbidity and Mortality Weekly Report Effectiveness of Face Mask or Respirator Use in Indoor Public Settings for Prevention of SARS-CoV-2 Infection-California, February-December 2021,” 2020, doi: 10.1093/cid/ciab640/6324500#supplementary-data.
- [37] M. N. Zahid, M. S. Moosa, S. Perna, and E. Bin Buti, “A review on COVID-19 vaccines: stages of clinical trials, mode of actions and efficacy,” *Arab Journal of Basic and Applied Sciences*, vol. 28, no. 1. Taylor and Francis Ltd., pp. 225–233, 2021. doi: 10.1080/25765299.2021.1903144.
- [38] H. L. Moline *ve diğerleri*, “Morbidity and Mortality Weekly Report Effectiveness of COVID-19 Vaccines in Preventing Hospitalization Among Adults Aged  $\geq 65$  Years-COVID-NET, 13 States, February-April 2021”, doi: 10.1101/2021.04.21.21255473v1.

- [39] J. Wang *ve diğerleri*, “Safety and immunogenicity of COVID-19 vaccination in patients with non-alcoholic fatty liver disease (CHESS2101): A multicenter study,” *J Hepatol*, vol. 75, no. 2, pp. 439–441, Aug. 2021, doi: 10.1016/j.jhep.2021.04.026.
- [40] S. M. Moghadas *ve diğerleri*, “COVID-19 deaths and hospitalizations averted by rapid vaccination rollout in the United States Key Points”, doi: 10.1101/2021.07.07.21260156.
- [41] G. Yamey *ve diğerleri*, “Funding the Development and Manufacturing of COVID-19 Vaccines.” [Online]. Available: <https://ssrn.com/abstract=3575660>
- [42] B. N. Ashraf, “Economic impact of government interventions during the COVID-19 pandemic: International evidence from financial markets,” *J Behav Exp Finance*, vol. 27, Sep. 2020, doi: 10.1016/j.jbef.2020.100371.
- [43] A. Stapleton, “Choosing not to follow rules that will reduce the spread of COVID-19,” *Journal of Contextual Behavioral Science*, vol. 17. Elsevier Inc., pp. 73–78, Jul. 01, 2020. doi: 10.1016/j.jcbs.2020.07.002.
- [44] I. Ayouni *ve diğerleri*, “Effective public health measures to mitigate the spread of COVID-19: a systematic review,” *BMC Public Health*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12889-021-11111-1.
- [45] D. Hu *ve diğerleri*, “More effective strategies are required to strengthen public awareness of COVID-19: Evidence from google trends,” *J Glob Health*, vol. 10, no. 1, Jun. 2020, doi: 10.7189/jogh.10.011003.
- [46] F. Bilgili *ve diğerleri*, “The Age Structure, Stringency Policy, Income, and Spread of Coronavirus Disease 2019: Evidence From 209 Countries,” *Front Psychol*, vol. 11, Feb. 2021, doi: 10.3389/fpsyg.2020.632192.
- [47] Z. Su *ve diğerleri*, “Vaccines are not yet a silver bullet: The imperative of continued communication about the importance of COVID-19 safety

