

**T.C.
PAMUKKALE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**

**MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE COVID-19
HASTALARININ BELİRLENMESİ**

YÜKSEK LİSANS TEZİ

MİNE AYAZ

DENİZLİ, TEMMUZ - 2021

**T.C.
PAMUKKALE ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**



**MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE COVID-19
HASTALARININ BELİRLENMESİ**

YÜKSEK LİSANS TEZİ

MİNE AYAZ

DENİZLİ, TEMMUZ - 2021

Bu tezin tasarımı, hazırlanması, yürütülmesi, arařtırmalarının yapılması ve bulgularının analizlerinde bilimsel etięe ve akademik kurallara özenle riayet edildiđini; bu çalışmanın doğrudan birincil ürünü olmayan bulguların, verilerin ve materyallerin bilimsel etięe uygun olarak kaynak gösterildiđini ve alıntı yapılan çalışmalara atfedildiđine beyan ederim.

MİNE AYAZ

ÖZET

**MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE COVID-19
HASTALARININ BELİRLENMESİ
YÜKSEK LİSANS TEZİ
MİNE AYAZ
PAMUKKALE ÜNİVERSİTESİ FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI
(TEZ DANIŞMANI:DOÇ. DR. ÖZCAN MUTLU)**

DENİZLİ, TEMMUZ - 2021

SARS-CoV-2, Aralık 2019'da Çin'in Hubei eyaletinde ortaya çıkan ve Dünya Sağlık Örgütü tarafından Mart 2020'de pandemi olarak ilan edilen COVID-19 salgınına yol açan yeni bir koronavirüs türüdür. Ortaya çıktığı ilk günden itibaren dünya çapında hızla yayılmış ve dört milyondan fazla insanın ölümüne neden olmuştur. COVID-19 salgınının daha fazla yayılmasını kontrol altına almak ve bulaşmayı durdurmak için SARS-CoV-2 ile enfekte hastaların mümkün olan en erken zamanda hızlı ve doğru teşhisi kritik önem taşımaktadır.

Bu çalışmada, makine öğrenimi (ML) algoritmalarını kullanarak tam kan sayımı sonuçlarından SARS-CoV-2 pozitif hastaların teşhis edilmesi amaçlanmaktadır. Çalışmada kullanılan veriler, COVID-19 salgınının yayılımıyla mücadele etmek amacıyla Brezilya, São Paulo'daki İsrail Albert Einstein Hastanesi tarafından paylaşılmış ve halka açık bir veri bilimi topluluğu tarafından yayınlanmıştır. Veri seti normal servis ve genel olmak üzere iki gruba ayrılmış ve K-En Yakın Komşu (KNN), Naif Bayes (NB) ve Destek Vektör Makinesi (SVM) makine öğrenme algoritmaları kullanılarak incelenmiştir. SVM normal servis grubu için % 92 ve genel grup için % 93,5 doğrulukla en iyi performansı göstermiştir. Elde edilen sonuçlar makine öğrenme algoritmaları ile oluşturulacak modellerle tam kan sayım sonuçlarını kullanarak SARS-CoV-2 pozitif hastalarını hızlı ve ucuz bir şekilde erken aşamalarda teşhis için alternatif bir yöntem olarak kullanılabilceğini göstermektedir.

ANAHTAR KELİMELELER: Salgın Hastalık, Pandemi, SARS-CoV-2, COVID-19, Makine Öğrenmesi, Tam Kan Sayımı.

ABSTRACT

IDENTIFYING COVID-19 PATIENTS WITH MACHINE LEARNING ALGORITHMS

MSC THESIS

MİNE AYAZ

**PAMUKKALE UNIVERSITY INSTITUTE OF SCIENCE
INDUSTRIAL ENGINEERING**

(SUPERVISOR:ASSOC. PROF. ÖZCAN MUTLU)

DENİZLİ, JULY 2021

SARS-CoV-2 is a novel coronavirus that emerged in Hubei province of China in December 2019 and caused the COVID-19 outbreak, which was declared a pandemic by the World Health Organization in March 2020. It has spread rapidly worldwide from the day when it first occurred and has caused the death more than four million people. It is critical to rapid and accurate diagnosis of SARS-CoV-2 infected patients as early as possible to control further spread of COVID-19 outbreak and stop contamination.

In this study, it is aimed to diagnose SARS-CoV-2 positive patients from complete blood count results using machine learning (ML) algorithms. The data used in this study were shared by Israelita Albert Einstein Hospital in São Paulo, Brazil, and published by a public data science community to struggle the spread of the co epidemic. The data set are divided into two groups: normal service and general group, and each group was analyzed by using K-Nearest Neighbors (KNN), Naive Bayes (NB) and Support Vector Machine (SVM) machine learning algorithms. SVM performed best with an accuracy of 92% for the normal service group and 93.5% for the general group. The results shows that models created using machine learning algorithms will be used as an alternative method to diagnose SARS-CoV-2 positive patients at early stages by making a fast and inexpensive screening using the outputs of complete blood count.

KEYWORDS: Epidemic, Pandemic, SARS-CoV-2, COVID-19, Machine Learning, Full Blood Count.

İÇİNDEKİLER

Sayfa

ÖZET.....	i
ABSTRACT	ii
İÇİNDEKİLER	iii
ŞEKİL LİSTESİ.....	iv
TABLO LİSTESİ	v
SEMBOL VE KISALTMALAR LİSTESİ	vi
ÖNSÖZ.....	vii
1. GİRİŞ.....	1
2. HASTALIKLAR, SALGINLAR, PANDEMİLER VE SARS CoV-2 PANDEMİSİ.....	5
2.1 Salgın Hastalıklar	5
2.2 Pandemiler.....	10
2.2.1 AIDS/HIV Pandemisi	12
2.2.2 COVID-19 Pandemisi.....	12
3. MAKİNE ÖĞRENME ALGORİTMALARI	18
3.1 Denetimsiz Öğrenme	18
3.2 Denetimli Öğrenme	19
3.2.1 K-En Yakın Komşu Algoritması	20
3.2.2 Naif Bayes Algoritması	22
3.2.3 Destek Vektör Makinesi (SVM).....	23
3.2.3.1 Doğrusal Destek Vektör Makinesi.....	24
3.2.3.2 Doğrusal Olmayan Destek Vektör Makineleri.....	26
3.3 Performans Değerlendirme Yöntemleri	28
3.3.1 Performans Değerlendirme Kriterleri	28
3.3.1.1 Doğruluk (Accuracy)	29
3.3.1.2 Kesinlik (Precision).....	30
3.3.1.3 Duyarlılık (Recall)	30
3.3.1.4 f-ölçütü (f-score)	30
3.3.1.5 İşlem Karakteristiği Eğrisi	31
3.3.2 k-Kat Çapraz Doğrulama (k-Fold Cross-Validation)	32
3.3.3 Izgara Arama (Grid Search).....	33
3.3.4 Veri Setinin Dengelenmesi	34
4. UYGULAMA	35
4.1 Veri Seti.....	35
4.2 Model.....	38
4.3 RapidMiner.....	39
4.4 KNN algoritması uygulama sonuçları	40
4.5 NB algoritması uygulama sonuçları	44
4.6 SVM algoritması uygulama sonuçları.....	46
4.7 KNN, NB ve SVM algoritmaları uygulama sonuçlarının karşılaştırılması 48	
5. SONUÇ VE ÖNERİLER	50
6. KAYNAKLAR.....	51
7. ÖZGEÇMİŞ.....	57

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1: Dünya geneli AIDS bağlantılı ölüm sayısı	12
Şekil 2.2: Dünya geneli COVID-19 ölüm vaka sayıları yayılım haritası	15
Şekil 3.3: KNN algoritmasının nasıl çalıştığını gösteren çizim.	20
Şekil 3.4: Doğrusal SVM Modeli.	24
Şekil 3.5: Çekirdek fonksiyonu ile yüksek boyutlu uzayda sınıflandırma.....	26
Şekil 3.6: Hata matrisi.....	29
Şekil 3.7: Performanslarına göre ROC Eğrileri	31
Şekil 3.8: 10-kat çapraz doğrulama şeması.....	33
Şekil 4.9: Normal servis grubu veri seti SARS-CoV-2 Test Sonucu pozitif ve negatif hasta sayısı.	37
Şekil 4.10: Genel grup veri seti SARS-CoV-2 Test Sonucu pozitif ve negatif hasta sayısı.....	38
Şekil 4.11: Normal servis grubu veri seti, KNN algoritması doğruluk ve f-ölçütü grafiği.	41
Şekil 4.12: Genel grup veri seti, KNN algoritması doğruluk ve f-ölçütü grafiği.....	43
Şekil 4.13: Normal servis grubu veri seti için algoritmaların ROC eğrileri.	48
Şekil 4.14: Genel grup veri seti için algoritmaların ROC Eğrileri.....	49

TABLO LİSTESİ

Sayfa

Tablo 2.1: Dünya tarihinde yaşanan önemli salgınlar.....	10
Tablo 3.2: Temel kernel fonksiyonları ve parametreleri.....	27
Tablo 4.3: Tam kan sayımı parametreleri.	36
Tablo 4.4: SARS-CoV-2 test sonuçlarına göre veri grupları.	37
Tablo 4.5: Normal servis grubu veri seti, KNN algoritması performans ölçümleri.	40
Tablo 4.6: Normal servis grubu veri seti, KNN algoritması karşılaştırma matrisi.	41
Tablo 4.7: Genel grup veri seti, KNN algoritması performans ölçümleri.	42
Tablo 4.8: Genel grup veri seti, KNN algoritması karşılaştırma matrisi.	43
Tablo 4.9: Normal servis grubu veri seti, NB algoritması karşılaştırma matrisi.	44
Tablo 4.10: Normal servis grubu veri seti, NB algoritması performans ölçümleri.	44
Tablo 4.11: Genel grup veri seti, NB algoritması karşılaştırma matrisi.	45
Tablo 4.12: Genel grup veri seti, NB algoritması performans ölçümleri.....	45
Tablo 4.13: Normal servis grubu veri seti, SVM algoritması karşılaştırma matrisi.	46
Tablo 4.14: Normal servis grubu veri seti, SVM algoritması performans ölçümleri.....	46
Tablo 4.15: Genel grup veri seti, SVM algoritması karşılaştırma matrisi.	47
Tablo 4.16: Genel grup veri seti, SVM algoritması performans ölçümleri.	47
Tablo 4.17: Normal servis grubu veri seti, algoritma performans ölçümlerinin karşılaştırması.....	48
Tablo 4.18: Genel grup veri seti, algoritma performans ölçümlerinin karşılaştırması.....	49

SEMBOL VE KISALTMALAR LİSTESİ

SEMBOLLER

C	:	Ceza Parametresi
D	:	Eğitim Veri Seti
N	:	Eğitim Veri Seti İçerisindeki Örnek Sayısı
x_i	:	Girdi Değişkeni
y_i	:	Çıktı Değişkeni
w	:	Destek Vektör Makinesinde Hiper Düzlemin Normali
b	:	Destek Vektör Makinesinde Sabit (bias) Değer
γ	:	Kernel Boyutu
k	:	En Yakın Komşu Sayısı

KISALTMALAR

PHEIC	:	Public Health Emergency of International Concern
SARS	:	Severe Acute Respiratory Syndrome
MERS	:	Middle East Respiratory Syndrome
AIDS	:	Acquired Immunodeficiency Syndrome
HIV	:	Human Immunodeficiency Virus
RT-PCR	:	Reverse Transcription Polymerase Chain Reaction
RDT	:	Rapid Diagnostic Test
BT	:	Bilgisayarlı Tomografi
NAAT	:	Nükleik Asit Amplifikasyon Testi
WHO	:	World Health Organization
DSÖ	:	Dünya Sağlık Örgütü
UNIADS	:	United Nations Programme on HIV/AIDS
AI	:	Artificial Intelligence
ML	:	Machine Learning
KNN	:	K-Nearest Neighbors (K-En Yakın Komşu)
SVM	:	Support Vector Machine (Destek Vektör Makinesi)
NB	:	Naive Bayes (Naif Bayes)
ROC	:	Receiver Operating Characteristic
AUC	:	Area Under Curve
SMOTE	:	Synthetic Minority Oversampling Technique

ÖNSÖZ

Tez çalışmamın da konusunu oluşturan içinde bulunduğumuz COVID-19 pandemi sürecinde samimiyetiyle yardım ve desteğini esirgemeyen danışman hocam Sayın Doç. Dr. Özcan MUTLU'ya, yüksek lisans eğitimimde emeği geçen tüm Pamukkale Üniversitesi Endüstri Mühendisliği Anabilim Dalı hocalarıma teşekkürlerimi sunarım.

Son olarak sevgisini ve dualarını hiçbir zaman esirgemeyen annem Ayşe AYZ, babam Ali AYZ'a, hayatımın her anında yanımda olduklarını hissettiren abim Mustafa AYZ ve kardeşlerim Özgür AYZ, Tayfun AYZ'a minnetlerimi sunarım.

1. GİRİŞ

31 Aralık 2019 tarihinde, Çin'in Wuhan şehrinde nedeni belirlenemeyen 27 zatürre vakası bildirilmiş, bildirilen vakaların canlı hayvan ticareti yapılan Huanan Deniz Ürünleri Pazarı ile bağlantılı olduğu tespit edilmiştir (Sohrabi ve diğ. 2020). Yapılan incelemeler neticesinde enfeksiyon etkeninin yeni bir koronavirüs türü olduğu anlaşılmış ve daha önce görülen SARS CoV virüsü ile benzerliği nedeniyle SARS-CoV-2 olarak adlandırılmıştır. Son bulgulara göre virüsün doğal yollarla ortaya çıktığı, orijinal kaynağının SARS ve MERS salgınlarındaki gibi yarasalar olduğu, insanlara geçişinin ise yarasalar aracılığıyla pangolinlerden gerçekleştiği düşünülmekle birlikte, konu hakkında araştırmalar devam etmektedir (Shield 2020).

Hastalığın büyük bir hızla ve kısa bir süre içerisinde Antarktika hariç dünyanın her kıtasına yayılması nedeniyle, Dünya Sağlık Örgütü (DSÖ) tarafından 30 Ocak 2020'de mevcut salgının Uluslararası Öneme Sahip Halk Sağlığı Acil Durumu (Public Health Emergency of International Concern, PHEIC) oluşturduğu bildirilmiş, 11 Mart 2020'de pandemi olarak ilan edilmiştir. 2021 yılı Temmuz ayı itibarıyla küresel boyutta COVID-19 bağlantılı, 185.038.806 resmi vaka, 4.006.882 ölüm vakası Dünya Sağlık Örgütü tarafından bildirilmiştir (WHO 2021).

Tüm dünyayı etkisi altına alan salgına karşı, bilim adamları ve tıp endüstrisi çalışanları tarafından enfekte hastaları çeşitli aşamalarda taramak, en iyi klinik çalışmaları bulmak, virüsün yayılımını kontrol altına almak, enfekte hastaları iyileştirmek için bir aşı geliştirmek, enfekte olmuş kişilerle teması takip etmek amacıyla yeni teknoloji çözümleri halen aranmaktadır.

Yakın zamanda yapılan araştırmalar neticesinde, daha iyi ölçek büyütme, hızlandırma işlem gücü, güvenilirlik ve hatta belirli sağlık hizmeti görevlerinde insandan daha iyi performans gösterme gibi yönleriyle Makine Öğrenimi ve Yapay Zekânın, çeşitli sağlık hizmeti sağlayıcıları tarafından kullanılan umut verici teknolojiler olduğu tespit edilmiştir. Bu nedenle, dünya çapında COVID-19 salgınıyla mücadele kapsamında birçok makine öğrenimi ve yapay zekâ çalışması gerçekleştirilmiştir (Lalmuanawma ve diğ. 2020).

Ahmad ve Muhammad Asad (2020), günümüzde devam eden COVID-19 salgının yayılımını önleme ve kontrol altında tutmada büyük öneme sahip vaka, iyileşen ve hayatını kaybeden kişi sayılarına ait geçmiş verileri kullanarak, gelecek günler için bu parametreler üzerinden tahminde bulunmuşlardır. Yapay sinir ağı yöntemi kullanılarak gerçekleştirilen bu uygulama ile elde edilen sonuçlar değerlendirildiğinde gerçek hayattan elde edilecek daha fazla parametre kullanılarak yapılacak çalışmaların; toplumun güvenliğini sağlama, daha doğru politikalar belirleme ve daha gerçekçi kararlar alma gibi süreçlerde devletlere yardımcı olacağı düşünülmektedir.

Wang ve diğerleri (2020), dünyada büyük bir etkiye sahip olan ve ne zaman biteceği bilinmeyen COVID-19 salgın trendini tahmin etmeye çalışmışlardır. Bunun için güncel COVID-19 epidemiyolojik verileri kullanarak, makine öğrenimi tabanlı bir zaman serisi tahmin modeli oluşturulmuştur. Toplanan verilerden, doğrulanmış ve iyileşen vakalar ile ölüm vakaları olmak üzere üç zaman serisi oluşturulmuş, bu üç dizinin her birinin bir zirveye sahip olduğu, diğer bir deyişle salgının ileride sona ereceği varsayılmıştır. Trend eğilimi dünya çapında ve belirlenen birkaç ülke bazında değerlendirilmiş ve bu değerlendirmeler salgının pik noktası olacağı varsayımına dayandırılmıştır. Ülkelerin salgına karşı aldığı önlemlerin, uygulanan politikaların vb. yapılan tahminleri etkileyebileceği düşünülmekle birlikte, elde edilen veriler doğrultusunda önerilen trend tahmin modelinin etkili olduğu savunulmaktadır.

Zheng ve diğerleri (2020), özellikle sağlık sistemi alt yapısının yetersiz olduğu ülke veya bölgelerde, ciddi hastalık riski yüksek olan hastaların tanımlaması ve acil olarak hastanelere kabul edilmeleri için önceliklerin belirlenmesinin önemli olduğunu belirtmişler, hastaların COVID-19 nedeniyle hastaneye yatış önceliğini dört biyolojik belirteç üzerinden, makine öğrenme algoritmaları yardımıyla değerlendirmişlerdir. Bu çalışma ile salgınlarda ve pandemilerde klinik bulguların hastaların durumunun hızlı ve verimli bir şekilde tanımlanması amacıyla sağlık hizmetleri yöneticileri tarafından pratikte kolaylıkla kullanılabilirliğini göstermişlerdir.

