



T.C.
KONYA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



**ARIMA VE DERİN ÖĞRENME MODELLERİ
KULLANILARAK TÜRKİYE'DEKİ COVID-19
VAKA VE VEFAT SAYILARININ TAHMİNİ**

Fatema NUSRAT

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Ocak-2022
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Fatema NUSRAT tarafından hazırlanan “ARIMA ve Derin Öğrenme Modelleri Kullanılarak Türkiye’deki COVID-19 Vaka ve Vefat Sayılarının Tahmini” adlı tez çalışması 20 / 01 / 2022 tarihinde aşağıdaki jüri tarafından oy birliği / ~~oy çokluğu~~ ile Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri

Başkan

Dr. Öğr. Üyesi Burak YILMAZ

Danışman

Doç. Dr. Ömer Kaan BAYKAN

Üye

Doç. Dr. Ahmet BABALIK

İmza

.....

.....

.....

Yukarıdaki sonucu onaylarım.

Prof. Dr. Saadettin Erhan KESEN
Enstitü Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Fatema NUSRAT

Tarih: 20 / 01 / 2022

ÖZET

YÜKSEK LİSANS TEZİ

ARIMA VE DERİN ÖĞRENME MODELLERİ KULLANILARAK TÜRKİYE'DEKİ COVID-19 VAKA VE VEFAT SAYILARININ TAHMİNİ

Fatema NUSRAT

**Konya Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı**

Danışman: Doç. Dr. Ömer Kaan BAYKAN

2022, 53 Sayfa

Jüri

Doç. Dr. Ömer Kaan BAYKAN

Doç. Dr. Ahmet BABALIK

Dr. Öğr. Üyesi Burak YILMAZ

Koronavirüs (Covid-19), 2019 yılında ortaya çıkan, hızlı bir şekilde yayılan, insan sağlığına etki eden dünya çapında büyük bir tehdittir. Çin'in Wuhan şehrindeki deniz ürünleri ve hayvan pazarlarından kaynaklandığı düşünülmektedir. Covid-19, bulaşıcı bir hastalıktır ve insandan insana bulaşmaktadır. 11 Mart 2020'de Dünya Sağlık Örgütü (World Health Organization- WHO), Covid-19 salgını küresel bir salgın olarak ilan etmiştir. Ülkelerin salgın kaynaklı vaka ve vefat sayılarını tahmin edebilmesi geleceğe yönelik planlama yapabilmeleri için büyük önem taşımaktadır. Bu tez çalışmasının amacı, geçmiş verilere dayanarak Türkiye'nin gelecekteki toplam Covid-19 vaka ve vefat sayılarını tahmin etmektir. Veri seti, Türkiye Cumhuriyeti Sağlık Bakanlığı'nın internet sitesindeki veriler kullanılarak oluşturulmuştur. Tahmin modelleri olarak, Uzun Kısa Süreli Bellek (Long Short Term Memory -LSTM), Çift Yönlü LSTM (Bidirectional LSTM- BiLSTM), Geçitli Tekrarlayan Birim (Gated Recurrent Unit - GRU) olmak üzere üç farklı derin öğrenme modeli ve Otoregresif Entegre Hareketli Ortalama (Autoregressive Integrated Moving Average- ARIMA) istatistiksel modeli kullanılmıştır. 11 Mart 2020 ile 31 Mayıs 2021 tarihleri arasındaki veriler modellerin eğitilmesi, test edilmesi için kullanılmış olup, 1 Haziran - 30 Haziran 2021 tarihleri arasındaki vaka ve vefat sayıları tahmin edilmiştir. Modellerin performansını değerlendirmek için Kök Ortalama Kare Hatası (Root Mean Square Error- RMSE) ve Ortalama Mutlak Yüzde Hatası (Mean Absolute Percentage Error- MAPE) kullanılmıştır. ARIMA modeli RMSE, MAPE ve tahmin değerleri açısından derin öğrenme tekniklerinden daha iyi performans göstermiştir. ARIMA modelinin ürettiği tahmin değerlerinin, Türkiye'nin gerçek vaka ve vefat sayıları ile daha uyumlu olduğu gözlenmiştir.

Anahtar Kelimeler: Covid-19, Çift Yönlü LSTM, Derin Öğrenme, Geçitli Tekrarlayan Birim, Otoregresif Entegre Hareketli Ortalama, Uzun Kısa Süreli Bellek

