



**KTO KARATAY ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
ELEKTRİK-ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI
ELEKTRİK VE BİLGİSAYAR MÜHENDİSLİĞİ BİLİM DALI
TEZLİ YÜKSEK LİSANS PROGRAMI**

**COVID-19 SALGINININ ETKİLERİNİN VERİ BİLİMİ YÖNTEMLERİYLE
İNCELENMESİ**

Yusuf Behzat KAVAK

Yüksek Lisans Tezi

**KONYA
Aralık 2021**

COVID-19 SALGINININ ETKİLERİNİN VERİ BİLİMİ YÖNTEMLERİYLE
İNCELENMESİ

Yusuf Behzat KAVAK

KTO Karatay Üniversitesi
Lisansüstü Eğitim Enstitüsü
Elektrik-Elektronik Mühendisliği Anabilim Dalı
Elektrik ve Bilgisayar Mühendisliği Bilim Dalı
Tezli Yüksek Lisans Programı

Yüksek Lisans Tezi

Tez Danışmanı: Dr. Öğr. Üyesi Semih YUMUŞAK

Konya
Aralık 2021

BİLDİRİM

Enstitü tarafından onaylanan Yüksek Lisans/Doktora tezimin tamamını veya herhangi bir kısmını basılı veya dijital biçimde arşivleme ve aşağıda belirtilen koşullar dahilinde erişime açma iznini KTO Karatay Üniversitesine verdiğimi bildiririm. Bu izinle, Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak ve gelecekteki çalışmalar (makale, kitap, lisans, patent vb.) için tezimin tamamının veya bir bölümünün kullanım hakları yalnızca bana ait olacaktır.

Tezimin bütünüyle kendi çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Telif hakkı bulunan ve sahiplerinden yazılı izinle kullanılması zorunlu olan kaynakları, yazılı izin alarak kullandığımı ve istenildiğinde izinlerin suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayımlanan “Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge” kapsamında, tezim, aşağıda belirtilen koşullar haricince, YÖK Ulusal Tez Merkezi ve KTO Karatay Üniversitesi Açık Erişim Sisteminde erişime açılır.

Enstitü / Fakülte Yönetim Kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.¹

Enstitü / Fakülte Yönetim Kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir.²

Tezimle ilgili gizlilik kararı verilmiştir.³⁴

Aralık 2021

Yusuf Behzat KAVAK

¹ MADDE 6(1) Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.

² MADDE 6(2) Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.

³ MADDE 7(1) Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.

⁴ MADDE 7(2) Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir.

ETİK BEYAN

KTO Karatay Üniversitesi Lisansüstü Eğitim Enstitüsü Tez Hazırlama ve Yazım Kurallarına uygun olarak Dr. Öğr. Üyesi Semih YUMUŞAK danışmanlığında tarafımdan üretilen bu tez/proje çalışmasında; sunduğum tüm veri, enformasyon, bilgi ve belgeleri bilimsel etik kuralları çerçevesinde elde ettiğimi, tüm değerlendirme, analiz, bulgu ve sonuçları bilimsel usullere uygun olarak sunduğumu, tez/proje çalışmasında yararlandığım kaynakların tümüne bilimsel normlara uygun biçimde atıfta bulunarak kaynak gösterdiğimi, tezimin/projemin kaynak gösterilen durumlar dışında özgün olduğunu bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

03 Aralık 2021

Yusuf Behzat KAVAK

Eđitim iin gittiđi İngiltere’de vefat eden arkadařım Ahmet Fatih IRKLI’nın anısına

TEŐEKKÖR

Beni her zaman destekleyip vizyonumu belirlemede büyük emekleri olan annem Zeynep ve babam Bayram'a teŐekkÖrlerimi sunarım. Beni yüksek lisans yapmaya ve bu alanı etkin kullanmak için beni cesaretlendiren danışmanım Dr. Öđr. Üyesi Semih YUMUŐAK'a saygı ve minnetlerimi sunarım.

Aralık, 2021

Yusuf Behzat Kavak

ÖZET

Yusuf Behzat KAVAK

Covid-19 Salgınının Etkilerinin Veri Bilimi Yöntemleriyle İncelenmesi

Yüksek Lisans Tezi

Konya, 2021

Covid-19 salgını, toplum üzerinde içinde ölümle sonuçlanabilen, ağır hastalıklar bırakan, toplumun sosyal çevresini doğrudan etkileyen dünya tarihinde şimdiye kadar görülmemiş yayılıma ulaşan toplumsal bir problem oluşturmaktadır. 2019 yılının aralık ayında Çin'in Hubei eyaletinde yayılmaya başlayan Covid-19 salgını tüm dünyaya çok kısa bir süre içinde yayılmıştır. Çin'deki salgının bastırılması için sokağa çıkma yasağı uygulanmaya, maske üretimleri ve bu salgın için gerekli makinalar üretilmeye başlanmıştır. İlk ortaya çıkmasından 3 ay gibi kısa bir süre içinde hemen hemen bütün ülkelerde vaka sayıları raporlanmaya başlanmıştır. Bu raporlar içindeki vaka sayıları, ölüm sayıları, ölüm oranları, aşılama sayıları gibi farklı verilerin pandeminin etkisini azaltmak, ilaçların verimliliğini ölçmek, aşı için gereken verilerin bulunabilmesi için toplanmıştır. Ülkeler, üniversiteler, araştırma kurumlarının ve hastanelerin iş birliği ile toplanan veriler bu tez çalışmasında kullanılmıştır. Veri bilimi ve makine öğrenmesi algoritmaları ile aşılama verimliliğinin gözlemlenmesi, ekonomik durumları farklılık gösteren ülkelerin içinde bulunduğu durum ve insan yoğunluğunun pandemideki etkisi gözlemlenmiştir. Aşı, vaka ve ölüm oranları arasındaki ilişki farklı algoritmalar ile modellenmiş ve bunun sonucunda ölüm/vaka oranlarının aşılama karşısında düşüş gösterdiği gözlemlenmiştir.

Anahtar Kelimeler

Covid-19, Veri Bilimi, Makina Öğrenmesi

ABSTRACT

Yusuf Behzat KAVAK

Examining Impacts of the Covid-19 Pandemic with Data Science Algorithms

Master's Thesis

Konya, 2021

Covid-19 is a pandemic that has reached an unprecedented spread in the world, which results in death, leave severe diseases on the society, and directly affect the social environment of the society. The Covid-19 epidemic, which started to spread in China's Hubei province in December 2019, spread all over the world in a very short period of time. In order to suppress the events in China, a curfew began to be implemented, mask production and the necessary machines for this epidemic began to be produced. Within a short period of 3 months after its first appearance, the number of cases started to be reported in almost all countries. Different data such as the number of cases, deaths, death rates, vaccination numbers in these reports were collected in order to reduce the effect of the pandemic, to measure the efficiency of drugs, and to find the data required for the vaccine. In this thesis, the analysis of the data collected with the cooperation of countries, universities, research institutions and hospitals with data science and machine learning algorithms, observation of the efficiency of vaccines, the situation of economically poor countries and the effect of human density on the pandemic have been observed. The relation between vaccine, case and death rates was modeled with different machine learning algorithms and as a result, it is observed that death/case rates decreased with vaccination.

Keywords

Covid-19, Data Science, Machine Learning

İÇİNDEKİLER

KABUL VE ONAY	i
BİLDİRİM	ii
ETİK BEYAN.....	iii
TEŞEKKÜR.....	v
ÖZET.....	vi
ABSTRACT.....	vii
İÇİNDEKİLER	viii
ŞEKİLLER DİZİNİ.....	x
SİMGELER DİZİNİ.....	xiii
KISALTMALAR DİZİNİ.....	xiv
1. GİRİŞ	1
2. KAYNAK TARAMASI	3
2.1. Covid-19 Virüsü ve İlgili Çalışmalar	3
2.1.1. Virüsün Tanımı.....	3
2.1.2. Hastalık Teşhisinin Yapılması.....	3
2.1.3. Laboratuvar.....	4
2.2. Makine Öğrenmesi ve Derin Öğrenmenin Bu Süreçte Kullanımı	4
2.2.1. Matematiksel Modeller ile Vaka Sayılarının Tahmini Üzerine Yapılan Çalışmalar	5
2.2.2. DNA ve RNA Modelleri Üzerine Yapılan Çalışmalar	7
2.2.3. Akciğer Tomografisi ve X-Ray Üzerinden Bilgisayarlı Görüntüleme Üzerine Çalışmalar	7
2.2.4. Aşıyla İlgili Çalışmalar.....	8
3. MAKİNE ÖĞRENMESİ YÖNTEMİ ve KAVRAMLARI	9
3.1. İstatistiksel Yöntemler.....	9
3.1.1. Popülasyon ve Örneklem Kümesi	9
3.1.2. Normal Dağılım	10
3.1.3. Merkezi Eğilim	10
3.1.4. Varyans ve Standart Sapma	10
3.1.5. Kovaryans ve Korelasyon.....	11

3.1.6. Merkezi Limit Teoremi.....	11
3.1.7. P-deđeri.....	11
3.1.8. Rastgele Deđişkenlerin Beklenen Deđeri	12
3.1.9. Koşullu Olasılık.....	12
3.1.10. Bayes Teoremi	12
3.2. Makine Öğrenmesi Yöntemleri	13
3.2.1. Lineer Regresyon ile Modelleme.....	13
4. VERİNİN TANIMI, TAHMİN YÖNTEMLERİNİN UYGULANMASI, VAKA ve ÖLÜM ORANLARININ ve AŞININ VERİMLİLİĞİNİN İNCELENMESİ	15
4.1. Verinin Tanımı ve Kaynağının İncelenmesi	15
4.2. Veri Bilimi ve Makine Öğrenmesi Algoritmalarının Uygulanması.....	15
5. SONUÇ	42
KAYNAKLAR	44
ÖZGEÇMİŞ	47

ŞEKİLLER DİZİNİ

Şekil 1. Dünya Toplam Vaka Sayılarının Bayes Regresyon Yöntemi ile Modellenmesi	16
Şekil 2. Dünya Toplam Vaka Sayılarının Polinomsal Regresyon Yöntemi ile Modellenmesi.....	17
Şekil 3. Dünya Toplam Vaka Sayılarının SVM Regresyon Yöntemi ile Modellenmesi	17
Şekil 4. Dünya Toplam Vaka Sayılarının Karar Ağaçları Regresyon Yöntemi ile Modellenmesi.....	18
Şekil 5. Dünya Toplam Vaka Sayılarının ElasticNet Regresyon Yöntemi ile Modellenmesi.....	18
Şekil 6. Dünya Toplam Vaka Sayılarının Lasso Regresyon Yöntemi ile Modellenmesi	19
Şekil 7. Dünya Toplam Vaka Sayılarının Random Forest Regresyon Yöntemi ile Modellenmesi.....	19
Şekil 8 Dünya Toplam Vaka Sayılarının Ridge Regresyon Yöntemi ile Modellenmesi	20
Şekil 9 Dünya Toplam Vaka Sayılarının XGBoost Regresyon Yöntemi ile Modellenmesi.....	20
Şekil 10 Günlük Artış Vaka	21
Şekil 11 Günlük Artış Ölüm	22
Şekil 12 Dünya Logaritmik Vaka Artışı	23
Şekil 13 Dünya Logaritmik Ölüm Sayısı Artışı.....	23
Şekil 14 Dünya Tanılı Hasta Sayısı	24
Şekil 15 Dünya Ölen Hasta Sayısı	25
Şekil 16 Dünya Zamana Göre Ölüm Oranı.....	25
Şekil 17 Toplam Aşılama Sayısı (4 Ağustos 2021)	26
Şekil 18 Nüfusa Göre Aşılama Yoğunluğu (4 Ağustos 2021).....	27
Şekil 19Avusturya Vaka Sayısı ve Günlük Artışı.....	28
Şekil 20 Belçika Vaka Sayısı ve Günlük Artışı	29
Şekil 21 Finlandiya Vaka Sayısı ve Günlük Artışı	30
Şekil 22 Almanya Vaka Sayısı ve Günlük Artışı.....	31
Şekil 23 İtalya Vaka Sayısı ve Günlük Artışı	32
Şekil 24 Hollanda Vaka Sayısı ve Günlük Artışı.....	33
Şekil 25 İsviçre Vaka Sayısı ve Günlük Artışı	34
Şekil 26 Türkiye Vaka Sayısı ve Günlük Artışı.....	35
Şekil 27 Birleşik Krallık Vaka Sayısı ve Günlük Artışı	36

Şekil 28. SVR Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği.....	37
Şekil 29. ElasticNet Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği.....	38
Şekil 30. Lasso Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği.....	38
Şekil 31. Linear Regresyon Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği.....	39
Şekil 32. Random Forest Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği.....	39
Şekil 33. Decision Tree Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği.....	40
Şekil 34. XGBoost Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği.....	40

TABLolar DİZİNİ

Tablo 1. Farklı Makine Öğrenmesi Modellerinin Günlük Vaka Tahminlerinin Hata Oranları	21
Tablo 2. Farklı Makine Öğrenmesi Modellerinin Aşı ve Vaka Girdileri ile Ölüm Oranı Tahminlerinin Hata Oranları	41

SİMGELER DİZİNİ

Simge	Açıklama
Ψ	Psi fonksiyonu
Φ_{nm}^{ab}	Kuvvet parametreleri matrisi

KISALTMALAR DİZİNİ

Kısaltma	Açıklama
dPbD	Parçalı Mahremiyet Tasarımı (Differential Privacy by Design)
PCR	Polimeraz Zincir Reaksiyonu
SARS	Ağır Akut Solunum Yolu Yetmezliği Sendromu
SIR	Susceptible-Infectious-Recovered
SVM	Destek Vektör Makinaları
WHO	Dünya Sağlık Örgütü
MERS	Middle East Respiratory Syndrome

1. GİRİŞ

Dünya üzerinde şimdiye kadar kara veba, kolera, ispanyol gribi, çiçek hastalığı, domuz gribi, kuş gribi, Hiv/Aids gibi virüs pandemiler olmuş bunların bazıları aşilar yoluyla bastırılmış bazıları zamanla son bulmuş diğlerleri de günlük yaşamımızdaki varlığını devam ettirmektedir.

Son olarak, 2019 yılının aralık ayında Çin'in Wuhan şehrinde ortaya çıkan ve yayılmaya başlayan COVID-19 (SARS-Cov-2) çok kısa bir süre içinde tüm dünyaya yayılmaya başlamıştır. Dünya Sağlık Örgütü (WHO) tarafından tüm dünyada pandemi ilan edilmiştir. Çin'de büyük bir vaka ve ölüm oranlarının ardından bunu ilk başta İtalya ve İran izlemiştir.

Pandemi süreci ilk başlarda maske ile yayılımın azaltılmaya çalışıldığı, PCR testlerinin elde edilmeye çalışıldığı, hastanelerde yatacak hastalar için solunum cihazları üretildikten sonra aşı çalışmalarının tamamlandığı ve insanlar üzerinde denemelerinin sürdüğü bir süreç olarak devam etmektedir.

Çin hastalık kontrol merkezini hastalığa yakalananların yüksek ateş, kuru öksürük, nefes darlığı, yorgunluk, tat ve koku kaybı, ishal ve damar sıkışıklığı gibi belirtiler gösterdiğini ama bazılarının da hiçbir belirti göstermeden atlattığını belirtiyor.

Her gün artan vaka ve ölüm oranları insan hayatını ve yaşamı doğrudan etkilemektedir. İnsanların hayatını bu denli etkileyen bir salgına karşı tüm dünyada sokağa çıkma yasakları, yurtdışı uçak giriş-çıkışlarının yasaklanması, çalışanların uzaktan çalışmaya geçmesi, restoran ve kafeler gibi insanların beraber bulunduğu sosyal ortamların kapatılması gibi sert kurallar uygulanmış ve uygulanmaya devam etmektedir. Bulaşıcı bir hastalığın hangi şartlar altında nerede çıkacağını bilebilmek fazlasıyla zorlu bir süreçtir. Ancak bölgesel verilerin düzenli toplanması, matematiksel modellerin doğru ve efektif bir şekilde kullanımı bu durumu kolaylaştırabilir.

Hastanelerde tespit edilen vaka sayıları her gün kayıt edilmekte, bu veriler ileri analizler için kullanılmaktadır. Bu kayıtlar dünya sağlık örgütü ve John Hopkins Üniversitesi gibi kurumlar tarafından toplanıp açık kaynak olarak yayınlanmaktadır (Bhardwaj, R., & Bangia, A., 2020). Bu veriler üzerinde makine öğrenmesi ve derin öğrenme

algoritmalarıyla ileriye yönelik alınacak kısıtlama kararları veya toplumu etkileyecek yöntemler üzerinde çalışılmıştır.

Covid-19 olarak ifade edilen virüsün aynı zamanda SARS-Cov-2 olarak anılmasının nedeni 2002 yılında ilk defa yayılan SARS virüsü ile %79,6 oranında benzerlik göstermesidir (S. Wang, Sun, Xu, Pei, & Lai, 2020).

Ülkemizde ilk başta yurt dışından gelen yolcuların termal kameralar ile kontrol edilmesi, yurt dışından dönenlerin belirli bir süre karantinaya alınması, havayollarının yurt dışı trafiğine kapatılması, tüm büyükşehirlerde hafta sonları karantina süreci, en sonunda evde karantina olmak üzere salgın kontrol edilmeye çalışılmaktadır.

Bu tez kapsamında şu ana kadar COVID-19 ve veri bilimi ile ilgili çalışmalar araştırılmış ve vaka, ölüm ve aşılama sayılarının ülkelerdeki durumu sistemli veriler üzerinden incelenip ilişkiler araştırılmıştır.

2. KAYNAK TARAMASI

2.1. Covid-19 Virüsü ve İlgili Çalışmalar

2.1.1. Virüsün Tanımı

Virüsler her ne kadar biyoloji alanında inceleniyor olsa da cansızdırlar. Kendilerini çevreleyen bir hücre zarı ve metabolizmaları olmadığı için cansızlar sınıfında anılmaktadır. Çoğalmak ve enerji sağlamak amacıyla bir canlı hücreye ihtiyaç duymaktadırlar. Bu hücreye konak adı verilmektedir.

Canlılar için bağışıklık sistemi virüslerle savaşan bir savaşçıdır. Hücrelerde var olabilmesi ve varlığını sürdürebilmesi bağışıklık sisteminin durumuyla doğrudan bir bağlantısı vardır. Hava yoluyla, doğrudan veya dolaylı yollarla, damlacıkla, yiyeceklerin üzerine vücut içerisine girerek, sivrisinek, kene, mantar gibi yollarla canlılara bulaşır. Aşıyla, hijyen kurallarına uyarak, dengeli ve düzenli beslenmeyle (bu sayede bağışıklık sistemi güçlü kalabilir), toplum içinde maske takmak gibi korunma yolları vardır.

2.1.2. Hastalık Teşhisinin Yapılması

Çin'in Hubei eyaletinin Wuhan şehrinde ortaya çıkan ve tüm dünyada yayılmaya başlayan, COVID-19 (SARS-Cov-19) çoğunlukla solunum yollarında açtığı hasarlar yüzünden Ağır Akut Solunum Yolu Yetmezliği Sendromu (SARS) olarak adlandırılmaktadır. 30 Ocak 2020'de Dünya Sağlık Örgütü'nün (WHO) acil olarak toplanmasının ardından pandemi ilan edilmiştir (Velavan & Meyer, 2020). Covid-19 2002 yılında Çin'in Guandong eyaletinde ortaya çıkan SARS ve 2012 yılında Ortadoğu'da ortaya çıkan MERS (Middle East Respiratory Syndrome) ile benzer özellikler göstermektedir (Fauci, Lane, & Redfield, 2020).

İlk başlarda zatürreye benzer ama bilinmeyen bir hastalık olarak gözlenen ama sonrasında benzer nedenlerle gelen hastaların ortak özelliklerinden yola çıkarak tanı konulmaya başlanmıştır. Bu özellikler vücut sıcaklığının 38°C ve üzerinde seyretmesi, düşük veya yüksek beyaz hücre sayımı veya düşük akyuvar sayısı, virüsün kapılmasından sonraki 3-5 gün arası hiçbir belirti göstermemesi fark edilmiştir. Bu hastalığın teşhisini yapan ekibin 425 benzer hastanın incelenmesi sonucu hastaların büyük bir çoğunluğunun (%55) 1

Ocak 2020'den önce Huanan Balık Pazarıyla (Huanan Seafood Wholesale Market) bağlantısı olduğu ortaya çıkarılmıştır (Li et al., 2020).

Wuhan'ı ziyaret eden veya ziyaret edenlerle bağlantısı olan herkesin potansiyel hasta olabileceği düşünülmüş ve 14 günlük karantina süreci uygulanmaya başlanmıştır. Solunum yoluyla ilgili olarak hastaneye başvuranlarda 3 farklı test uygulanmış. Bu testler tamamen izole ortamda gerçekleşmiştir. Test sonuçlarına göre bir şekilde hastalıkla ilişki kurulması, 2 tane RT-PCR (Reverse Transcription Polymerase Chain Reaction) testinin koronavirus için pozitif çıkması ve genetik yapı dizisinin koronavirus yapısıyla eşleşmesi hastada virüsün olduğunun ispatı kabul edilmiştir (Basu, S., & Campbell, R. H., 2020).