- measures,” *Brain, Behavior, and Immunity - Health*, vol. 12. Elsevier Inc., Mar. 01, 2021. doi: 10.1016/j.bbih.2021.100204.
- [48] X. Tang ve diğerleri, “SARS-CoV-2 infection induces beta cell transdifferentiation,” *Cell Metab*, vol. 33, no. 8, pp. 1577-1591.e7, Aug. 2021, doi: 10.1016/j.cmet.2021.05.015.
- [49] S. A. SeyedAlinaghi ve diğerleri, “Characterization of SARS-CoV-2 different variants and related morbidity and mortality: a systematic review,” *European Journal of Medical Research*, vol. 26, no. 1. BioMed Central Ltd, Dec. 01, 2021. doi: 10.1186/s40001-021-00524-8.
- [50] J. Lamptey ve diğerleri, “Genomic and epidemiological characteristics of sars-cov-2 in Africa,” *PLoS Negl Trop Dis*, vol. 15, no. 4, Apr. 2021, doi: 10.1371/JOURNAL.PNTD.0009335.
- [51] S. Kannan, P. Shaik, S. Ali, and A. Sheeza, “Evolving biothreat of variant SARS-CoV-2 – molecular properties, virulence and epidemiology,” 2021.
- [52] B. Roy, J. Dhillon, N. Habib, and B. Pugazhandhi, “Global variants of COVID-19: Current understanding,” *Journal of Biomedical Sciences*, vol. 8, no. 1, pp. 8–11, Jul. 2021, doi: 10.3126/jbs.v8i1.38453.
- [53] R. Sanjuán, M. R. Nebot, N. Chirico, L. M. Mansky, and R. Belshaw, “Viral Mutation Rates,” *J Virol*, vol. 84, no. 19, pp. 9733–9748, Oct. 2010, doi: 10.1128/jvi.00694-10.
- [54] R. Sanjuán and P. Domingo-Calap, “Mechanisms of viral mutation,” *Cellular and Molecular Life Sciences*, vol. 73, no. 23. Birkhauser Verlag AG, pp. 4433–4448, Jul. 08, 2016. doi: 10.1007/s00018-016-2299-6.
- [55] S. A. SeyedAlinaghi ve diğerleri, “Characterization of SARS-CoV-2 different variants and related morbidity and mortality: a systematic review,” *European Journal of Medical Research*, vol. 26, no. 1. BioMed Central Ltd, Dec. 01, 2021. doi: 10.1186/s40001-021-00524-8.

- [56] P. Paul *ve diğerleri*, “Morbidity and Mortality Weekly Report Genomic Surveillance for SARS-CoV-2 Variants Circulating in the United States,” 2020, doi: 10.1101/2020.10.07.20208504v2.full.
- [57] K. Tao *ve diğerleri*, “The biological and clinical significance of emerging SARS-CoV-2 variants,” *Nature Reviews Genetics*, vol. 22, no. 12. Nature Research, pp. 757–773, Dec. 01, 2021. doi: 10.1038/s41576-021-00408-x.
- [58] S. Weigang *ve diğerleri*, “Within-host evolution of SARS-CoV-2 in an immunosuppressed COVID-19 patient as a source of immune escape variants,” *Nat Commun*, vol. 12, no. 1, Dec. 2021, doi: 10.1038/s41467-021-26602-3.
- [59] M. Haghani, M. C. J. Bliemer, F. Goerlandt, and J. Li, “The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review,” *Saf Sci*, vol. 129, Sep. 2020, doi: 10.1016/j.ssci.2020.104806.
- [60] T. Y. Akintunde *ve diğerleri*, “Bibliometric analysis of global scientific literature on effects of COVID-19 pandemic on mental health,” *Asian J Psychiatr*, vol. 63, Sep. 2021, doi: 10.1016/j.ajp.2021.102753.
- [61] M. M. Hossain, “Current status of global research on novel coronavirus disease (COVID-19): a bibliometric analysis and knowledge mapping,” *F1000Res*, vol. 9, p. 374, May 2020, doi: 10.12688/f1000research.23690.1.
- [62] N. Bragazzi, M. Mansour, A. Bonsignore, and R. Ciliberti, “The Role of Hospital and Community Pharmacists in the Management of COVID-19: Towards an Expanded Definition of the Roles, Responsibilities, and Duties of the Pharmacist,” *Pharmacy*, vol. 8, no. 3, p. 140, Aug. 2020, doi: 10.3390/pharmacy8030140.
- [63] M. Haghani, M. C. J. Bliemer, F. Goerlandt, and J. Li, “The scientific literature on Coronaviruses, COVID-19 and its associated safety-related research dimensions: A scientometric analysis and scoping review,” *Saf Sci*, vol. 129, Sep. 2020, doi: 10.1016/j.ssci.2020.104806.