Ahamad ve diğerleri (2020), salgın süreçlerinde büyük öneme sahip tedavi ve izolasyon ihtiyaçları konusunda hızlı kararlar almaya yardımcı olmak için denetimli makine öğrenimi algoritmalarından Karar Ağacı (Decision Trees), Rassal Orman (Random Forest), GBM (Gradient Boosting Machine) ve XGBoost (Extreme Gradient

Boosting) kullanarak COVID-19 hastalık tanılarını yüksek doğrulukla tahmin eden bir model geliştirmişlerdir. Model, temel hasta bilgileri (yaş ve cinsiyet), seyahat öyküsü ve klinik semptomlar (ateş, öksürük ve burun akıntısı ve zatürre) gibi ayrıntıları içermektedir. Seyahat geçmişinin enfeksiyonun görülmesi ile bağlantılı en önemli özelliklerden biri olduğuna, diğer bazı semptomların da COVID-19 salgını ile ilgili olduğunu ancak kas ağrısı ve önceki SARS salgınında belirgin olarak gözlenen ishal rahatsızlığı semptomlarının daha az görüldüğüne dikkatleri çekmişlerdir.

Öztürk ve diğerleri (2020), salgının daha fazla yayılmasını önlemek ve salgından etkilenen hastaları hızla tedavi etmek için pozitif vakaların olabildiğince erken tespiti amacıyla, göğüs X-ray görüntülerini kullanarak COVID-19 vakalarının tespiti için derin öğrenmeye dayalı otomatik yeni bir model önermişlerdir. Önerilen model ikili sınıflar için % 98.08 ve çok sınıflı durumlar için % 87.02 sınıflandırma doğruluk oranına sahiptir. Elde edilecek daha fazla görüntü ile oluşturulacak iyi eğitilmiş bir derin öğrenme modelinin, sağlık uzmanlarına tanı koyma sürecinde yardımcı olabilecek etkili bir yaklaşım olduğu öne sürülmüştür.

Hâlihazırda etkili bir tedavi yönteminin olmaması ve enfeksiyonun yayılımının gün geçtikçe artması nedeniyle bulaşmayı ortadan kaldırmak için alınabilecek en etkili önlem enfekte bireylerin izolasyonunu sağlamaktır. Enfekte bireylerin hızlı ve doğru teşhisi, bulaşmayı önlemek için izolasyon kararının alınmasında oldukça önemlidir (Ahamad ve diğ. 2020). Enfeksiyonun etiyolojik teşhisinde solunum yolu örnekleri üzerinden çalışılan RT-PCR testi kullanılmaktadır. SARS-CoV-2 enfeksiyonunun tanısında altın standart olarak kabul edilen RT-PCR testi, numunenin alınmasından hastanın bilgilendirilmesine kadar belli bir zaman gerektirmekte ve bazı tesisler için kullanımı maliyetli olabilmektedir. Ayrıca göğüs BT tarama sonuçları ile karşılaştırıldığında, kullanılan standart tanı testlerinin doğruluğu % 80 oranındadır ve BT göğüs taramalarındaki radyografik modeller, RT-PCR testi ile karşılaştırıldığında daha yüksek hassasiyet ve duyarlılık göstermektedir. Ancak DSÖ'ye göre göğüs taramaları erken aşamalarda nispeten düşük bir pozitif tespit oranına sahiptir ve BT'nin maliyetli bir süreç olmasının yanında genellikle büyük çapta hizmet veren sağlık merkezlerinde buldukları için kolayca erişilebilir değildirler (Lippi ve diğ. 2020).

Sarmış (2020), sađlık kuruluşlarının salgın durumlarında karşılaştığı en önemli sorunlarından biri olan hastalığın hızlı tanısı üzerine yapılan çalışmaları incelemiştir. SARS-CoV-2 enfeksiyonunun teşhisi için genel olarak burun ve boğazın arkasından sürüntü toplama şeklinde elde edilen örnekler üzerinden çalışılan RT-PCR testlerinin, enfeksiyonun doğru teşhisi amacıyla eş zamanlı olarak antikor testleri ile desteklenmesini önerirken, rutin olarak uygulanan tam kan sayımı veya bazı basit biyokimya testlerin hastalığın teşhisi ve seyri hakkında hızlı bir şekilde kararların alınmasında bilgi verici olacağını savunmaktadır.

Yapılan çalışmalar değerlendirilerek bu çalışma kapsamında COVID-19 salgını ile mücadeleye katkı sağlamak amacıyla tam kan sayımı sonuçları üzerinden SARS-CoV-2 pozitif hastalar tahmin edilmeye çalışılmış, çalışma sırasında K-En Yakın Komşu, Naif Bayes ve Destek Vektör Makinesi makine öğrenimi algoritmalarından yararlanılmıştır.

2. HASTALIKLAR, SALGINLAR, PANDEMİLER VE SARS CoV-2 PANDEMİSİ

2.1 Salgın Hastalıklar

Hastalıklar hayatla aynı zamanda, insanın ortaya çıkışından öncede var olmuştur (Bayat 2016). İnsan ve hayvan kalıntılarında hastalık bulgularını inceleyen, hastalıkların evrimsel sürecini irdeleyen paleopatoloji bilim dalı (Pelin ve Polat 1997), milyonlarca yıllık kalıntıların patolojisini analiz imkânı sunmuştur. Yeryüzünün ilk devirlerinde görülen hastalıkların araştırılması, bulunan hayvan kalıntılarının incelenmesiyle mümkün olmuştur. Günümüz hastalıklarına ait bazı izler de milyonlarca yıl öncesine ait fosillerden elde edilmiştir. Henüz paleopatolojik araştırması yapılmamış çok sayıda buluntu mevcut olduğu gibi, bunlara devamlı olarak yeni materyaller de eklenmektedir.

Hayvanlar, milyonlarca yıldır tabiatın zor şartlarına karşı biyolojik bir güç olan içgüdülerini kullanarak hayatta kalmış, nesillerinin devamlılığını sağlamış ve hastalıklarına çareler üretmişlerdir. Hayvanların biyolojik bir güç olarak sahip oldukları içgüdü, insanda yerini zekâyâ bırakmaktadır. İnsanlar, doğada yaşayan hayvanların faydalı ve zararlı bitkileri ayırt edebildiğini fark etmiş, hasta olan hayvanların hangi türden bitkileri yediğini gözlemlemiş ve bu bitkileri kendi hastalıklarının tedavisinde kullanmışlardır.

Medeniyetlerin ortaya çıkışına da öncülük eden yerleşik hayata geçiş ile birlikte bazı hayvan türleri ehlileştirilmiş ve insanlarla iç içe yaşamaya başlamıştır. Bu durum daha önce hayvanlarda görülen birçok hastalık türünün insanlarda da görülmesine neden olmuştur. Bunun yanında tarımsal üretimle birlikte ekilebilir alanlar oluşturulmuştur. Sulama ihtiyacı için açılan kanallar ve göletler sivrisineklerin artışına sebep olmuş, sonuç olarak sıtma (malarya) hastalığı günümüzde de görülen yaygın hastalıklardan biri haline gelmiştir. İnsanların sürekli olarak barınabileceği alanlar oluşturması, topluluk halinde yaşamaya başlaması bulaşıcı hastalıkları ve bunların yayılımını da arttırmıştır (Bayat 2016).

Salgın hastalıklar ve pandemi süreci; sosyal, politik ve ekonomik bir bağlamda virüslerin ortaya çıkışı, yayılışı ve canlı organizmalar üzerindeki etkileri gibi gelişmelerle birlikte ortaya çıkmaktadır (Choonora 2020).

İnsanlık tarihi süresince değişen üretim yöntemleri insan ile doğa arasındaki ilişkinin temel belirleyicilerinden olmuştur. İnsanın varoluşundan itibaren üretim faaliyetleri devam etmekle birlikte insan ile doğa arasındaki ilişkinin kırılma noktası, insanın hayatını devam ettirmesine gerekenden daha fazlası için üretim faaliyetlerine yönelmesiyle gerçekleşmiştir (Kılıçbay 1989).

Tarihi avcı-toplayıcı döneme kadar dayanan salgın hastalıkların ortaya çıkış ve yayılma sürecinde yürütülen ekonomik faaliyetlerin ve bu faaliyetler için kullanılan aletlerin oldukça büyük etkisi olmuştur (Beyaz 2021). Başlarda saban kullanımı ve sulama kanallarının oluşturulması ile başlayan tarım faaliyetleri birçok sorunu da beraberinde getirmiştir. Özellikle verimli toprakların tarıma elverişli arazi kazandırmak amacıyla sürülmesi bakterilerin çoğalmaları ve hızlı bir şekilde yeryüzüne yayılmalarına olanak sağlamış, sonuç olarak daha önce hiç bilinmeyen çeşitli hastalıklar gün yüzüne çıkmıştır (Ponting 2000). Tarımsal üretimin yanında insanların hayvanlar ile yakın temas içerisinde olması ve avladıkları hayvanların etini çiğ tüketmesi parazitlerin; topluluklar halinde yaşamaları ise salgın hastalıkların hızlı bir şekilde yayılmasına olanak sağlamıştır (Sarı 2007).

Ne var ki kapitalizmin ortaya çıkışıyla birlikte doğa-toplum-ekonomi arasındaki ilişki alt üst olmuştur. Günümüzde de devam eden bu anlayışın iki temel yanığı vardır. Bu yanıklardan ilki ekonomik büyümenin doğal bir sınırının olmadığı; ikinci yanığı ise doğanın ve doğal varlıkların insanlığa katkısının bir bedelinin olmadığıdır. Bu yanığın bir sonucu olarak tüm canlı türlerini etkileyen iklim değişimi, ekolojik dengenin kırılması, biyolojik çeşitlilikteki azalma, canlı türlerinin giderek yok olması, salgın hastalıkların hızla yayılması gibi birçok küresel sorun ortaya çıkmakta ve çıkmaya da devam etmektedir. (Beyaz 2021).

Birçok salgın hastalık insanlık tarihi boyunca dünyanın her yerinde devletlerin ve milletlerin hayatlarını derinden etkilemiş ve tarihin çeşitli dönemlerinde meydana gelen bu salgınlar milyonlarca insanın ölümüne neden olmuştur (Yıldız 2014).

Orta çağda nedeni anlaşıl原因amayan ve bu nedenle günah işleme, tanrılara karşı görevlerini ihmal etme gibi davranışların ilahi bir cezası olduğuna inanılan genelde plague (bela, felaket) olarak adlandırılan bulaşıcı hastalıklar, dünya nüfusunu büyük ölçüde azaltırken, dünya tarihinin şekillenmesinde de önemli role sahip olmuştur.

M.S. II. yüzyılda büyük olasılıkla kızamık ve çiçek salgını olan Antonine Plague, Roma İmparatorluğu'nda büyük nüfus kayıplarına ve ekonomik sıkıntılara neden olmuştur. İmparatorluk yapısının bozulmasına neden olan bulaşıcı hastalıklar, Roma İmparatorluğu'nun yıkılışında da etkili olmuştur (Akpınar 2012).

Tüm dünyayı etkileyen, milyonlarca insanın yok olmasına ve göçlere neden olan, çalışma yaşamını alt üst eden, tarım ve hayvancılığı geriletken, kıtlığa yol açan, devletleri ekonomik boyutta zayıflatan salgın hastalıklardan en önemlilerinden biri de, Batıda “black death”, İslam âleminde “taûn”, Türklere göre de “kıran/ölet” olarak adlandırılan vebadır.

Doğu ile Batı medeniyetleri arasında köprü görevi üstlenen yüzyıllar boyu süren salgınlardan kurtulamayan Anadolu da görülen ilk büyük salgın olan Justinianus vebası da, VI. yüzyıl ortalarında Akdeniz'e kıyısı olan ülkeleri etkilemiş ve nüfusun önemli bir kısmını da yok etmiştir. Verilen kayıplar hakkında kesin bir bilgi bulunmamakla birlikte, sadece İstanbul'da günlük 16 bin kişinin hayatını kaybetmesi salgının büyüklüğü hakkında fikir vermektedir (Bayat 2016).

1853 yılı Şubat ayında, San Francisco'dan gelen bir tüccar gemisi, gemide hastalık olduğunu belirten sarı bayrağı ile Honolulu limanında görülmüştür. Çiçek hastalığı olduğu tespit edilen bu salgın nedeniyle aynı yıl Mayıs ayı içerisinde daha fazla vaka görülmeye başlanmış, Ekim 1853'te salgın zirve noktasına ulaşmış ve 1854 yılı Ocak ayının ortasına kadar da seyrini sürdürmüştür (Byrne 2008). XVIII. yüzyılda bulaşıcı çiçek hastalığı nedeniyle Avrupa'da 60 milyon fazla kişi hayatını kaybetmiş, milyonlarcası da sakat kalmıştır. Osmanlı devletinde çiçek salgınına karşı üretilen deneysel yöntem İngiltere İstanbul elçisinin eşi Lady Montagu'nün vasıtasıyla İngiltere'de uygulanmış ve çiçek aşısı bu uygulama sayesinde bulunmuştur (Bayat 2016). Türkiye'de en son çiçek hastalığı 1957 yılında görülmüş ve çiçek aşısına 1980 yılında son verilmiştir. Çiçek hastalığının kökünün 1979 yılında dünyadan kazındığı

iddia edilmekle birlikte çiçek virüsünün biyolojik bir silah olarak kullanılması ihtimalini göz ardı etmek mümkün görünmemektedir (Akpınar 2012).

Cristoph Colomb'un Amerika'yı keşfinden önce yaklaşık 100 milyon yerli insanın (İnka, Maya ve Aztek uygarlıklarında) bu kıtada yaşadığına inanılmaktaydı. Bulaşıcı hastalıklara karşı bağışıklığı olmayan, yüzlerce yıl varlığını sürdürmüş yerliler, Avrupa'dan gelen kızamık, grip, difteri, tifo, dizanteri ve çiçek hastalığı gibi birçok bulaşıcı hastalık nedeniyle zaman içerisinde yok olmuştur. Yaşanan bu olay tarihe yön vermesi açısından toplumların birbirine üstünlük sağlamasında salgınların önemini vurgulamaktadır (Demirel 2019).

Göçebe su kuşları ve kanatlı hayvanların virüsü olarak da bilinen influenza, hayvanlarla iç içe yaşanmaya başlanması ve hayvanların evcilleştirilerek günlük hayata dâhil edilmesiyle birlikte insana geçiş fırsatı bulmuştur. Yapılan çalışmalar influenza virüsünün 1510, 1557 ve 1580 yıllarında Avrupa'da salgınlara neden olduğunu, 1580 yılındaki salgının ise Afrika'ya kadar yayılarak, bilinen ilk kıtasal atlama yapan salgın olduğunu göstermektedir (Kara 2010).

Tarihteki bilinen en büyük salgın, bir yılda yaklaşık 50 milyon veya daha fazla kişiyi öldüren 1918-1919 grip salgınıdır (Byrne 2008). Salgının başlangıcı İspanya'da olmamasına rağmen İspanyol gribi olarak bilinmesinin nedeni, İspanya'nın, I. Dünya Savaşı'na katılmamış olması ve uygulanan sansür sebebiyle diğer Avrupa devletlerinin salgından söz etmemesi, ilk defa İspanyol basınında salgının gündeme getirilmesine bağlanmaktadır (Aslan 2020). Ortaya çıkışının I. Dünya Savaşı'nın (1914-1918) sonlarına denk gelmesi ve yüksek bulaşma kabiliyetine sahip olması nedeniyle hızla yayılmıştır. H1N1 tipi virüsün ölümcül bir alt türü olan bu bulaşıcı hastalık akut solunum sıkıntısı sebebiyle hastalığa yakalanan kişilerin büyük bir oranda hayatını kaybetmesine neden olmuştur.

1997 yılında influenza H5N1 virüs türü kuşlar arasında hızla yayılmaya başlamış, insanlarda da vakalar görülmüştür. "Avian İnfluenza" ya da "Kuş Gribi" olarak da bilinen bu salgın henüz kişiden kişiye bulaşma kabiliyetine sahip değildir (Byrne 2008). Hücrelere yeterli oksijen gitmediği için solunum sıkıntısı yaşanmasına, vücudun zamanla morarmasına da neden olmaktadır, bu yüzden "Mavi Ölüm" olarak da bilinmektedir (Aslan 2020). 2007 yılı Nisan ayı itibarıyla, 291 kişide H5N1 vakası

tespit edilmiş olup, bu vakaların 172'si ölümlle sonuçlanmıştır. Salgınla birlikte milyonlarca kuş ya virüs nedeniyle ölmüş ya da salgınları sona erdirmek amacıyla insanlar tarafından itlaf edilmiştir (Byrne 2008).

Ülkemiz gibi ılıman iklim tipine sahip ülkelerde özellikle Aralık-Mart aylarında grip vakaları daha yoğun olarak görülmektedir. 2009 yılına gelindiğinde Amerika'nın Yeni Meksika eyaletinde sezonun son günlerinde olunmasına rağmen grip vakalarında anlamlandırılmayan bir artış gözlemlenmiştir. İki çocuk hastada influenza benzeri klinik bulguların bildirilmesi vakaların nedeninin anlaşılmasını sağlamıştır. Dünya'nın nüfus yoğunluğu en fazla olan şehirlerinden Meksiko City'de ortaya çıkan ve "Domuz Gribi" olarak adlandırılan bu yeni tip İnfluenza A, H1N1 tipi virüs çok kısa sürede tüm dünyayı etkisi altına almıştır. Yakın ve korumasız şekilde kümes hayvanları ve domuzlarla temasta bulunulması halinde görülen bu yeni tür özellikle çocukları ve yetişkinleri etkilemiştir (Kara 2010).