2.1.3. Laboratuvar

Dünya Sağlık Örgütü'nün MERS-Cov virüsü zamanında elde ettiği veriler, ilk çalışmalar için kullanılmıştır (World Health, 2020). Covid-19 tanısı, vakanın burun deliklerinden ve genizden alınan sürüntüdeki RNA'nın, SARS-Cov-2 RNA'sıyla eşleşmesi ile belirlenmektedir. Örneklerin analiz koşulları kadar toplanma şartları da önemlidir. Örnekler toplandıktan sonra hızlı bir şekilde laboratuvara verilmelidir. Laboratuvardaki örnekler analiz sürecine kadar 2-8°C'de saklama dolaplarında bekletilmelidir (Togay & Yılmaz, 2020).

2.2. Makine Öğrenmesi ve Derin Öğrenmenin Bu Süreçte Kullanımı

Makine öğrenmesi ve derin öğrenme pandemi sürecinde geleceğe yönelik tahmin yapılması, akciğer tomografisi görüntülerinin sınıflandırılması, bölgesel uygulanan kısıtlamaların faydalarının görüntülenmesi, ulusların birbirleriyle olan sınırlarının kapatılması, sokağa çıkma yasaklarının etkisinin görüntülenmesi gibi birçok alanda uygulanmıştır.

Virüsün yayılımının matematiksel modelleri, tomografi görüntüleri, bölgesel etkilerin araştırılması için kaggle gibi açık kaynaklı platformlarda yarışmalar düzenlenmiştir (kaggle, 2020), halkın bilgilenebilmesi için düzenli olarak televizyonlardan veriler yayınlanmıştır.

Covid-19 ile ilgili yapılan 11 adet çalışmanın tarandığı bir çalışmada (Chen & See, 2020) yapay zekanın tanı koyma, halk sağlığı, karar verme ve terapi olmak üzere 4 alanda

kullanıldığı tespit edilmiştir. Alanın daha da büyümesi, yapay zekanın etik kurallarının artması ve büyük veri ile beraber kullanılırsa gözetleme, klinik ve hastanelerdeki sıradan işlemlerin yapılması ve hasta yönetiminde de kullanılabileceği tespit edilmiştir.

2.2.1. Matematiksel Modeller ile Vaka Sayılarının Tahmini Üzerine Yapılan Çalışmalar

Virüsün yayılımının matematiksel modeli üzerine yapılan bir araştırmada (Z. Zhang & Jain, 2020) konvolüsyonel tanıma bağlı diferansiyel değişkenlerin gerçek problemler için matematiksel modeller kadar önemli olduğu ortaya konulmuştur. Bu çalışmada ebola ile koronavirus gen dizilimi karşılaştırılıp matematiksel model olarak birbirinden farklı bir yayılım yöntemi olduğu ve bunun nedeninin yayılma yöntemleri olduğu öne sürülmektedir.

Günlük vaka ve ölüm sayılarından kısa dönemli tahminlerde bulunmaya çalışan başka bir çalışmada (Zhao et al., 2020) vaka sayılarından poisson dağılımı kullanarak ve zaman aralıkları için gauss dağılımı ile modellediği sistem ile modelden elde ettiği vaka sayısının kısa süreli zaman aralığında sabit kaldığını varsayarak etkili çoğaltma katsayısını tahmin etmeye çalışmış, mevcut iletim hızının (ivmesinin) sabit kalacağını veya belirli bir yüzdelik farkla değişeceğini varsayarak sonlu dağılımı kullanarak gelecekteki veriyi tahmin etmeye çalışmıştır.

İnsanların kurallara uymadığı durumlarda çevresine nasıl zarar verebileceği üzerine yapılan bir matematiksel model çalışmasında (Ferhat, 2020) lineer 2 ve çok boyutlu matris dönüşümler üzerinde insanların birbirlerini nasıl etkileyebileceği bir formülasyon üzerinde anlatılmaya çalışılmıştır. Bu çalışmaya göre ülkelerin hava yollarının ve kara yollarının kapatılmasının önemi vurgulanmaya çalışılmıştır. Asıl dikkat çekilmeye çalışılan konu olarak da bir grup insanın hava gibi herkesin ortak kullandığı ortak bir şey vasıtasıyla yayılan salgın esnasında, yan yana dahi bulunmanın bu süre içerisinde birbirini ne kadar etkileyebileceği matematiksel bir şekilde ortaya konulmaya çalışılmıştır.

Küçük veri kümeleri ile Covid-19 vakaları üzerinde derin öğrenme algoritmalarından hangisinin en verimli algoritmayı bulmaya çalışan bir başka çalışmada (Zeroual, Harrou, Dairi, & Sun, 2020) tekrarlanan sinir ağları (RNN), uzun kısa süreli bellek (LSTM), iki yönlü uzun kısa süreli bellek (BiLSTM), kapılı tekrarlanan birimler (GRUs) ve değişkenli oto enkoder (VAE) algoritmaları uygulanarak küresel çaptaki salgının etkileri

gözelemlenmeye çalışılmıştır. Bu algoritmalarından en iyi performansı değişkenli oto enkoder vermiştir.

Facebook takımının ortaya çıkardığı prophet modeli ise bu salgın için en çok kullanılan modellerden biridir. Tahminde bulunabilmek için lojistik regresyon metoduyla verideki en hızlı yükselen tepe noktalarını eğittikten sonra işlenmiş veriyi prophet metoduna tekrar yükleyerek yeni bir model üretmektedir (P. Wang, Zheng, Li, & Zhu, 2020).

Ülkelerin virüs bulaştırma kaynaklarını ve sonrasında demografik yapı üzerinde yayılımını araştıran bir çalışmada (Ankaralı, Ankaralı, & Erarslan, 2020) İtalya gibi Avrupa'nın merkezi sayılabilecek bir ülkenin 53 ülkeye bulaştırdığı, bunu 47 ülkeye bulaştıran Çin, 21 ülkeye bulaştıran Fransa ve 12 ülkeye bulaştıran İran takip etmekte olduğu yazılmıştır. Yine bu çalışmaya göre toplam vaka sayısının, ülkenin ortanca yaşının bir yıl arttığı durumda 0,216 kat, şehirliklik %1 arttığında 0,015 arttığı belirtilmektedir. Günlük ölümlerde ise ortanca yaşın 1 yıl arttığı durumda 0,139 kat, şehirleşme %1 arttığında ise 0,048 kat arttığı tespit edilmiştir.

Virüs belirlenmesi ve tanınım konulması sağlık sistemi için büyük bir zorluk oluşturmaktadır. Bunun sağlanabilmesi için küresel ölçekte tecrübe ve kaynakların düzenli olarak paylaşılması gerekmektedir. Ama dijital çağda veri güvenliği büyük bir endişeye sebep vermektedir. Bu hassasiyeti gidermek üzerine yola çıkılan bir çalışmada (Ulhaq & Burmeister, 2020) tasarlanmış parçalı gizlilik (dPbD) olarak isimlendirilen bir yapının (framework) tanıtımını yapılmıştır ve bu modelin makine öğrenmesi algoritmalarında kullanım yolları araştırılmıştır. Tahmin netliğini bozmayacak en iyi sistem ile bulmaya çalışılmıştır. Makine öğrenmesi, derin öğrenme, yapay zeka alanlarında modellerin geliştirilebilmesi için veriye ihtiyaç vardır.

Mart ayında Şangay'da yapılan bir araştırmada (Sun et al., 2020) geçmişe yönelik 336 tane koronovirüs hasta verisi üzerine bir analizde, veri kümesi eğitim verisi ve test verisi olarak ikiye ayrılmıştır. 336 hastadan oluşan veri kümesinde 310 tanesi ağır olmayacak bir şekilde hastalığı geçirirken 26 tanesi ağır bir hastalık süreci geçirdiği raporlanmıştır. 26 ağır geçiren hastanın 10'u daha hastaneye gelmeden semptomları çok ağırlı olmaya başlamıştır. Bu vakalardan 74 tanesi Wuhan'dan, 4 tanesi İran'dan, geri kalanlar ise Çin'in diğer bölgelerinde görülmüştür. Bu vakaların yaş ortalaması 48 ve hastalığı ağır şekilde seyredenlerin ortanca değeri 65'tir. Bu yüzden 65 yaş kritik bir seviye olarak

belirlenmiştir. Yine bu hastalar üzerinde yapılan bir incelemede %79'unun yüksek tansiyonu, %29'unun diyabeti, %17'sinin koroner damar hastalığı ve %4'ünün tüberkülozu olduğu ortaya çıkmıştır.

2.2.2. DNA ve RNA Modelleri Üzerine Yapılan Çalışmalar

DNA sarmallarını sınıflandırılması bioinformatik alanında önemli bir yere sahip, bu sayede araştırmacılar gen analizi yapıp muhtemel hastalıkları tespit edebiliyorlar. DNA sınıflandırması için derin öğrenme alanındaki konvolüsyonel sinir ağları (CNN), derin sinir ağları (DNN) N-gram olasılıksal model algoritmalarını kullandı (X. Zhang, Beinke, Al Kindhi, & Wiering, 2020). Bu çalışmada farklı büyüklüklerde 4 farklı virüs veri seti kullanıldı. Hepatit-C, influenza ve korona, bağışıklık sistemi virüsü, SARS-Cov-2 virüsü veri setleri inceleniyor. Buradan, DNA dizilerinden içindeki bilgi açısından zengin özellikleri hesaplamak için Levenshtein mesafesi ve rastgele oluşturulmuş DNA alt dizilerine dayanan yeni bir yöntem keşfedilmiştir.

2.2.3. Akciğer Tomografisi ve X-Ray Üzerinden Bilgisayarlı Görüntüleme Üzerine Çalışmalar

PCR testlerinin ulaşım zorluğunun yüksek ve yanlış-negatif yüzdesinin yüksek olması nedeniyle yapılan bir çalışmada (Yang & Yan, 2020) başka sağlık alanında kullanılmak üzere yapay zeka destekli görüntü tanıma sistemlerinin hastaya tanı konulurken kullanılmasının öneminden bahsedilmektedir. Göğüs X-ray görüntüsünde radyoloji uzmanlarının hassasiyet doğru tanı oranının %67,1 olduğunu bu oranın bilgisayarlı görüntü sistemleriyle artırılabilirliği iddia edilmiştir. (Giovagnoni, 2020)

MobileNet, DenseNet, Xception, ResNet, InceptionV3, InceptionResNetV2, VGGNet, NASNet derin öğrenme tabanlı altyapılar arasında X-ray ve CT görüntülerini kullanarak (Kassani, Kassani, Wesolowski, Schneider, & Deters, 2020) en doğru Covid-19 tanısını koyan modelin tespit edilmesi üzerine yapılan bir çalışmada DenseNet121 %99 bir oranla birinci, ResNet50 ile LightGBM'nin birlikte kullanıldığı yapı %98 oranla ikinci en doğru tahmini yapmıştır.

Göğüs X-ray görüntüsünün çok seviyeli eşikleme (Multi-Level Thresholding) işlemine tabii tutulduğu çalışmada (Mahdy, Ezzat, Elmousalami, Ella, & Hassanien, 2020) gri

görüntüsünü birçok farklı parçaya ayırmaktadır. Bunun amacı görüntüdeki yoğunluğu en yüksek olan alanların belirlenmesidir. Bu işlemden sonra destek vektör makinaları algoritmasıyla sınıflandırmaya tabi tutup tanıyı koyması beklenmiştir.

X-Ray görüntüleri üzerinden derin öğrenme algoritmalarından evrimsel sinir ağı ile otomatik sınıflandırma yapılan bir çalışmada (Yasar&Ceylan, 2020) 2 tane eğitim ve test veri seti oluşturup Covid-19 hastası olan ve olmayan gruplar üzerinde araştırma yapılmıştır. Bu çalışmada 23 ve 54 katmanlı mimarilere sahip 2 farklı CNN modeli elde edilmiş ve doğruluk değeri, hassaslık, f-1 değerleri üzerinden doğrulama yapılmıştır.

Akciğer tomografi görüntüleriyle oluşturulan eğitim ve test verileriyle yapılan bir başka çalışmada (Yasar&Ceylan, 2020) Covid-19'un erken dönem teşhisi CNN modeliyle bulunmaya ve otomatik sınıflandırma yapılmaya çalışılmıştır. Bu çalışmada k-en yakın komşular (k-NN) ve destek vektör makinaları (SVM) algoritmalarıyla derin öğrenmenin doğruluk değerleri karşılaştırılmıştır.

Akciğer tomografi görüntüleri ile derin öğrenme yaptıkları başka bir çalışmalarında (Yasar&Ceylan) verilerin modellendiği ve veri yapısı kurularak yanlış-negatif, yanlış-pozitif ve resimdeki eksikliklerin düşük değerlerine bakılmıştır. 24 katmanlı bir CNN mimarisine yürütülen çalışma derin öğrenmeyle sınıflandırma parametrelerini kullanarak hastalık tanısı tahmini üzerinde çalışılmıştır.

2.2.4. Aşıyla İlgili Çalışmalar

Öğrenme aktarımı (transfer learning) yöntemiyle hızlı bir şekilde ilaç modeli tasarlamaya çalışan bir çalışmada (S. Wang et al., 2020) 4.9 milyon ilaç benzeri molekül verisini, 1.5 milyar molekül verisinin toplandığı Amerikan Gıda ve İlaç Dairesi onaylı ZINC15 veri merkezinden alarak potansiyel anti-SARS-Cov-2 ilaçları üzerine denemeler yapmaktadır. Bu denemelerden en yüksek doğruluk değerine sahip 7 tane ilaç deney ispatı için kullanılmıştır.

3. MAKİNE ÖĞRENMESİ YÖNTEMİ ve KAVRAMLARI

Makine öğrenmesinin bir üst alanı olan yapay zeka kavramı yeni bir alan olmamakla birlikte 1950'li yıllarda Alan Turing'in oluşturduğu bir düşünce testine dayanmaktadır (A. Turing, 1950). Eğer makine bir insanı, kendisinin insan olduğuna ikna edebiliyorsa makinenin zekası olduğu kabul edilmektedir. Bu test tasarımında perde arkasındaki bir makinenin cevaplarıyla bir insan gibi davranması ve insan olduğuna inandırması gerekiyor.

Makine öğrenmesi ise veri setindeki veriler ile anlamlı varsayımsalar çıkarmaya ve tahminlerde bulunulmasına, bir takım istatistiksel yöntemlerle yardım eden bir çeşit algoritmadır. Kendi içerisindeki yöntemlere, istatistiksel analizlere, istenilen sonucun değerine, veri setinin içindeki özellik sayılarına göre yapacağı işlemler değişiyor. Tahmin sırasındaki en büyük destekçileri veri setleri, modeller ve istatistiksel yöntemlerdir.

3.1. İstatistiksel Yöntemler

İstatistik, belirli bir amaç doğrultusunda toplanan verilerin, düzenlenmesi, işlenmesi ve yorumlanmasına yönelik yöntemlerin tümünün yapıldığı ve bu yöntemlerin araştırıldığı bir alandır. Bunların grafikler halinde tablolar ile sunulması bu tablolar sonucu karar alma karar alma sürecine etki etmesi de istatistiğin etkili kullanım alanlarından biridir. Olasılık değeri göz önünde bulundurularak, sonucun henüz bilinmediği durumların sonucunu tahmin etmek için de kullanılır (Shemesh, Polak, Lundin, Sollid, & Yaari, 2020).

3.1.1. Popülasyon ve Örneklem Kümesi

Popülasyon bir gruptaki veya kümedeki elemanların veya hepsine verilen isimdir. Örnek vermek gerekirse herhangi bir öğrenci topluluğu Türkiye'deki öğrenci topluluğunun bir popülasyonudur. Konya'daki plastik üreten şirketler de Konya'daki bütün şirketlerin bir popülasyonudur. Konya'daki bütün plastik üreten şirketlerle ilgili veriye ulaşamadığımız durumda başvuracağımız şey ise bir örneklem kümesi olacaktır. Örneklem kümesi popülasyonun bir alt kümesidir.

3.1.2. Normal Dağılım

Olasılık dağılımı, bir olay veya deneydeki sonuçların ihtimalini gösteren bir fonksiyondur. Elinizdeki veri setinde bir tane özellik olduğunu varsayalım. Bu özellik bir değerdir ve bu özelliğin olasılık dağılımı fonksiyonu alabileceğin değerlerin ihtimallerini gösterir. Normal dağılım ise bir olasılık dağılımıdır ve şekli bir çan eğrisine benzer (Maghdid et al., 2020). Çünkü düşüncede değerlerin yayılımı ortalama olarak gerçekleşecek ve en fazlası ortada olacaktır.

3.1.3. Merkezi Eğilim

Merkezi eğilim veri setindeki verilerin olasılık dağılımlarının merkezdeki özellikleri ile ilgilidir. Bunlar ortanca, ortalama ve moddur.

Ortalama (Mean): Veri setindeki verilerin ortancasıdır.

Ortanca (Median): Veriler aşağıdan yukarıya sıralandığı zamandaki ortada kalan veridir.

Mod (Mode): En çok tekrarlanan değerdir.

3.1.4. Varyans ve Standart Sapma

Varyans değerinin değişim ölçüsüdür. Veri setindeki her elemanın ortalamadan çıkarıldıktan sonra karesinin alınması ve tüm karelerin toplanması ve veri sayısına bölünmesiyle hesaplanır.

$$\frac{\sum(x_i - \text{mean})}{N} \quad (1)$$

X_i , o sırada gözlenen veriyi, mean, ortalama değeri, N ise veri sayısını temsil etmektedir.

Standart sapma ise değerlerin nasıl dağıldığını incelenmiştir (Leys, Ley, Klein, Bernard, & Licata, 2013). Standart sapma varyansın kare kökü alınarak hesaplanır..

3.1.5. Kovaryans ve Korelasyon

Kovaryans nicel bir ölçüdür. İki farklı verinin birbirleri ile kaç farklı şekilde eşleştiğini ölçülmüştür. 2 farklı özelliğin ortalamalarının birbirlerine olan farklılıklarını kıyaslanmıştır (Price, 1970).

$$Cov(X, Y) = E[(X - \mu_x)(Y - \mu_y)] \quad (2)$$

E beklenen sonuç, m ise ortalama değeri temsil eder.

Korelasyon, her değişkenin standart sapması ile kovaryansın normalleştirilmesidir.

$$Corr(x, y) = E \left[\frac{(x - \mu_x)}{\sigma_x} \frac{(y - \mu_y)}{\sigma_y} \right] \quad (3)$$

Bu normalizasyon sayesinde eski çıktılar yeni korelasyon değerleri ile yer değiştirir. Bu değerler her zaman 0 ile 1 arasındadır.

3.1.6. Merkezi Limit Teoremi

Doğa bilimlerinde ve sosyal bilimleri gibi birçok alanda rastgele bir değer normal değeri bilinmediğinde normal dağılım kullanılır. Merkezi limit teoremi normal dağılımın neden birçok alanda kullanıldığını gösterir. Merkezi limit teoremine göre bir dağılımdan ne kadar çok örnek alınır, örnek ortalamaları nüfus dağılımına bakılmaksızın normal bir dağılımı doğru eğilim gösterecektir. Normal dağılımı bu kadar önemli kılan şey ise ortalama ve standart sapma ile kolayca hesaplanıyor olmasıdır. Yani bir veri kümesinin ortalamasını ve standart sapmasını biliyorsak hemen hemen bütün özellikleri hesaplanabilmektedir (Jones, 2004).

3.1.7. P-değeri

P-değeri, rastgele bir değişkenin aldığı değer olasılığıdır. Bir A değişkenimiz ve bir x değerimiz olduğunu varsayalım. x'in p-değeri, A'nın x değerini veya gözlemlenme şansı aynı veya daha az olan herhangi bir değeri alma olasılığıdır.

3.1.8. Rastgele Değişkenlerin Beklenen Değeri

Rastgele bir değişkenin beklenen değeri, değişkendeki diğer değerlerin ağırlıklı ortalamasıdır. Ağırlıklı ortalama burada özel bir değer anlamına gelmektedir (Hill & Kertz, 1982).

Aralıklı ve sürekli rastgele değişkenler için beklenen değer farklı olarak hesaplanır.

- Aralıklı rastgele değişkenler sonlu sayıda çok fazla değer alır. Bir yıldaki yağmurlu günlerin sayısı bir aralıklı rastgele değişkendir.
- Sürekli rastgele değişkenler sonsuz sayıda değer alırlar. Evden iş yerine giderken geçirilen zaman sürekli rastgele değişkene bir örnektir. Bunu saatlerle mi, dakikalarla mı ölçeceğinize göre sonuç değişebilir.