- [64] B. Gharizadeh *ve diğerleri*, “Navigating the Pandemic Response Life Cycle: Molecular Diagnostics and Immunoassays in the Context of COVID-19 Management,” *IEEE Rev Biomed Eng*, vol. 14, pp. 30–47, 2021, doi: 10.1109/RBME.2020.2991444.
- [65] G. Nageshwaran, R. C. Harris, and C. El Guerche-Seblain, “Review of the role of big data and digital technologies in controlling COVID-19 in Asia: Public health interest vs. privacy,” *Digital Health*, vol. 7. SAGE Publications Inc., 2021. doi: 10.1177/20552076211002953.
- [66] S. Kleitman *ve diğerleri*, “To comply or not comply? A latent profile analysis of behaviours and attitudes during the COVID-19 pandemic,” *PLoS One*, vol. 16, no. 7 July 2021, Jul. 2021, doi: 10.1371/journal.pone.0255268.
- [67] M. S. Keshavan, “Pandemics and psychiatry: Repositioning research in context of COVID-19,” *Asian J Psychiatr*, vol. 51, Jun. 2020, doi: 10.1016/j.ajp.2020.102159.
- [68] H. Blain *ve diğerleri*, “Atypical clinical presentation of COVID-19 infection in residents of a long-term care facility,” *Eur Geriatr Med*, vol. 11, no. 6, pp. 1085–1088, Dec. 2020, doi: 10.1007/s41999-020-00352-9.
- [69] F. Sameni *ve diğerleri*, “COVID-19 and Skin Manifestations: An Overview of Case Reports/Case Series and Meta-Analysis of Prevalence Studies,” *Frontiers in Medicine*, vol. 7. Frontiers Media S.A., Oct. 29, 2020. doi: 10.3389/fmed.2020.573188.
- [70] Y. Alimohamadi, M. Sepandi, M. Taghdir, and H. Hosamirudsari, “Determine the most common clinical symptoms in COVID-19 patients: A systematic review and meta-analysis,” *Journal of Preventive Medicine and Hygiene*, vol. 61, no. 3. Pacini Editore S.p.A., pp. E304–E312, Oct. 06, 2020. doi: 10.15167/2421-4248/jpmh2020.61.3.1530.
- [71] R. da Rosa Mesquita *ve diğerleri*, “Clinical manifestations of COVID-19 in the general population: systematic review,” *Wiener Klinische Wochenschrift*, vol.

- 133, no. 7–8. Springer, pp. 377–382, Apr. 01, 2021. doi: 10.1007/s00508-020-01760-4.
- [72] X. Gao *ve diğerleri*, “Improving the early diagnosis of suspected patients with COVID-19: A retrospective study of 106 patients,” *J Infect Dev Ctries*, vol. 14, no. 6, pp. 547–553, Jun. 2020, doi: 10.3855/jidc.12992.
- [73] L. Garibyan and N. Avashia, “Polymerase chain reaction,” *Journal of Investigative Dermatology*, vol. 133, no. 3, pp. 1–4, 2013, doi: 10.1038/jid.2013.1.
- [74] K. M. Akter *ve diğerleri*, “RT-PCR Mediated Identification of SARS-CoV-2 Patients from Particular Regions of Bangladesh and the Multi-factorial Analysis Considering Their Pre and Post Infection Health Conditions,” *Biotechnology Journal International*, pp. 43–56, Dec. 2020, doi: 10.9734/bji/2020/v24i630121.
- [75] K. H. Song *ve diğerleri*, “Dynamics of viral load and anti-SARS-CoV-2 antibodies in patients with positive RT-PCR results after recovery from COVID-19,” *Korean Journal of Internal Medicine*, vol. 36, no. 1, pp. 11–14, Jan. 2021, doi: 10.3904/kjim.2020.325.
- [76] T. Ishige *ve diğerleri*, “Highly sensitive detection of SARS-CoV-2 RNA by multiplex rRT-PCR for molecular diagnosis of COVID-19 by clinical laboratories,” *Clinica Chimica Acta*, vol. 507, pp. 139–142, Aug. 2020, doi: 10.1016/j.cca.2020.04.023.
- [77] P. Niu *ve diğerleri*, “Three Novel Real-Time RT-PCR Assays for Detection of COVID-19 Virus.”
- [78] J. Lim and J. Lee, “Current laboratory diagnosis of coronavirus disease 2019,” *Korean Journal of Internal Medicine*, vol. 35, no. 4. Korean Association of Internal Medicine, pp. 741–748, 2020. doi: 10.3904/KJIM.2020.257.