Tablo 2.1'de (LePan 2020) görüleceği üzere zaman içerisinde bazıları kısa bazıları da uzun süren fakat her seferinde çok sayıda can kaybına neden olan salgın hastalıklar zuhur etmiş, insanoğlu da bunlara karşı önlemler almaya ve bunlardan kurtulmaya çalışmıştır (Turan ve Çelikyay Hamza 2020).

Tablo 2.1: Dünya tarihinde yaşanan önemli salgınlar.

Salgın	Zaman Dilimi	Tip/Çıkış Kaynağı	Tahmini Ölüm Sayısı
Antonine Salgını	165-180	Çiçek veya Kızamık (öngörülen)	5 milyon
Justinian Vebası (1. Veba Salgını)	541-542	Yersinia Pestis Bakterisi/ Fare, Pire	30-50 milyon
Japonya Çiçek Salgını	735-737	Variola Majör Virüsü	1 milyon
Kara Veba	1347-1351	Yersinia Pestis Bakterisi / Fare, Pire	200 milyon
Yeni Dünya Çiçek Salgını	1520 ve sonrası	Variola Majör Virüsü	56 milyon
İtalyan Vebası	1629-1631	Yersinia Pestis Bakterisi / Fare, Pire	1 milyon
Londra Büyük Vebası	1665	Yersinia Pestis Bakterisi	100 bin
Kolera Pandemileri (1-6)	1817-1923	V. Cholera Bakterisi	1 milyondan fazla
Üçüncü Veba Salgını	1885	Yersinia Pestis Bakterisi / Fare, Pire	12 milyon
Sarı Humma	1880'lerin sonu	Virüs / Sivrisinekler	100-150 bin
Rus Gribi	1889-1890	H2N2 Virüs	1 milyon
İspanyol Gribi	1918-1919	H1N1 Virüs / Domuz	40-50 milyon
Asya Gribi	1957-1958	H2N2 Virüs	1.1 milyon
Hong Kong Gribi	1968-1970	H3N2 Virüs	1 milyon
HIV/AIDS	1981-günümüz	Virüs / Şempanze	25-35 milyon
SARS	2002-2003	Koronavirüs / Yarasa, Vahşi Kedi	770 bin
Domuz Gribi	2009-2010	H1N1 virüs / Domuz	200 bin
Ebola	2014-2016	Ebolavirüs / Vahşi Hayvanlar	11 bin
MERS	2015-günümüz	Koronavirüs / Yarasa, Deve	850 bin
COVID-19	2019-günümüz	Koronavirüs / (muhtemelen) Pangolin	4 milyondan fazla (WHO 2021)

2.2 Pandemiler

Bir ülke veya bölgede bulaşıcı hastalığa maruz kalmış insanların her yıl belli bir sayıda olması “endemi”, vaka sayıları yıllarca birbirine yakın seyrederken sayılarda ani artışların görülmesi “epidemi, salgın” olarak adlandırılır. Bu salgın durumunun bulunduğu ülke veya bölgenin sınırlarını aşarak başka ülke, kıta hatta dünyayı etkiler hale gelmesi ise “pandemi” durumu olarak tanımlanmaktadır (Öztek 2020).

Salgın durumu bilinenden farklı yeni bir virüsle oluşuyor, kişiden kişiye ve kolay bir şekilde bulaşabiliyorsa küresel bir tehdit olarak Dünya Sağlık Örgütü (DSÖ) gündeminde ele alınmaktadır. Karşılaşılan salgının pandemi olup olmadığına DSÖ karar vermektedir.

Bir hastalık veya tıbbi durumun yaygın olarak görülmesi ve çok sayıda insanın ölümüne neden olması pandemi olarak nitelendirilmesi için tek başına yeterli değildir. Aynı anda bulaşıcı özellik de gösteriyor olması gereklidir. Örneğin yaygın olarak görülen kanser hastalığı, çok sayıda ölüme sebep olmasına rağmen bulaşıcı bir yapıya sahip olmaması nedeniyle pandemi olarak kabul edilmemektedir.

Dünya Sağlık Örgütüne göre pandemi ilan edilebilmesi için salgının aşağıdaki üç ana unsuru taşıyor olması gerekmektedir:

- Daha önce görülmemiş olması
- İnsanlara bulaşıyor ve tehlikeli bir hastalık yaratıyor olması
- İnsanlar arasında kolayca yayılması ve bu yayılımın devamlılık göstermesi (Aslan 2020).

Tarihte milyonlarca insanın hayatını kaybetmesine neden olan çok sayıda pandemi ve epidemi görülmüştür.

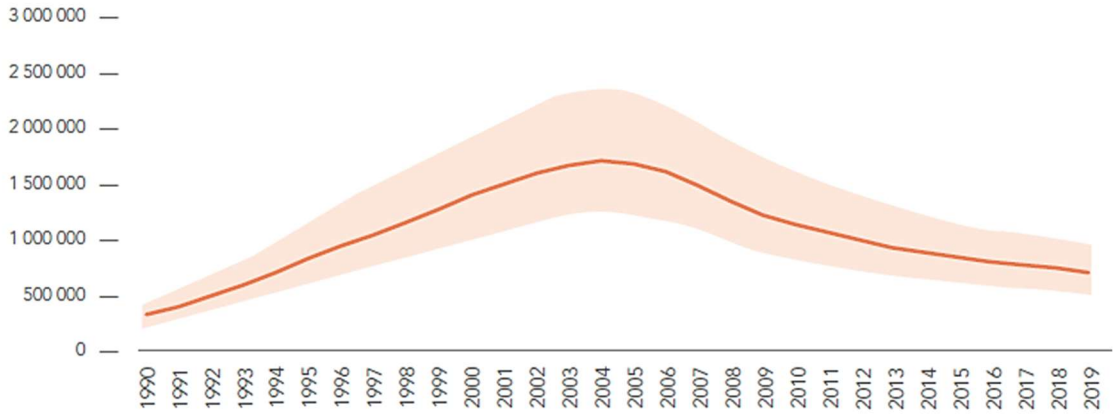
- Kara veba
- Kolera
- Grip (İnfluenza)
- İspanyol gribi
- Hong Kong gribi
- Domuz gribi
- Tifo

Günümüzde etkileri hala devam eden pandemiler de mevcuttur.

- AIDS/ HIV
- COVID-19 (SARS-CoV-2) (Öztek 2020).

2.2.1 AIDS/HIV Pandemisi

AIDS (Acquired Immuno Deficiency Syndrome) için etiyolojik bir ajan olan HIV (Human Immunodeficiency Virus) virüsü ilk olarak 1981 yılında teşhis edilmiş, zamanla tüm dünyayı kasıp kavuran, önemli bulaşıcı hastalıklardan biri haline gelmiştir. İnsanda bağışıklık sistemi yetmezliğine neden olan, daha çok cinsel yolla bulaşan bir hastalıktır. HIV virüsü bağışıklık sistemine nüfuz ederek kişinin enfeksiyonlara karşı direncini yok eder ve kişiyi savunmasız hale getirerek ölümüne sebep olur (Aslan 2020). UNAIDS 2020 güncellenmiş raporuna göre HIV ile yaşayan 38 milyon kişi bulunurken, bunlardan 25,4 milyonu şu anda tedavi görmekte iken 12,6 milyon kişi hala tedavi beklemektedir. 2019 yılında küresel anlamda tahmini 1.7 milyon kişide yeni enfeksiyon tanımlanmışken, 690.000 kişi de AIDS'e bağlı hayatını kaybetmiştir. Şekil 2.1'de Dünya genelinde 1990-2019 yılları arasında AIDS virüsü bağlantılı can kayıplarının sayısı görülmektedir. UNAIDS'in amacı 2020 yılında AIDS bağlantılı ölüm sayısını 500.000'in altına düşürmektir (UNAIDS 2020).



Şekil 2.1: Dünya geneli AIDS bağlantılı ölüm sayısı

2.2.2 COVID-19 Pandemisi

Koronavirüsler zarflı yüzeylelerinde glikoproteinden oluşan çubuksu uzantıları bulunan, tek zincirli yapıda, pozitif polariteli, zoonotik RNA virüslerden meydana gelen geniş bir ailedir (Laçinel Gürlevik 2020). Zarflı yüzey çıkıntılarının taç benzeri görünümü nedeniyle Latince “Corona”, yani “Taç” anlamı göz önünde bulundurularak Uluslararası Virüs Taksonomisi Komitesi (International Committee on Taxonomy of

Viruses, ICTV) tarafından Corona Virus (Taçlı virüs) olarak adlandırılmıştır (Huizen 2020).

Koronavirüsü birçok hayvan türünde (fare, kuş, tavuk, hindi, beyaz balina, tavşan, kedi, köpek, domuz) ciddi solunum yolu, sindirim sistemi, kalp-damar rahatsızlıklarına neden olabilmektedir. İnsanlarda görülmesi halinde basit düzeyde soğuk algınlığından bronşite, ağır akut solunum sendromuna (Severe Acute Respiratory Syndrome, SARS), çoklu organ yetmezliği gibi ciddi hastalıklara yol açabilmekle birlikte can kayıplarına da neden olabilmektedir.

İnsan koronavirüsleri çoğunlukla solunum yolu rahatsızlıklarına yol açan patojenlerdir.

Bilinen 7 farklı insan koronavirüsü bulunmaktadır.

1. HCoV-229E
2. HCoV-OC43
3. SARS CoV
4. HCoV-NL63 (Haven koronavirüsü)
5. HKU1
6. MERS-CoV
7. SARS CoV-2 (COVID-19)

İlk defa 2002 yılında Çin'de karşılaşılan yeni Ağır Akut Solunum Sendromu Koronavirüsünün (Severe Acute Respiratory Syndrome Coronavirus, SARS-CoV) mink kedileri ve yarasalardan insanlara bulaştığı düşünülmektedir. Dünya Sağlık Örgütü tarafından hızlı ve kapsamlı bir plan dâhilinde çalışmalar yürütülmüş 2003 yılı Temmuz ayına kadar tüm dünyada varlığı sona ermiştir. 29 ülkede görülen SARS-CoV virüsü 8.096 vakaya ve 774 kişinin ölümüne neden olmuştur.

Diğer bir koronavirüs türü olan Orta Doğu Solunum Sendromu Koronavirüsü (Middle East Respiratory Syndrome Coronavirus, MERS-CoV) Suudi Arabistan'ın Cidde şehrinde bulunan bir hastaneye ağır zatürre ve böbrek yetmezliği şikâyetleriyle başvuran bir kişinin balgamından alınan örneğin incelenmesiyle tespit edilmiştir. Bu yeni tür virüsün de tek hörgüçlü develerden insanlara bulaştığı bilinmektedir.

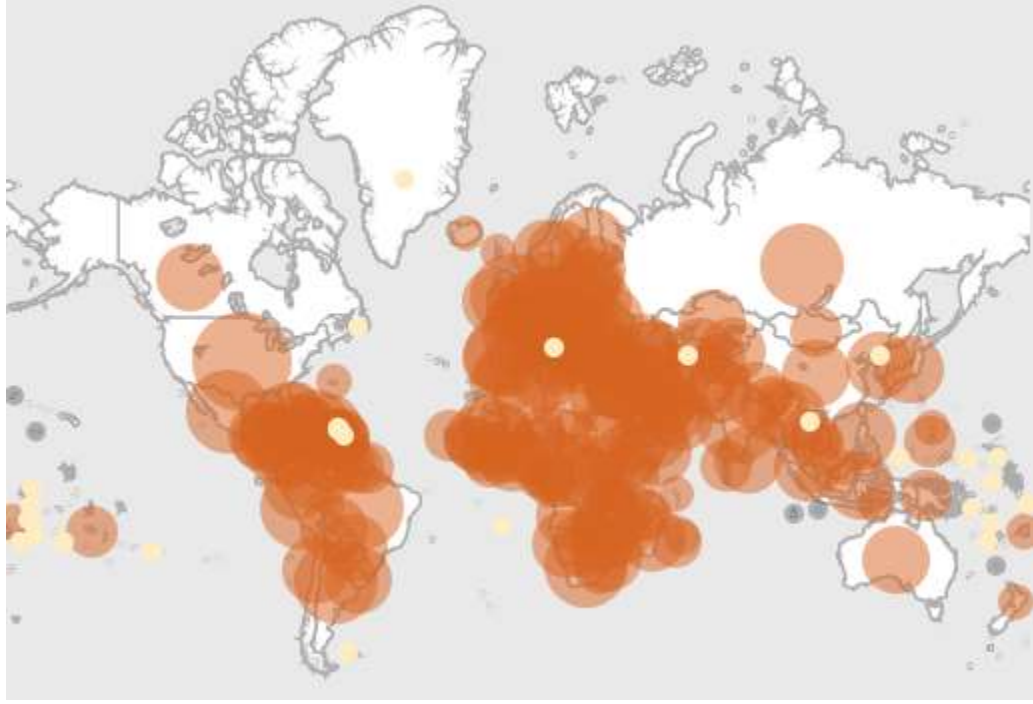
31 Aralık 2019 tarihinde Çin'in Hubei eyaleti, Wuhan şehrinde, nedeni belirlenemeyen zatürre vakaları bildirilmiştir. Şehirdeki toptan balık ve canlı hayvan satışı yapılan pazar çalışanlarında vakaların kümелendiği belirtilmiştir. Virüs ilk ortaya çıktığı zaman insanlarda daha önce tanımlanmamış yeni bir koronavirüs anlamında “novel” Coronavirus, 2019-nCoV olarak adlandırılmıştır (Laçinel Gürlevik 2020). Sonraları 2003 yılında yaşanan SARS salgını ile benzerliğinden dolayı bu yeni koronavirüsün adı ICTV tarafından “SARS-CoV-2” olarak ilan edilmiş ve sebep olduğu hastalık DSÖ tarafından “COVID-19” olarak açıklanmıştır (WHO 2020).

COVID-19 salgını 30 Ocak'ta Dünya Sağlık Örgütü tarafından, “uluslararası boyutta halk sağlığı acil durumu” olarak belirlenmiştir. Salgının ilk görüldüğü yer olarak bilinen Çin dışında 113 ülkede de karşılaşılmaya üzerine, virüsün yayılımı ve etkileri sebebiyle 11 Mart 2020 tarihinde küresel salgın (pandemi) ilan edilmiştir.

Enfeksiyon zinciri; bulaş kaynağı, bulaşma yolu ve enfeksiyona duyarlı kişilerden oluşmaktadır. Hayvan kökenli olduğu düşünülmekle birlikte, SARS-CoV-2 virüsünün kaynağı hala netlik kazanmamış ve araştırılmaya devam edilmektedir. Virüs insandan insana damlacık, temas ve kimi durumlarda aerosol yoluyla bulaşmakta olup; semptomatik olduğu gibi asemptomatik şekilde de seyrebilmektedir.

Toplumun tümü salgına karşı duyarlıdır. Salgınla karşılaşma ihtimali yönünden en riskli meslek grubu sağlık çalışanlarıdır (T.C. Sağlık Bakanlığı 2020). Herkes COVID-19 ile yaş fark etmeksizin ciddi bir şekilde hastalanabilir veya hayatını kaybedebilir fakat 60 yaş ve üzeri kişiler ile yüksek tansiyon, kalp ve akciğer rahatsızlıkları, diyabet, obezite veya kanser gibi altta yatan sağlık sorunları olan kişiler COVID-19 açısından hassas gruptadırlar (WHO 2020).

Temmuz ayı itibariyle dünyada 185.038.806 SARS-CoV-2 vakası rapor edilirken, 4.006.882 ölüm bildirilmiştir. Virüsün Dünya üzerindeki yayılımı da Şekil 2.2'de görülmektedir (WHO 2021).



Şekil 2.2: Dünya geneli COVID-19 ölüm vaka sayıları yayılım haritası

Dünya Sağlık Örgütü'nün Çin Halk Cumhuriyeti ile ilgili düzenlediği SARS-CoV-2 raporuna göre hayatını kaybeden kişiler çoğunlukla ileri yaşlarda ya da hipertansiyon, diyabet, kanser gibi kronik rahatsızlıklara sahiptir. Enfeksiyonun yaygın belirtileri ateş, öksürük, solunum yetersizliği ve nefes darlığıdır. Hastalığın daha ileri safhalarında zatürre, ağır akut solunum sıkıntısı, böbrek yetmezliği gibi ciddi rahatsızlıklar veya hayatını kaybetme durumları görülebilmektedir (T.C. Sağlık Bakanlığı 2020).

Salgının yayılımını durdurmanın en etkili yolu, enfekte bireyleri olabildiğince erken tespit etmektir. SARS-CoV-2'nin varlığını tespit etmek ve doğrulamak amacıyla moleküler bir test türü olan Ters Transkripsiyon Polimeraz Zincir Reaksiyonu (Reverse Transcription Polymerase Chain Reaction, RT-PCR) kullanılmaktadır. Bir NAAT (Nükleik Asit Amplifikasyon Testi) türü olan ve virüse ait genetik materyalin tespit edilebilir seviyelere yükseltildiği bu yöntem sayesinde, bir pamuklu çubuk yardımıyla burun veya boğazdan alınan örnekler üzerinden enfeksiyonun varlığı tespit edilmektedir. Diğer bir yöntem olan hızlı antijen testleri (Rapid Diagnostic Test, RDT), burun veya boğazdan alınan numuneler sayesinde viral proteinleri (antijenleri) tespit etmektedir. Bu testler, PCR'e göre daha ucuz ve sonuçları daha hızlı sunmakla birlikte çoğunlukla doğruluk oranları düşüktür.

Kişiler COVID-19'u semptomatik veya asemptomatik olarak geçirebilmektedir. Antikor testleri, kişi enfeksiyona bağlı semptom göstermese bile kişinin öncesinde enfeksiyon geçirip geçirmediğini gösterebilmektedir. Serolojik testler olarak da bilinen ve sıklıkla kan numunesi üzerinde çalışılan bu testler, enfeksiyona karşılık olarak üretilen antikorları tespit etmektedir. Birçok insanda antikorlar, günler veya haftalar sonra gelişmeye başlamaktadır bu sebeple serolojik testler enfeksiyon veya hastalığın erken tanı veya teşhisi amacıyla kullanılmamaktadır (WHO 2020).