Aralıklı rastgele değişkenler şu formülle hesaplanır:

$$E[x] = \sum x_i p(x_i) \quad (4)$$

x_i , x 'in alacağı değerleri, $p(x_i)$ ise x_i 'nin alacağı değerlerin ihtimalini temsil eder.

Sürekli rastgele değişkenler aynı mantıkla ama farklı bir yöntemle hesap edilir.

$$E[x] = \int_{x_{min}}^{x_{max}} x f(x) dx \quad (5)$$

3.1.9. Koşullu Olasılık

Olasılık kısaca olayların olma ihtimalini hesaplar ve 0 ile 1 arasında bir değer alır. Bir A olayının ihtimali $p(A)$ şeklinde gösterilir ve beklenen ihtimal sayısının tüm sayılara bölünmesiyle hesaplanır. Koşullu olasılık ise A olayının olma olasılığını, A ile bağlantısı olan daha önce meydana gelmiş başka bir olayın ihtimalidir. $P(A|B)$ ile gösterilir.

3.1.10. Bayes Teoremi

Bayes teoremine göre, A olayının olasılığı daha önce meydana gelen B olayının ihtimali ile hesaplanabilir. Ve B olayının ihtimali de daha önce meydana gelen A olayının ihtimali ile hesaplanabilir.

$$p(A|B) = \frac{p(A) \cdot p(B|A)}{p(B)} \quad (6)$$

3.2. Makine Öğrenmesi Yöntemleri

Bu bölümde makine öğrenmesi algoritmalarının bazı çeşitlerini covid-19'un dünya üzerindeki etkileri SIR (suspected, infected, recovered) modeli ile incelenmiştir.

3.2.1. Lineer Regresyon ile Modelleme

Veri kümesinin niceliksel ve niteliksel özelliklerine göre tanımlanır. Niceliksel özellikler bir insanın kilosu, şehrin sıcaklık değeri gibi sayısal özelliklerdir. Niteliksel özellikler ise sınıflandırma şeklinde kategorilere ayırır. “Kadın mı erkek mi?”, “Canlı mı cansız mı?” gibi özelliklerine bakılır. Lineer regresyonu sayısal değerler ile yapılmaktadır. Veri setindeki bağımlı ve bağımsız değişkenlerden bağımlı olan tahmin unsuru, bağımsız olan tahmin edilmeye çalışılır.

Lineer regresyon tahmin unsuru ile tahmin edilmeye çalışılan veriler arasında lineer bir ilişki olup olmadığını varsa bu ilişkiyi anlayıp tahminleri bulmaya çalışır. İki çeşit lineer regresyon vardır. Basit lineer regresyon ve çoklu lineer regresyon. Basit lineer regresyonda Bağımlı olan Y ile bağımsız olan X arasında bir bağlantı aranır. Matematiksel model şu şekildedir.

$$Y \approx \beta_0 + \beta_1 \times X \quad (7)$$

B₁ ve B₂ model katsayısı veya parametreleridir. Y değeri model tahmin sonucunu verir. Gerçek değer ile tahmin sonucu farkından ortalama mutlak değer ve ortalama kare değer ile modelin doğruluğunu ölçümlenebilir.

Doğrusal olmayan durumlar için lineer yöntemi kullanılmaz. Bu salgın logaritmik olarak hızlı bir şekilde arttığı için doğrusal değildir ve lineer yöntem bu durum için uygun değildir. Bunun için lineer regresyonun SVM (Destek Vektör Makinaları), Polinom Regresyon, Bayes Regresyon yöntemleri kullanılmış ve bu yöntemlerle dünyadaki hastalık tanısı koyulan hastalar üzerinden incelenmiştir. Bu yöntemler lineer regresyonun alt dalları olup formülasyon yönünden farklı özelliklere sahiptir.

Destek vektör makinaları sınıflandırma için kullanılan bir yöntemdir. Bu yöntemin asıl amacı sınıflandırma için çekilen çizginin en iyisini bulmaya çalışmasıdır. Burada

sınıflandırma değil regresyon yapılacağı için destek vektör makinelerinin daha az bilinen destek vektör makinaları regresyon yöntemi kullanılarak bir tahmin elde edilmeye çalışılmıştır.

Çoğu lineer regresyon modelde, asıl hedef ortalama kare hatalar toplamını minimize etmektir. Bu durumda tahminin doğruluğundan belli bir miktarda çıkmış olunabilir. Başka bir ifadeyle bu yöntem ne kadar hatanın uygulanan yöntem için kabul edilebilir olduğuna dair mevcut veri setine göre bir çerçeve çizmesidir.

4. VERİNİN TANIMI, TAHMİN YÖNTEMLERİNİN UYGULANMASI, VAKA ve ÖLÜM ORANLARININ ve AŞININ VERİMLİLİĞİNİN İNCELENMESİ

4.1. Verinin Tanımı ve Kaynağının İncelenmesi

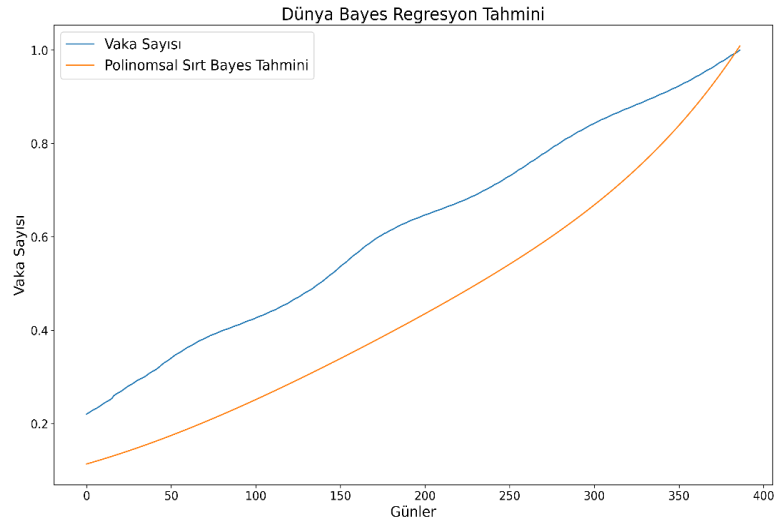
John Hopkins üniversitesinin önderliğinde toplanmaya ve açık veri halinde sunulan veriler kullanılmıştır. Verinin içinde ISO 3166-1 alpha-3 – 3 harfli ülke kodları, coğrafi olarak bulunduğu kıta, toplam vaka ve ölüm sayıları, yoğun bakımda yatan Covid-19 kaynaklı hasta sayıları, test sayıları, aşı uygulama sayıları, yaş dağılımları gibi veri setleri bulunmaktadır (Github User Content, t.y.).

4.2. Veri Bilimi ve Makine Öğrenmesi Algoritmalarının Uygulanması

Dünya geneli vaka sayısı grafiğinin bayes regresyon, polinomsal regresyon ve destek vektör makinesi regresyonu kullanılarak tahmini gerçekleştirilmiştir (Github, t.y.). Şekil 1’de bayes regresyon tahmini ile gerçek veri grafiklerinin karşılaştırması görülmektedir. Söz konusu iki grafik arasındaki ortalama mutlak hata ve ortalama kare hata değerleri hesaplanmıştır. Hesaplamalara göre iki veri grafiği arasındaki ortalama mutlak hata 0.0129 ve ortalama kare hata 0.0003 olarak gerçekleşmiştir. Söz konusu sonuçlar kabul edilebilir doğrulukta tahminlerin yapıldığını göstermektedir. Şekil 2’de ise polinomsal regresyon tahmin sonuçları yer almaktadır. Şekil 2 incelendiğinde polinomsal regresyonun tahmini ile gerçek veri grafiklerinin tutarlı artışlar gösterdiği anlaşılmaktadır. Söz konusu grafikler arasındaki ortalama mutlak hata 0.0129 ve ortalama kare hata 0.0003 olarak hesaplanmıştır. Hata hesaplamaları ile bayes ve polinomsal regresyonların aynı olduğu gözlemlenmiştir. Şekil 3’de Dünya geneli vaka sayısının destek vektör makine ile oluşturulmuş tahmin grafiği yer almaktadır. Tahmin grafiği ile gerçek veri grafiği arasındaki ortalama mutlak hata 0.3006 ve ortalama kare hata 0.1138 olarak gerçekleşmiştir. Bu hata değerleri ile destek vektör makineleri en yüksek hata değerlerine sahip tahmin algoritması olmuştur. Şekil 9’da dünya geneli vaka sayılarının XGBoost ile oluşturulmuş tahmin grafiği yer almaktadır. Tahmin grafiği ile gerçek veri grafiği arasındaki ortalama mutlak hata 0.0127 ve ortalama kare hata 0.0003 olarak gerçekleşmiştir. Şekil 4’de dünya geneli vaka sayılarının karar ağaçları ile oluşturulmuş tahmin grafiği yer almaktadır. Tahmin grafiği ile gerçek veri grafiği arasındaki ortalama mutlak hata 0.0132 ve ortalama kare hata 0.0003 ile düşük hata değerlerine sahip tahmin algoritmalarıdır. Şekil 5 ve Şekil 6’daki ElasticNet ve Lasso ile oluşturulmuş tahmin

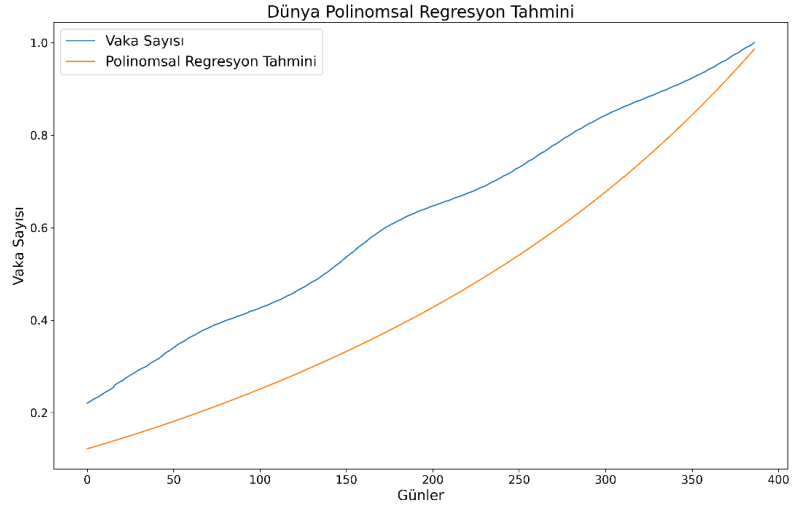
algoritmaları aynı hata değerlerine sahip olup ortalama mutlak hata 0.2862 ve ortalama kare hata 0.1048'dir. Şekil 8'de Ridge ile oluşturulmuş grafik de ortalama kare hata 0.0639 ve ortalama kare hatanın 0.0052 olduğu gözlemlenmiştir. Şekil 7'de Random Forest ile oluşturulmuş grafik de ortalama mutlak hata 0.0117 ve ortalama kare hata 0.0003 olduğu gözlemlenmiştir. Tez çalışması kapsamında gerçekleştirilen deneyler sonuçları arasında genel bir karşılaştırma yapıldığında en başarılı tahminlerin Random Forest regresyonu kullanılarak gerçekleştirildiği görülmektedir.

Dünya genelindeki vaka artışı Şekil 1'de gerçek vaka sayıları verisinin normalize edilerek bayes regresyon ile bir tahmin yapılmıştır.



Şekil 1. Dünya Toplam Vaka Sayılarının Bayes Regresyon Yöntemi ile Modellenmesi

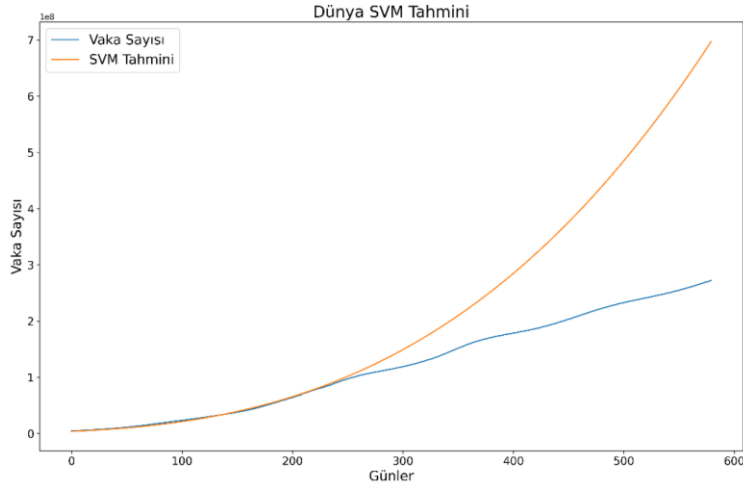
Bu grafikte vaka sayısı grafiğine çok yakın bir grafiği sırt bayes yöntemiyle yakaladığını ve bu grafiğin doğru değere yaklaştığını gözlemlenmiştir. Burada 5. dereceden polinom denklemi oluşturulup grafik hazırlanmıştır.



Şekil 2. Dünya Toplam Vaka Sayılarının Polinomsal Regresyon Yöntemi ile Modellenmesi

Dünya genelindeki vaka artışı Şekil 2’de normalize olan veri üzerinden polinomsal regresyon algoritma yöntemiyle bir yaklaşım yapılmıştır. Bu yöntem de Şekil 1’deki bayes ile yakın bir değer vermektedir. Grafik verinin zamana göre eğitimine göre oluşturulmaktadır. Burada 4. dereceden bir polinom denklemi oluşturulmuş ve gerçek veriye uygun bir model oluşturduğu gözlemlenmiştir.

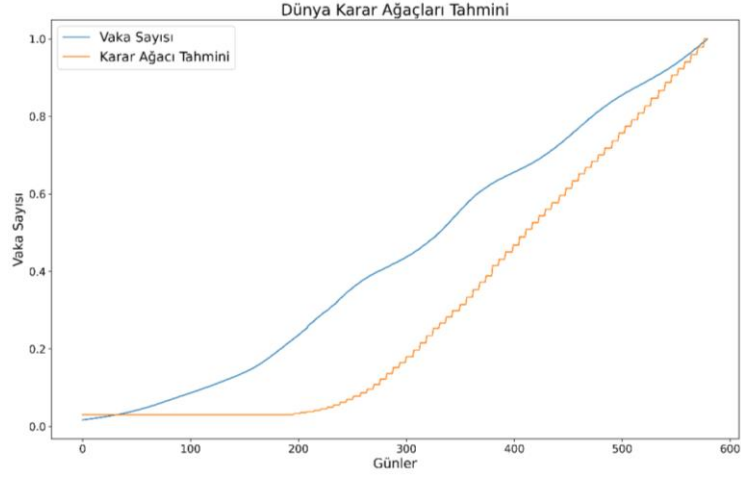
Dünya genelindeki vaka artışı Şekil 3’deki Svm algoritma tahmini incelenmiştir.



Şekil 3. Dünya Toplam Vaka Sayılarının SVM Regresyon Yöntemi ile Modellenmesi

Dünya genelindeki vaka artışı Şekil 3’deki Svm algoritma tahmini incelenmiştir. Destek vektör makinaları algoritması tam olarak veriyi izleyemediği gözlemlenmektedir.

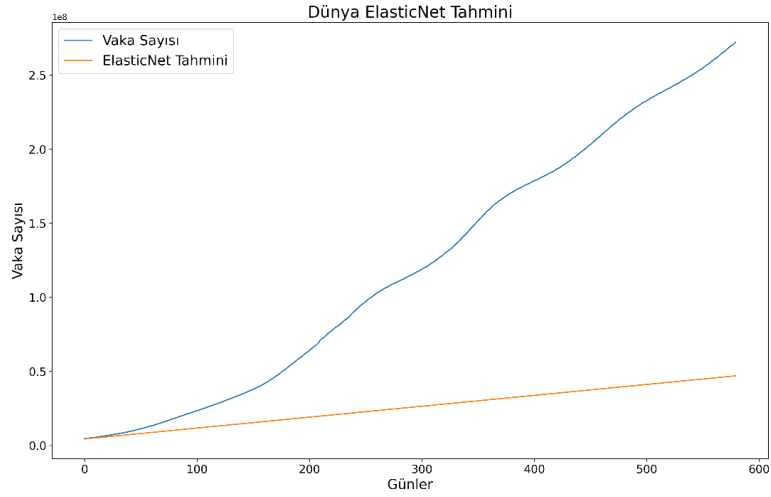
Dünya genelindeki vaka artışı Karar Ağacı algoritması ile tahminlenmiş sonucu Şekil 5’de gösterilmektedir.



Şekil 4. Dünya Toplam Vaka Sayılarının Karar Ağacı Regresyon Yöntemi ile Modellenmesi

Karar ağacı algoritma tahmini vaka sayısını gerçek değerleri modelin eğitimi sonunda yakaladığı gözlemlenmiştir.

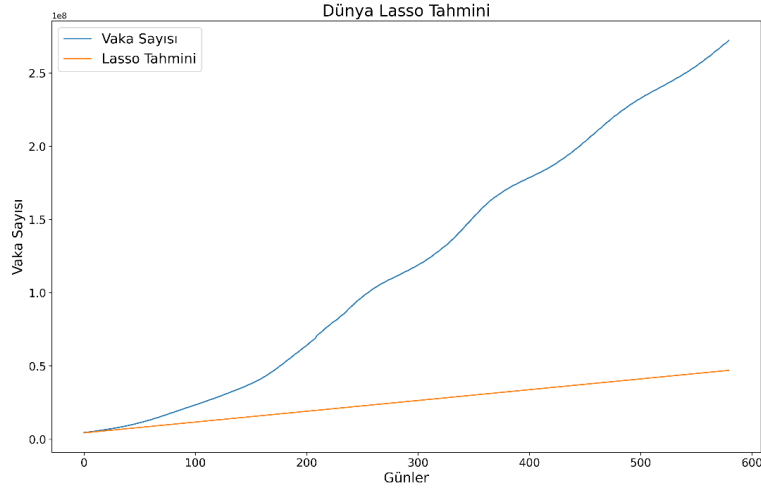
Dünya genelindeki vaka artışı ElasticNet algoritması ile tahminlenmiş sonucu Şekil 5’de gösterilmektedir.



Şekil 5. Dünya Toplam Vaka Sayılarının ElasticNet Regresyon Yöntemi ile Modellenmesi

ElasticNet algoritmasının grafiği ile vaka sayısı arasındaki fark ElasticNet model grafiğinin yanlışlığını göstermektedir.

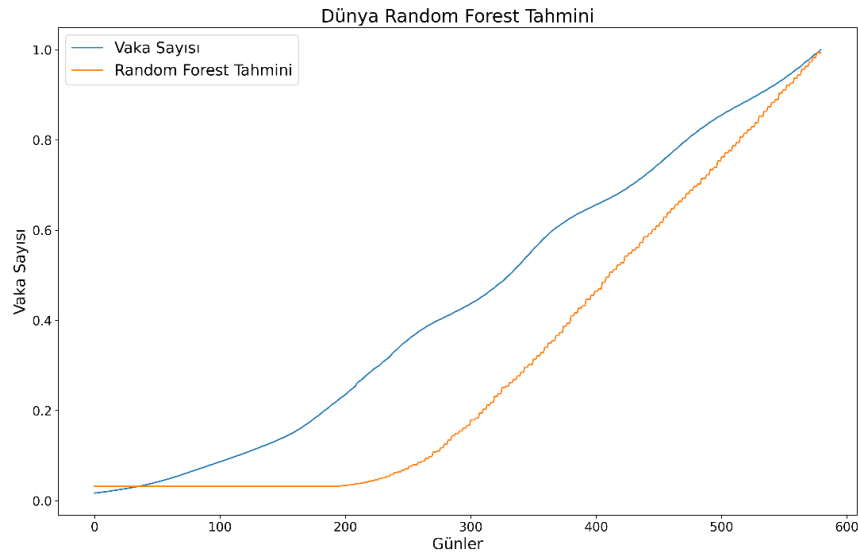
Dünya genelindeki vaka artışı Şekil 6'deki Lasso algoritma tahmini incelenmiştir.



Şekil 6. Dünya Toplam Vaka Sayılarının Lasso Regresyon Yöntemi ile Modellenmesi

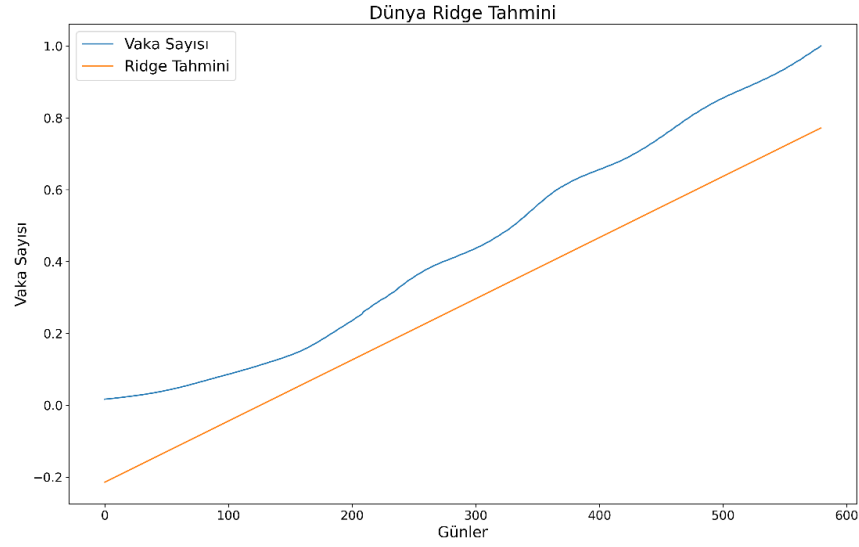
Lasso algoritmasının çizdiği grafik de Şekil 5'deki ElasticNet grafiği benzer oranda yanlışlık yaptığı gözlemlenmiştir.