- [79] Y. Uwamino *ve diğ erleri*, “Evaluation of the usability of various rapid antibody tests in the diagnostic application for COVID-19,” *Ann Clin Biochem*, vol. 58, no. 3, pp. 174–180, May 2021, doi: 10.1177/0004563220984827.
- [80] N. Eshghifar, A. Busheri, R. Shrestha, and S. Beqaj, “Evaluation of analytical performance of seven rapid antigen detection kits for detection of SARS-CoV-2 virus,” *Int J Gen Med*, vol. 14, pp. 435–440, 2021, doi: 10.2147/IJGM.S297762.
- [81] S. Y. Chen *ve diğ erleri*, “Multicenter evaluation of two chemiluminescence and three lateral flow immunoassays for the diagnosis of COVID-19 and assessment of antibody dynamic responses to SARS-CoV-2 in Taiwan,” *Emerg Microbes Infect*, vol. 9, no. 1, pp. 2157–2168, Jan. 2020, doi: 10.1080/22221751.2020.1825016.
- [82] G. Zhang, S. Nie, Z. Zhang, and Z. Zhang, “Longitudinal change of severe acute respiratory syndrome coronavirus 2 antibodies in patients with coronavirus disease 2019,” *Journal of Infectious Diseases*, vol. 222, pp. 183–188, 2020, doi: 10.1093/infdis/jiaa229.
- [83] S. Domeracki, R. N. Clapp, K. Taylor, C. M. Lu, H. Lampiris, and P. D. Blanc, “Cycle Threshold to Test Positivity in COVID-19 for Return to Work Clearance in Health Care Workers,” *J Occup Environ Med*, vol. 62, no. 11, pp. 889–891, Nov. 2020, doi: 10.1097/JOM.0000000000001996.
- [84] A. Kimball *ve diğ erleri*, “Morbidity and Mortality Weekly Report Asymptomatic and Presymptomatic SARS-CoV-2 Infections in Residents of a Long-Term Care Skilled Nursing Facility-King County, Washington, March 2020,” 2019. [Online]. Available: <https://www.cdc.gov/mmwr>
- [85] R. Duncan and B. Williamson, “Blood evidence: how DNA is revolutionizing the way we solve crimes,” *Journal of Clinical Investigation*, vol. 112, no. 9, pp. 1268–1268, Nov. 2003, doi: 10.1172/jci20207.



- [86] D. feng Lv ve diğeri, “Dynamic change process of target genes by RT-PCR testing of SARS-Cov-2 during the course of a Coronavirus Disease 2019 patient,” *Clinica Chimica Acta*, vol. 506, pp. 172–175, Jul. 2020, doi: 10.1016/j.cca.2020.03.032.
- [87] D. feng Lv ve diğeri, “Dynamic change process of target genes by RT-PCR testing of SARS-Cov-2 during the course of a Coronavirus Disease 2019 patient,” *Clinica Chimica Acta*, vol. 506, pp. 172–175, Jul. 2020, doi: 10.1016/j.cca.2020.03.032.
- [88] T. Ishige ve diğeri, “Highly sensitive detection of SARS-CoV-2 RNA by multiplex rRT-PCR for molecular diagnosis of COVID-19 by clinical laboratories,” *Clinica Chimica Acta*, vol. 507, pp. 139–142, Aug. 2020, doi: 10.1016/j.cca.2020.04.023.
- [89] Y. Zhang ve diğeri, “Discrimination of False Negative Results in RT-PCR Detection of SARS-CoV-2 RNAs in Clinical Specimens by Using an Internal Reference,” *Virol Sin*, vol. 35, no. 6, pp. 758–767, Dec. 2020, doi: 10.1007/s12250-020-00273-8.
- [90] M. R. Hasan ve diğeri, “Detection of SARS-CoV-2 RNA by direct RT-qPCR on nasopharyngeal specimens without extraction of viral RNA,” *PLoS One*, vol. 15, no. 7 July, Jul. 2020, doi: 10.1371/journal.pone.0236564.
- [91] A. J. Johnson, N. Karabatsos, and R. S. Lanciotti, “Detection of Colorado Tick Fever Virus by Using Reverse Transcriptase PCR and Application of the Technique in Laboratory Diagnosis,” 1997. [Online]. Available: <https://journals.asm.org/journal/jcm>
- [92] G. Baier, D. Telford, and E. Gulbins, “35-Improved specificity of RT-PCR amplifications using nested cDNA primers,” 1993.
- [93] V. M. Corman ve diğeri, “Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR,” *Eurosurveillance*, vol. 25, no. 3, Jan. 2020, doi: 10.2807/1560-7917.ES.2020.25.3.2000045.