ABSTRACT

MS THESIS

FORECASTING OF COVID-19 CONFIRMED AND DEATH CASES IN TURKEY USING ARIMA AND DEEP LEARNING MODELS

Fatema NUSRAT

**Konya Technical University
Institute of Graduate Studies
Department of Computer Engineering**

Advisor: Assoc. Prof. Dr. Ömer Kaan BAYKAN

2022, 53 Pages

**Jury
Assoc. Prof. Dr. Ömer Kaan BAYKAN
Assoc. Prof. Dr. Ahmet BABALIK
Asst. Prof. Dr. Burak YILMAZ**

The coronavirus (Covid-19) is a major worldwide threat that emerged in 2019, spread rapidly, and affects human health. It is thought to originate from seafood and animal markets of Wuhan city, China. Covid-19 is a contagious disease and is transmitted from person to person. On March 11, 2020, the World Health Organization (WHO) declared the Covid-19 outbreak a global pandemic. It is of great importance for countries to be able to forecast the number of confirmed and death cases caused by the epidemic so that they can plan for the future. The aim of this thesis work is to forecast Covid-19 total confirmed and death cases of Turkey based on previous data. The dataset was created using the data on the website of the Ministry of Health of the Republic of Turkey. Three different deep learning models as Long Short Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), and Autoregressive Integrated Moving Average (ARIMA) statistical model was used as prediction models. Data from March 11, 2020, to May 31, 2021, were used to train and test the models, and the number of confirmed and death cases between June 1 and June 30, 2021, was forecasted. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were used to evaluate the performance of the models. The ARIMA model outperformed deep learning techniques in terms of RMSE, MAPE and forecasting values. The forecasted values produced by the ARIMA model are more compatible with the actual number of confirmed and death cases in Turkey.

Keywords: Autoregressive Integrated Moving Average, Bidirectional LSTM, Covid-19, Deep Learning, Gated Recurrent Unit, Long Short Term Memory

ÖNSÖZ

Öncelikle, “ARIMA ve Derin Öğrenme Modelleri Kullanılarak Türkiye’deki COVID-19 Vaka ve Vefat Sayılarının Tahmini” adlı bu çalışmayı tamamlamam için bana büyük güç ve cesaret verdiği için Allah'a (SWT) teşekkür etmek isterim.

Daha sonrasında bu çalışmanın bu hale gelebilmesi için bana değerli tavsiyeleri ile rehberlik eden danışmanım Doç. Dr. Ömer Kaan BAYKAN'a minnet ve şükranlarımı sunuyorum.

Son olarak araştırma çalışmalarım ve tez yazımda sürekli manevi desteklerini esirgemeyen anne babama, aile bireylerime, arkadaşlarıma teşekkür ederim.

Fatema NUSRAT
KONYA-2022

İÇİNDEKİLER

ÖZET	iv
ABSTRACT.....	v
ÖNSÖZ	vi
İÇİNDEKİLER	vii
SİMGELER VE KISALTMALAR	viii
1. GİRİŞ	1
2. KAYNAK ARAŞTIRMASI	3
3. MATERYAL VE YÖNTEM.....	7
3.1. Veri Seti	7
3.2. Derin Öğrenme Modelleri.....	8
3.2.1. Uzun kısa süreli bellek (LSTM)	8
3.2.2. Çift yönlü uzun kısa süreli bellek (BiLSTM)	10
3.2.3. Geçitli tekrarlayan birim (GRU).....	11
3.3. Zaman Serisi Analiz Tekniği:	13
3.3.1. Otoresif entegre hareketli ortalama (ARIMA):	13
3.4. Performans Değerlendirmesi	14
3.5. Önerilen Yöntem.....	14
4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA.....	18
4.1. Derin Öğrenme Yöntemleri ile Vefat Sayılarının Tahmini	19
4.2. ARIMA Modeli ile Vefat Sayılarının Tahmini.....	28
4.3. Derin Öğrenme Yöntemleri ile Vaka Sayılarının Tahmini.....	32
4.4. ARIMA Yöntemi ile Vaka Sayılarının Tahmini.....	41
5. SONUÇLAR VE ÖNERİLER.....	48
5.1 Sonuçlar	48
5.2 Öneriler	49
KAYNAKLAR	50

SİMGELER VE KISALTMALAR

Simgeler

b	: Bias
C	: Kesişim noktası
$\tilde{C}_t$	: Hafıza hücrecini güncelle
C_t	: Geçerli bellek hücresi
ε_t	: Rastgele Hata
f_t	: Unutma kapısı
h_t	: Gizli durum
$\vec{h}_t$	: İleri Gizli durumu
$\bar{h}_t$	: Geri gizli durumu
$\tilde{h}_t$	: Aday gizli durumu
i_t	: Giriş kapısı
N	: Gerçek test örneği sayısı
o_t	: Çıkış kapısı
r_t	: Sıfırlama kapısı
W	: Ağırlık
y_t	: Güncel zaman serisi
y	: Gerçek sonuç
$\hat{y}$	: Tahmin sonucu
z_t	: Güncelleme kapısı

Kısaltmalar

ARIMA	: Otoresif Entegre Hareketli Ortalama (Autoregressive Integrated Moving Average)
ACF	: Otokorelasyon Fonksiyonu (Autocorrelation Function)
AR	: Otoresif (Autoregressive)
BiLSTM	: Çift Yönlü LSTM (Bidirectional LSTM)
DL	: Derin Öğrenme (Deep Learning)
GRU	: Geçitli Tekrarlayan Birim (Gated Recurrent Unit)
LSTM	: Uzun Kısa Süreli Bellek (Long Short Term Memory)
MA	: Hareketli Ortalama (Moving Average)
MAPE	: Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error)
PACF	: Kısmi Otokorelasyon Fonksiyonu (Partial Autocorrelation Function)
RMSE	: Kök Ortalama Kare Hata (Root Mean Square Error)
WHO	: Dünya Sağlık Örgütü (World Health Organization)

1. GİRİŞ

Tarih boyunca bulaşıcı hastalıklar felaketlere yol açmış olup, tüm dünyada bulaşıcı hastalıklar hızla artmaktadır. Son on yılda, H1N1 ve MERS gibi 20'den fazla bulaşıcı ajanın neden olduğu bazı salgın hastalıklar ortaya çıkmıştır (Balkhair, 2020). SARS-CoV ve MERS-CoV gibi koronavirüslerin ortaya çıkışı, 21. yüzyılda halk sağlığı sistemlerine küresel zorluklar getirmiştir (De Wit ve ark., 2016).