Ancak antikor testleri farklı amaçlarla kullanılabilir:

1. NAAT test sonucunun negatif olduğu fakat epidemiyolojik olarak COVID-19 ile ilintisi bulunan olgularda serolojik testlerin uygulanması tanıyı destekleyebilir.
2. PCR testlerinin hassasiyetini saptamak amacıyla yararlanılabilir.
3. Hastalığı asemptomatik olarak seyreden kişilerin belirlenmesi mümkün olabilir.
4. Serolojik testler devam eden salgının detaylı incelenmesine imkân sağlar, salgının atak hızı ve şiddetinin geçmişte karşılaşılan benzer olay ya da gelişmelerle birlikte değerlendirilmesini sağlar.
5. Belirli bir popülasyon içerisindeki seroprevalans oranı belirlenebilir (Stowell ve Guarner 2020).

Enfeksiyonun erken teşhisinin yanı sıra doğru teşhisi de oldukça önemlidir. Teşhis hatalarının oluşturacağı klinik ve ekonomik sonuçlar, söz konusu salgınlar olduğunda, özellikle COVID-19 gibi pandemi düzeyinde ise, tartışmasız büyük çapta yankı oluşturacaktır. Yanlış-pozitif veya yanlış-negatif test sonuçlarının oluşturulması sadece hastanın bireysel sağlığını tehlikeye atmakla kalmayacak, aynı zamanda salgını kontrol altına almak için ulusal ve uluslararası otoriteler tarafından oluşturulan halk sağlığı politikalarının, acil durum planlarının ve kısıtlayıcı önlemlerin etkililiğini de azaltabilecektir. Son kanıtlara göre, SARS-CoV-2 enfeksiyonunun tespiti için kullanılan RT-PCR testlerinin birçoğunun doğruluğu optimumdan (<% 100) düşüktür (Lippi ve diğ. 2020). Bu nedenle, tanısal hata riskini en aza indirmek, özellikle şüpheli hastalarda COVID-19 teşhisi için, RT-PCR test sonuçları klinik ve epidemiyolojik kanıtlar (maruz kalma olasılığı, belirtiler, semptomlar, özellikle diğer solunum hastalıkları),

göğüs BT (Computed Tomography, CT) ve laboratuvar bulguları gibi çalışmalarla desteklenmelidir (Ahamad ve diğ. 2020).

Bu çalışma kapsamında da, COVID-19 salgınıyla mücadele sürecinde enfeksiyonun erken aşamalarda teşhisi amacıyla klinik uygulamalarda evrensel olarak sıklıkla kullanılan ve çıktıları hızlı tetkik edilebilen tam kan sayımı sonuçları kullanılarak SARS-CoV-2 pozitif hastalar tahmin edilmeye çalışılmıştır.

3. MAKİNE ÖĞRENME ALGORİTMALARI

Teknolojik alanda yaşanan gelişmeler, büyük miktarda veri artışını da beraberinde getirmiştir. Artan veri miktarı, geleneksel analiz yöntemlerinin uygulamada yetersiz kalmasına neden olmuş, veri analizi için yeni yöntemlere ihtiyaç duyulmuştur. Bu sürecin bir sonucu olarak ortaya çıkan yöntemlerden biri de yapay zekânın (Artificial Intelligence, AI) bir alt sınıfı olan makine öğrenmesidir (Machine Learning, ML). Makine öğrenmesi, verilerdeki örüntüleri otomatik olarak algılayan ve ortaya çıkan bu örüntüleri geleceği tahmin etmek için kullanan bir dizi yöntemler bütünüdür.

Gerçek hayatta makine öğrenmesi üzerine çok sayıda çalışma yapılmıştır. Belge sınıflandırma, e-posta spam filtreleme, görüntü, el yazısı, yüz ve aktivite tanıma, hava durumu tahmini, öneri sistemleri (Netflix, Spotify, Youtube, Amazon vb.) günlük yaşamda sıkça karşılaştığımız uygulamalardan sadece bir kısmıdır.

Makine öğrenimi, genel olarak denetimsiz (unsupervised, descriptive) ve denetimli (supervised, predictive) öğrenme olmak üzere iki alt başlıkta incelenmektedir.

3.1 Denetimsiz Öğrenme

Denetimsiz öğrenme modelinde her bir giriş değişkeni için karşılık gelen bir çıktı değişkeni bulunmamaktadır. Kullanıcı yönlendirmesi olmadan, etiketsiz veriler üzerinden öğrenme gerçekleşmektedir. Amaç mevcut veriler hakkında bilgi edinmek, veri yapısını keşfetmektir. Veri seti sistem tarafından işlenerek gruplandırılmaktadır. Belli bir çıktı veya sınıf yönlendirmesi olmadığından, küme sayısında bir kısıtlama yoktur. Kümeleme (Clustering), denetimsiz öğrenmede uygulamada en sık kullanılan tekniktir.

3.2 Denetimli Öğrenme

Pratikte en yaygın kullanılan makine öğrenimi formu denetimli öğrenme yöntemidir. Bu yöntemde amaç, etiketli bir veri kümesi $D = \{(x_i, y_i)\}_{i=1}^N$ verildiğinde, girdilerden (x_i) çıktıya (y_i) bir eşleştirme fonksiyonu yaratmaktır. Burada D modelin eğitilmesi için kullanılan veri setini, N ise veri seti içerisinde bulunan örnek sayısını ifade etmektedir.

Bir veri setinde girdiler (veya özellikler) görüntü, e-posta mesajı, moleküler şekil, grafik gibi sayısal veya kategorik büyüklüklere sahip karmaşık yapıları nesnelere olabilmektedir. Benzer şekilde çıktılar da sayısal veya kategorik türlerde olabilmekte ve çıktı değişkeninin türüne göre de sınıflama ve regresyon modelleri olarak ikiye ayrılmaktadır. Çıktının evet/hayır, pozitif/negatif gibi kategorik özelliklere sahip olması halinde sınıflama veya örüntü tanıma modeli; yaş, ağırlık, gelir seviyesi gibi sayısal değerlere sahip olması durumunda regresyon modeli olarak adlandırılmaktadır.

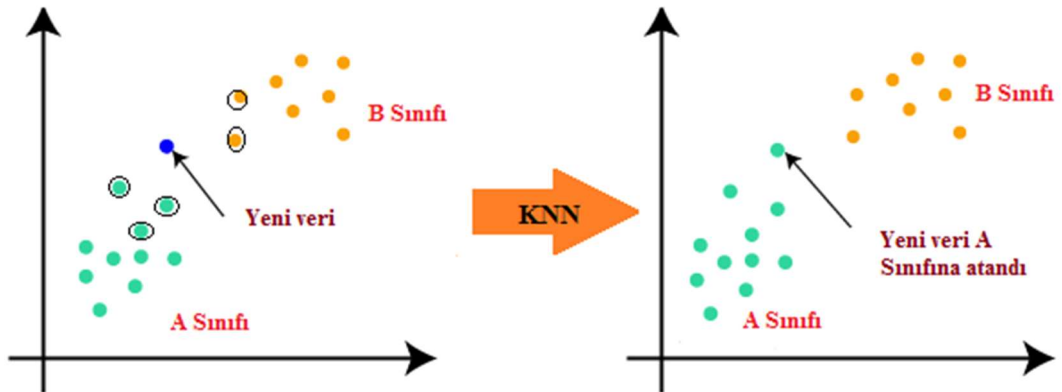
Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler şunlardır:

1. K-En Yakın Komşu (K-Nearest Neighbor, KNN)
2. Naif Bayes (Naive Bayes)
3. Destek Vektör Makineleri (Support Vector Machine, SVM)
4. Karar Ağaçları (Decision Trees)
5. Rastgele Orman (Random Forest)
6. Lineer Regresyon (Linear Regression)
7. Lojistik Regresyon (Logistic Regression) (Murphy 2012).

3.2.1 K-En Yakın Komşu Algoritması

K-En Yakın Komşu (KNN) yöntemi 1950'lerin başlarında istatistikçiler tarafından tanımlanmış (Chakrabarti ve diğ. 2009) ancak büyük boyutlarda veri setlerinde yoğun hesaplama gerektirdiğinden 1960'lı yıllara kadar beklenen ilgiyi görmemiştir. 1960'lı yıllarda hesaplama kabiliyetinin artmasıyla birlikte KNN yöntemi uygulamada kabul görmüş, günümüzde de özellikle örüntü tanıma (pattern recognition) alanında yaygın olarak kullanılmaktadır (Han ve diğ. 2012).

KNN sınıflandırıcısı uygulanması oldukça basit olmakla birlikte iyi bir parametre seçimi ve yeterli etiketli eğitim verisine sahip olması koşuluyla oldukça iyi sonuçlar vermektedir. Bellek tabanlı öğrenme veya örnek tabanlı öğrenme tekniği olarak bilinen KNN yöntemin temeli benzetim yoluyla, yani karşılaştırarak öğrenmeye dayanmaktadır (Murphy 2012). Etiketli bilinmeyen bir test verisi verildiğinde, KNN eğitim veri kümesindeki en yakın k komşuya bakarak çoğunluk durumundaki etikete göre atama yapmaktadır. Burada yakınlık ile ifade edilen örneğin Öklid mesafesi gibi bir uzaklık ölçüsüdür (Han ve diğ. 2012). Örneğin Şekil 3.3'teki gibi A ve B olmak üzere iki sınıfımız ve yeni bir verimiz olsun, bu yeni verinin sınıflardan hangisinde yer alacağına karar verilmesi gerekmektedir. Öncelikli olarak değerlendirme sırasında bakılacak komşu sayısının belirlenmesi gerekmektedir. $k = 5$ alınacak olursa, yeni veri için en yakın 5 komşu arasındaki uzaklıklar hesaplanır ve en yakın da bulunan komşuların çoğunluk sınıflarına göre atama yapılır (Jaiswal 2018).



Şekil 3.3: KNN algoritmasının nasıl çalıştığını gösteren çizim.

En yakın komşu sınıflandırıcıları kullanılarak sayısal çıktılar için tahmin yürütmek de mümkündür. Sayısal bir değişken için tahmin yapılırken, bilinmeyen dizinin k en yakın komşularının ortalama değerini dönüştürür.

KNN yöntemi, test örneklerini sınıflandırırken son derece yavaş olabilmektedir. Bunun nedeni bu tarz örnek tabanlı sınıflandırıcıların test verisi ile karşılaşana kadar eğitim verisini sadece ön işlem yaparak bekletmesidir. Test verisi verildiğinde, bekletilen eğitim verisi üzerinden genelleme yapıldığı için, eğitim aşaması daha hızlı gerçekleşirken test aşaması daha uzun sürmektedir (Han ve diğ. 2012).

KNN yönteminin uygulama aşamasında dikkat edilmesi gereken ve performansını etkileyen i) uzaklık ölçüsü, ii) komşu sayısı ve iii) verileri ağırlıklandırma gibi önemli parametreler bulunmaktadır.

Uzaklık ölçüsü için Öklid (Euclidean), Manhattan, Minkowski, Chebyshev ve Dilca gibi birçok ölçü bulunmaktadır (Taşcı ve Onan 2016). En yaygın kullanılan uzaklık ölçüsü Öklid uzaklığıdır. İki nokta veya grup, diyelim ki $x = (x_1, x_2, \dots, x_n)$ ve $y = (y_1, y_2, \dots, y_n)$ olsun, bu iki nokta arasındaki Öklid mesafesi (d), Denklem (3.1)'de verildiği şekilde hesaplanmaktadır.

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.1)$$

Dikkat edilmesi gereken diğer bir kavram, sınıflandırma için kaç komşunun seçileceğine karar veren k parametresidir. Uygun k değerinin seçimi, KNN algoritmasının tanısal performansı üzerinde önemli etkiye sahiptir. Büyük k değerleri küçük örüntülerin göz ardı edilmesine neden olacaktır. (Zhang 2016). Uygulamalarda k parametresi için genel olarak kullanılan değerler 3, 5 ve 7'dir. En iyi k değeri, verinin boyutuna ve niteliğine göre değişmektedir (Khan ve diğ. 2002).

En yakın komşu sınıflandırıcılar, her bir komşuya eşit ağırlık atayarak mesafeye dayalı karşılaştırmalar yapmaktadır. Bu durum, gürültülü veri grupları ile yapılan çalışmalarda, uygulamanın başarısını etkilemektedir (Han ve diğ. 2012). Ağırlıklandırma yöntemiyle, sınıflandırılmakta olan örneğe daha yakın olan komşu

örneklerin daha uzak komşulara göre daha ağırlıklı etki göstermesini sağlamak amaçlanmıştır. Bu yöntemde, bir komşunun sınıflandırmaya etkisi, sınıflandırılacak yeni örneğe olan uzaklığı ile ters orantılıdır. Her bir komşunun ağırlığı, komşular arası uzaklık (d) olmak üzere $1/d^2$ şeklinde hesaplanmaktadır (Taşcı ve Onan 2016).

3.2.2 Naif Bayes Algoritması

Naif Bayes (NB), temeli Bayes teoremine dayanan en basit sınıflandırıcılardan biridir. Bayes teoremi, olasılık dağılımı içinde rastgele bir değişken için koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi göstermektedir ve matematiksel olarak iki olay arasındaki koşullu olasılık Denklem (3.2)'deki gibi hesaplanmaktadır.

$$P(A|B) = \frac{P(A).P(B|A)}{P(B)} \quad (3.2)$$

Denklem (3.3)'te verilen; $P(A)$, A 'nın marjinal olasılığını; $P(B)$, B 'nin marjinal olasılığını; $P(A|B)$, B olayının bilindiğinde A olayının koşullu olasılığını; $P(B|A)$, A olayı bilindiğinde B olayının koşullu olasılığını ifade etmektedir.

NB sınıflandırıcısı, naif bir varsayımla tahminde kullanılacak tüm özelliklerin kendi aralarında birbirinden bağımsız fakat tahmin edilecek özelliklerle bağımlı olduğunu kabul etmektedir. Bu varsayım ile NB sınıflandırıcısı, Bayes teoreminden türetilen haliyle Denklem (3.3)'te verilen şekliyle hesaplanmaktadır.

$$P(C_i|X) = \frac{P(C_i).P(X|C_i)}{P(X)} \quad (3.3)$$

Burada X , birbirinden bağımsız n adet ($x_1, x_2 \dots x_n$) özellik ve C_i , ($C_1, C_2 \dots C_m$) şeklinde m adet sınıf içermektedir. Denklem (3.4)'te bulunan $P(X)$ olasılığı her sınıf için eşittir ve NB sınıflandırıcısı en yüksek olasılık değerine sahip sınıfa göre karar vermektedir. Yani sınıf değerlendirmesi en yalın haliyle Denklem (3.4)'e göre yapılmaktadır.

$$\max_{C_i} \{P(C_i).P(X|C_i)\} \quad (3.4)$$

Gerçek hayatta özniteliklerin birbirinden bağımsızlığı söz konusu olmasa da, NB sınıflandırıcısının kullanıldığı birçok alanda başarılı tahminler yürüttüğü bilinmektedir (Karabatak 2015).

3.2.3 Destek Vektör Makinesi (SVM)

Destek Vektör Makinesi (SVM), 1963 yılında Vladimir Vapnik ve Alexey Chervonenkis tarafından tanımlanan, istatistiksel öğrenme kuramına dayalı bir denetimli makine öğrenme algoritmasıdır (Akpınar 2017). Başlangıçta iki sınıflı doğrusal verilerin sınıflandırılması amacıyla tasarlanmış matematiksel model, sonrasında çok sınıflı ve doğrusal olmayan verilerin sınıflandırılması amacıyla geliştirilmiştir (Vapnik, 1995, Vapnik 2000).

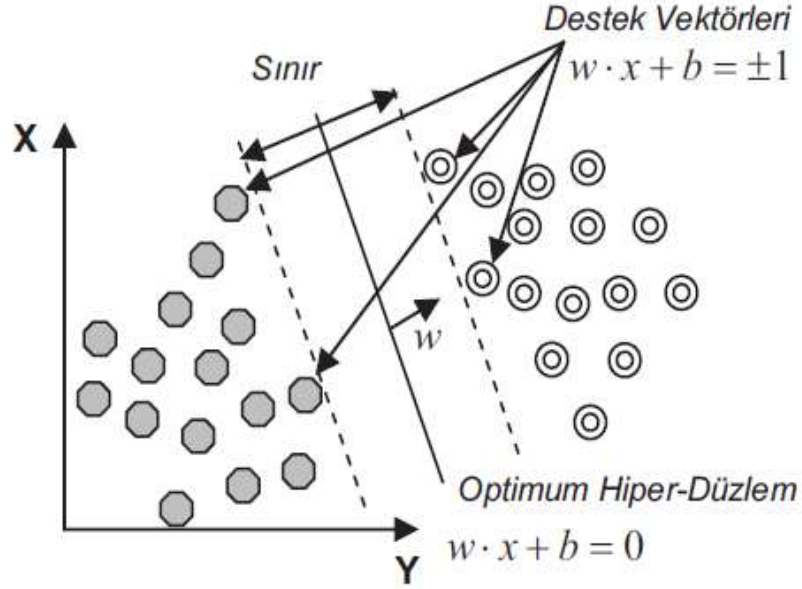
SVM, denetimli makine öğrenimi algoritmaları arasında en yüksek performanslılarından biridir. Genel olarak sınıflandırma işlemleri için tercih edilmekle birlikte regresyon analizinde de sıkça kullanılmaktadır. Gerçek hayatta, SVM algoritmaları metin kategorizasyonu, görüntü tanıma ve biyoinformatik alanlarında, haber hikâyelerinin sınıflandırılması, elle yazılmış rakam tanıma ve kanser dokusu örnekleri üzerine yapılan birçok çalışmada yaygın olarak kullanılmaktadır (Kowalczyk 2017).