Dünya genelindeki vaka artışı Şekil 7'deki Random Forest algoritma tahmini incelenmiştir.



Şekil 7. Dünya Toplam Vaka Sayılarının Random Forest Regresyon Yöntemi ile Modellenmesi

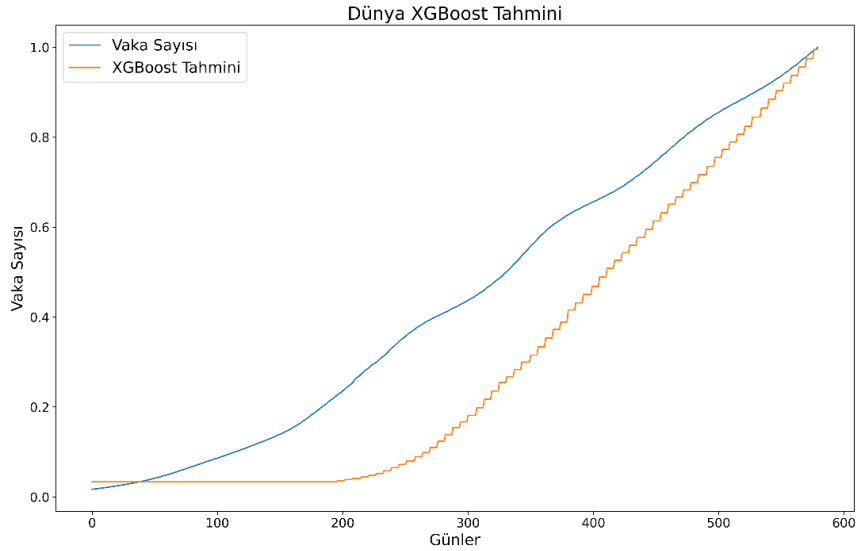
Random Forest algoritmasının gerçek değerleri modelin sonunda yakaladığı gözlemlenmiştir.



Şekil 8 Dünya Toplam Vaka Sayılarının Ridge Regresyon Yöntemi ile Modellenmesi

Dünya genelindeki vaka artışı Şekil 8’de Ridge algoritma tahmini incelenmiştir. Ridge algoritması gerçek değerleri paralel takip ettiği gözlemlenmiştir.

Dünya genelindeki vaka artışı Şekil 9’da XGBoost algoritma tahmini incelenmiştir. XGBoost algoritmasının gerçek değerleri modelin sonunda yakaladığı gözlemlenmiştir.

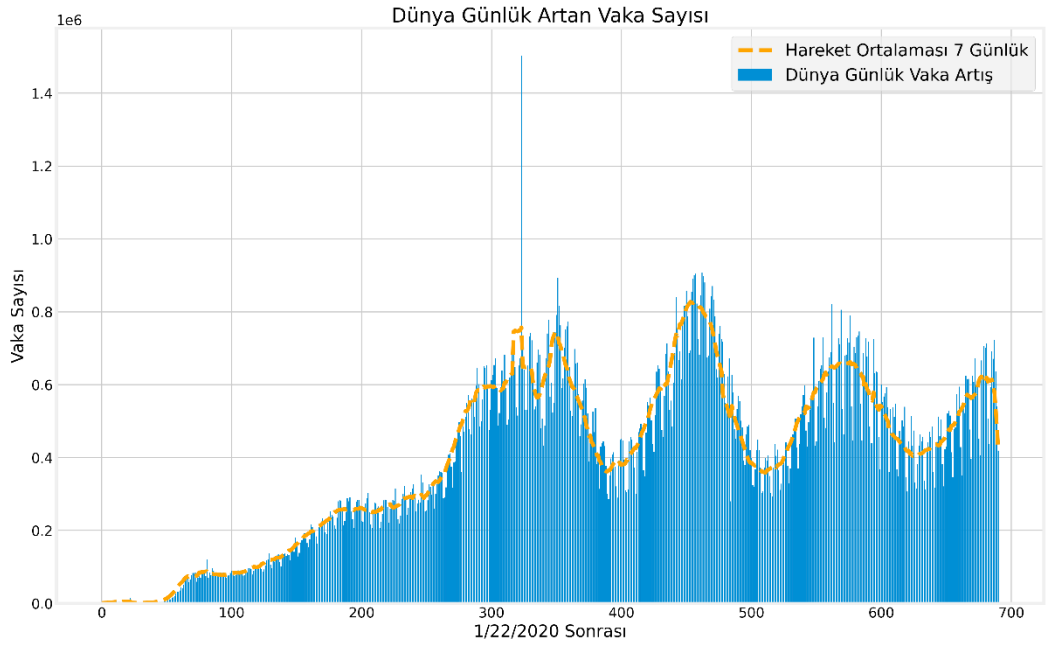


Şekil 9 Dünya Toplam Vaka Sayılarının XGBoost Regresyon Yöntemi ile Modellenmesi

Tablo 1. Farklı Makine Öğrenmesi Modellerinin Günlük Vaka Tahminlerinin Hata Oranları

Modeller	Ortalama Mutlak Hata	Ortalama Kareler Hatası	Kök Ortalama Kare Hatası
Random Forest Reg.	0.0117	0.0003	0.016
XGBoost Regresyonu	0.0127	0.0003	0.018
Polinomsal Regresyon	0.0129	0.0003	0.017
Bayes Regresyonu	0.0129	0.0003	0.017
Karar Ağaç Regresyonu	0.0132	0.0003	0.019
Ridge Regresyonu	0.0639	0.0052	0.072
LASSO Regresyonu	0.2862	0.1048	0.324
ElasticNet Regresyonu	0.2862	0.1048	0.324
SVM Regresyonu	0.3006	0.1138	0.337

Zamana bağlı olarak vaka sayılarının lineer olarak arttığı grafiklerden gözlemlenmiştir.

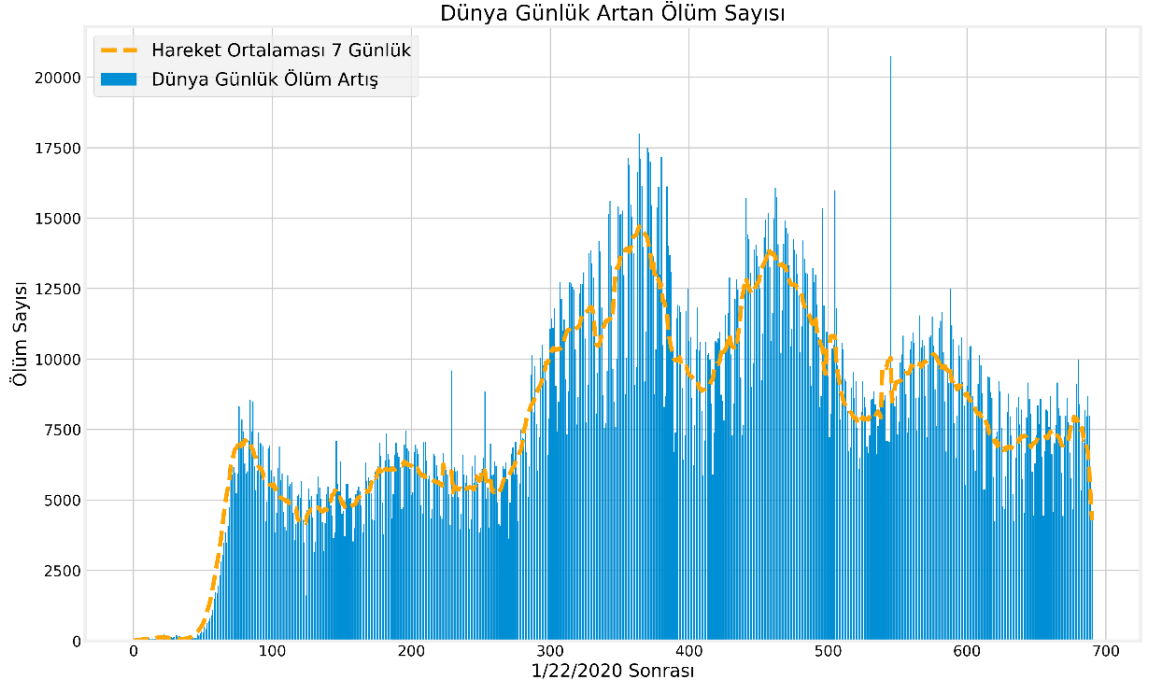


Şekil 10 Günlük Artış Vaka

Dünya genelinde vaka artışını gösteren Şekil 10 incelendiğinde, ilk 300 günlük süreçte genel olarak hızlı yükseliş gösteren bir artış trendi yaşandığı görülmektedir. Bu ilk 300 günlük sürecin sonunda dünya genelinde birinci salgın dalgasının yaşandığı söylenebilir. 300 ile 400 günler arasında vaka sayılarında kısmi bir azalış gözlemlenmiştir. Bu aşamadan sonra ikinci bir artış dalgası yaşandığı görülmektedir. Bu artış trendinin tepe dönüş noktasının birinci dalganın tepe dönüş noktasından daha yukarıda olduğu anlaşılmaktadır. Daha sonra benzer bir azalış-artış trendi daha yaşanarak üçüncü dalganın

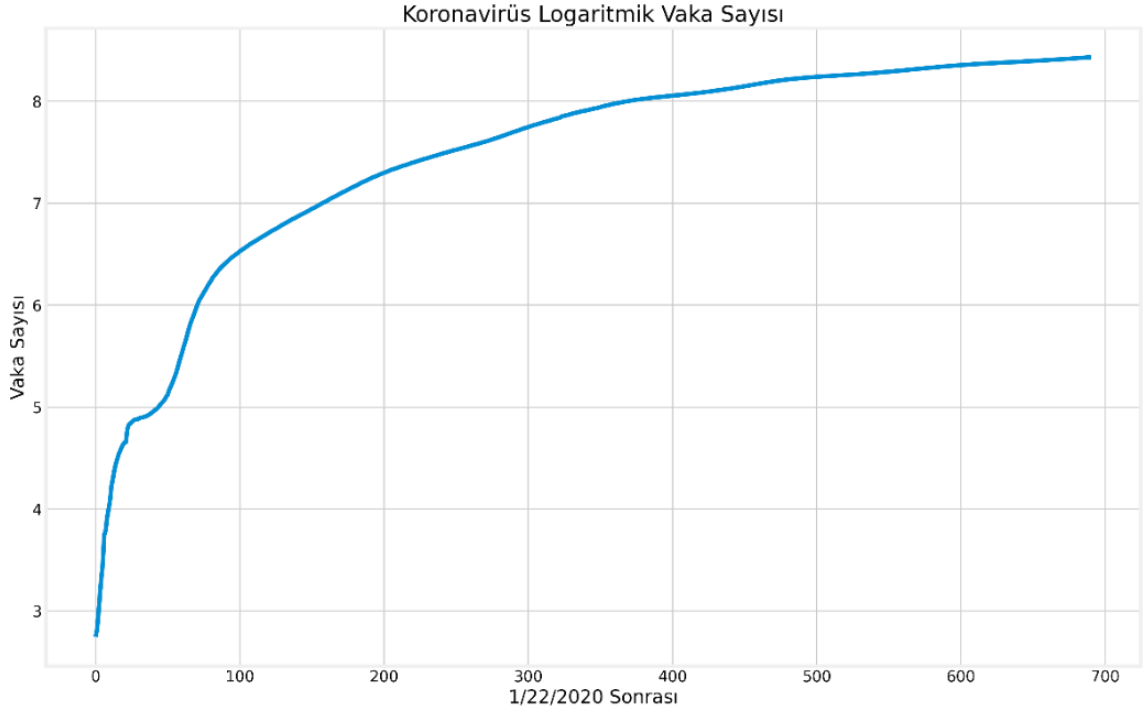
yaşandığı görülmektedir. Bu kapsamda hali hazırda üçüncü dalganın tepe noktalarına doğru yaklaştığımız değerlendirilmesinin yapılması mümkündür.

Şekil 11’de ölüm oranları vaka oranlarıyla korelasyon halinde birbirini takip etmektedir. Bu grafikte zamanla gelişen tedavi yöntemleri ve aşının etkisi görülebilmektedir. Tüm dünyada çok farklı aşılama yüzdeleri ve sayıları ölüm oranlarını bölgesel olarak değiştirmektedir.



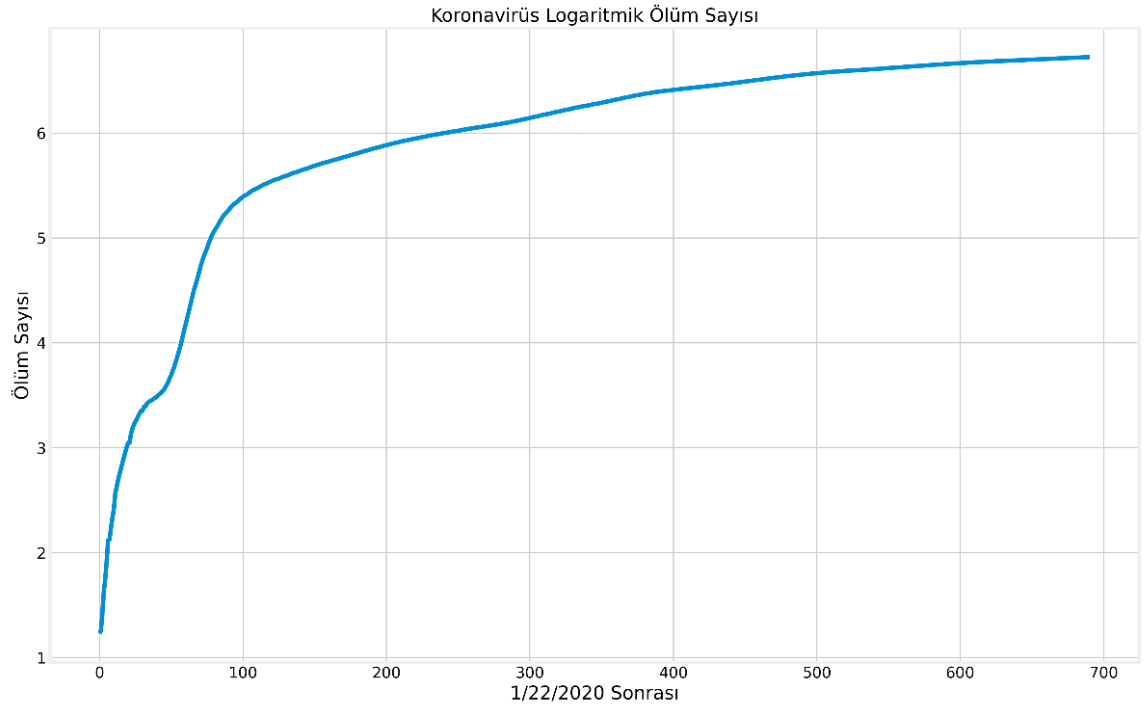
Şekil 11 Günlük Artış Ölüm

Şekil 12’de verilen dünya koronavirüs logaritmik vaka sayısındaki artış grafiği incelendiğinde logaritmik olarak yavaşladığı gözlenmiştir.



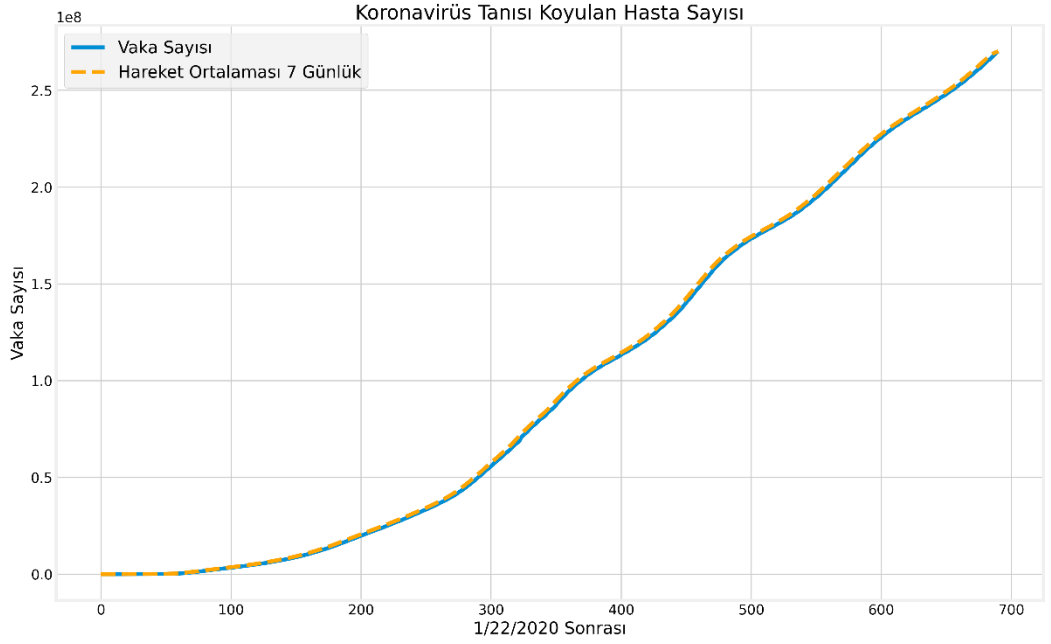
Şekil 12 Dünya Logaritmik Vaka Artışı

Şekil 13’de verilen dünya koronavirüs logaritmik ölüm sayısındaki artış grafiği incelendiğinde logaritmik olarak yavaşlandığı gözlenmiştir.



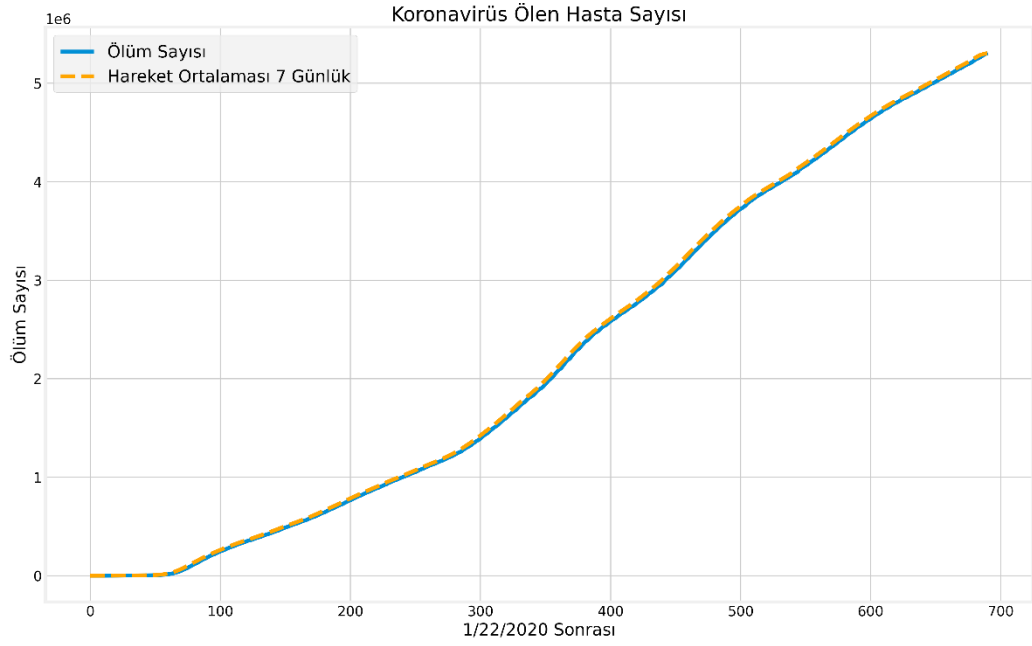
Şekil 13 Dünya Logaritmik Ölüm Sayısı Artışı

Dünya genelinde korona virüs tanısı konulan toplam hasta sayısını gösteren Şekil 14 incelendiğinde, grafiğin toplam hasta sayısını göstermesinden kaynaklı olarak sürekli artış trendi içinde olduğu söylenebilir. Buna göre, vaka artış dalgaları yaşandığı dönemlerde toplam hasta sayısının artış eğrisinin eğitimi artarken vaka sayısının azalmaya girdiği dönemlerde artış eğrisinin eğiminin azaldığı görülmektedir.