Yeni koronavirüs (Covid-19) olarak adlandırılan koronavirüs, akut solunum yolu hastalığı olup insandan insana bulaşmaktadır (Gorbalenya ve ark., 2020). Covid-19'un, Aralık 2019'da, Çin'in Hubei eyaleti, Wuhan şehrindeki deniz ürünleri ve hayvan pazarlarından kaynaklandığı düşünülmektedir (Pang ve ark., 2020).

Covid-19, öncelikle Çin'de hızla yayılan bulaşıcı bir hastalıktır. Çin sağlık yetkilileri, hastalığı karakterize etmek ve kontrol etmek, hastalık şüphesi olan kişileri izole etmek, temaslıların yakından izlenmesi, hastalardan epidemiyolojik ve klinik veri toplanması ve teşhis ve tedavi prosedürlerinin geliştirilmesi için acil araştırmalar yapmıştır. Gerçek zamanlı RT-PCR testi adı verilen Covid-19 için bir tanı testi geliştirilmiştir (Wang ve ark., 2020a).

Covid-19'un en yaygın semptomları, daha az yaygın semptomları ve ciddi semptomları olmak üzere üç tür semptomu vardır. En sık görülen semptomlar ateş, öksürük, yorgunluk, tat veya koku kaybıdır. Daha az görülen semptomlar boğaz ağrısı, baş ağrısı, ağrı ve sızılar, ishal, deride kızarıklık, el ve ayak parmaklarında renk değişikliği, kırmızı veya tahriş olmuş gözlerdir. Ciddi semptomlar nefes almada zorluk, nefes darlığı, konuşma, hareket kaybı veya kafa karışıklığı, göğüs ağrısıdır (Wang ve ark., 2020b). Koronavirüs, kardiyovasküler hastalık, diyabet, kronik solunum yolu hastalığı, kanser hastalıklarına sahip insanları ve 60 yaş üstü insanları daha çok tehdit etmektedir (Paules ve ark., 2020). Mutasyonlara bağlı olarak semptomlar değişebilmektedir.

Covid-19 kaynaklı vefat oranı, yaklaşık %2,92 olup, tipik gripten daha fazladır. Grip kaynaklı vefat oranı %0,05 ila %0,1 aralığındadır. Bu, Covid-19'un yaklaşık 30 ila 60 kat daha öldürücü olduğu anlamına gelmektedir (McCarty ve DiNicolantonio, 2020). Artan yaşla birlikte vefat oranı da artmaktadır. Erken Çin raporlarına göre 80 yaş üstü yaşlı hastalarda Covid-19 kaynaklı vefat oranı 3 kat daha yüksektir (Wu ve McGoogan, 2020). İtalya'da, Covid-19 ile ilgili tüm vefatların %83'ü, 70 yaşın üzerindeki bir yaş grubunda rapor edilmiştir. Kore'de genel vefat oranı %0,9 olmasına rağmen, 80 yaş ve

üzerindekilerde vefat oranı %9,3'tür. ABD için de benzer sonuçlar rapor edilmiştir (Bulut ve Kato, 2020).

Dünya Sağlık Örgütü (WHO) direktörü, 11 Mart 2020'de Covid-19'u pandemi olarak ilan etmiştir (World Health Organization, 2020). Aralık 2019'dan Haziran 2021'e kadar, teyit edilen Covid-19 vaka ve vefat vakalarının sayısı dünya genelinde sırasıyla 183.446.208 ve 3.979.114'e ulaşmıştır (European Centre for Disease Prevention and Control, 2021).

Türkiye Cumhuriyeti Sağlık Bakanlığı'na göre ilk Covid-19 vakası 11 Mart 2020'de (Zorlu, 2020), doğrulanmış ve ilk vefat 15 Mart 2020'de gerçekleşmiştir (Zontur, 2020). Mart 2020'den Mayıs 2021'e kadar Türkiye'de Covid-19 vaka ve vefat sayıları sırasıyla 5249404 ve 47527'ye ulaşmıştır (T.C. SAĞLIK BAKANLIĞI, 2020).

Salgın hastalık kaynaklı vaka ve vefat sayılarının hızla arttığı günümüzde, ülkeler, kısa, orta ve uzun vadede oluşabilecek vaka ve vefat sayılarını doğru şekilde tahmin etmek istemektedir. Gelecekteki vaka ve vefat sayılarının doğru şekilde tahmin edilebilmesi, ülke ve toplumu etkileyen pek çok kararın önceden planlanabilmesine, tedbirler alınmasına olanak sağlayacaktır.