SVM algoritmasında temel amaç, Şekil 3.4'te görüldüğü gibi iki sınıfı en uygun şekilde ayıran bir hiper düzlem veya bir karar yüzeyi tanımlamaktır. Sınır genişliğini belirleyen noktalar **destek vektörleri**, destek vektörleri arasındaki uzaklığı maksimuma taşıyarak en uygun ayrımı sağlayan hiper düzlem **optimum hiper düzlem** olarak adlandırılır (Kavzoğlu ve Çölkesen 2010).

Bazı veriler doğrusal olarak ayrılabilirken bazı verilerin doğrusal olarak ayrılması mümkün değildir. SVM, çekirdek fonksiyonları sayesinde doğrusal olmayan verilerin ayrımı için de çözüm sağlamaktadır (Metlek ve Kayaalp 2020).

3.2.3.1 Doğrusal Destek Vektör Makinesi

Doğrusal SVM, verileri Şekil 3.4'te verildiği gibi sınır genişliğini sınırlandıran destek vektörleri arasındaki mesafeyi maksimuma çıkararak iki sınıfa ayıran bir optimum hiper düzlem belirlemeyi amaçlamaktadır (Kavzoğlu ve Çölkesen 2010).



Şekil 3.4: Doğrusal SVM Modeli.

X , n örneklemeden oluşan eğitim veri seti olmak üzere Denklem (3.5)'teki gibi ifade edilebilir.

$$X = \{(x_1 y_1), \dots, (x_n y_n), x_i \in R^+, y_i \in \{-1, +1\}\} \quad (3.5)$$

Buradaki x_i giriş vektörlerini ve y_i bu giriş vektörlerine karşılık gelen etiketleri ifade etmektedir. Bir doğru denklemi şeklinde hiper düzlem denklemi Denklem (3.6)'daki gibi ifade edilebilir.

$$w \cdot x_i + b = 0, \quad w \in R^+, b \in R \quad (3.6)$$

Buradaki w hiper düzlemin normali/ağırlık vektörü ve b sabit (bias) değeridir. Tüm verilerin hesaplama kolaylığı açısından karar sınırından en az 1 uzaklıkta olduğunu varsayılırsa kısıtlar Denklem (3.7) ve (3.8)'deki gibi ifade edilebilir.

$$y_i = +1 \text{ için } w \cdot x_i + b \geq +1 \quad (3.7)$$

$$y_i = -1 \text{ için } w \cdot x_i + b \leq -1 \quad (3.8)$$

Denklem (3.7) ve (3.8) düzenlenerek tek bir formda Denklem (3.9)'daki gibi ifade edilebilir.

$$y_i(w \cdot x_i + b) \geq 1, \quad i = 1, 2, \dots, n \quad (3.9)$$

Destek vektör makinelerinde genel amaç, destek vektörleri arasındaki mesafeyi maksimuma çıkarmaktır. Mesafeyi maksimum yapmak için w değerini minimum yapmak gerekmektedir. Amaç fonksiyonu Denklem (3.10)'daki gibi ifade edilebilir.

$$\min \left[\frac{1}{2} \|w\|^2 \right] \quad (3.10)$$

Buna bağlı kısıtlar;

$$y_i(w \cdot x_i + b) \geq 1 \text{ ve } y_i \in \{-1, +1\} \quad (3.11)$$

Burada amaç fonksiyonu için verilerin doğrusal olarak ayrılabilirdiği kabul edilmekte ve belirlenen doğrunun esneme payı bulunmamaktadır. Ancak bazı durumlar da verilerin yapıları nedeniyle doğrusal olarak ayrılması mümkün değildir. Bu gibi durumlar da doğrusal olarak ayrılamayan verileri işlemek için, sisteme esneme payını (soft margin) arttırmak amacıyla yapay değişken “ ε ” (slack variable) eklenmektedir. Doğrusal olarak ayrılamayan verilen için Denklem (3.9), formüle ε değişkeni eklenerek Denklem (3.12)'deki gibi genelleştirilerek ifade edilebilir.

$$y_i(w \cdot x_i + b) \geq 1 + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad \varepsilon_i \geq 0 \quad (3.12)$$

Burada $\varepsilon_i = 0$ ise x_i örnekleri doğru olarak sınıflandırılmıştır, $0 < \varepsilon_i < 1$ ise x_i örnekleri doğru olarak sınıflandırılmıştır ancak sınırın içindedir, $1 < \varepsilon_i$ ise x_i örnekleri yanlış olarak sınıflandırılmıştır.

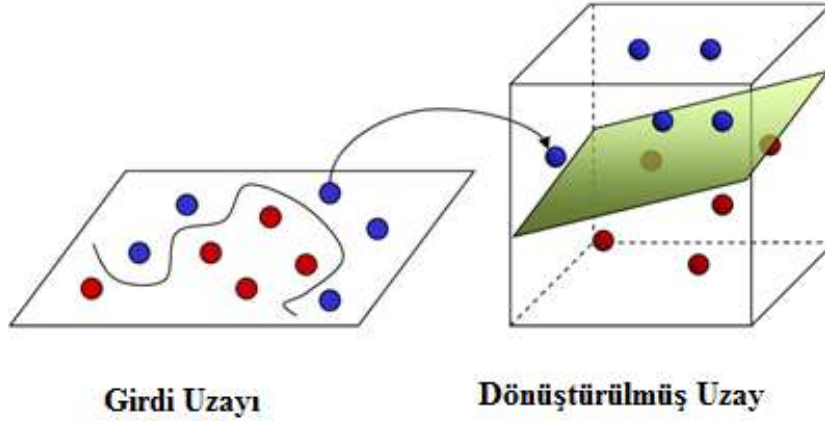
Daha önce oluşturulan amaç fonksiyonu düzenlenirse yeni amaç fonksiyonu Denklem (3.13)'teki gibi ifade edilebilir.

$$\min \left[\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \varepsilon_i \right] \quad (3.13)$$

Buradaki C parametresi, iki düzlem arasındaki mesafeyi maksimum da tutarken esneme payını kabul edilebilir seviyede tutmak için kullanılan ceza faktörü olarak tanımlanmaktadır (Alpaydın 2004, Metlek ve Kayaalp 2020).

3.2.3.2 Doğrusal Olmayan Destek Vektör Makineleri

Verilerin doğrusal olarak ayrılmadığı durumlarda, veriyi doğrusal olmayan bir fonksiyona uydurmak yerine, verileri daha yüksek boyutlu bir alanda eşleştirmek çözüm olabilmektedir. SVM algoritmasında, doğrusal ayrılabilir olmayan bir veriyi doğrusal ayrılabilir bir veriye dönüştürmek için çekirdek fonksiyonlardan yararlanır. Veri, bir çekirdek fonksiyonu yardımıyla iki boyutlu ise üç boyutlu bir özellik uzay formunda Şekil 3.5'teki gibi haritalanır (Moreira 2011).



Şekil 3.5: Çekirdek fonksiyonu ile yüksek boyutlu uzayda sınıflandırma.

Özellik uzayındaki doğrusal işlem, orijinal girdi uzayındaki doğrusal olmayan işleme eşdeğerdir. Yeni özellik uzayında verilerin doğrusal olarak ayrılabilir olmaması durumu göz önünde bulundurularak esnek yapıya sahip hiper düzlem kullanılmaktadır.

Literatürde kullanılan birçok çekirdek fonksiyon bulunmaktadır. Bunlardan bazıları Doğrusal (Linear), Polinom (Polynomial), Radyal Tabanlı (RB/Radial Based), Sigmoid, Laplace, Anova, Bessel çekirdek fonksiyonlarıdır. En popüler çekirdek fonksiyonlar Tablo 3.2’de matematiksel ifadeleriyle verilmiştir (Metlek ve Kayaalp 2020).

Tablo 3.2: Temel kernel fonksiyonları ve parametreleri.

Kernel Fonksiyonu	Matematiksel İfadesi
Doğrusal	$K(x_i, x_j) = x_i \cdot x_j$
Polinom	$K(x_i, x_j) = (x_i \cdot x_j + 1)^d$
Radyal	$K(x_i, x_j) = e^{-\gamma \ x_i - x_j\ ^2}$
Sigmoid	$K(x_i, x_j) = \tanh(\alpha \cdot x_i \cdot x_j + 1)$

SVM algoritmasının uygulanmasında tüm destek vektör makineleri için c ceza parametresinin belirlenmesi, uygulanacak çekirdek fonksiyonun seçimi ve fonksiyon türüne bağlı hiper parametre değerlerinin belirlenmesi oldukça önemlidir. Genellikle çekirdek fonksiyon seçimi için önerilen yaklaşım en çok kullanılan ve en iyi performansı sergileyen RBF türünü denemektir. Parametre değerleri için tüm problemler için kullanılabilecek optimal bir değer yoktur, bu nedenle veri setine uygun parametre değerlerini belirlemek için ızgara arama (grid search) ve çapraz doğrulama (cross validation) yaklaşımlarının kullanımı önerilmektedir (Kowalczyk 2017). Her bir parametre için alt sınır, üst sınır ve belirli bir değer aralığı belirlenir, en iyi sınıflandırma performansını gösteren değerler optimal hiper parametre değerleri olarak kabul edilir (Ayhan ve Erdoğmuş 2014).

3.3 Performans Değerlendirme Yöntemleri

Denetimli öğrenme makine öğrenmesi yöntemleri kullanılarak gerçekleştirilecek tahmin/sınıflandırma çalışmalarında kullanılacak algoritmanın, var ise algoritmanın parametreleri ve/veya hiper parametrelerinin seçimi, performans değerlendirilmesinde kullanılacak ölçütler ve veri setinin yapısı oldukça önemlidir.

3.3.1 Performans Değerlendirme Kriterleri

Kurulan modelin veya kullanılan algoritmanın kendisine verilen görevi ne kadar iyi yerine getirip getiremediğini görebilmek için performansının değerlendirilmesi gerekmektedir. Makine öğrenmesinde performans değerlendirmesi önemli bir husus olup, bu amaçla birçok performans değerlendirme kriteri geliştirilmiştir. En sık kullanılan performans değerlendirme ölçütleri; Doğru Pozitif (DP), Yanlış Pozitif (YP), Doğru Negatif (DN), Yanlış Negatif (YN), Doğruluk (accuracy), Kesinlik (precision), Duyarlılık (recall), f-ölçütü (f-score), Alıcı İşlem Karakteristiği Eğrisi (Receiver Operating Characteristic Curve, ROC), Eğri Altında Kalan Alan (Area Under Curve, AUC) olarak karışımıza çıkmaktadır.

Performans kıyaslaması için genellikle doğruluk değeri kullanılmaktadır. Tek başına doğruluk değerinin kullanılması performans başarısının ölçümünde yetersiz olacaktır (Balaban ve Kartal 2015).

Bazı performans değerlendirme kriterlerinin hesaplanmasında hata matrisi (confusion matrix) kullanılmaktadır. Şekil 3.6'da, hata matrisi ve bu matris kullanılarak hesaplanan birkaç performans ölçütü gösterilmektedir.

		Gerçek	
		Pozitif	Negatif
Tahmin	Pozitif	DP (Doğru Pozitif)	YP (Yanlış Pozitif)
	Negatif	YN (Yanlış Negatif)	DN (Doğru Negatif)

Şekil 3.6: Hata matrisi

Şekil 3.6’da belirtilen;

Doğru Pozitif (DP): Tahmin pozitif ve gerçekte pozitif olan (doğru tahmin)

Yanlış Pozitif (YP): Tahmin pozitif fakat gerçekte negatif olan (yanlış tahmin)

Doğru Negatif (DN): Tahmin negatif ve gerçekte negatif olan (doğru tahmin)

Yanlış Negatif (YN): Tahmin negatif fakat gerçekte pozitif olan (yanlış tahmin) sayısını ifade etmektedir.

3.3.1.1 Doğruluk (Accuracy)

Doğruluk değeri, doğru sınıflandırılmış pozitif ve negatif örnek sayısının toplam pozitif ve negatif örnek sayısına bölünmesi ile hesaplanır ve Denklem (3.14)’de gösterildiği şekilde ifade edilmektedir.

$$Doğruluk = \frac{DP + DN}{DP + YP + DN + YN} \quad (3.14)$$

3.3.1.2 Kesinlik (Precision)

Kesinlik değeri, doğru sınıflandırılmış pozitif örnek sayısının tahmin edilen toplam pozitif örnek sayısına bölünmesi ile hesaplanır ve Denklem (3.15)'te gösterildiği şekilde ifade edilmektedir.

$$Kesinlik = \frac{DP}{DP + YP} \quad (3.15)$$

3.3.1.3 Duyarlılık (Recall)

Duyarlılık değeri, doğru sınıflandırılmış pozitif örnek sayısının model içerisindeki toplam pozitif örnek sayısına bölünmesiyle elde edilir, doğru pozitif oranı (DPO) olarak da adlandırılır ve Denklem (3.16)'da gösterildiği şekilde ifade edilmektedir.

$$Duyarlılık = \frac{DP}{DP + YN} \quad (3.16)$$

3.3.1.4 f-ölçütü (f-score)

f-ölçütü, kesinlik ve duyarlılık değerlerinin harmonik ortalaması alınarak hesaplanır. Denklem (3.17)'de verildiği şekilde ifade edilmektedir (Fawcett 2006).

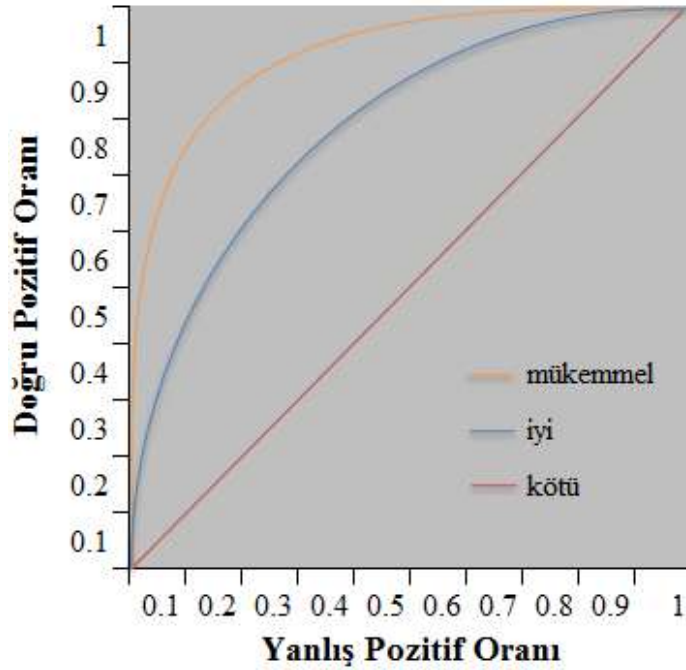
$$f - ölçütü = 2 * \frac{Kesinlik * Duyarlilik}{Kesinlik + Duyarlilik} \quad (3.17)$$

3.3.1.5 İşlem Karakteristiği Eğrisi

İşlem Karakteristiği (Receiver Operating Characteristic, ROC) Analizi, II. Dünya Savaşı yıllarında radarda tespit edilen sinyallerin doğru tanımlanması amacıyla geliştirilmiştir. 1967 yılında Leo Lusted tarafından tıpta kullanımı önerilmiş ve 1969 yılında medikal görüntüleme cihazlarıyla ilgili karar verme süreçlerinde kullanılmaya başlanmıştır. Sonraki yıllarda tıpta farklı klinik durumlarda uygulanan tanı testlerinin

performansının değerlendirilmesinde, günümüzde ise istatistiksel modellerin performansının değerlendirilmesi ile geniş kullanım alanına sahip bir yöntemdir.

ROC grafikleri, Şekil 3.7’de görüldüğü gibi, Doğru Pozitif (DP) oranının Y ekseninde ve Yanlış Pozitif (YP) oranının X ekseninde çizildiği iki boyutlu grafiklerdir.



Şekil 3.7: Performanslarına göre ROC Eğrileri

Doğru Pozitif Oranı (DPO) ve Yanlış Pozitif Oranı (YPO) hesaplanırken hata matrisi kullanılmakta ve Denklem (3.18) ve (3.19)’de gösterildiği şekilde ifade edilmektedir.

$$\text{Doğru Pozitif Oranı} = \frac{DP}{DP + YN} \quad (3.18)$$

$$\text{Yanlış Pozitif Oranı} = \frac{YP}{YP + DN} \quad (3.19)$$

ROC Eğrisinin altında kalan alan, AUC (Area Under Curve) olarak adlandırılmakta ve algoritmanın sınıflandırma başarısını göstermektedir. AUC, 0 ile 1.0 arasında değerler almakla birlikte, gerçekçi bir model 0.5’ten küçük bir değer almamalıdır. Eğri altında kalan alan aşağıda verilen derecelendirmeler kullanılarak yorumlanabilir.

- .90-1.00 = mükemmel
- .80-.90 = iyi
- .70-.80 = orta
- .60-.70 = zayıf
- .50-.60 = başarısız (Ertorsun ve diğ. 2009).

3.3.2 k-Kat Çapraz Doğrulama (k-Fold Cross-Validation)

Uygulama için kullanılacak model üzerinde eğitimin gerçekleştirilebilmesi ve güvenilir bir tahmin yapabilmek için yeterli veriye sahip olmamız gerekmektedir. Elimizde yeteri kadar verinin olmaması, kurulan modelin performansını da etkileyecektir. Bunun için az miktarda veriye sahip olunan durumlarda basit ama popüler bir çözüm olan, çapraz doğrulama yöntemi kullanmaktır (Murphy 2012).

Çapraz doğrulama (cross-validation), elde bulunan mevcut veriyi iki parçaya ayırarak öğrenme algoritmalarının performansını karşılaştırmak ve değerlendirmek için kullanılan bir yöntemdir. İkiye ayrılan verinin bir parçası modeli eğitmek için kullanılırken, diğer parçası ise modelin geçerliliğini denetlemek için kullanılmaktadır.

k-kat çapraz doğrulama (k-fold cross-validation) yöntemi, çapraz doğrulama yönteminin veri setini k sayıda eşit parçaya ayırdığı şeklindedir. Her defasında k parçadan bir tanesi test için kullanılırken, geriye kalan k-1 parça eğitim için kullanılmaktadır. (Han ve diğ. 2012). İşlem, Şekil 3.8’de görüldüğü gibi, her k parça test için bir kez kullanılıncaya kadar devam etmektedir (Çınar 2017). Genel olarak, doğrulama tahmini için 10-kat çapraz doğrulama kullanımı önerilmektedir (Han ve diğ. 2012).