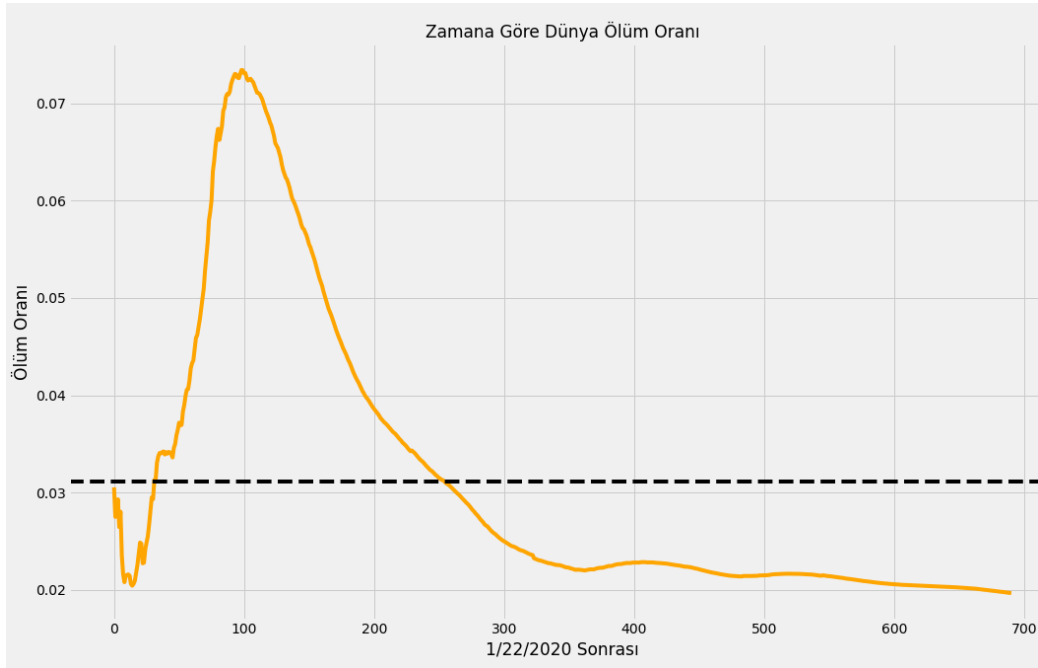
Şekil 14 Dünya Tanılı Hasta Sayısı

Dünya genelinde korona virüs yüzünden ölen hasta sayısı Şekil 15 grafiği incelendiğinde Şekil 14 grafiği ile benzer bir grafik çizmektedir. Vaka sayıları ölüm sayıları ile benzer şekilde ilerleme görülmektedir.

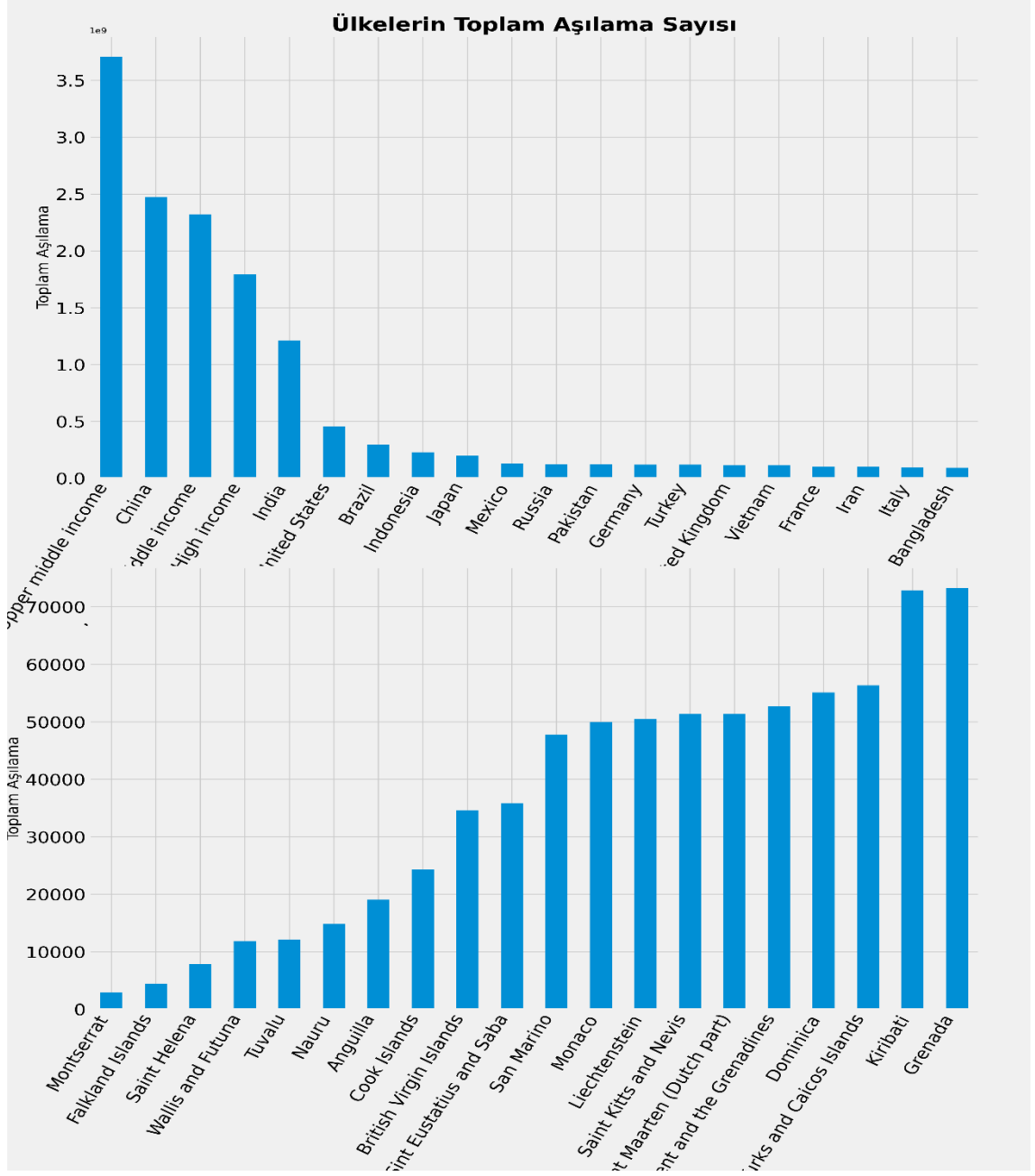


Şekil 15 Dünya Ölen Hasta Sayısı

Şekil 16’da verilen zamana göre ölüm oranı grafiği incelendiğinde vaka başına ölüm oranlarının düştüğü gözlemlenmiştir. Grafikte görüleceği üzere ölüm oranlarının %3.4’den %2.1’lere kadar düşmesinde aşı uygulamalarının artması ve tedavi yöntem ve araçlarının gelişmesi görülmektedir.

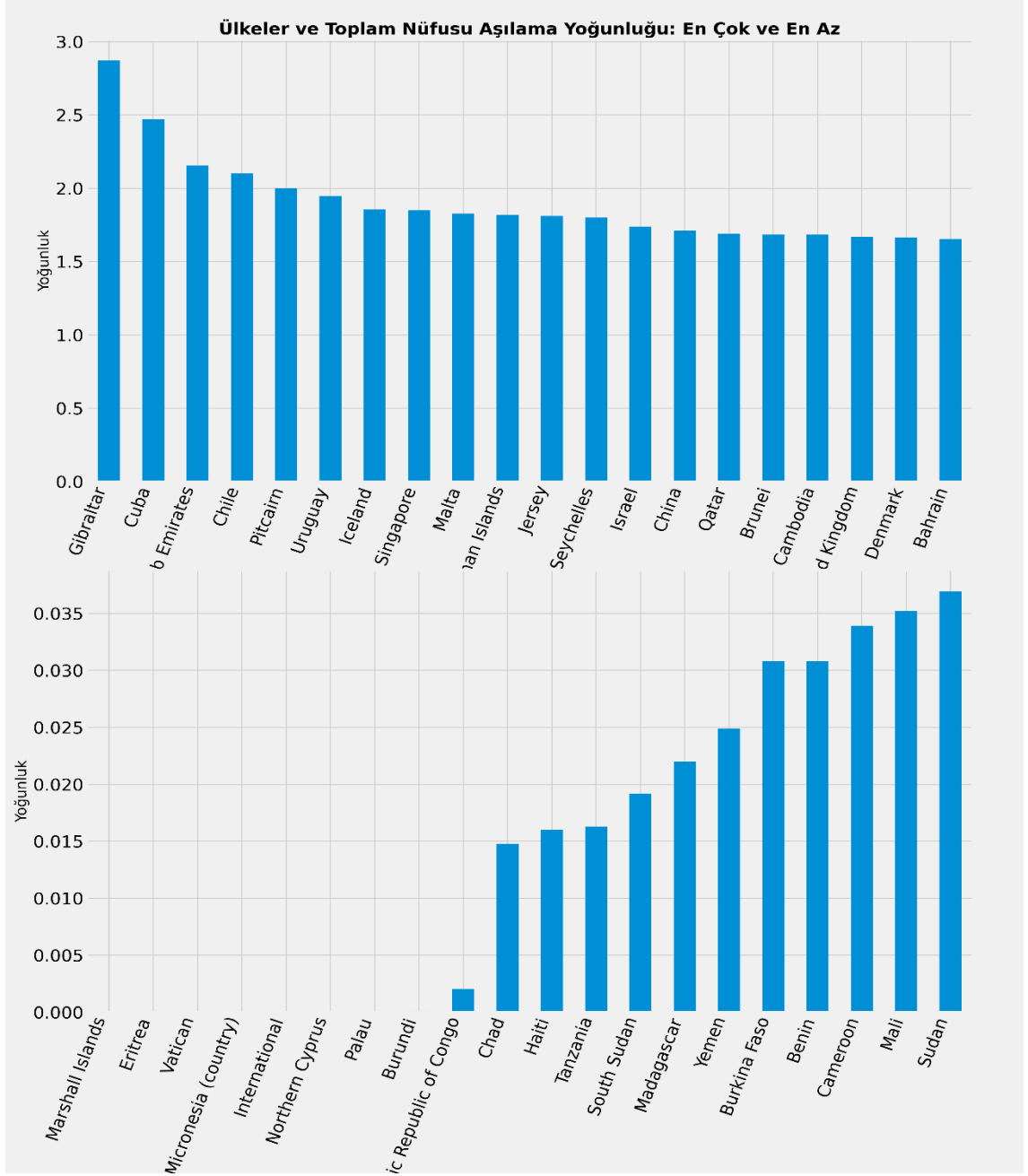


Şekil 16 Dünya Zamana Göre Ölüm Oranı



Şekil 17 Toplam Aşılama Sayısı (4 Ağustos 2021)

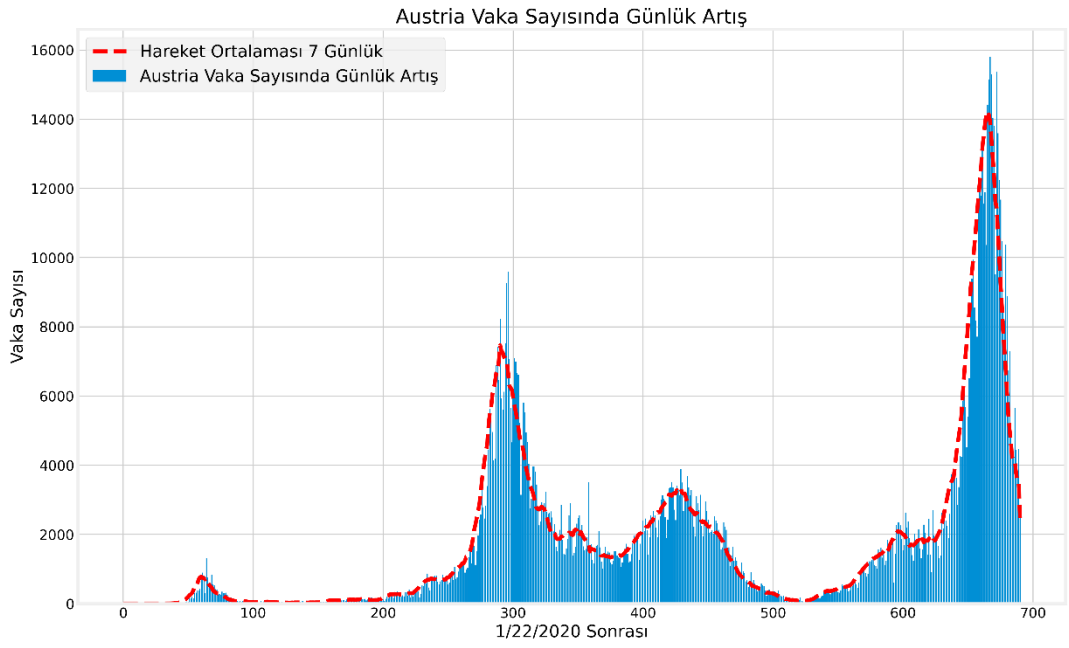
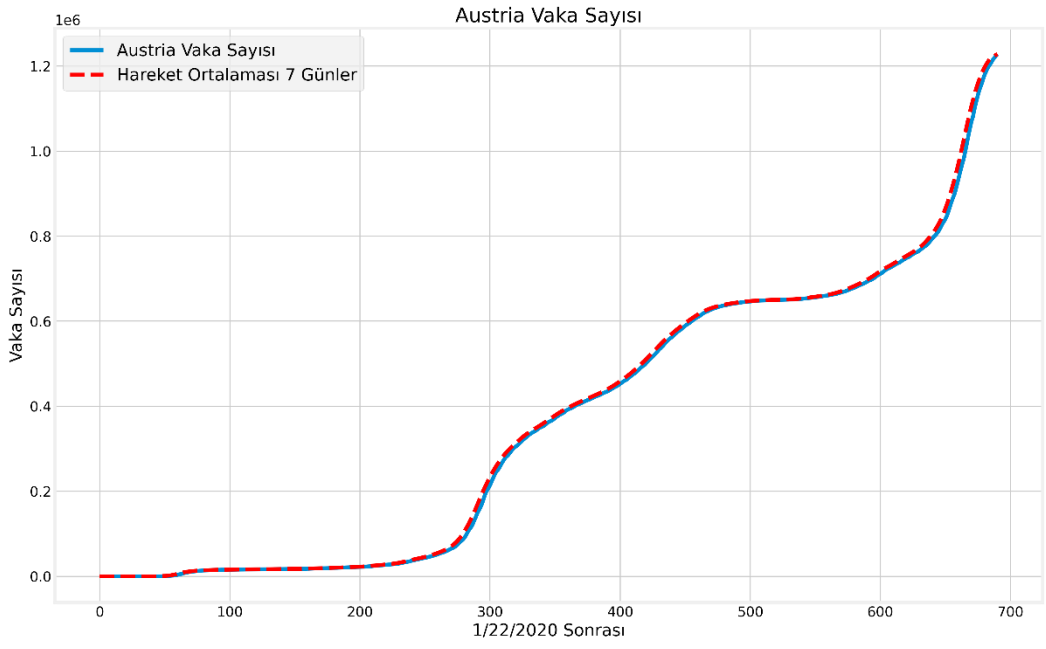
Tüm dünyadaki aşılama sayıları Şekil 17’de gösterilmektedir. Dünyanın nüfus bakımından en büyük 3 ülkesi Çin, Hindistan ve Amerika Birleşik Devletleri ilk 3 sırayı paylaşırken geri kalan sıralamada ekonomik ve sosyal olaylardan dolayı değişiklik göstermektedir. Şekil 11’in alttaki grafiği incelendiğinde ise fakir ülkelerin ve halklarının aşuya ulaşmaktaki zorluğu gözlemlenmektedir.



Şekil 18 Nüfusa Göre Aşılama Yoğunluğu (4 Ağustos 2021)

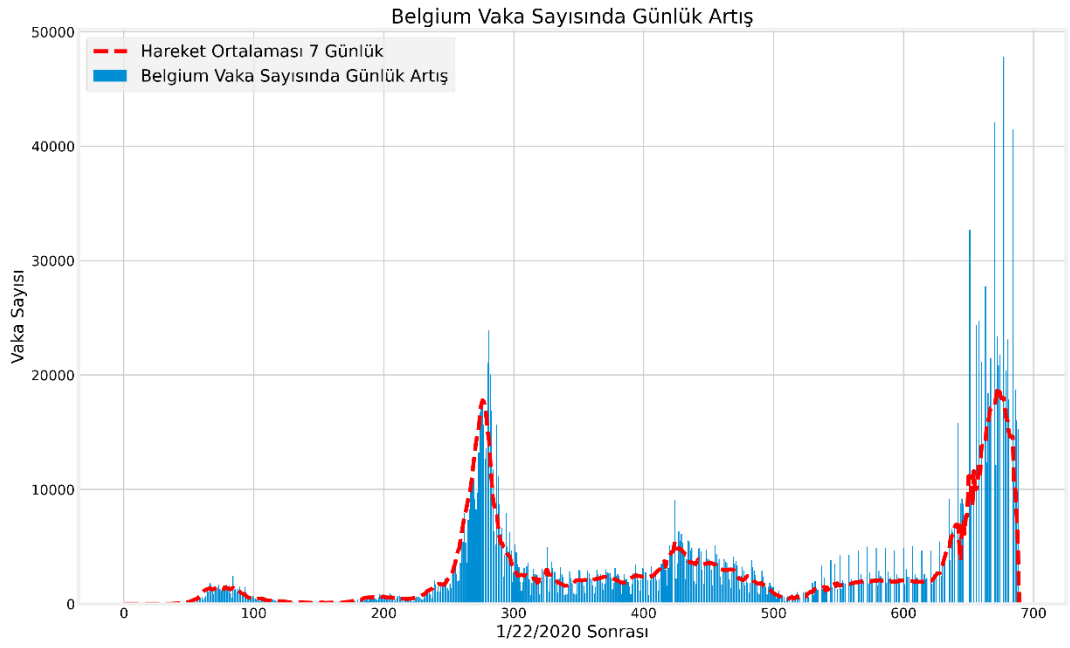
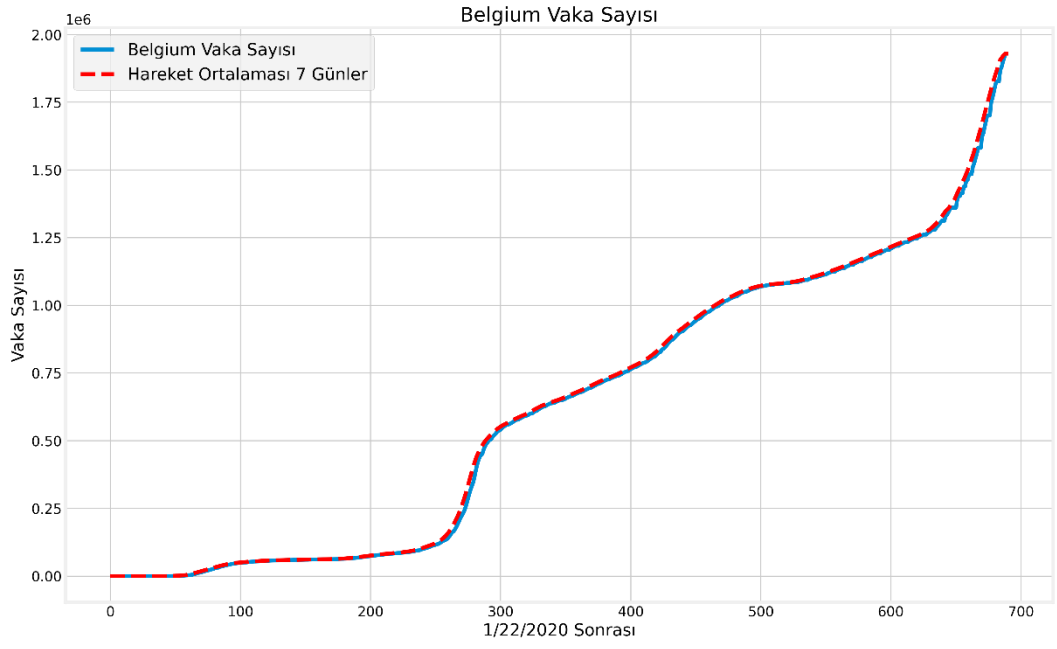
Şekil 18’de aşılama yoğunluğu grafikleri incelendiğinde nüfus olarak düşük ama maddi olarak güçlü ülkelerin aşı yoğunluğunun daha yüksek olduğu gözlemlenmiştir.

Şekil 19’da Avusturya’nın vaka sayıları ve günlük artışları gözlemlenmiştir.