Bu tez çalışmasının temel amacı, 11 Mart 2020 ile 31 Mayıs 2021 tarihleri arasındaki verilere dayanarak kısa dönemli (1 Haziran - 30 Haziran 2021 tarihleri arasında) Türkiye'nin toplam Covid-19 hasta ve vefat sayılarını tahmin etmektir. Tahmin modelleri olarak, Uzun Kısa Süreli Bellek (LSTM), Çift Yönlü LSTM (Çift Yönlü LSTM- BiLSTM) ve Geçitli Tekrarlayan Birim (GRU) olmak üzere üç farklı derin öğrenme modeli ve istatistiksel model olarak Otoregresif Entegre Hareketli Ortalama (ARIMA) modeli kullanılmıştır. Model performanslarını değerlendirmek için Kök Ortalama Kare Hatası (RMSE), Ortalama Mutlak Yüzde Hatası (MAPE) kullanılmıştır. Elde edilen tahmin sonuçları, gerçek değerlere yakıp olup kabul edilebilir düzeydedir.

Bu tez çalışmanın geri kalanı şu şekilde düzenlenmiştir. İkinci bölümde kaynak araştırması sonuçları sunulmuştur. Üçüncü bölümünde, tez kapsamında kullanılan materyal ve yöntemler açıklanmıştır. Dördüncü bölümünde, deneysel çalışmalar ve sonuçları sunulmuştur. Beşinci bölümde tez çalışması sonuçlarının genel bir değerlendirmesi yapılmış ve gelecekte yapılabilecek çalışmalar hakkında bilgi verilmiştir.

2. KAYNAK ARAŞTIRMASI

Birçok araştırmacı, Covid-19 tahmini üzerine çok sayıda araştırma yapmıştır. Gelecekteki koşulları tahmin etmek için farklı yaklaşımlar kullanılmıştır. Bu bölümde farklı derin öğrenme, makine öğrenmesi, zaman serisi modelleri ve matematiksel tekniklerle yapılmış bazı çalışmaları açıklanmıştır.

Roosa ve ark. (2020) çalışmalarında, Çin'deki Covid-19 teyitli vakaları tahmin etmek için lojistik büyüme modeli (Logistic Growth Model -LGM), Richards büyüme modeli (Richards Growth Model) ve alt salgın dalga modeline (Sub-epidemic Wave Model) dayalı üç fenomenolojik model önermiştir. Çin Ulusal Sağlık Komisyonu web sitesinden 22 Ocak 2020'den 5 Şubat 2020'ye kadar olan verileri kullanmışlar.

Chen ve ark. (2020) makalelerinde, 22 Ocak - 6 Nisan 2020 tarihleri arasında ABD'de teyit edilmiş Covid-19 verilerinin günlük yeni ve kümülatif vakalarını kullanmış ve 5 parametrelili bir lojistik büyüme modelini yeniden yapılandırmıştır.

Chaudhry ve ark. (2020) çalışmalarında, Pakistan'da hızla artan Covid-19 vaka sayısını tahmin etmek için basit bir hareketli ortalama modeli önermiştir. Veriler Ulusal Sağlık Enstitüsü (National Institute of Health -NIH), Pakistan ve Dünya Sağlık Örgütü'nden (World Health Organization -WHO) toplanmıştır.

Alzahrani ve ark. (2020) çalışmalarında, Suudi Arabistan'da önümüzdeki dört hafta içinde teyit edilen Covid-19 vakalarını tahmin etmek için Otoregresif Entegre Hareketli Ortalama (ARIMA) modelini önermiştir. Otoregresif Model, Hareketli Ortalama, Otoregresif Hareketli Ortalama ve ARIMA olmak üzere dört model kullanmışlar. Veriler, Suudi Sağlık Bakanlığı'nın resmi web sitesinden toplanmıştır. 2 Mart 2020'den 20 Nisan 2020'ye kadar olan verileri kullanılmıştır. ARIMA modelinin RMSE ve R2 puanlarının sonucu sırasıyla 21.17 ve 0.99'dur.

Gupta ve Pal (2020), çalışmalarında, Hindistan'daki Covid-19 salgınının eğilimlerini analiz etmek ve tahmin etmek için Otomatik Regresif Entegre Hareketli Ortalama (Auto-Regressive Integrated Moving Average -ARIMA) modelini ve Üstel Düzeltme yöntemini (Exponential Smoothing Methods) uygulamıştır. Veriler, John Hopkins Üniversitesi'nden alınmıştır ve 30 Ocak 2020 ile 24 Mart 2020 arasındaki veriler kullanılmıştır.

Benvenuto ve ark. (2020) çalışmalarında, Covid-19 prevalansının ve insidansının epidemiyolojik eğilimini tahmin etmek için Otomatik Regresif Entegre Hareketli Ortalama (ARIMA) modelini kullanmıştır. Johns Hopkins Üniversitesi'nin

resmi web sitesinde yer alan 20 Ocak 2020'den 10 Şubat 2020'ye kadar olan günlük Covid-19 yaygınlık verilerini kullanmışlardır.

Dehesh ve ark. (2020) çalışmalarında, 22 Ocak 2020 - 1 Mart 2020 tarihleri arasında Johns Hopkins Üniversitesi resmi web sitesinden toplanan günlük teyit edilmiş Covid-19 vakalarını tartışmıştır. Çin, İtalya, Güney Kore, İran ve Tayland olmak üzere beş ülkenin teyit edilmiş vakalarının gelecek 17 gündeki değerlerini tahmin etmek için ARIMA modelini kullanmıştır.