	1. Parça	2. Parça	3. Parça	4. Parça	5. Parça	6. Parça	7. Parça	8. Parça	9. Parça	10. Parça
1. Adım	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
2. Adım	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
3. Adım	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
4. Adım	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
5. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim
6. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim	Eğitim
7. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim	Eğitim
8. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim	Eğitim
9. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test	Eğitim
10. Adım	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Eğitim	Test

Şekil 3.8: 10-kat çapraz doğrulama şeması.

3.3.3 Izgara Arama (Grid Search)

Hiperparametreler, eğitim sürecinden bağımsız, eğitimden önce kullanıcı tarafından belirlenen parametrelerdir. Izgara arama (Grid search) yöntemi ise, hiperparametre optimizasyonu (hyperparameter optimization) amacıyla kullanılan en temel yaklaşımdır. Kullanıcı tarafından belirlenmek istenen hiperparametreler için olası farklı değerler tanımlanmaktadır. Bu değerlerin tüm kombinasyonları için, algoritmanın en iyi performansı gösterdiği hiperparametre seti belirlenmektedir. Kombinasyonun doğruluğunu test etmek için çapraz doğrulama yöntemi kullanılmaktadır (Stalfort 2019).

Izgara arama yöntemi, küçük boyutlu veri setlerinde yüksek verimlilik sağlamaktadır. Yöntem, kolaylıkla uygulanabilmekle birlikte, tüm parametreleri optimize etmeye çalışmak oldukça zaman alacaktır. Bunun yerine; problem hakkında sahip olunan ön bilgiler kullanılarak, hiper parametreler için olası değer aralıkları belirlenerek zamandan tasarruf sağlanabilecektir (Çarkacı 2018).

3.3.4 Veri Setinin Dengelenmesi

Çoğu zaman veri setinde sınıflara ait gözlem sayıları arasında büyük fark bulunur. Bu tarz veri setlerine dengesiz veri seti denir. Günlük hayatta özellikle tıbbi verilerde bu problem daha yaygındır. Bu türde veri kümeleriyle yapılan çalışmalarda öğrenme aşamasında örneklem sayısının fazla olduğu sınıf (çoğunluk sınıfı) daha baskın olurken örneklem sayısının az olduğu sınıflara (azınlık sınıfı) ait gözlem değerlerinin sınıflandırılmasında bazı tutarsızlıklar görülebilmektedir. Sınıflara ait gözlem sayılarını eşitlemek için iki yöntemden birisi kullanılır. i) çoğunluk sınıflarındaki gözlemlerin rastgele silinerek tüm sınıflardaki gözlem sayısını azınlık sınıfı sayısına indirmek (alt örnekleme, undersampling), ii) azınlık sınıfındaki yer alan gözlemlerin sayısını arttırarak (aşırı örnekleme, oversampling) sınıflar arasında denge oluşturulmaya çalışılır. Uygulamada daha çok aşırı örnekleme yöntemi tercih edilmektedir. Aşırı örnekleme yöntemleri içinde en popüler yöntem SMOTE (Synthetic Minority Oversampling Technique) yöntemidir. SMOTE yönteminde, az sayıda gözleme sahip sınıflar için sentetik gözlemler oluşturularak sınıflar arasında denge sağlanır. SMOTE yönteminde öncelikle azınlık sınıfından rastgele bir örnek seçilir ardından seçilen bu örneğin k komşusuna bakılarak sentetik örnekler oluşturulur. Bu yöntemde azınlık sınıfındaki örnekler kopyalanarak değil en yakın k komşularına bakılarak aşırı örneklenmiş hale getirilmektedir (Blagus ve Lusa 2013).

4. UYGULAMA

Çalışma kapsamında, COVID-19 salgını ile mücadele sürecinde, hastalığın erken ve doğru teşhisi amacıyla bireylerin tam kan sayımı sonuçlarına göre SARS-CoV-2 pozitif hastaların tahmini amaçlanmıştır. Bu tahmin çalışması sırasında test sonuçlarının pozitif ve negatif olarak sınıflandırılmasında KNN, NB ve SVM algoritmalarından yararlanılmıştır.

4.1 Veri Seti

Bu çalışmada kullanılan tüm veriler, SARS-CoV-2 salgınının yayılımıyla mücadele kapsamında halka açık bir veri kaynağından (www.kaggle.com) elde edilmiştir.

Bu veri seti, São Paulo, Brezilya'daki İsrail Hastanesi Albert Einstein'da görülen ve hastaneye ziyaret sırasında SARS-CoV-2 RT-PCR ve ek laboratuvar testleri yapmak için numuneler toplanan hastaların bilgilerini içermektedir. Tüm veriler, uluslararası uygulamalar ve tavsiyeler doğrultusunda anonimleştirilmiştir. Tüm klinik veriler hastane tarafından, ortalama sifra ve birim standart sapmaya sahip olacak şekilde standartlaştırılmıştır. Yayınlanan tam veri seti, 28 Mart ve 3 Nisan 2020 arasında test edilen 5644 bireysel hastayı içermektedir (Data4u 2020).

Yayınlanan veriler, yaş (yüzdeler grup), RT-PCR SARS-CoV-2 test sonucu ve Tablo 4.3'te verilen standart tam kan sayımı parametreleri: Hct (Hemotokrit), Hgb (Hemoglobin), PLT (Platelet, Trombosit sayısı), MPV (Mean platelet volume, Ortalama trombosit hacmi), RBC (Red blood cells, Kırmızı kan hücresi, Eritrosit sayısı), Lym (Lenfosit), MCHC (Mean corpuscular hemoglobin concentration, Ortalama korpusküler hemoglobin konsantrasyonu), WBC (White blood cell, Beyaz kan hücresi, Lökosit Sayısı), Baso (Bazofil), Neu (Nötrofil), MCH (Mean corpuscular hemoglobin, Ortalama korpusküler hemoglobin), Eos (Eozonofil), MCV (Mean corpuscular volume, Ortalama Eritrosit Hacmi), Mono (Monosit) ve RBCDW (Red blood cell distribution width, Eritrosit dağılım genişliği) (Lappin 2002) ve diğer test sonuçlarını içermektedir.

Tablo 4.3: Tam kan sayımı parametreleri.

No	Özellik	Açıklama
1	Hct	Hemotokrit
2	Hgb	Hemoglobin
3	PLT	Trombosit sayısı
4	MPV	Ortalama trombosit hacmi
5	RBC	Kırmızı kan hücresi, Eritrosit sayısı
6	Lym	Lenfosit
7	MCHC	Ortalama korpusküler hemoglobin konsantrasyonu
8	WBC	Beyaz kan hücresi
9	Baso	Bazofil
10	Neu	Nötrofil
11	MCH	Ortalama korpusküler hemoglobin
12	Eos	Eozonofil
13	MCV	Ortalama eritrosit hacmi
14	Mono	Monosit
15	RBCDW	Eritrosit dağılım genişliği

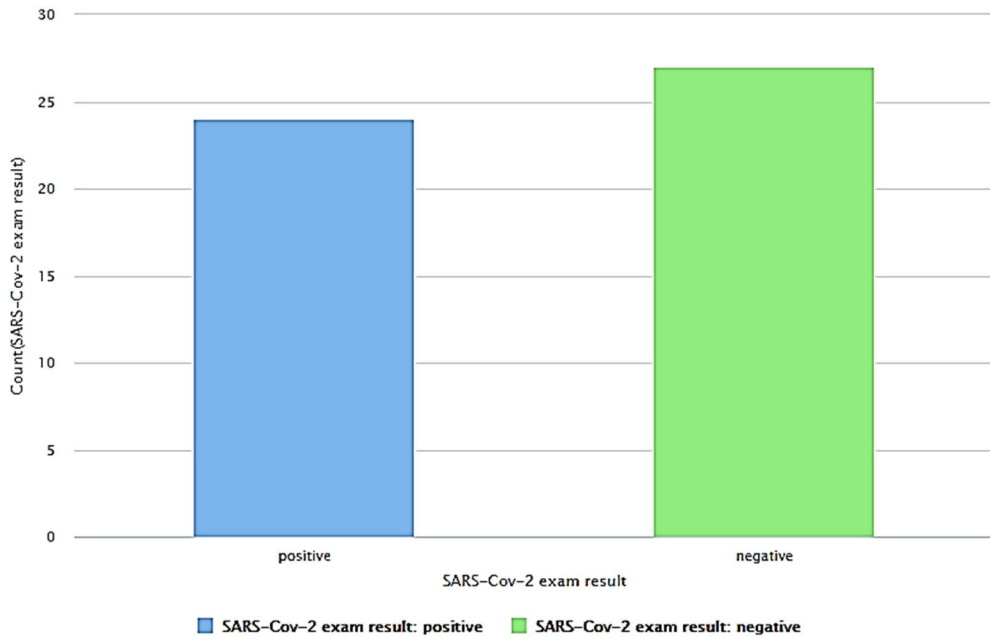
Veri seti içerisinde, 5644 hastadan 603 tanesi için yukarıda belirtilen tam kan sayımı parametre sonuçları bulunmaktadır. Ancak bazı parametreler için sonuçlarda eksiklikler bulunmaktadır bu nedenle hesaplama kolaylığı sağlamak ve yanlış sonuçlar oluşmasını engellemek amacıyla tüm parametre sonuçları bulunan 510 hastaya ait veriler analize dâhil edilmiştir.

Testlerin, enfeksiyonun hangi evresinde gerçekleştirildiğine dair zaman bilgisine sahip olunmaması nedeniyle tam kan sayımı sonuçları; Genel (Hastaneye kabul edilmeyen), Normal Servis, Yarı Yoğun Bakım Ünitesi ve Yoğun Bakım Ünitesi olmak üzere hastaların hastaneye kabul durumlarına göre dört gruba ayrılarak Tablo 4.4'te gösterilmiştir. Yarı Yoğun Bakım Ünitesi ve Yoğun Bakım Ünitesi'nde olan hastalar, çalışmanın amacı hastalığa erken tanı koymak olduğu için analize dâhil edilmemiştir. Yaş bilgisi ve tüm kan testi değerlerine sahip olmayan hasta sonuçları da veri setinden çıkarılmıştır.

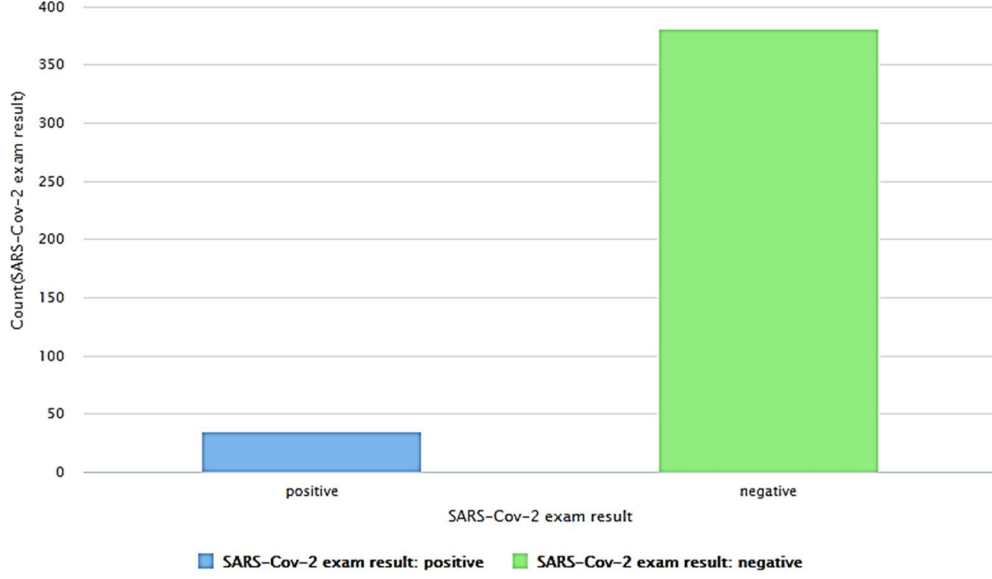
Tablo 4.4: SARS-CoV-2 test sonuçlarına göre veri grupları.

Hasta Sayısı	Genel	Normal Servis	Yarı Yoğun Bakım Ünitesi	Yoğun Bakım Ünitesi	Toplam
SARS-CoV-2 Negatif	381 (% 92)	27 (% 53)	21 (% 75)	8 (% 53)	437 (% 86)
SARS-CoV-2 Pozitif	35 (% 8)	24 (% 47)	7 (% 25)	7 (% 47)	73 (% 14)

Tam kan sayımı değerlerinin tümüne sahip 510 hastanın, 51'i normal servis grubunda kabul edilirken (RT-PCR SARS-CoV-2 testi 24'ü pozitif ve 27'si negatif), 416'sı genel grupta değerlendirilmiştir (RT-PCR SARS-CoV-2 testi 35'i pozitif ve 381'i negatif). Her iki hasta grubu için model tahminleri ayrı ayrı rapor edilecektir. Ayrıca Şekil 4.9'da görüldüğü üzere normal servis grubunda SARS CoV-2 Test Sonucu Pozitif (24) ve Negatif (27) hasta sayıları arasında bir denge bulunurken, Şekil 4.10'da görüldüğü gibi genel grupta SARS CoV-2 Test Sonucu Pozitif (35) ve Negatif (381) hasta sayıları arasında büyük bir dengesizlik bulunmaktadır.



Şekil 4.9: Normal servis grubu veri seti SARS-CoV-2 Test Sonucu pozitif ve negatif hasta sayısı.



Şekil 4.10: Genel grup veri seti SARS-CoV-2 Test Sonucu pozitif ve negatif hasta sayısı.

Bu sebeple çalışma kapsamında sadece toplam doğruluğa odaklanılmamış, daha bilgilendirici olması adına testler duyarlılık (recall), kesinlik (precision), doğruluk (accuracy), f-ölçütü (f-score) ve AUC (Area Under Curve) ölçüleri göz önünde bulundurularak değerlendirilmiştir. Ayrıca dengesiz veri kümeleri için uygulama da sıkça kullanılan SMOTE yöntemi çalışmaya dâhil edilerek genel grup veri seti için modellerde uygulanmıştır.

4.2 Model

Uygulama normal servis grubu ve genel gruptaki hastalara ait veriler; KNN, NB ve SVM algoritmaları kullanılarak incelenmiş ve elde edilen sonuçlar belirlenen performans ölçütlerine (doğruluk, AUC, f-ölçütü, duyarlılık, kesinlik) göre değerlendirilmiştir.

İlk adımda her iki veri grubu için de KNN algoritması; k parametresi 1-7 değer aralığı ve normal/ağırlıklı oylama durumuna göre performans çıktılarını optimize etmek amacıyla ızgara arama yöntemi kullanılarak incelenmiştir. KNN algoritması için standart olarak Öklid uzaklığı kullanılmıştır. İkinci adımda her iki veri grubu için de NB algoritması kullanılarak oluşturulan model için performans çıktıları incelenmiştir. Üçüncü adımda ise her iki veri grubu için de SVM algoritması, RBF

çekirdek fonksiyonu; $c (2^{-5}, 2^{10})$ ve $\gamma (2^{-10}, 2^5)$ deęer aralıklarına göre performans çıktılarını optimize etmek amacıyla ızgara arama yöntemi kullanılarak incelenmiştir. Son adımda her üç model için de elde edilen performans çıktıları karşılaştırmalı olarak incelenmiştir. Oluşturulan her üç algoritma için de model 10-kat çapraz doğrulama yöntemi kullanılarak, RapidMiner programı yardımıyla oluşturulmuştur.

4.3 RapidMiner

RapidMiner, makine öğrenmesi, veri ve metin madencilięi, tahmin edici analiz ve iş analizi amaçlarına yönelik olarak geliştirilmiş bir yazılım aracıdır. İş ve ticaret dünyasında kullanıldığı gibi araştırma ve eğitim alanlarında da sıklıkla kullanılmaktadır. RapidMiner bir topluluk yazılımıdır (community software). Topluluk üyeleri tarafından ve akademik çalışmalar için ücretsiz olarak kullanılabilir (Seker 2016).