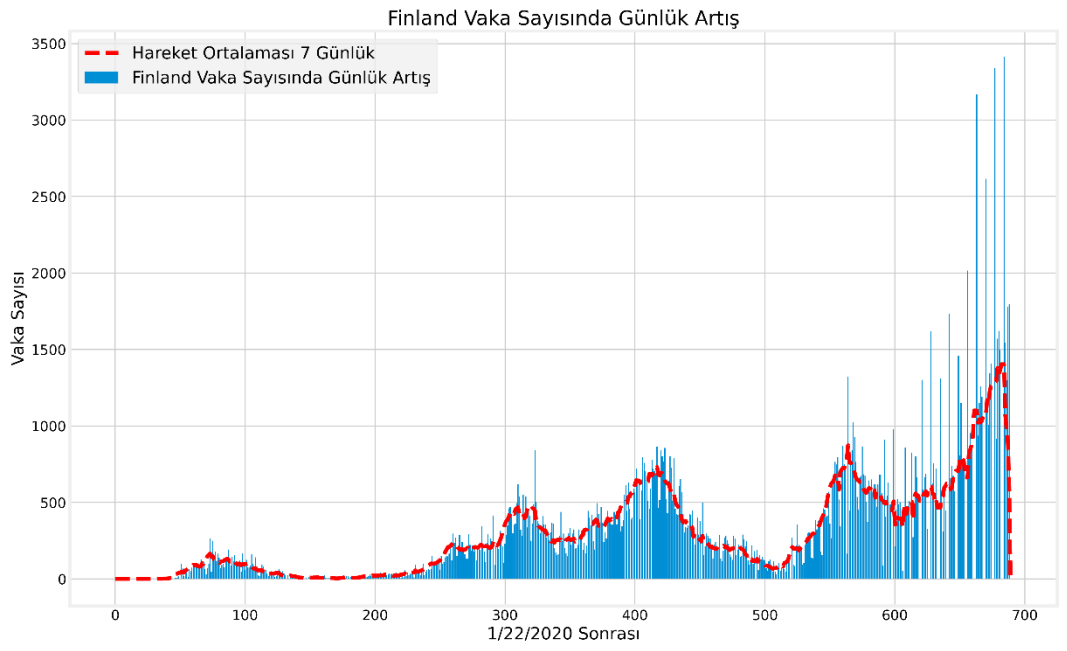
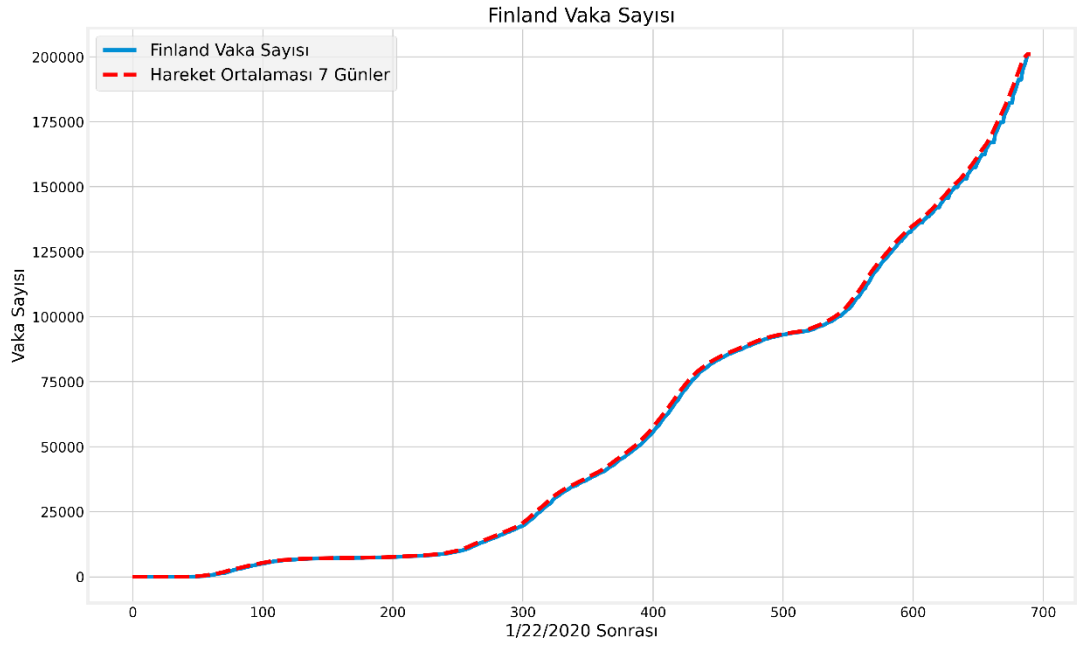
Şekil 19 Avusturya Vaka Sayısı ve Günlük Artışı

Şekil 20’de Belçika’nın vaka sayıları ve günlük artışları gözlemlenmiştir.



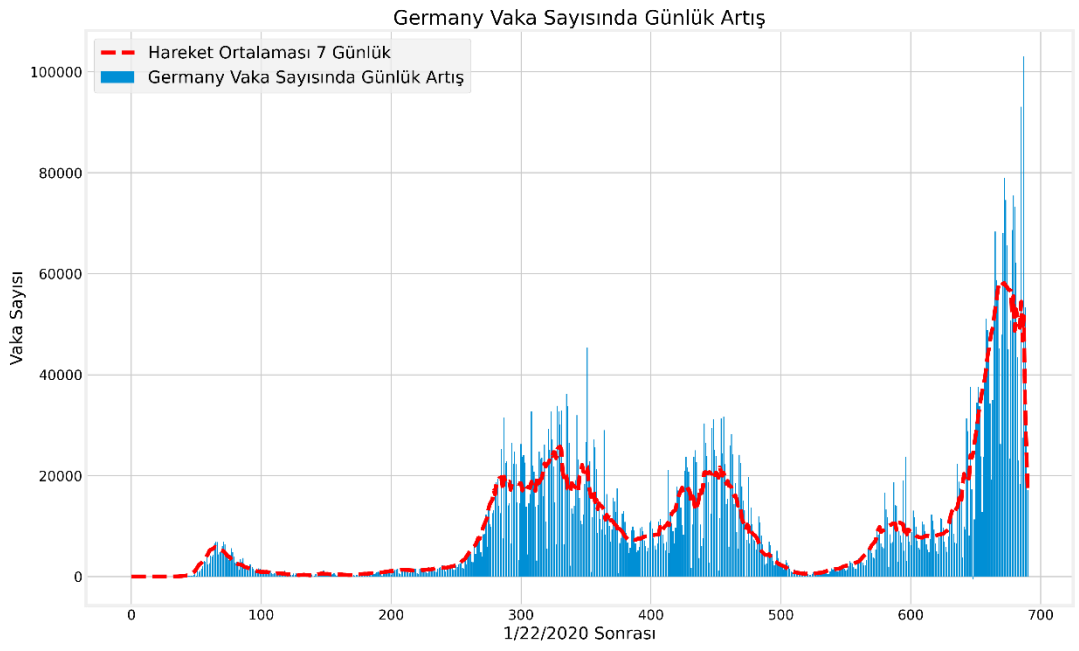
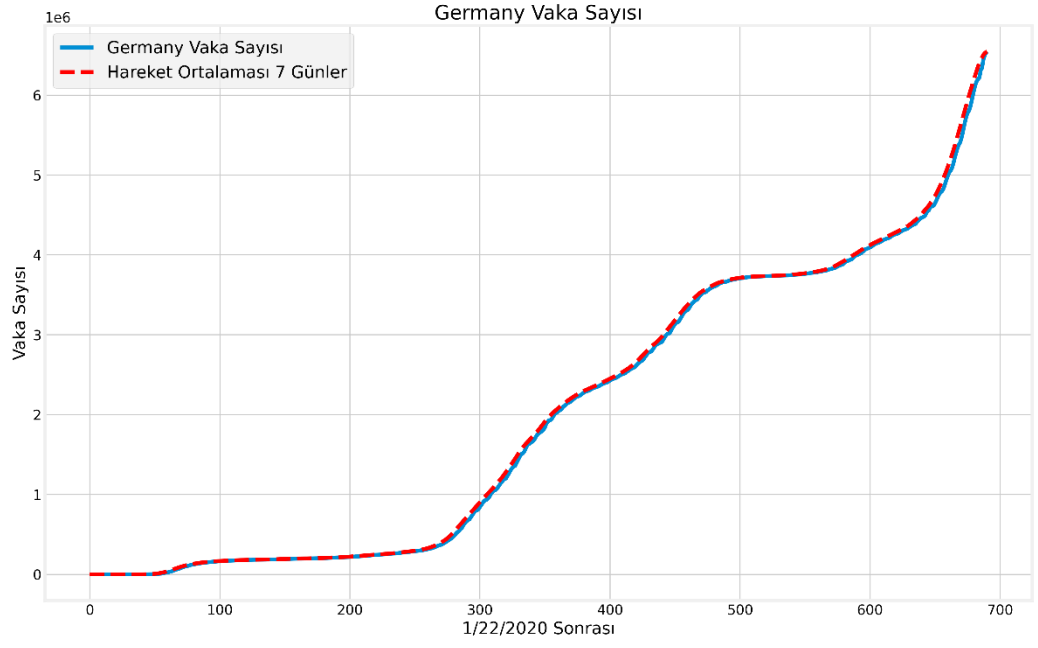
Şekil 20 Belçika Vaka Sayısı ve Günlük Artışı

Şekil 21’de Finlandiya’nın vaka sayıları ve günlük artışları gözlemlenmiştir.



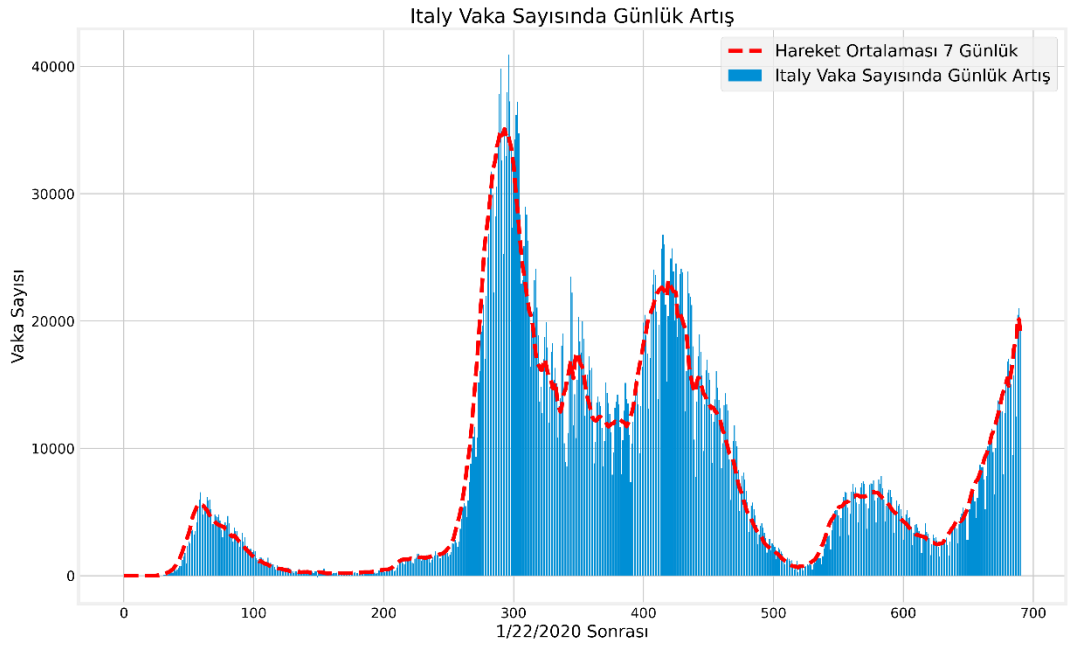
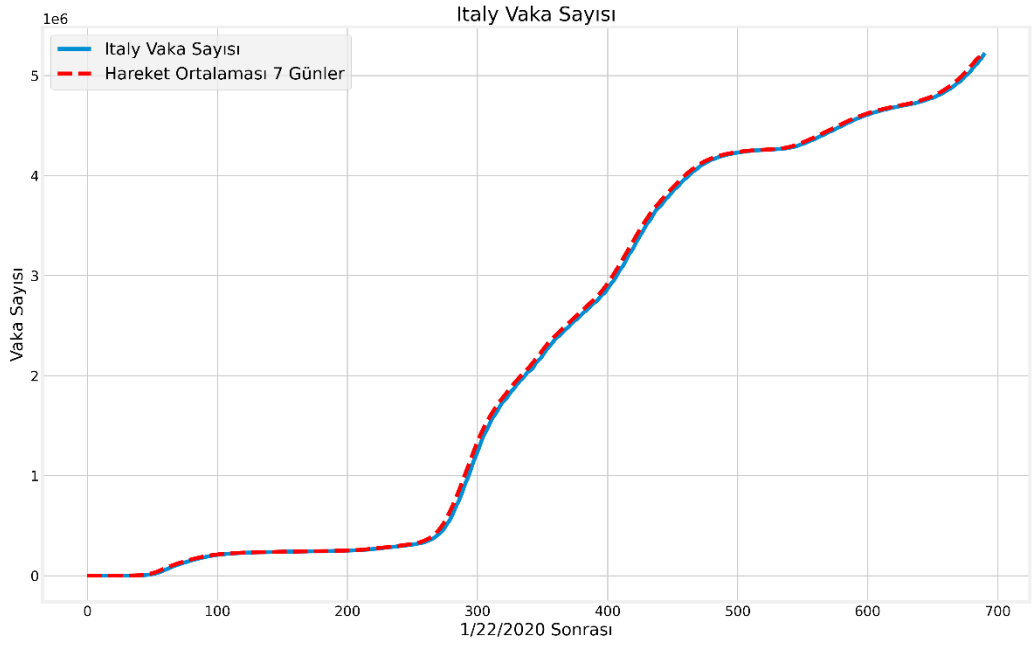
Şekil 21 Finlandiya Vaka Sayısı ve Günlük Artışı

Şekil 22’de Almanya’nın vaka sayıları ve günlük artışları gözlemlenmiştir.



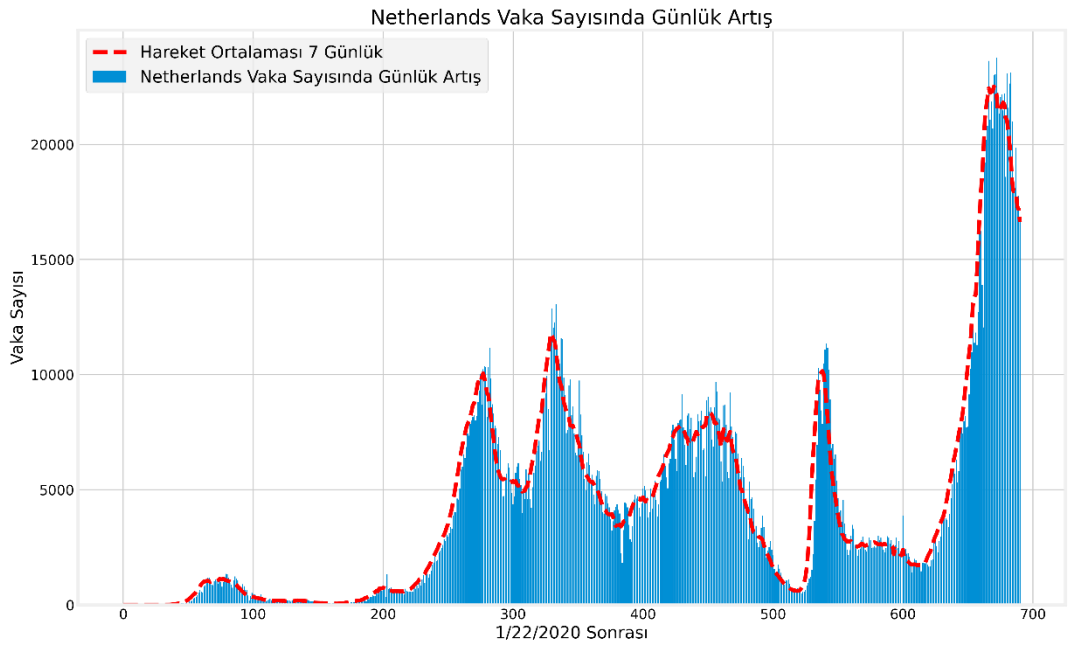
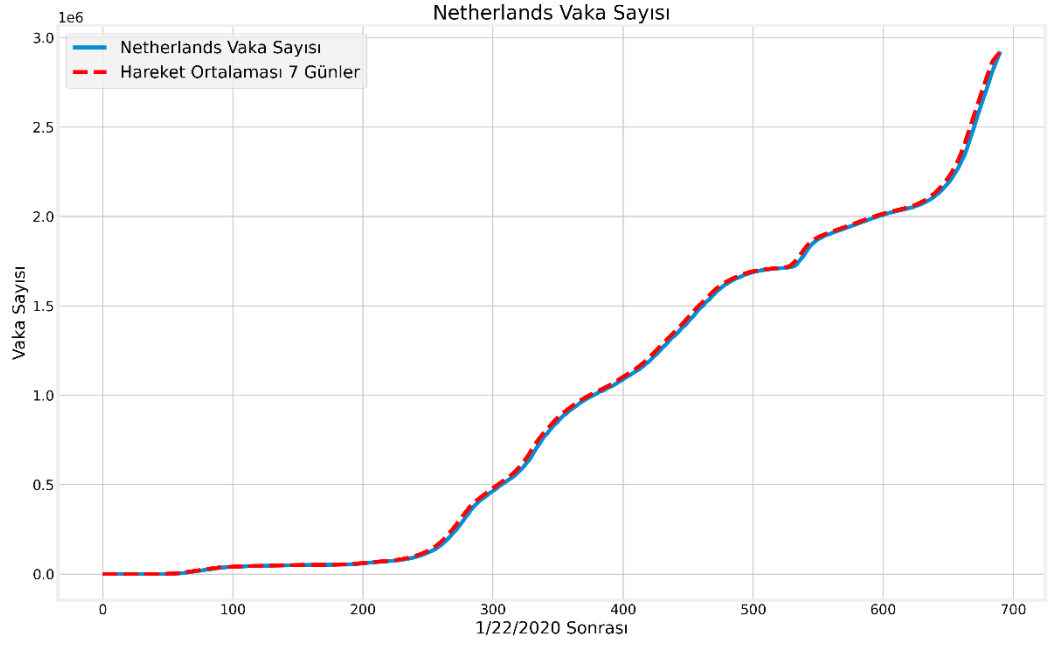
Şekil 22 Almanya Vaka Sayısı ve Günlük Artışı

Şekil 23'de İtalya'nın vaka sayıları ve günlük artışları gözlemlenmiştir.



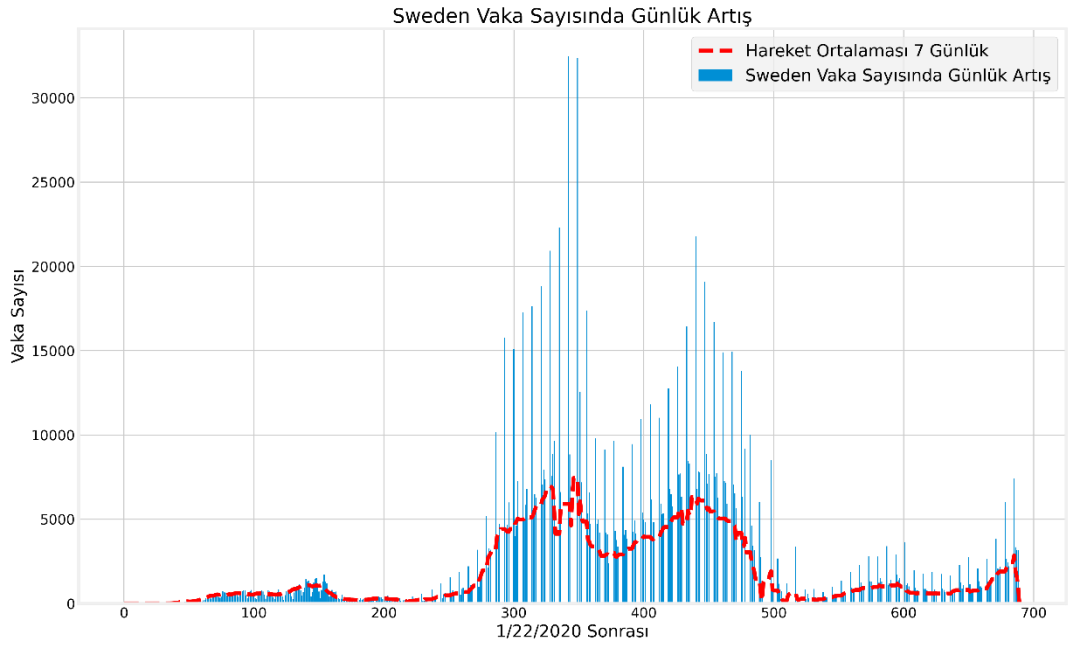
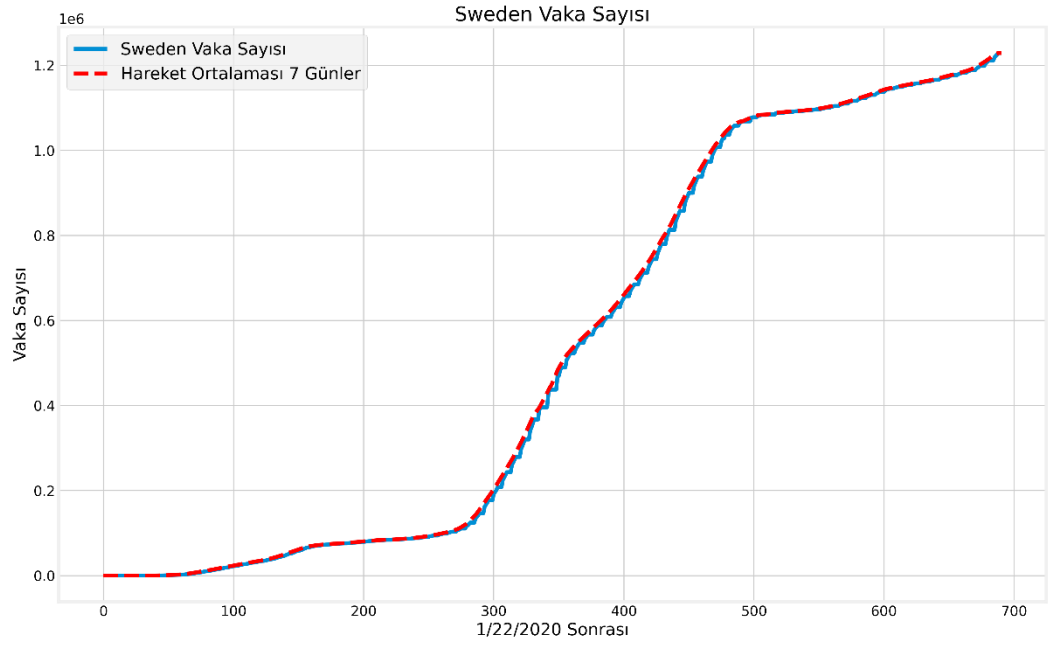
Şekil 23 İtalya Vaka Sayısı ve Günlük Artışı

Şekil 24’de Hollanda’nın vaka sayıları ve günlük artışları gözlemlenmiştir.