- [94] J. Winer, C. Kwang, S. Jung, I. Shackel, and P. M. Williams, “Development and Validation of Real-Time Quantitative Reverse Transcriptase-Polymerase Chain Reaction for Monitoring Gene Expression in Cardiac Myocytes in Vitro,” 1999. [Online]. Available: <http://www.idealibrary.comon>
- [95] S. A. Bustin and T. Nolan, “RT-QPCR testing of SARS-COV-2: A primer,” *Int J Mol Sci*, vol. 21, no. 8, Apr. 2020, doi: 10.3390/ijms21083004.
- [96] T. Ishige *ve diğerleri*, “Highly sensitive detection of SARS-CoV-2 RNA by multiplex rRT-PCR for molecular diagnosis of COVID-19 by clinical laboratories,” *Clinica Chimica Acta*, vol. 507, pp. 139–142, Aug. 2020, doi: 10.1016/j.cca.2020.04.023.
- [97] M. R. Hasan *ve diğerleri*, “Detection of SARS-CoV-2 RNA by direct RT-qPCR on nasopharyngeal specimens without extraction of viral RNA,” *PLoS One*, vol. 15, no. 7 July, Jul. 2020, doi: 10.1371/journal.pone.0236564.
- [98] Y. Zhang *ve diğerleri*, “Discrimination of False Negative Results in RT-PCR Detection of SARS-CoV-2 RNAs in Clinical Specimens by Using an Internal Reference,” *Virol Sin*, vol. 35, no. 6, pp. 758–767, Dec. 2020, doi: 10.1007/s12250-020-00273-8.
- [99] G. Lippi, A. M. Simundic, and M. Plebani, “Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19),” *Clinical Chemistry and Laboratory Medicine*, vol. 58, no. 7. De Gruyter, pp. 1070–1076, Jul. 01, 2020. doi: 10.1515/cclm-2020-0285.
- [100] Woloshin, S, Patel, N, and Kesselheim, “False Negative Tests for SARS-CoV-2 Infection — Challenges and Implications,” 2020.
- [101] L. Wynants, “Supplementary material Systematic review and critical appraisal of prediction models for diagnosis and prognosis of COVID-19 infection Contents,” 2020.

- [102] Wynants *ve diğerleri*, “Prediction models for diagnosis and prognosis of covid-19 systematic review and critical appraisal,” 2020.
- [103] L. Li *ve diğerleri*, “Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy,” *Radiology*, vol. 296, no. 2, pp. E65–E71, Aug. 2020, doi: 10.1148/radiol.2020200905.
- [104] Gozes *ve diğerleri*, “Rapid AI Development Cycle for the Coronavirus (COVID-19) Pandemic Initial Results for Automated Detection & Patient Monitoring using Deep Learning CT Image Analysis,” 2020.
- [105] Ozturk *ve diğerleri*, “Automated detection of COVID-19 cases using deep neural networks with X-ray images,” 2020.
- [106] Mei *ve diğerleri*, “Artificial intelligence-enabled rapid diagnosis of patients with COVID-19,” 2020.
- [107] MICHAEL B. WEINSTOCK, ANA ECHENIQUE, and JOSHUA W. RUSSELL, “Chest X-Ray Findings in 636 Ambulatory Patients with COVID-19 Presenting to an Urgent Care Center A Normal Chest X-Ray Is no Guarantee,” 2020.
- [108] Fan *ve diğerleri*, “Hematologic parameters in patients with COVID-19 infection,” 2020.
- [109] D. Ferrari, A. Motta, M. Strollo, G. Banfi, and M. Locatelli, “Routine blood tests as a potential diagnostic tool for COVID-19,” *Clin Chem Lab Med*, vol. 58, no. 7, pp. 1095–1099, Jul. 2020, doi: 10.1515/cclm-2020-0398.
- [110] V. Formica *ve diğerleri*, “Complete blood count might help to identify subjects with high probability of testing positive to SARS-CoV-2,” *Clinical Medicine, Journal of the Royal College of Physicians of London*, vol. 20, no. 4, Jul. 2020, doi: 10.7861/CLINMED.2020-0373.