Chintalapudi ve ark. (2020), İtalya'nın Covid-19 kayıtlı ve iyileşen vaka sayılarını ARIMA modelini kullanarak tahmin etmeye çalışmıştır. İtalya Sağlık Bakanlığı web sitesinde yer alan, 2020 yılı Şubat ayı ortasından Mart ayı sonuna kadar olan veriler kullanılmıştır. Covid-19 kayıtlı ve iyileşen vakaların tahmin çalışmasında, RMSE değeri sırasıyla 514,74 ve 186,85 olarak bulunmuştur.

Abdumajeed ve ark. (2020) çalışmalarında, Nijerya'daki Covid-19 vakalarının tahmini için bir yöntem önermiştir. Günlük Covid-19 vaka sayısını tahmin etmek için ARIMA, Prophet ve Holt-Winters modelleri uygulanmıştır. 27 Şubat 2020'den 5 Nisan 2020'ye kadar olan veriler Nijerya Hastalık Kontrol Merkezi (Nigeria Center for Disease Control –NCDC) ve Wikipedia web sitelerinden toplanmıştır.

Maleki ve ark. (2020) çalışmalarında, TP–SMN–AR modelleri olarak adlandırılan iki parçalı ölçekli karışım normal dağılımlarına dayanan Otoregresif zaman serisi modelleri sunmuştur. Doğrulanmış ve iyileşen Covid-19 vakalarının zaman serisi verilerini analiz etmişler.

Chakraborty ve Ghosh (2020) çalışmalarında, Kanada, Fransa, Hindistan, Güney Kore ve İngiltere için yeni Covid-19 vakalarının gerçek zamanlı tahminini ve risk değerlendirmesi sonuçlarını sunmuştur. Otoregresif entegre hareketli ortalama modeline ve Dalgacık tabanlı tahmin modeline dayalı hibrit bir model önermiştir. Ayrıca, farklı ülkeler için vaka ölüm oranlarını önemli ölçüde etkileyen temel nedensel değişkenleri bulmak için optimal bir regresyon ağacı algoritması uygulanmıştır. Veri kümeleri, Global Change Data Lab'den toplanmıştır. Hindistan ve İngiltere için, sırasıyla 30 Ocak 2020 - 4 Nisan 2020 ve 31 Ocak 2020 - 4 Nisan 2020 tarihleri arasında teyit edilen günlük vakalar değerlendirilmiştir. Kanada, Fransa ve Güney Kore için günlük Covid-19 vaka verileri sırasıyla 20 Ocak 2020 - 4 Nisan 2020, 25 Ocak 2020 - 4 Nisan 2020 ve 26 Ocak - 4 Nisan 2020 dönemleri için alınmıştır.

Ardabili ve ark. (2020) çalışmalarında, Covid-19 teyitli vakaları tahmin etmek için Çok Katmanlı Algılayıcı (Multi-layer perceptron -MLP) ve Uyarlanabilir Ağ

Tabanlı Bulanık Çıkarım Sistemi (Adaptive network-based fuzzy inference system - ANFIS) adlı makine öğrenme modellerini kullanmıştır. 30 gün boyunca, toplam vakalar için İtalya, Almanya, İran, ABD ve Çin olmak üzere beş ülkeye ait veriler worldometers web sitesinden toplanmıştır.

Dil ve ark. (2020) çalışmalarında, Doğu Akdeniz bölgesindeki yedi ülkeye ve Pakistan'a odaklanmış, doğrulanmış Covid-19 vakalarını tahmin etmek için bir yöntem sunmuştur. Veriler, Pakistan Ulusal Sağlık Hizmetleri Düzenlemeleri ve Koordinasyonu Covid-19 panosundan ve Avrupa Hastalık Önleme ve Kontrol Merkezi'nden günlük olarak toplanmıştır. Toplanan veriler 14 Şubat - 26 Nisan 2020 tarihleri aralığındadır.

Al-Qaness ve ark. (2020) çalışmalarında, Çin'deki teyit edilmiş Covid-19 vakalarının sayısını tahmin etmek için sürü zekası algoritmalarını kullanan gelişmiş bir uyarlanabilir nöro-bulanık çıkarım sistemi (Adaptive Neuro-Fuzzy Inference System – ANFIS) önermiştir. Önerilen modelin performansı, önümüzdeki on günün teyit edilmiş vakalarını tahmin etmek için Dünya Sağlık Örgütü Covid-19 salgınının resmi verileri kullanılarak değerlendirilmiştir. Önerilen modelin RMSE ve R2 skoru sırasıyla 5779 ve 0.96'dır.

Chimmula ve Zhang (2020) çalışmalarında, Kanada'daki Covid-19 vakalarını tahmin etmek için derin öğrenme tabanlı bir Uzun Kısa Süreli Bellek (Long Short Term Memory –LSTM) modelini kullanmıştır. Elde edilen sonuçlar, İtalya ve ABD gibi diğer ülkelerin tahmin sonuçlarıyla da karşılaştırmıştır. Johns Hopkins Üniversitesi ve Kanada Sağlık Otoritesinden toplanan veriler, 31 Mart 2020'ye kadar teyit edilen vaka sayısını kapsamaktadır.