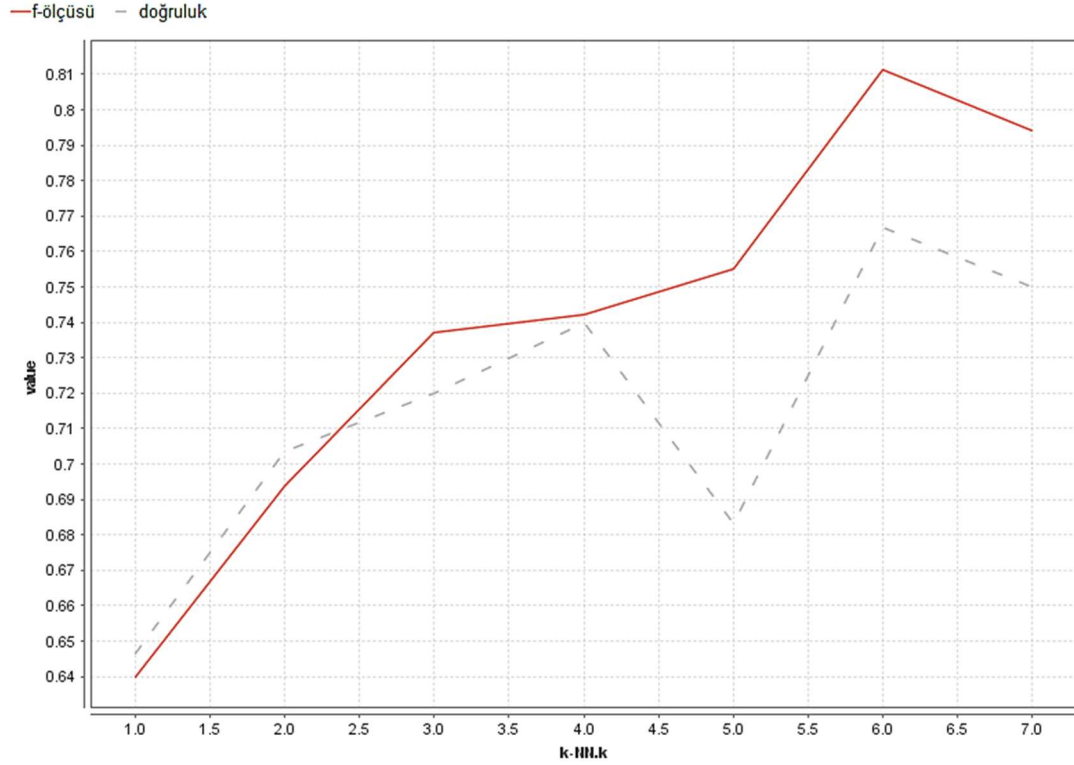
4.4 KNN algoritması uygulama sonuçları

Tablo 4.5'te Normal servis grubu veri seti için KNN algoritmasının, k parametresi 1-7 değer aralığında ve normal/ağırlıklı oylama durumlarına göre performans başarımları değerleri verilmektedir. En yüksek doğruluk (0,767), k=6 değeri için, ağırlıklı oylama kullanılarak elde edilmiştir. AUC (0,917), f-ölçütü (0,811), kesinlik (0,725) ve duyarlılık (0,967) değerleri de sınıflandırma başarısını desteklemektedir.

Tablo 4.5: Normal servis grubu veri seti, KNN algoritması performans ölçümleri.

Sıra	k	Ağırlık Durumu	Duyarlılık	Kesinlik	f-ölçütü	AUC	Doğruluk
1	1	Normal	0,667	0,680	0,694	0,500	0,707
2	2	Normal	0,900	0,623	0,706	0,597	0,627
3	3	Normal	0,883	0,708	0,777	0,703	0,750
4	4	Normal	1,000	0,633	0,765	0,797	0,710
5	5	Normal	0,933	0,702	0,774	0,861	0,747
6	6	Normal	0,967	0,640	0,755	0,889	0,703
7	7	Normal	0,950	0,648	0,758	0,881	0,723
8	1	Ağırlıklı	0,650	0,615	0,640	0,450	0,647
9	2	Ağırlıklı	0,717	0,680	0,694	0,667	0,703
10	3	Ağırlıklı	0,833	0,683	0,737	0,725	0,720
11	4	Ağırlıklı	0,850	0,742	0,742	0,850	0,740
12	5	Ağırlıklı	0,933	0,675	0,755	0,822	0,683
13	6	Ağırlıklı	0,967	0,725	0,811	0,917	0,767
14	7	Ağırlıklı	0,950	0,702	0,794	0,850	0,750

Şekil 4.11'de normal servis grubu veri seti için KNN algoritması, en iyi performans başarısını gösteren ağırlık oylama yöntemi ve belirlenen k değer aralığına göre doğruluk ve f-ölçütü değerlerinin değişimi grafiksel olarak gösterilmiştir.



Şekil 4.11: Normal servis grubu veri seti, KNN algoritması doğruluk ve f-ölçütü grafiği.

Tablo 4.6’da, Normal servis grubu veri seti için en yüksek performans başarısını gösteren değerler için karşılaştırma matrisi verilmiştir. Karşılaştırma matrisine göre KNN algoritması 51 hasta verisinin 39 (23+16) tanesini doğru olarak sınıflandırmıştır.

Tablo 4.6: Normal servis grubu veri seti, KNN algoritması karşılaştırma matrisi.

Sınıflandırıcı	Pozitif	Negatif
KNN	Pozitif	11
	Negatif	16

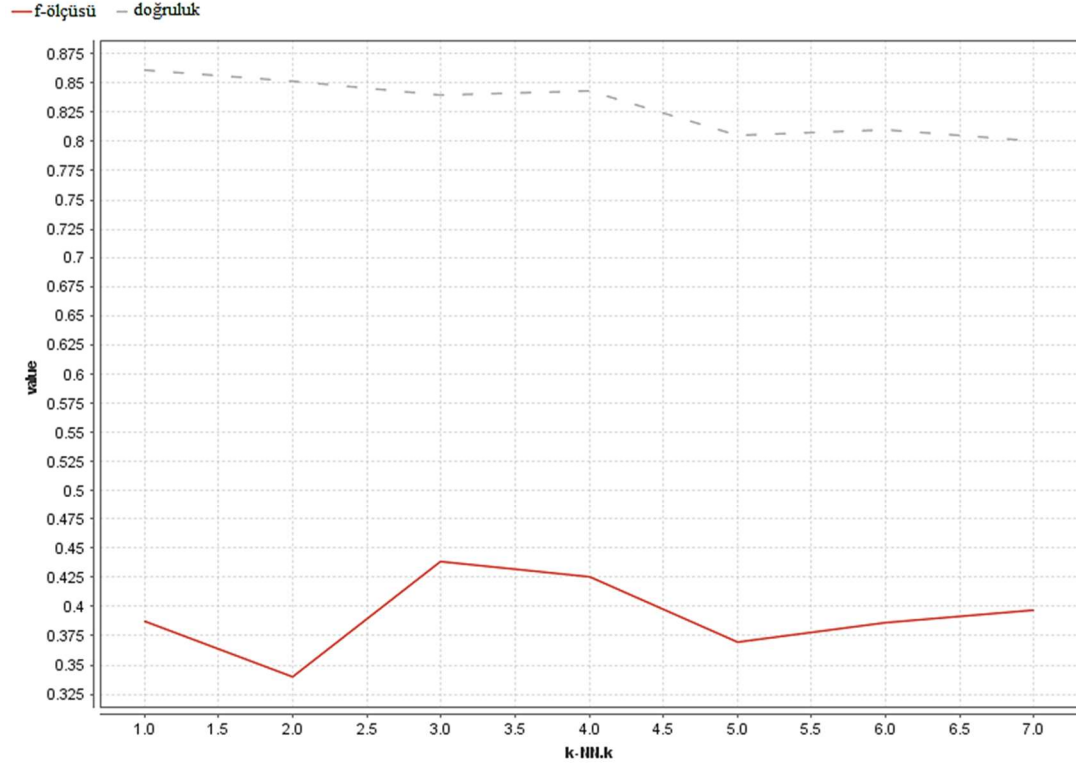
Tablo 4.7’de Genel grup veri seti için KNN algoritmasının, k parametresi 1-7 değer aralığında ve normal/ağırlıklı oylama durumlarına göre performans başarımları değerlendirilmiştir. Genel grup veri setinde pozitif ve negatif sınıflardaki veri sayısının dengesiz dağılımı, SMOTE yöntemi kullanılarak, pozitif sınıf için örnek sayısı sentetik olarak artırılarak giderilmiştir. SMOTE yönteminin dâhil edilmesi ile KNN modelinde en yüksek doğruluk (0,861), k=1 değeri için, ağırlıklı oylama

kullanılarak elde edilmiştir. Doğruluk oranına kıyasla AUC (0,500), f-ölçütü (0,387), kesinlik (0,320) ve duyarlılık (0,525) değerleri oldukça düşüktür.

Tablo 4.7: Genel grup veri seti, KNN algoritması performans ölçümleri.

Sıra	k	Ağırlık Durumu	Duyarlılık	Kesinlik	f-ölçütü	AUC	Doğruluk
1	1	Normal	0,375	0,220	0,286	0,500	0,844
2	2	Normal	0,717	0,256	0,371	0,747	0,798
3	3	Normal	0,633	0,260	0,360	0,795	0,807
4	4	Normal	0,775	0,284	0,408	0,831	0,803
5	5	Normal	0,708	0,283	0,393	0,827	0,812
6	6	Normal	0,867	0,272	0,408	0,842	0,781
7	7	Normal	0,833	0,263	0,397	0,840	0,784
8	1	Ağırlıklı	0,525	0,320	0,387	0,500	0,861
9	2	Ağırlıklı	0,458	0,292	0,340	0,752	0,851
10	3	Ağırlıklı	0,725	0,320	0,438	0,803	0,839
11	4	Ağırlıklı	0,650	0,343	0,425	0,828	0,843
12	5	Ağırlıklı	0,692	0,257	0,369	0,833	0,806
13	6	Ağırlıklı	0,683	0,279	0,386	0,850	0,810
14	7	Ağırlıklı	0,783	0,270	0,397	0,836	0,801

Şekil 4.12’de Genel grup veri seti için KNN algoritması, en iyi performans başarısını gösteren ağırlık oylama yöntemi ve belirlenen k değer aralığına göre doğruluk ve f-ölçütü değerlerinin değişimi grafiksel olarak gösterilmiştir.



Şekil 4.12: Genel grup veri seti, KNN algoritması doğruluk ve f-ölçütü grafiği.

Tablo 4.8’de, Genel grup veri seti için en iyi performans başarısı gösteren değerler için karşılaştırma matrisi verilmiştir. Karşılaştırma matrisine göre KNN algoritması 416 hasta verisinin 358 (18+340) tanesini doğru olarak sınıflandırmıştır.

Tablo 4.8: Genel grup veri seti, KNN algoritması karşılaştırma matrisi.

Sınıflandırıcı	Pozitif	Negatif
KNN	Pozitif	41
	Negatif	340

4.5 NB algoritması uygulama sonuçları

Tablo 4.9’da, Normal servis grubu veri seti için NB algoritması için elde edilen karşılaştırma matrisi verilmiştir. Karşılaştırma matrisine göre NB algoritması 51 hasta verisinin 46 (22+24) tanesini doğru olarak sınıflandırmıştır.

Tablo 4.9: Normal servis grubu veri seti, NB algoritması karşılaştırma matrisi.

Sınıflandırıcı	Pozitif	Negatif
NB	Pozitif	3
	Negatif	24

Tablo 4.10’da, Normal servis grubu veri seti için NB algoritması için elde edilen en iyi performans başarımları değerleri verilmektedir. Doğruluk (0.903) değerine göre iyi bir performans gösteren algoritma için AUC (0.917), f-ölçütü (0,898), kesinlik (0,880) ve duyarlılık (0,917) değerleri de algoritma başarısını desteklemektedir.

Tablo 4.10: Normal servis grubu veri seti, NB algoritması performans ölçümleri.

Sınıflandırıcı	Duyarlılık	Kesinlik	f-ölçütü	AUC	Doğruluk
NB	0,917	0,880	0,898	0,917	0,903

Genel grup veri setinde pozitif ve negatif sınıflardaki veri sayısının dengesiz dağılımı, SMOTE yöntemi kullanılarak, pozitif sınıf için örnek sayısı sentetik olarak artırılarak giderilmeye çalışılmıştır. Ancak NB modelinde SMOTE yönteminin dâhil edilmesi ile daha düşük performans sonuçları elde edilmiştir. Bu nedenle NB algoritması için genel grup veri seti SMOTE yöntemi uygulanmadan elde edilen sonuçlara göre değerlendirilmiştir.

Tablo 4.11’de, Genel grup veri seti için NB algoritması için elde edilen karşılaştırma matrisi verilmiştir. Karşılaştırma matrisine göre NB algoritması 416 hasta verisinin 366 (20+346) tanesini doğru olarak sınıflandırmıştır.

Tablo 4.11: Genel grup veri seti, NB algoritması karşılaştırma matrisi.

Sınıflandırıcı	Pozitif	Negatif
NB	Pozitif	20
	Negatif	346

Tablo 4.12’de, Genel grup veri seti için NB algoritması için elde edilen en iyi performans başarımları değerleri verilmektedir. Doğruluk (0,880) değerine göre iyi bir performans gösteren algoritma için AUC (0,862) değeri de algoritma başarısını desteklerken, f-ölçütü (0,444), kesinlik (0,364) ve duyarlılık (0,571) değerleri oldukça düşüktür.

Tablo 4.12: Genel grup veri seti, NB algoritması performans ölçümleri.

Sınıflandırıcı	Duyarlılık	Kesinlik	f-ölçütü	AUC	Doğruluk
NB	0,571	0,364	0,444	0,862	0,880

4.6 SVM algoritması uygulama sonuçları

Tablo 4.13'te, Normal servis grubu veri seti için SVM algoritmasının, c ($2^{-5}, 2^{10}$) ve γ ($2^{-10}, 2^5$) değer aralıklarına göre en iyi performansı gösteren karşılaştırma matrisi verilmiştir. Karşılaştırma matrisine göre SVM algoritması 51 hasta verisinin 47 (23+24) tanesini doğru olarak sınıflandırmıştır.

Tablo 4.13: Normal servis grubu veri seti, SVM algoritması karşılaştırma matrisi.

Sınıflandırıcı	Pozitif	Negatif	
SVM	Pozitif	23	3
	Negatif	1	24

Tablo 4.14'te, Normal servis grubu veri seti için SVM algoritması için belirlenen parametre değer aralıklarına göre en iyi performans başarımları değerlendirilmiştir. En yüksek Doğruluk (0,920), ($2^1, 2^{-7}$) değerleri kullanılarak elde edilmiştir. AUC (0,967) ve f-ölçütü (0,933) değerleri de sınıflandırma başarısını desteklemektedir.

Tablo 4.14: Normal servis grubu veri seti, SVM algoritması performans ölçümleri.

Sınıflandırıcı	Duyarlılık	Kesinlik	f-ölçütü	AUC	Doğruluk
SVM	0,967	0,917	0,933	0,967	0,920

Tablo 4.15'te, Genel grup veri seti için SVM algoritmasının, c ($2^{-5}, 2^{10}$) ve γ ($2^{-10}, 2^5$) değer aralıklarına göre en iyi performansı gösteren karşılaştırma matrisi verilmiştir. Genel grup veri setinde pozitif ve negatif sınıflardaki veri sayısının dengesiz dağılımı, SMOTE yöntemi kullanılarak, pozitif sınıf için örnek sayısı sentetik olarak artırılarak giderilmiştir. SMOTE yönteminin dâhil edilmesi ile elde edilen karşılaştırma matrisine göre SVM algoritması 416 hasta verisinin 387 (18+369) tanesini doğru olarak sınıflandırmıştır.

Tablo 4.15: Genel grup veri seti, SVM algoritması karşılaştırma matrisi.

Sınıflandırıcı	Pozitif	Negatif
SVM	Pozitif	18
	Negatif	369

Tablo 4.16’da, Genel grup veri seti için SVM algoritması için belirlenen parametre değer aralıklarına göre en iyi performans başarımları değerlendirilmiştir. En yüksek Doğruluk (0,935), $(2^0, 2^{-3})$ değerleri kullanılarak elde edilmiştir. AUC (0,895) değeri de algoritma başarısını desteklerken, f-ölçütü (0.554), kesinlik (0,60) ve duyarlılık (0,514) değerleri oldukça düşüktür.

Tablo 4.16: Genel grup veri seti, SVM algoritması performans ölçümleri.

Sınıflandırıcı	Duyarlılık	Kesinlik	f-ölçütü	AUC	Doğruluk
SVM	0,514	0,600	0,554	0,895	0,935

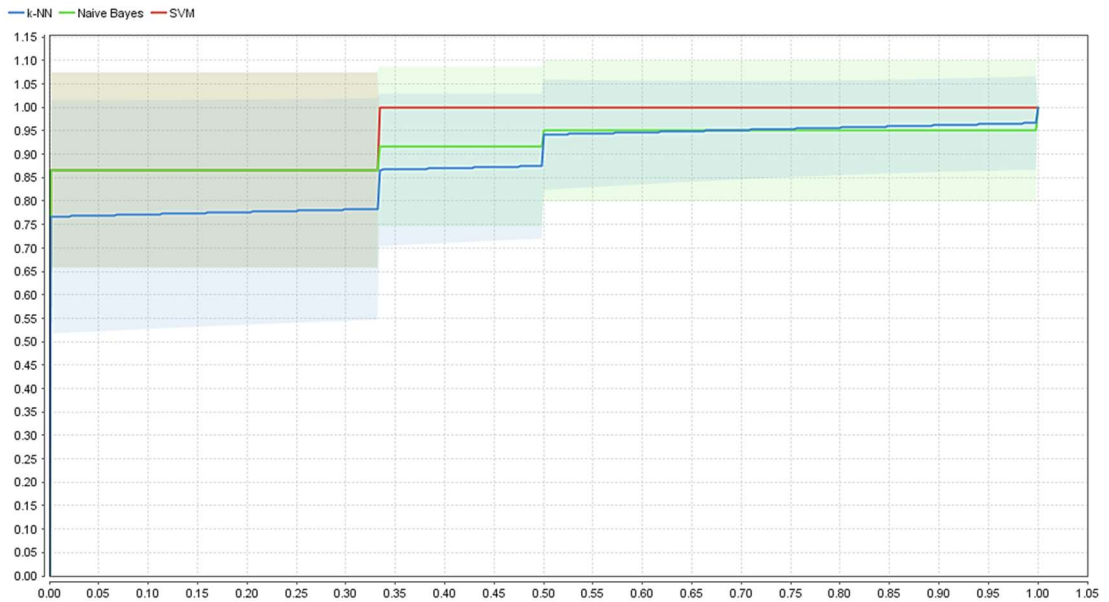
4.7 KNN, NB ve SVM algoritmaları uygulama sonuçlarının karşılaştırılması

Tablo 4.17 incelendiğinde, normal servis grubu için doğruluk (0,920) değerine göre SVM algoritması en iyi performansı göstermiştir. AUC (0,967), f-ölçütü (0.933), kesinlik (0,917) ve duyarlılık (0,967) değerleri de algoritma başarısını desteklemektedir.

Tablo 4.17: Normal servis grubu veri seti, algoritma performans ölçümlerinin karşılaştırması.