- [111] J. Wu *ve diğeri*, “Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results,” 2020, doi: 10.1101/2020.04.02.20051136.
- [112] F. Soares *ve diğeri*, “A novel specific artificial intelligence-based method to identify COVID-19 cases using simple blood exams,” 2020, doi: 10.1101/2020.04.10.20061036.
- [113] A. As *ve diğeri*, “Artificial intelligence driven assessment of routinely collected healthcare data is an effective screening test for COVID-19 in patients presenting to hospital,” 2020, doi: 10.1101/2020.07.07.20148361.
- [114] M. Kukar *ve diğeri*, “COVID-19 diagnosis by routine blood tests using machine learning,” *Sci Rep*, vol. 11, no. 1, p. 10738, Dec. 2021, doi: 10.1038/s41598-021-90265-9.
- [115] M. Alpaslan, S. Özkaçmaz, Y. Dadali, and İ. Dündar, “BİLİMSEL ARAŞTIRMA MAKALESİ Covid-19 Tanılı Hastalarımızın Bilgisayarlı Tomografi Sonuçları: Tipik ve Atipik Bulgular Computed Tomography Results of Patients with Covid-19: Typical and Atypical Find-ings.”
- [116] S. LEZKI, H. SÖNMEZ, A. BEKKI, L. TERLEMEZ, and C. BAL, “VERİ MADENCİLİĞİ,” *T.C. ANADOLU ÜNİVERSİTESİ YAYINI*, 2019.
- [117] K. A. Bhavsar, J. Singla, Y. D. Al-Otaibi, O. Y. Song, Y. Bin Zikria, and A. K. Bashir, “Medical diagnosis using machine learning: a statistical review,” *Computers, Materials and Continua*, vol. 67, no. 1. Tech Science Press, pp. 107–125, 2021. doi: 10.32604/cmc.2021.014604.
- [118] G. Koren, G. Nordon, K. Radinsky, and V. Shalev, “Machine learning of big data in gaining insight into successful treatment of hypertension,” *Pharmacol Res Perspect*, vol. 6, no. 3, Jun. 2018, doi: 10.1002/prp2.396.
- [119] S. Kimeswenger *ve diğeri*, “Artificial neural networks and pathologists recognize basal cell carcinomas based on different histological patterns,”

- Modern Pathology*, vol. 34, no. 5, pp. 895–903, May 2021, doi: 10.1038/s41379-020-00712-7.
- [120] X. Zeng and G. Luo, “Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection,” *Health Inf Sci Syst*, vol. 5, no. 1, Dec. 2017, doi: 10.1007/s13755-017-0023-z.
- [121] T. Inoue *ve diğerleri*, “XGBoost, a Machine Learning Method, Predicts Neurological Recovery in Patients with Cervical Spinal Cord Injury,” *Neurotrauma Rep*, vol. 1, no. 1, pp. 8–16, Jan. 2020, doi: 10.1089/neur.2020.0009.
- [122] S. Shenoy *ve diğerleri*, “Artificial intelligence in differentiating tropical infections: A step ahead,” *PLoS Negl Trop Dis*, vol. 16, no. 6, Jun. 2022, doi: 10.1371/JOURNAL.PNTD.0010455.
- [123] M. Groezinger, D. Huppert, R. Strobl, and E. Grill, “Development and validation of a classification algorithm to diagnose and differentiate spontaneous episodic vertigo syndromes: results from the DizzyReg patient registry,” *J Neurol*, vol. 267, pp. 160–167, Dec. 2020, doi: 10.1007/s00415-020-10061-9.
- [124] A. Väänänen, K. Haataja, K. Vehviläinen-Julkunen, and P. Toivanen, “AI in healthcare: A narrative review,” *F1000Res*, vol. 10, p. 6, Oct. 2021, doi: 10.12688/f1000research.26997.2.
- [125] G. Hyunjung, L. Ling, and P. Hui, “Modelling global public health strategies in COVID-19 pandemic using deep reinforcement learning,” 2020, doi: 10.21203/rs.3.rs-31226/v1.
- [126] “<https://www.worldometers.info/coronavirus/>,” Apr. 2023.
- [127] F. Zhou *ve diğerleri*, “Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study,” *The Lancet*, vol. 395, no. 10229, pp. 1054–1062, Mar. 2020, doi: 10.1016/S0140-6736(20)30566-3.