Zeroual ve ark. (2020) çalışmalarında, günlük teyit edilen ve iyileşen Covid-19 vaka sayılarının tahmini için derin öğrenme modellerine dayalı tahmin yöntemleri önermiştir. Tekrarlayan Sinir Ağı (Recurrent Neural Network -RNN), Uzun Kısa Süreli Bellek (Long Short-Term Memory -LSTM), Çift Yönlü LSTM (Bidirectional LSTM - BiLSTM), Değişken Otomatik Kodlayıcı (Variational Auto Encoder -VAE) ve Geçitli Tekrarlayan Birim (Gated Recurrent Unit -GRU) olmak üzere beş derin öğrenme modeli kullanılmıştır. Bu çalışmada kullanılan veriler, İtalya, İspanya, Fransa, Çin, ABD ve Avustralya olmak üzere altı ülkeye aittir.

Elsheikh ve ark. (2021), Suudi Arabistan'daki toplam teyit edilen vaka sayısını, toplam iyileşen vaka sayısını ve toplam vefat sayısını tahmin etmek amacıyla bir çalışma gerçekleştirmiştir. Çalışmada, LSTM ağı, ARIMA modeli, doğrusal olmayan

otoregresif yapay sinir ağıları (NARANN) modeli kullanılarak, sonuçları karşılaştırılmıştır.

Shim ve ark. (2020) çalışmalarında, Güney Kore'deki Covid-19'un bulaşma potansiyelini incelemiştir. Çalışmada, Ampirik raporlama gecikme dağılımı kullanılarak ve genelleştirilmiş büyüme modeli simüle edilerek tahmin işlemleri yapılmıştır. Kore Hastalık Kontrol ve Önleme Merkezleri'nden, kamuya açık olan, 20 Ocak 2020 ile 26 Şubat 2020 arasındaki günlük teyit edilmiş vaka verileri kullanılmıştır.

Shastri ve ark. (2020) çalışmalarında, Hindistan ve ABD için Covid-19 teyitli ve ölüm vakalarını tahmin etmek için derin öğrenme tabanlı modeller kullanmıştır. Çalışmada, yığılmış LSTM, Çift Yönlü LSTM ve Evrişimli LSTM, her iki ülkenin bir ay sonraki vaka sayılarını tahmin etmek için kullanılmıştır. Veriler, Hindistan ve ABD'nin resmi kurumlarından toplanmıştır.

Naudé (2020) çalışmasında, Covid-19'a karşı yapay zeka yöntemlerine dayalı bir yöntem sunmuştur. Bu çalışma, erken uyarılar, izleme ve tahmin, veri panoları, teşhis ve prognoz, tedaviler ve sosyal kontrol olmak üzere altı alanda Covid-19 ile mücadeleye katkıda bulunmuştur.

Ayyoubzadeh ve ark. (2020), doğrusal regresyon ve uzun kısa süreli bellek kullanarak İran'daki pozitif Covid-19 vakalarının sayısını tahmin etmeyi çalışmıştır. Çalışmada kullanılan veriler, Google Trends web sitesinden 15 Şubat 2020 ile 18 Mart 2020 tarihleri arasında toplanmıştır.

Tamang ve ark. (2020) çalışmalarında, Covid-19 kaynaklı vaka ve ölüm vakalarının sayısını tahmin etmek için yapay sinir ağı tabanlı bir eğri uyurma modeli önermiştir. Veriler Hindistan, ABD, Fransa ve İngiltere, Çin ve Güney Kore için DSÖ'nün web sitesinden toplanmıştır.

Moftakhar ve ark. (2020) çalışmalarında, İran'daki günlük Covid-19 vakalarının tahmini için Yapay Sinir Ağları ve ARIMA olmak üzere iki farklı yöntem kullanmıştır. 19 Şubat - 30 Mart 2020 tarihleri arasındaki veriler kullanılarak yöntemler eğitilmiş ve önümüzdeki otuz gün boyunca günlük yeni Covid-19 enfekte vaka sayısı tahmin edilmiştir. Veriler, İran Sağlık Bakanlığı'nın günlük raporlarından ve John Hopkins Üniversitesi yayınlarından toplanmıştır.

3. MATERYAL VE YÖNTEM

Bu tez çalışmasında, Covid-19 salgınının gelişimini tahmin etmek amacıyla 3 farklı derin öğrenme modeli ile istatistiksel bir yöntem incelenmiştir. Deneysel çalışmalarda, LSTM, BILSTM ve GRU olmak üzere üç derin öğrenme modeli ve istatistiksel bir yöntem olan ARIMA yöntemi kullanılmıştır. Model performansını değerlendirmek için RMSE, MAPE hata metrikleri kullanılmıştır. Bu tez kapsamında toplam teyit edilmiş vaka sayısı ve toplam vefat sayısı tahmini için iki kategoride çalışmalar gerçekleştirilmiştir. Deneysel çalışmalarda, 11 Mart 2020 ile 31 Mayıs 2021 arasındaki veriler kullanılarak modeller eğitilmiş, doğrulanmış ve 1 Haziran ile 30 Haziran 2021 arasında oluşacak vaka sayısı ve vefat sayısı tahmin edilmeye çalışılmıştır. 3. bölümde, veri seti ile çalışmada kullanılan yöntemler açıklanmıştır.