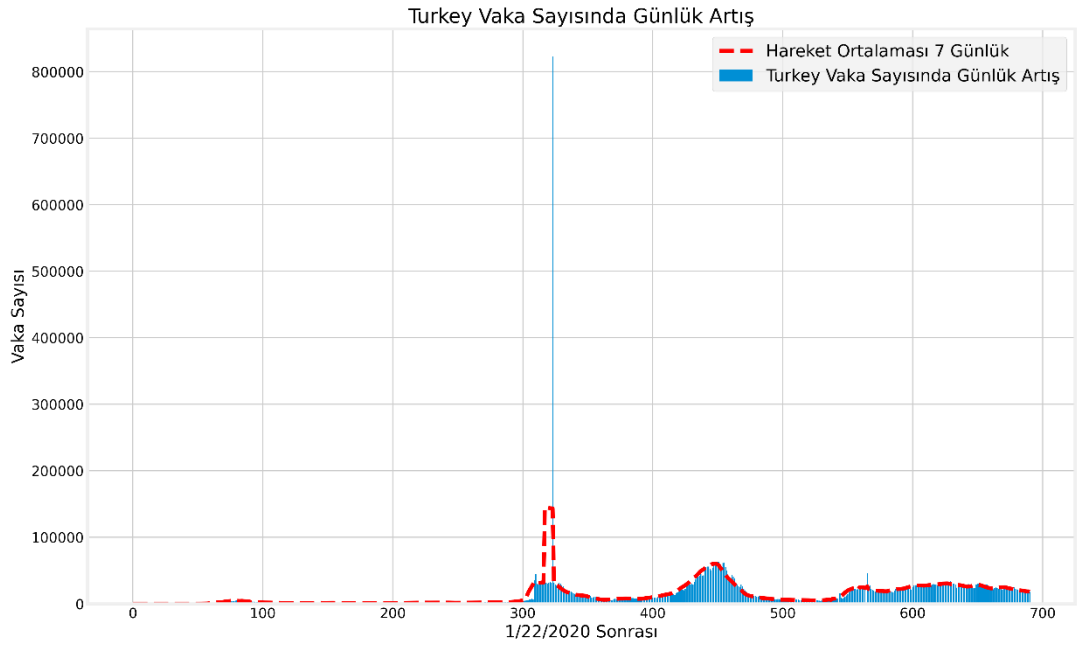
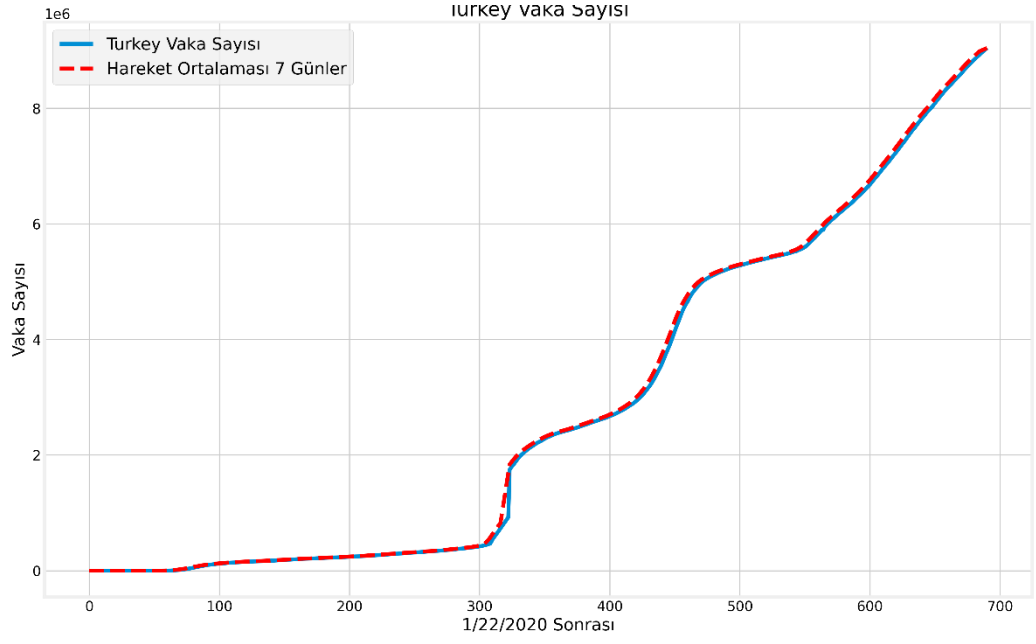
Şekil 24 Hollanda Vaka Sayısı ve Günlük Artışı

Şekil 25’de İsviçre’nin vaka sayıları ve günlük artışları gözlemlenmiştir.



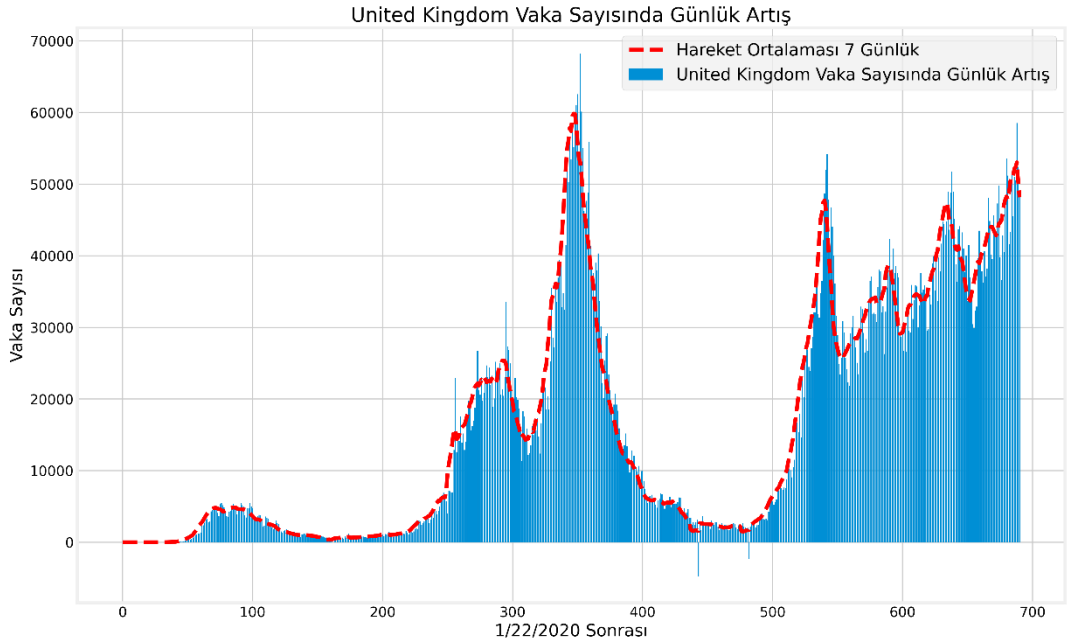
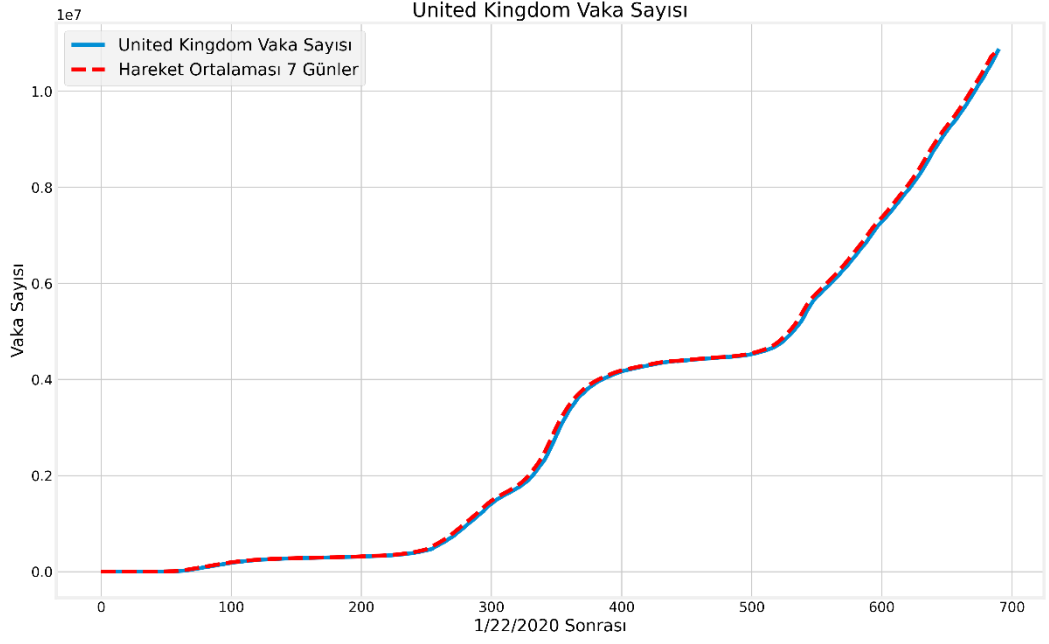
Şekil 25 İsviçre Vaka Sayısı ve Günlük Artışı

Şekil 26'da Türkiye'nin vaka sayıları ve günlük artışları gözlemlenmiştir.



Şekil 26 Türkiye Vaka Sayısı ve Günlük Artışı

Şekil 27’de Birleşik Krallığın vaka sayıları ve günlük artışları gözlemlenmiştir.



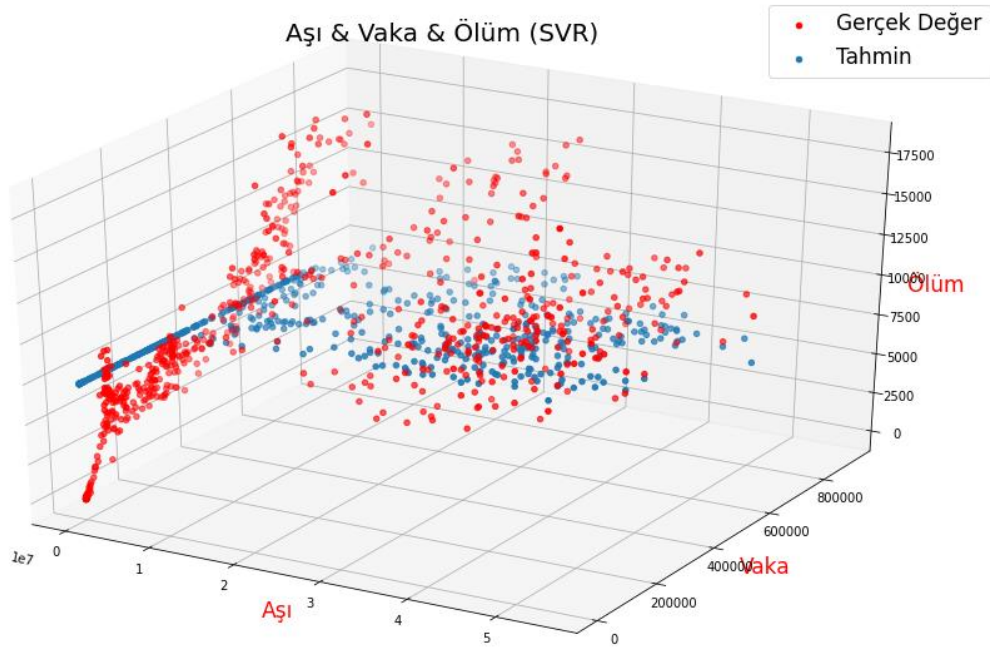
Şekil 27 Birleşik Krallık Vaka Sayısı ve Günlük Artışı

Şekil 19'dan Şekil 27'ye kadar olan vaka sayıları ve günlük vaka sayısı artış grafikleri incelendiğinde Şekil 26'daki Türkiye grafiği hariç tüm grafiklerin benzer seyrettiği görülmektedir. Bu tablolardan bütün ülkelerdeki vaka sayılarının nüfustan bağımsız olarak sosyo-kültürel ve ekonomik benzerlik gösteren ülkelerdeki vaka sayı grafiklerinin birbirine çok benzediği görülmekte ve sayıların nüfusa oranla değiştiği görülmektedir. Şekil 20'de görüleceği üzere, 300-400 günleri arasında Türkiye'nin bir günlük pik yapan

değerinin ise o zamana kadar açıklanmayan verilerin hepsinin o gün içine dahil edilerek açıklanmasından kaynaklanmaktadır.

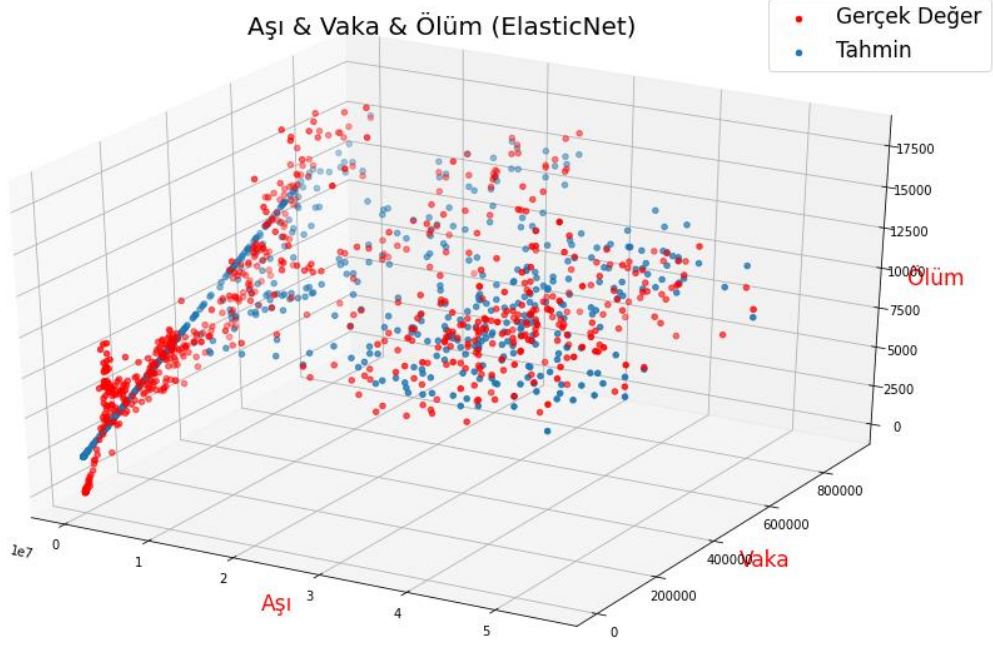
Şekil 28 ile Şekil 34 arasında 3 boyutlu olarak Aşı-Vaka-Ölüm değerleri görselleştirilmiştir. Her bir grafik kırmızı noktalar ile gerçek değerleri, mavi noktalar ile de farklı algoritmaların tahmin değerlerini göstermektedir.

Şekil 28'deki SVR modeli grafiği ölüm sayılarının arttığı zamanlarda yanlış tahmin yürüttüğünü göstermektedir.

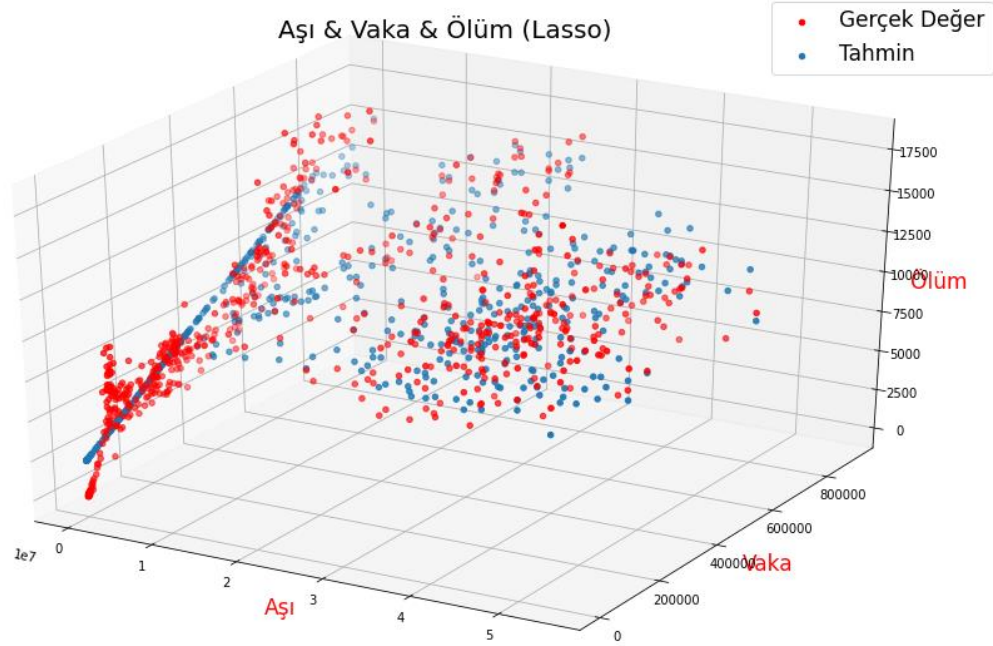


Şekil 28. SVR Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği

Şekil 29'de ElasticNet modeli tahmin ile aşı, vaka ve ölüme bağlı 3 boyutlu grafiği gözlemlenmiştir. ElasticNet Şekil 28'deki SVR algoritmasının aksine vaka artışı olduğu zamanlarda daha tutarlı tahminler yapabilmektedir.