- [128] A. P. Khawaja *ve diğ erleri*, “Associations with covid-19 hospitalisation amongst 406,793 adults: the UK Biobank prospective cohort study,” 2020, doi: 10.1101/2020.05.06.20092957.
- [129] L. J. Ahrenfeldt, “Burden and prevalence of risk factors for severe COVID-19 disease in the ageing European population-A SHARE-based analysis,” 2020, doi: 10.21203/rs.3.rs-73657/v1.
- [130] B. Wang, R. Li, Z. Lu, and Y. Huang, “49-Does comorbidity increase the risk of patients with COVID-19,” 2020.
- [131] A. Alimadadi, S. Aryal, I. Manandhar, P. B. Munroe, B. Joe, and X. Cheng, “Artificial intelligence and machine learning to fight covid-19,” *Physiological Genomics*, vol. 52, no. 4. American Physiological Society, pp. 200–202, Apr. 01, 2020. doi: 10.1152/physiolgenomics.00029.2020.
- [132] Mitchell Thomas, *Machine Learning*. New Delhi: McGraw Hill Education India, 1997.
- [133] J. F. R. T. Trevpr Hastie, *The Elements of Statistical Learning*. 2001.
- [134] T. Vanrullen, C. Malagôn, and L. F. Chaari, “50-Rise of Data Mining,” 2011.
- [135] I. H. Witten, E. Frank, M. A. Hall, and M. Kaufmann, “Data Mining Practical Machine Learning Tools and Techniques Third Edition M< Contents,” 2005.
- [136] P. N. , S. M. , & K. Tan, *Introduction to Data Mining*. **Pearson Education India**, 2016.
- [137] Y. Freund and R. E. Schapire, “Journal of Computer and System Sciences s SS1504 journal of computer and system sciences,” 1997.
- [138] J. H. Friedman, “Greedy Function Approximation: A Gradient Boosting Machine,” 2001.

- [139] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [140] Q. Meng ve diğerleri, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree,” 2017.
- [141] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: unbiased boosting with categorical features,” 2018. [Online]. Available: <https://github.com/catboost/catboost>
- [142] D. W. T. H. R. T. Gareth James, *An Introduction to Statistical Learning*. 2013.
- [143] J. B. MacQueen, “Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,” 1967.
- [144] P. J. R. Leonard Kaufman, *Finding Groups in Data An Introduction to Cluster Analysis*. Wiley Interscience, 1990.
- [145] M. , K. H. P. , S. J. , & X. X. Ester, “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” *In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, 1996.
- [146] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, “OPTICS: Ordering Points To Identify the Clustering Structure,” 1999.
- [147] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data Clustering: A Review,” 2000.
- [148] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic, 1981.
- [149] U. Von Luxburg, “A tutorial on spectral clustering,” *Stat Comput*, vol. 17, no. 4, pp. 395–416, Dec. 2007, doi: 10.1007/s11222-007-9033-z.

- [150] T. Zhang, R. Ramakrishnan, and M. Livny, “BIRCH: An Efficient Data Clustering Method for Very Large Databases.”
- [151] D. Comaniciu and P. Meer, “Mean shift: A robust approach toward feature space analysis,” *IEEE Trans Pattern Anal Mach Intell*, vol. 24, no. 5, pp. 603–619, May 2002, doi: 10.1109/34.1000236.
- [152] B. J. Frey and D. Dueck, “Clustering by Passing Messages Between Data Points,” 2007.
- [153] Christopher M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [154] D. H. Fisher, “Machine Learning Knowledge Acquisition Via Incremental Conceptual Clustering,” 1987.
- [155] T. Kohonen, “The Self-Organizing Map,” *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990, doi: 10.1109/5.58325.
- [156] M. R. Berthold *ve diğerleri*, “KNIME: The Konstanz Information Miner. In Studies in Classification, Data Analysis, and Knowledge Organization,” 2008.
- [157] F. Cabitza *ve diğerleri*, “Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests,” *Clin Chem Lab Med*, vol. 59, no. 2, pp. 421–431, Feb. 2021, doi: 10.1515/cclm-2020-1294.



## ÖZGEÇMİŞ

Abdullah Türker TOKU ilk, orta ve lise öğrenimini Osmaniye’de tamamladı. Küçük yaşlardan itibaren uzak doğu sporları ve basketbola ilgi duydu ve bu alanlarda başarılar elde etti. Ankara Üniversitesi Elektronik Haberleşme, Anadolu Üniversitesi Uluslararası İlişkiler, Yönetim Bilişim Sistemleri ve Karabük Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği (tezsiz) programlarını bitirdi. Aynı zamanda İstanbul Üniversitesi Felsefe programlarında öğrenimine devam etmektedir. 2015 yılında kendi soyadını taşıdığı Bilgi Teknolojileri ve Danışmanlık şirketini kurdu. Bir çok kamu ve özel sektöre halen danışmanlık yapmaktadır.