Sınıflandırıcı	Duyarlılık	Kesinlik	f-ölçütü	AUC	Doğruluk
KNN	0,967	0,725	0,811	0,917	0,767
NB	0,917	0,880	0,898	0,917	0,903
SVM	0,967	0,917	0,933	0,967	0,920

Şekil 4.13'de belirtildiği üzere; her üç algoritmanın ROC Eğrisi karşılaştırıldığında normal servis grubu veri seti için, sınıflandırma performansı açısından SVM algoritmasının daha başarılı olduğu görülmektedir.



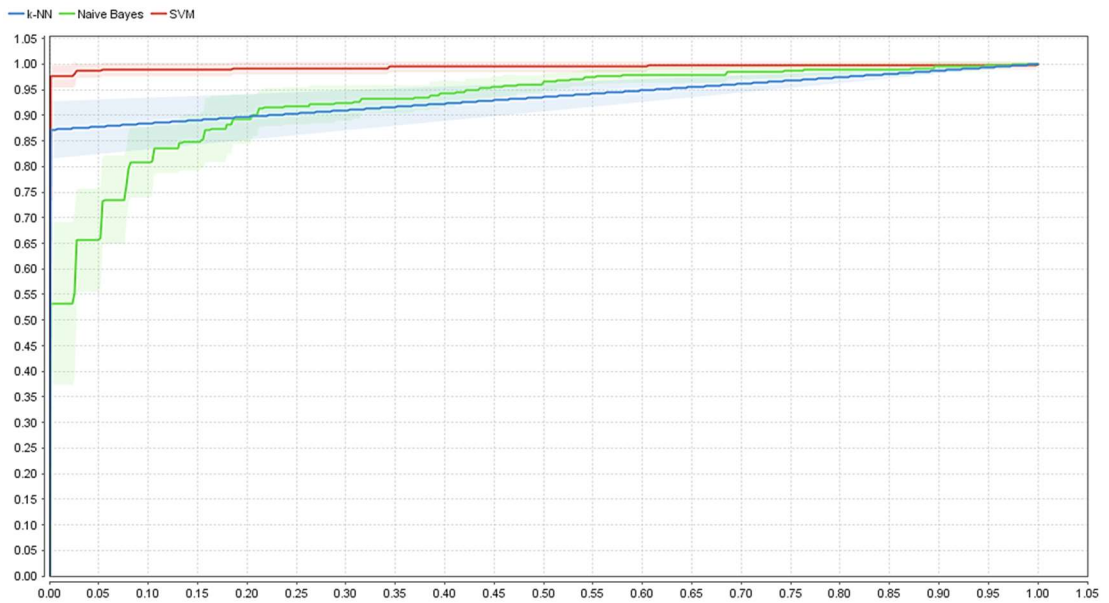
Şekil 4.13: Normal servis grubu veri seti için algoritmaların ROC eğrileri.

Tablo 4.18 incelendiğinde, genel grup için doğruluk (0.935) değerine göre SVM algoritması en iyi performansı göstermiştir. AUC (0,895) değeri algoritma başarısını desteklerken f-ölçütü (0.554), kesinlik (0,60) ve duyarlılık (0,514) değerleri oldukça düşüktür.

Tablo 4.18: Genel grup veri seti, algoritma performans ölçümlerinin karşılaştırması.

Sınıflandırıcı	Duyarlılık	Kesinlik	f-ölçütü	AUC	Doğruluk
KNN	0,525	0,725	0,387	0,500	0,861
NB	0,571	0,364	0,444	0,862	0,880
SVM	0,514	0,600	0,554	0,895	0,935

Şekil 4.14'te belirtildiği üzere; her üç algoritmanın ROC Eğrisi karşılaştırıldığında genel grup veri seti için, sınıflandırma performansı açısından SVM algoritmasının daha başarılı olduğu görülmektedir.



Şekil 4.14: Genel grup veri seti için algoritmaların ROC Eğrileri.

5. SONUÇ VE ÖNERİLER

İlk olarak Aralık 2019'da Çin'in Wuhan kentinde ortaya çıkan yeni koronavirüs türü SARS-CoV-2, ortaya çıktığı ilk zamanlardan bu yana yayılımı ve gelişimiyle günlük yaşamda, halk sağlığı ve küresel ekonomi üzerinde tehdit oluşturmaya devam etmektedir. Günümüzde farklı amaçlara yönelik tedaviler ve aşılar mevcut olmakla birlikte etkili bir tedavinin var olmaması nedeniyle bulaşmayı durdurmak ve salgının yayılımı azaltmak için uygulanabilecek en etkili strateji enfekte bireylerin izolasyonunu sağlamaktır. Bu çalışma kapsamında, COVID-19 salgınıyla mücadele sürecinde klinik uygulamalarda evrensel olarak sıklıkla kullanılan ve çıktıları hızlı tetkik edilebilen tam kan sayımı sonuçları kullanılarak SARS-CoV-2 pozitif hastalar makine öğrenme algoritmaları yardımıyla tahmin edilmeye çalışılmıştır. Normal servis grubu ve genel grup olmak üzere iki farklı gruba ayrılan veri seti, KNN, NB ve SVM makine öğrenme algoritmaları kullanılarak, üç farklı model geliştirilerek incelenmiştir. Oluşturulan modeller arasında en iyi performansı normal servis grubu için % 92 ve genel grup için % 93,5 doğrulukla SVM göstermiştir.

Veri güvenliği, mahremiyet ve etik değerler gibi nedenlerle COVID-19 için kamuya açık veri miktarı oldukça azdır. Model çıktıları değerlendirildiğinde gerçek yaşamdan elde edilecek daha kapsamlı veri kümeleri ile eğitilecek makine öğrenmesi modellerinin, sağlık hizmeti sağlayıcılarına tam kan sayımı çıktılarını kullanarak hızlı ve ucuz bir tarama yaparak SARS-CoV-2 pozitif hastaları erken aşamalarda teşhis imkânı sunacağı öngörülmektedir.

6. KAYNAKLAR

Ahamad, M. M., Aktar, S., Rashed-Al-Mahfuz, M., Uddin, S., Liò, P., Xu, H., Summers, M. A., Quinn, J. M. W. and Moni, M. A., "A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients", *Expert Systems with Applications*, 160, 113661, (2020).

Ahmad I. and Muhammad Asad S., "Predictions of Coronavirus COVID-19 Distinct Cases in Pakistan Through an Artificial Neural Network", *Epidemiology and Infection*, 148 (222), 1-10, (2020).

Akpınar, H., "Bulaşıcı Hastalıkların Yayılımının Tahmininde Deterministik Modellerin Kullanılması", *Öneri Dergisi*, 10 (38), 97, (2012).

Akpınar, H., *DATA -Veri Madenciliği Veri Analizi*, Papatya Bilim Yayınevi, (2017).

Alpaydın, E., *Introduction to Machine Learning*, MIT Press, (2004).

Ayhan, S. ve Erdoğan, Ş., "Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi", *Eskişehir Osmangazi Üniversitesi İBFF Dergisi*, 9 (1), 175-198, (2014).

Aslan, R., "Tarihten Günümüze Epidemiler: Pandemiler ve Covid-19", *Göller Bölgesi Aylık Ekonomi ve Kültür Dergisi*, 8 (85), 37, (2020).

Balaban, M. E., ve Kartal, E., *Veri Madenciliği ve Makine Öğrenmesi Temel Algoritmaları ve R Dili ile Uygulamaları (Birinci Baskı)*, İstanbul: Çağlayan Kitabevi, (2015).

Bayat, A. H., *Tıp Tarihi*, İstanbul: Zeytinburnu Belediyesi Kültür Yayınları, (2016).

Beyaz, Z., "Kapitalizmin Pandemi Sömürsü Bağlamında Salgın Hastalıklara Bir Bakış", *Çalışma ve Toplum Dergisi*, 68 (1), 181-204, (2021).

Blagus, R. and Lusa, L., "SMOTE for high-dimensional class-imbalanced data", *BMC Bioinformatics*, 14 (106), (2013).

Byrne, J. P., *Encyclopedia of Pestilence, Pandemics, and Plagues*, London: Greenwood Press, (2008).

Choonara, J., "Socialism in a Time of Pandemics", *International Socialism Journal*, 2 (166), 3, (2020).

Chakrabarti, S., Cox, E., Frank, E., Güting, R. H., Han, J., Jiang, X., Kamber, M., Lightstone, S. S., Nadeau, T. P., Neapolitan, R. E., Pyle, D., Refaat, M., Schneider, M., Teorey, T. J. and Witten, I. H., *Data Mining: Know It All (First Edition)*, United States: Morgan Kaufmann Publishers, (2009).

Çarkacı, N., “Derin Öğrenme Uygulamalarında Hiper Parametre Seçim Yöntemleri”, (25 Nisan 2021), <https://medium.com/deep-learning-turkiye/derin-ogrenme-uygulamalarinda-model-dogrulama-ve-hiper-parametre-secim-yontemleri-823812d95f3>, (2018).

Çınar, A., C., “k-kez çapraz doğrulama yöntemi”, (25 Mayıs 2021), <https://www.ahmetcevahircinar.com.tr/2017/03/25/k-kez-capraz-dogrulama-yontemi/>, (2017).

Data4u, E., “Hospital Israelita Albert Einstein, Sao Paulo, Brazil, Diagnosis of Covid-19 and its clinical spectrum”, (12 Şubat 2021), <https://www.kaggle.com/einsteindata4u/covid19>, (2020).

Demirel, M., “Eskiçağ Ön Asya Toplumlarında Görülen Salgın Hastalıklar”, Yüksek Lisans Tezi, *Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Tarih Anabilim Dalı*, Ankara, 38, (2019).

Ertorsun, A. D., Bağ, B., Uzar, G. ve Turanoğlu, M. A., ROC (Receiver Operating Characteristic) Eğrisi Yöntemi ile Tanı Testlerinin Performanslarının Değerlendirilmesi, *XIII. Öğrenci Sempozyumu*, Ankara, (2009).

Fawcett, T., “An introduction to ROC analysis”, *Pattern recognition letters*, 27 (8), 861–874, (2006).

Han, J., Kamber, M. and Pei, J., *Data Mining Concepts and Techniques (Third Edition)*, United States: Morgan Kaufmann Publishers, (2012).

Huizen, J., “Why is it called coronavirus?”, (29 Mayıs 2021), <https://www.medicalnewstoday.com/articles/why-is-it-called-coronavirus>, (2020).

Jaiswal, S., “K-Nearest Neighbor(KNN) Algorithm for Machine Learning”, (30 Mayıs 2021), <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>, (2018).

Kara A., “Domuz Gribi (Pandemik İnfluenza 2009, İnfluenza 2009 H1N1)”, *Çocuk Sağlığı ve Hastalıkları Dergisi*, 53 (1): 42-58, (2010).

Karabatak, M., “A new classifier for breast cancer detection based on Naïve Bayesian”, *Measurement*, 72, 32-36, (2015).

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

Khan, M., Ding, Q. and Perrizo, W., “K-Nearest Neighbor Classification on Spatial Data Streams Using P-Trees”, (eds: Chen MS., Yu P. S., Liu B.), *Advances in Knowledge Discovery and Data Mining. PAKDD 2002. Lecture Notes in Computer Science*, (2336), Springer, Berlin, Heidelberg, (2002).

Kılıçbay, M. A., “Çevrenin Çerçevesi”, *Türkiye Günlüğü*, 3, 35-39, (1989).

Kowalczyk, A., *Support Vektor Machines Succinctly*, Morrisville: Syncfusion, (2017).

Laçinel Gürlevik, S., "Coronaviruses and New Coronavirus SARS-CoV-2", *Journal Of Pediatric Infection*, 14 (1), 46–48, (2020).

Lalmuanawma, S., Hussain, J. And Chhakchhuak, L., “Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review”, *Chaos, Solitons & Fractals*, 139, 110059, (2020).

Lappin T. R., Maxwell A. P. and Jhonstone P. G., “EPO’s Alter Ego: Erythropoietin has multiple actions”, *Stem Cells*, 20 (6), 485-492, (2002).

LePan, N., Visualizing the History of Pandemics, (30 Nisan 2021), <https://www.visualcapitalist.com/history-of-pandemics-deadliest/>, (2020).

Lippi, G., Simundic A. M. and Plebani, M., “Potential preanalytical and analytical vulnerabilities in the laboratory diagnosis of coronavirus disease 2019 (COVID-19)”, *Clinical Chemistry and Laboratory Medicine (CCLM)*, 58 (7), 1070–1076, (2020).

Metlek, S. ve Kayaalp, K., *Makine Öğrenmesinde, Teoriden Örnek Matlab Uygulamalarına Kadar Destek Vektör Makineleri*, Ankara: İksad Yayınevi, (2020).

Moreira, C., “Learning to rank academic experts,” Doktora tezi, *Technical University of Lisbon*, (2011).

Murphy, K. P., *Machine Learning: A Probabilistic Perspective*, MA: MIT Press, Cambridge, (2012).

Öztek, Z., “Pandemi Mücadelesi ve Yan Kazanımlar”, *Sağlık ve Toplum Dergisi*, 30 (Özel Sayı) , 6, (2020).

Öztürk, T., Talo, M., Yıldırım E. A., Baloglu, U. B., Yıldırım, Ö. ve Acharya, U. R., “Automated detection of COVID-19 cases using deep neural networks with X-ray images”, *Computers in Biology and Medicine* , 121, 103792, (2020).

Pelin, C. ve Polat, O., “Paleopatoloji”, *Ankara Patoloji Bülteni*, 14 (1), 80-83, (1997).

Ponting, C., *Dünyanın Yeşil Tarihi - Çevre ve Büyük Uygarlıkların Çöküşü*, (Çev: Ayşe Başcı Sander), İstanbul: Sabancı Üniversitesi, 197, (2000).

Sarı, N., *Tıp Tarihi ve Etiği Ders Kitabı*, İstanbul: Cerrahpaşa Tıp Fakültesi Yayınları, (2007).

Sarmış, A., “Laboratory Diagnosis of COVID-19: Review of Literature”, *Medical Research Reports*, 3 (1), 3-10, (2020).

Seker, S. E., “Rapid Miner ile K-NN uygulaması”, (24 Nisan 2021), <http://bilgisayarkavramlari.com/2016/05/22/rapid-miner-ile-k-nn-uygulamasi/>, (2016).

Shield, C., "Coronavirus: From bats to pangolins, how do viruses reach us?", (24 Mayıs 2021), <https://www.dw.com/en/coronavirus-from-bats-to-pangolins-how-do-viruses-reach-us/a-52291570>, (2020).

Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, N., Kerwan, A., Al-Jabir, A., Iosifidis, C. and Agha, R., “World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)”, *International Journal of Surgery*, 76, 71-76, (2020).

Stalfort, J., “Hyperparameter tuning using Grid Search and Random Search: A Conceptual Guide”, (25 Nisan 2021), <https://medium.com/@jackstalfort/hyperparameter-tuning-using-grid-search-and-random-search-f8750a464b35>, (2019).

Stowell, S. and Guarner, J., “Role of serology in the coronavirus disease 2019 pandemic”, *Clinical Infectious Diseases*, 71 (8), (2020).

Taşcı, E. ve Onan, A., “K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi”, *Akademik Bilişim 2016*, Türkiye, (2016).

T.C. Sağlık Bakanlığı, COVID-19 (SARS-CoV2 Enfeksiyonu) Genel Bilgiler, Epidemiyoloji ve Tanı, Bilimsel Danışma Kurulu Çalışması, (30 Nisan 2021), <https://covid19.saglik.gov.tr/Eklenti/39551/0/covid-19rehberigenelbilgilerepidemiolojivetanipdf.pdf>, (2020).

Turan, A. ve Çelikyay Hamza, H., “Türkiye’de KOVID-19 ile Mücadele: Politikalar ve Aktörler”, *Uluslararası Yönetim Akademisi Dergisi*, 3 (1), 3, (2020).

UNAIDS, Global AIDS Update 2020, (30 Nisan 2021), <https://www.unaids.org/en/resources/documents/2020/global-aids-report>, (2020).

Vapnik, V.N., *The Nature of Statistical Learning Theory*, New York: Springer-Verlag, (1995).

Vapnik, V.N., *The Nature of Statistical Learning Theory*, 2. Baskı, New York: Springer-Verlag, (2000).

Yıldız, F., “19. Yüzyıl’da Anadolu’da Salgın Hastalıklar (Veba, Kolera, Çiçek, Sıtma) ve Salgın Hastalıklarla Mücadele Yöntemleri”, Yüksek Lisans Tezi, *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü, Tarih Anabilim Dalı*, Denizli, 1, (2014).

Zhang, Z., “Introduction to machine learning: K-nearest neighbors”, *Annals of Translational Medicine*, 4 (11), 218–218, (2016).

Zheng Y., Zhu Y., Ji M., Wang R., Liu X., Zhang M., Liu J., Zhang X., Qin C. H., Fang L. and Ma S., “A Learning-Based Model to Evaluate Hospitalization Priority in COVID-19 Pandemics”, *Patterns*, 1 (6), 1-21, (2020).

Wang, P., Zheng X., Li, J. and Zhu B., “Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics”, *Chaos, Solitons and Fractals*, 139, 110058, (2020).

WHO, “ACT-Accelerator one year on”, (25 Nisan 2021), <https://www.who.int/news/item/23-04-2021-act-accelerator-one-year-on>, (2021).

WHO, “Basic Documents”, (30 Nisan 2021), https://apps.who.int/gb/bd/pdf_files/BD_49th-en.pdf, (2020).

WHO, “Coronavirus disease (COVID-19)”, (30 Nisan 2021), <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/coronavirus-disease-covid-19>, (2020).

WHO, “Naming the coronavirus disease (COVID-19) and the virus that causes it”, (29 Mayıs 2021), [https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-\(covid-2019\)-and-the-virus-that-causes-it](https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it), (2020).

WHO, “R&D Blueprint and COVID-19”, (30 Nisan 2021), <https://www.who.int/teams/blueprint/covid-19>, (2021).

WHO, “WHO Coronavirus (COVID-19) Dashboard”, (1 Mayıs 2021), <https://covid19.who.int/>, (2021).