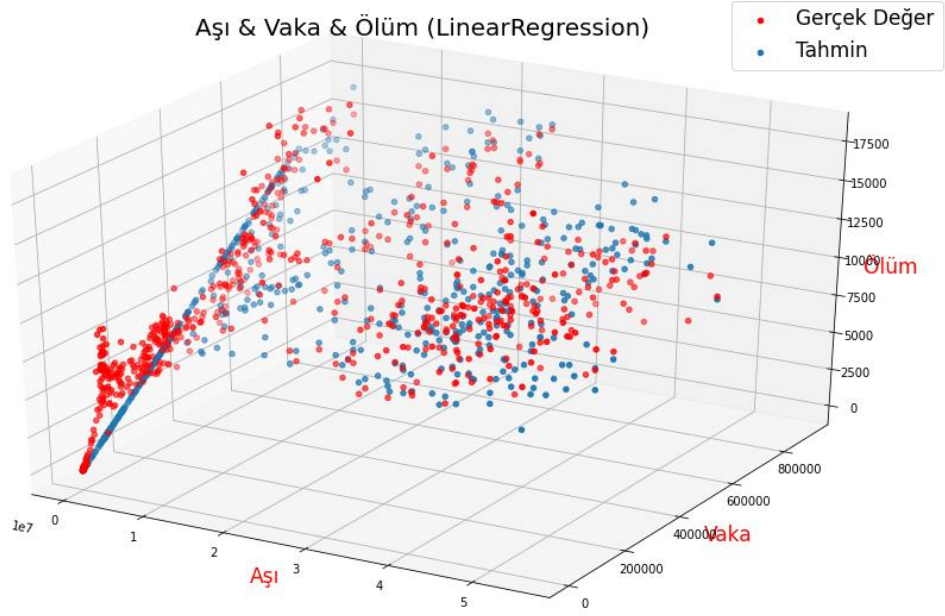
Şekil 29. ElasticNet Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği



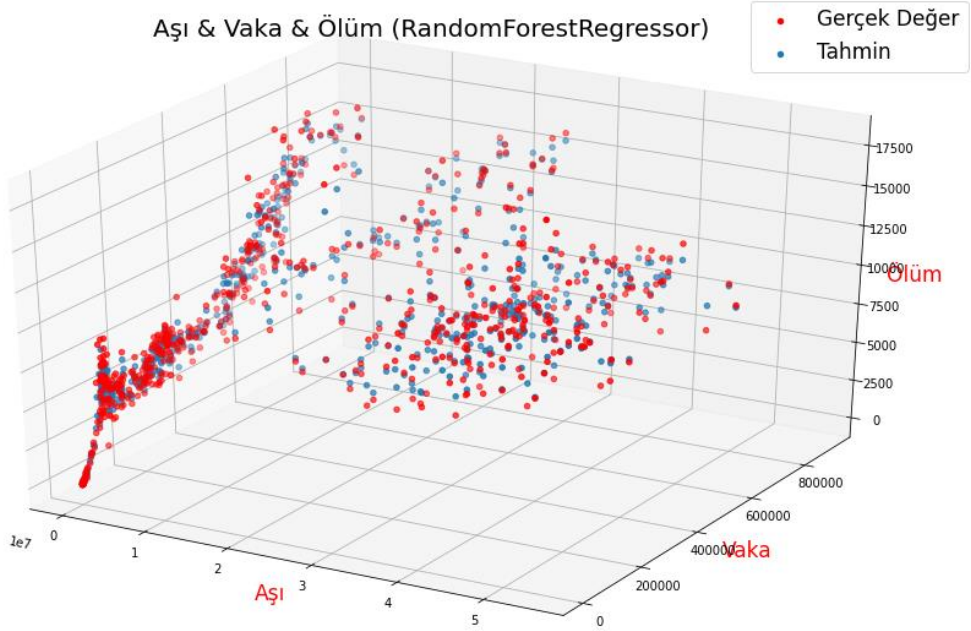
Şekil 30. Lasso Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği

Şekil 30'daki Lasso modelinin vakaların arttığı durumlarda yüksek hata değerleri verdiği gözlemlenmiştir.

Şekil 31'deki Linear Regresyon modelinin en büyük hata değerlerine sahip olduğu gözlemlenmiştir.



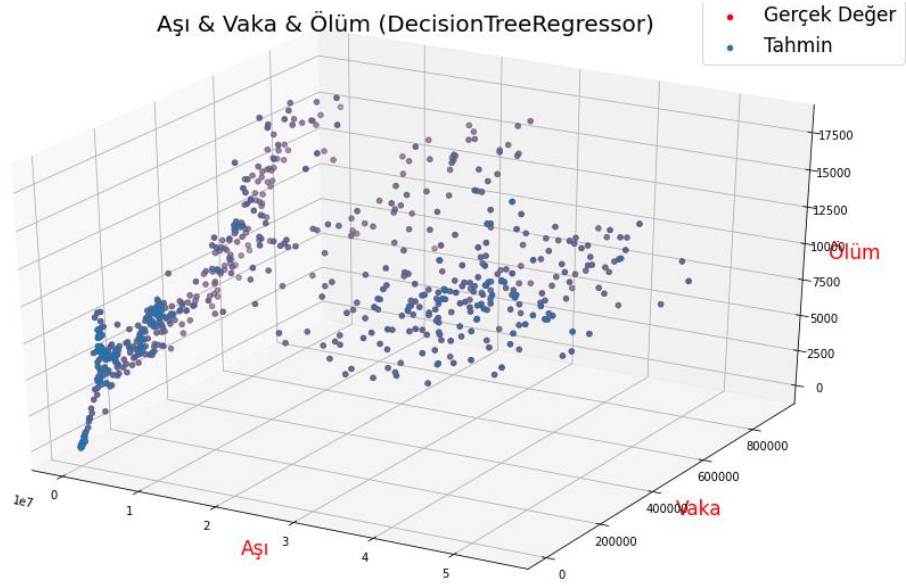
Şekil 31. Linear Regresyon Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği



Şekil 32. Random Forest Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği

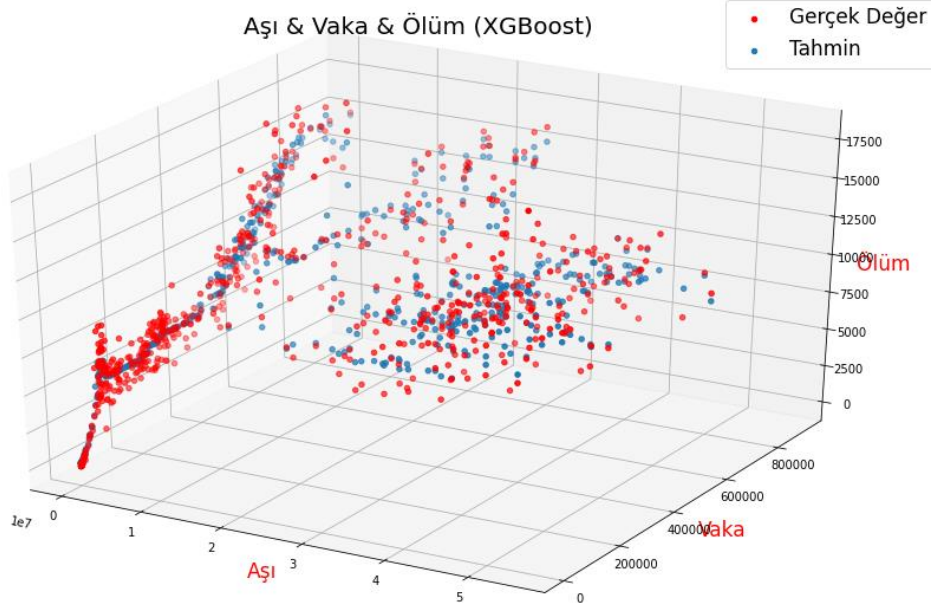
Şekil 32'de Random forest modelinin hata değerlerinin düşük olduğu gözlemlenmiştir.

Şekil 33’de Decision tree modelinin gerçek değerlere yakın tahminler yaptığı gözlemlenmiştir.



Şekil 33. Decision Tree Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği

Şekil 34’de XGBoost modeli, verinin artmasıyla doğru tahminler yaptığı gözlemlenmiştir.



Şekil 34. XGBoost Algoritması ile Aşı ve Vaka Sayıları Kullanılarak Ölüm Sayısının Tahmin Grafiği

Şekil 28-34 arası incelendiğinde aşılamanın artmasıyla beraber vakalar karşısındaki ölüm oranlarının azaldığı grafiklerden görülebilmekle birlikte aşı, vaka ve ölüm arasındaki ilişkinin farklı makine öğrenmesi yöntemleriyle tahmin edilebildiği gösterilmiştir. Aşının 0 ve 0'a yakın olduğu durumda lineer regresyon grafiğinde görüleceği üzere ölüm ve vaka arasındaki lineer bir ilişki olduğu grafikten gözlemlenmiştir.

Şekil 28'de SVR algoritmasıyla yüksek ölüm oranlarının düşük doğrulukta tespit edilebildiği görülmektedir. Şekil 29'da ElasticNet algoritması aşının 0 olduğu durumda lineer bir tahmin göstermiş ve vaka ölüm oranlarını daha yakın tespit edebildiği gözlemlenmiştir. Şekil 30'de Lasso ve Şekil 21'de lineer regresyon algoritmalarında benzer şekilde lineer tahminlerle gerçekleştiği gözlemlenmiştir. Şekil 32'de Random forest ve Şekil 33'de Decision tree ağaç tabanlı algoritmaların lineer yöntemlere göre daha iyi olduğu gözlemlenmiştir.

Yukarıda görsel çıktıları verilen algoritmaların hata oranları Tablo 2'de listelenmektedir.

Tablo 2. Farklı Makine Öğrenmesi Modellerinin Aşı ve Vaka Girdileri ile Ölüm Oranı Tahminlerinin Hata Oranları

Modeller	Ortalama Mutlak Hata	Ortalama Kareler Hatası	Kök Ortalama Kare Hatası
Polinomsal Regresyon	0.063	0.015	0.123
XGBoost Regresyonu	0.067	0.010	0.101
Random Forest Regresyonu	0.074	0.012	0.110
SVM Regresyonu	0.083	0.012	0.110
Karar Ağaç Regresyonu	0.087	0.016	0.128
LASSO Regresyonu	0.188	0.058	0.241
ElasticNet Regresyonu	0.188	0.058	0.241
Lineer Regresyonu	0.468	0.271	0.521

Hata oranları incelendiğinde Polinomsal Regresyon modelinin en düşük Ortalama Mutlak Hata'ya, Random Forest ve SVM regresyonu yöntemlerinin en düşük Ortalama Kare Hata'ya, XGBoost regresyonun ise en düşük Kök Ortalama Kare Hatasına sahip olduğu gözlenmektedir. Lineer regresyon modelinin diğer tüm regresyon modellerine göre daha kötü sonuç verdiği gözlenmiştir.

5. SONUÇ

2019 yılının Aralık ayından beri devam etmekte olan Covid-19 salgını tüm dünyayı ekonomik, sosyal, kültürel olarak etkilemiş, 200 milyondan fazla kişiye bulaşmış ve 4 milyondan fazla kişinin ölümüne sebep olmuştur. Salgınla dünya ilk defa yüzleşmektedir ve bunun nedeni gelişen teknolojiler sayesinde binlerce kilometrelik mesafeleri bize yakınlaştıran sistemler ve bu virüsün insandan insana soluduğumuz hava ile bulaşmasıdır. İnternetin, bilgisayarların ve dijital aletlerin hayatın her alanında kullanıldığı ve her şeyin veri merkezlerinde saklanabildiği bu çağda verinin değeri her geçen gün artmakta ve bu artış artarak devam etmektedir. Bu salgında üniversitelerin ve bilim merkezlerinin, salgın nedeniyle insanlar üzerinde etkili olan ve olabilecek olan verileri çok dikkatli bir şekilde toplanmış ve bu verilerin organizasyonları sayesinde çok hızlı bir şekilde büyük veri setlerinin oluşturulması sağlanmıştır.

Veri kümeleri başlarda sosyal mesafenin etkinliğini, ülkelerin hastanelerinin doluluk oranlarını, burada kullanılan ilaçların etkinliğini, sokağa çıkma yasaklarının verimliliğini vs. gibi birçok etkeni ölçmek için veri bilimciler tarafından kullanılmıştır. Özellikle aşının şimdiye kadar bulunan aşı sürelerinden (yaklaşık olarak 4 yıl) çok daha kısa bir sürede (2020, Aralık) bulunmasında veri kaynaklarının çok büyük rolü olmuştur. 2010 yılında ciddi ivme kazanan ve fiyatları düşmeye başlayan hesaplama cihazlarının (yani bilgisayarların) performans olarak güçlenmeye başlaması ve tüm dünyada bilim insanları ve yazılımcıların yapay zeka, makine öğrenmesi ve veri bilimi gibi alanlara yönelmesiyle bilgi yükünün artması da verinin bu pandemi sürecindeki önemini arttırmıştır.

Bu tez çalışmasında, Avrupa ülkeleri ve Türkiye dahilinde yapılan makine öğrenmesi ve veri bilimine dayalı analizlerde nüfusa bağlı olmaksızın virüs dağılımının benzer şekilde ilerlediği görülmüştür.

Veri bilimi yöntemleri salgın sürecinde verimli bir şekilde tüm dünyada kullanılmaya çalışılmış ve etkinliğini kanıtlamıştır. Aşıların bulunma sürecinden kapanmaların verimliliğinin ölçülmesine kadar birçok alanda kullanılmıştır.

Bu tez kapsamında regresyon tekniklerinin verinin doğru toplanması durumunda geleceğe yönelik gerçeğe yakın tahminler üretebileceği gözlemlenmiştir. Tüm dünyanın anlık verileri ortak bir veri kümesinde toplanabildiği ve bunun insanlık için kullanıma sunulduğu gözlemlenmiştir.

Ayrıca, test ve aşı sayılarının ülkelerin ekonomik durumuna bağlı olarak değiştiği, düşük ekonomik değerlere sahip ülkelerin sağlık ve veri altyapısının yetersiz kaldığı, yüksek ekonomik değerlere sahip ülkelerin de iyi altyapılara sahip olduğu gözlemlenmiştir.

Dünya veya ülkeler genelindeki vaka sayının tahmin için regresyon yöntemlerinin yanı sıra bir boyutlu verilerin analizinde kullanılan derin öğrenme yöntemlerinden yararlanılmasının gelecek çalışmalar için önemli bir alternatif olduğu değerlendirilmektedir. Bu kapsamda bir boyutlu evrişimli sinir ağları, evrişimli tekrarlayan sinir ağları, uzun-kısa süreli bellek ağları ve yerel olarak bağlanmış rastgele ağların önemli derin öğrenme modelleri olduğu değerlendirilmektedir. Ayrıca söz konusu yöntemlerin birlikte kullanıldığı hibrit modellerin geliştirilmesi de tahmin sonuçlarının geliştirilmesinde önemli bir alternatif olarak görülmektedir.

KAYNAKLAR

- A. M. Turing, I.—Computing Machinery and Intelligence, *Mind*, Volume LIX, Issue 236, October 1950, Pages 433–460, <https://doi.org/10.1093/mind/LIX.236.433>
- Ankaralı, H., Ankaralı, S., & Erarslan, N. J. A. K. T. B. D. (2020). Covid-19, SARS-CoV2, Enfeksiyonu: Güncel Epidemiyolojik Analiz ve Hastalık Seyrinin Modellemesi. 25(Supplement 1), 1-22.
- Basu, S., & Campbell, R. H. (2020). Going by the numbers: Learning and modeling COVID-19 disease dynamics. *Chaos, Solitons & Fractals*, 138, 110140.
- Bhardwaj, R., & Bangia, A. (2020). Data driven estimation of novel COVID-19 transmission risks through hybrid soft-computing techniques. *Chaos, solitons, and fractals*, 140, 110152. <https://doi.org/10.1016/j.chaos.2020.110152>
- Chen, J., & See, K. C. (2020). Artificial Intelligence for COVID-19: Rapid Review. *J Med Internet Res*, 22(10), e21476. doi:10.2196/21476
- Fauci, A. S., Lane, H. C., & Redfield, R. R. (2020). Covid-19 — Navigating the Uncharted. 382(13), 1268-1269. doi:10.1056/NEJMe2002387
- Ferhat, Ş. J. M. A. Ü. F. B. D. (2020) Covid-19'un Lineer Hücresel Dönüşümler Yardımıyla Matematiksel Modellenmesi. 8(2), 763-768.
- Giovagnoni, A. J. L. R. M. (2020). Facing the COVID-19 emergency: we can and we do. 1-2.
- Github User Content. (t.y.). Owid COVID Dataset. Erişim adresi: <https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/owid-covid-data.csv>
- Github. (t.y.). Yusuf Behzat Kavak Yüksek Lisans Tezi Kaynak Kodları. Erişim adresi: <https://github.com/yusufbehzat/yukseklisans-tezi>
- Hill, T. P., & Kertz, R. P. J. T. A. o. P. (1982). Comparisons of stop rule and supremum expectations of iid random variables. 10(2), 336-345.
- Jones, G. L. J. P. s. (2004). On the Markov chain central limit theorem. 1, 299-320.
- Kassani, S. H., Kassani, P. H., Wesolowski, M. J., Schneider, K. A., & Deters, R. J. a. p. a. (2020). Automatic Detection of Coronavirus Disease (COVID-19) in X-ray and CT Images: A Machine Learning-Based Approach.
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. J. J. o. E. S. P. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. 49(4), 764-766.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., . . . Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. 382(13), 1199-1207. doi:10.1056/NEJMoa2001316
- Maghdid, H. S., Zrar Ghafoor, K., Safaa Sadiq, A., Curran, K., Rawat, D. B., & Rabie, K. J. a. e.-p. (2020). A Novel AI-enabled Framework to Diagnose Coronavirus COVID 19 using Smartphone Embedded Sensors: Design Study. arXiv:2003.07434. <https://ui.adsabs.harvard.edu/abs/2020arXiv200307434M>

- Mahdy, L. N., Ezzat, K. A., Elmousalami, H. H., Ella, H. A., & Hassanien, A. E. (2020). Automatic X-ray COVID-19 Lung Image Classification System based on Multi-Level Thresholding and Support Vector Machine. 2020.2003.2030.20047787. doi:10.1101/2020.03.30.20047787 %J medRxiv
- Price, G. R. J. N. (1970). Selection and covariance. 227(5257), 520-521.
- S. Latif *et al.*, "Leveraging Data Science to Combat COVID-19: A Comprehensive Review," in *IEEE Transactions on Artificial Intelligence*, vol. 1, no. 1, pp. 85-103, Aug. 2020, doi: 10.1109/TAI.2020.3020521.
- Shemesh, O., Polak, P., Lundin, K. E. A., Sollid, L. M., & Yaari, G. (2020). Machine learning analysis of naïve B-cell receptor repertoires stratifies celiac disease patients and controls. 2020.2011.2009.371336. doi:10.1101/2020.11.09.371336 %J bioRxiv
- Sun, L., Song, F., Shi, N., Liu, F., Li, S., Li, P., . . . Shi, Y. (2020). Combination of four clinical indicators predicts the severe/critical symptom of patients infected COVID-19. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, 128, 104431-104431. doi:10.1016/j.jcv.2020.104431
- Togay, A., & Yilmaz, N. J. T. E. A. H. D. (2020). SARS-CoV-2'nin Laboratuvar Tanısı. 30, 70-75.
- Ulhaq, A., & Burmeister, O. J. a. e.-p. (2020). COVID-19 Imaging Data Privacy by Federated Learning Design: A Theoretical Framework. arXiv:2010.06177. <https://ui.adsabs.harvard.edu/abs/2020arXiv201006177U>
- Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Tropical medicine & international health : TM & IH*, 25(3), 278-280. doi:10.1111/tmi.13383
- Wang, P., Zheng, X., Li, J., & Zhu, B. (2020). Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics. *Chaos, Solitons & Fractals*, 139, 110058. doi:<https://doi.org/10.1016/j.chaos.2020.110058>
- Wang, S., Sun, Q., Xu, Y., Pei, J., & Lai, L. (2020). A Transferable Deep Learning Approach to Fast Screen Potent Antiviral Drugs against SARS-CoV-2. 2020.2008.2028.271569. doi:10.1101/2020.08.28.271569 %J bioRxiv
- World Health, O. (2020). *Laboratory testing of 2019 novel coronavirus (2019-nCoV) in suspected human cases: interim guidance, 17 January 2020*. Geneva: World Health Organization.
- Yang, W., & Yan, F. (2020). Patients with RT-PCR-confirmed COVID-19 and Normal Chest CT. 295(2), E3-E3. doi:10.1148/radiol.2020200702
- Yasar, H., & Ceylan, M. (2021). A new deep learning pipeline to detect Covid-19 on chest X-ray images using local binary pattern, dual tree complex wavelet transform and convolutional neural networks. *Applied Intelligence*, 51(5), 2740-2763.
- Yasar, H., & Ceylan, M. (2021). A novel comparative study for detection of Covid-19 on CT lung images using texture analysis, machine learning, and deep learning methods. *Multimedia Tools and Applications*, 80(4), 5423-5447.

- Yasar, H., & Ceylan, M. (2021). Deep Learning–Based Approaches to Improve Classification Parameters for Diagnosing COVID-19 from CT Images. *Cognitive Computation*, 1-28.
- Zeroual, A., Harrou, F., Dairi, A., & Sun, Y. (2020). Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study. *Chaos, Solitons & Fractals*, 140, 110121. doi:<https://doi.org/10.1016/j.chaos.2020.110121>
- Zhang, X., Beinke, B., Al Kindhi, B., & Wiering, M. J. a. e.-p. (2020). Comparing Machine Learning Algorithms with or without Feature Extraction for DNA Classification. arXiv:2011.00485. <https://ui.adsabs.harvard.edu/abs/2020arXiv201100485Z>
- Zhang, Z., & Jain, S. (2020). Mathematical model of Ebola and Covid-19 with fractional differential operators: Non-Markovian process and class for virus pathogen in the environment. *Chaos, Solitons & Fractals*, 140, 110175. doi:<https://doi.org/10.1016/j.chaos.2020.110175>
- Zhao, H., Merchant, N. N., McNulty, A., Radcliff, T., Cote, M. J., Fischer, R., . . . Ory, M. G. (2020). COVID-19: Short term prediction model using daily incidence data. 2020.2011.2023.20237024. doi:10.1101/2020.11.23.20237024 %J medRxiv

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Yusuf Behzat KAVAK

EĞİTİM DURUMU

Lisans Öğrenimi : 2016, KTO Karatay Üniversitesi, Mühendislik Fakültesi,
Elektrik-Elektronik Mühendisliği

Yüksek Lisans Öğrenimi :

Bildiği Yabancı Diller : İngilizce

Tarih: 03 Aralık 2021