3.1. Veri Seti

Bu tez çalışmasında kullanılan Covid-19 verileri Türkiye Cumhuriyeti Sağlık Bakanlığı'nın internet sitesinden toplanmıştır (TR Ministry of Health, 2021). Veriler günlük olarak güncellenmekte ve tarihsel olarak zaman serisi formatında sunulmaktadır. Bu çalışmada, 11 Mart 2020 ile 31 Mayıs 2021 tarihleri arasındaki veriler kullanılmıştır. Toplanan veriler için, Excel'de yeni bir CSV dosyası oluşturulmuştur. Dosyada, tarih, vaka sayıları ve vefat sayıları olmak üzere üç çeşit veri bulunmaktadır. Covid-19 veri seti %70 eğitim ve %30 test setine bölünmüştür. Modelleri eğitmek için 11 Mart 2020'den 17 Ocak 2021'e kadar olan veriler kullanılmıştır. Kalan veriler modelin performansını test etmek için kullanılmıştır. Covid-19 veri setine ilişkin bazı bilgiler Çizelge 3.1'de açıklanmıştır.

Çizelge 3.1. Türkiye Covid-19 veri setinin açıklaması

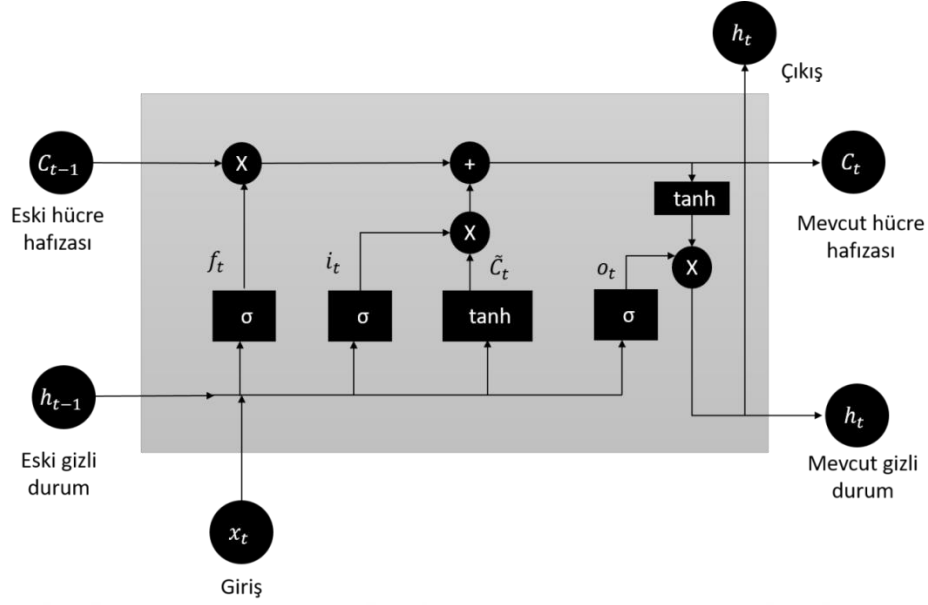
	Özellik	Değer
1	Açıklama	Zaman serisi verileri
2	[Min, Max]	Vaka sayıları [1, 5249404] Vefat sayıları [0, 47527]
3	Davranış	Durağan olmayan
4	Başlangıç tarihi	11 Mart 2020
	Bitiş tarihi	31 Mayıs 2021
5	Öznitellikler	Tarih, vaka sayıları, vefat sayıları
6	Kayıp veri	-

3.2. Derin Öğrenme Modelleri

Derin öğrenme teknikleri dünyada farklı uygulama alanlarında kullanılmaktadır. Bu bölümde, Covid-19 salgını tahmin etmek için kullanılan üç derin öğrenme yöntemi hakkında bilgi verilmiştir.

3.2.1. Uzun kısa süreli bellek (LSTM)

Derin öğrenmede yaygın olarak kullanılan RNN'nin en yaygın varyantlarından biri LSTM olarak bilinir. Hochreiter ve Schmidhuber tarafından önerilmiştir (Hochreiter ve Schmidhuber, 1997). RNN, uzun vadeli bağımlılıkları yakalama zorluğuna sahiptir ve bazı durumlarda veri dizilemeyi modellemek için uygun olmayabilmektedir (Yu ve ark., 2019). LSTM, 'uzun vadeli bağımlılıkları' öğrenme yeteneğine sahiptir. Sıralı verilerle ilgili çeşitli öğrenme problemleri için etkili ve ölçeklenebilir bir modeldir. LSTM ile RNN'de bulunan kaybolan gradyan problemine çözüm önerilmiştir. LSTM ağları el yazısı tanıma, dil modelleme ve çeviri, konuşma tanıma gibi alanlarda kullanılmıştır (Greff ve ark., 2016). Bu çalışmada, zaman serisi verilerine dayalı tahmin için LSTM ağı kullanılmıştır. LSTM'nin mimarisi, iç mekanizmaları Şekil 3.1'de gösterilmiştir. LSTM, bellek hücresi, bir giriş kapısı, bir çıkış kapısı ve bir unutma kapısı olarak adlandırılan üç kapıdan oluşur ve kapılar bilgi akışının kontrolü için kullanılmaktadır. Kapılar yardımıyla hangi bilgilerin saklanılacağına, ekleneceğine veya unutulacağına karar verilir. Şekil 3.1'de gösterilen bağlantılara dayalı olarak, LSTM matematiksel denklemleri aşağıdaki gibi ifade edilir:



Şekil 3.1. LSTM mimarisi (Olah, 2015)

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3.3)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (3.4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.5)$$

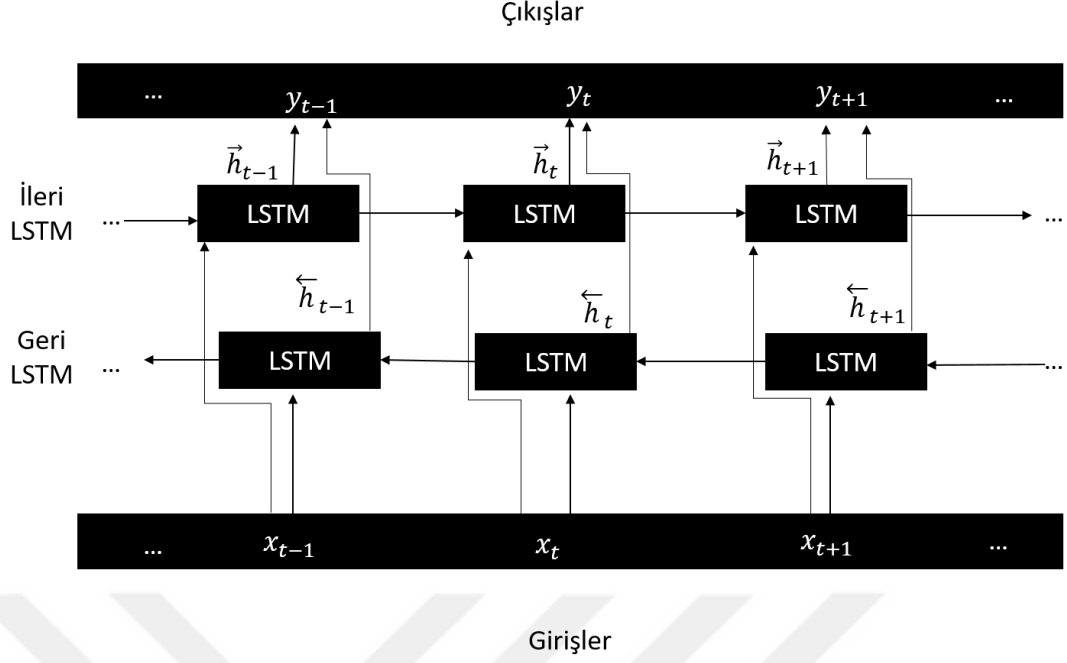
$$h_t = o_t * \tanh(C_t) \quad (3.6)$$

Denklemlerde, f_t , unutma kapısını, i_t , giriş kapısını, o_t , çıkış kapısını, x_t , giriş vektörünü, W değerleri ağırlıkları, b değerleri biasları, C_t , hücre durumunu, h_t , LSTM biriminin çıkış vektörü olarak da bilinen gizli durum vektörünü, σ , sigmoid aktivasyon fonksiyonunu, $\tanh$, hiperbolik tanjant aktivasyon fonksiyonunu göstermektedir. Denklem 3.1'de, f_t , unutma kapısı ile gerekli olmayan, hücreden çıkarılacak bilgiler ve saklanması gereken bilgilere karar verilir. Önceki gizli katmandan (h_{t-1}) ve girdi katmanından (x_t) gelen bilgiler sigmoid aktivasyon fonksiyonundan geçirilir. Denklem 3.2'de, i_t hücre durumunu güncelleyen giriş kapısı olarak kabul edilir. Hangi bilgilerin güncelleneceğine karar verilir. Önceki gizli katmandan (h_{t-1}) ve girdi katmanından (x_t) gelen bilgiler sigmoid aktivasyon fonksiyonundan geçirilir. Denklem 3.3'te, gizli katmandan (h_{t-1}) ve giriş katmanından (x_t) gelen bilgiler, ağı düzenlemeye yardımcı olmak için $\tanh$ aktivasyon fonksiyonundan geçirilir. $\tanh$ fonksiyonu, önem

düzeylerine karar vermek, geçen değerleri ağırlıklandırmak için kullanılır. Denklem 3.4'te, C_t mevcut bellek hücresi durumu, önceki hücre durumu C_{t-1} , unutma kapısı f_t , giriş kapısı i_t ve $\tilde{C}_t$ değerleri kullanılarak hesaplanır. Son olarak denklem 3.5, 3.6 ile çıkış kapısı o_t ve LSTM biriminin çıkışı h_t hesaplanır. Sigmoid katman, hücre durumunun hangi kısımlarının çıktıya ulaşacağına karar verir. tanh fonksiyonu ile C_t mevcut bellek hücresi durumundan yeni değerler elde edilir. Denklemlerde kullanılan $t - 1$ ve t önceki ve şimdiki zaman adımlarıdır.

3.2.2. Çift yönlü uzun kısa süreli bellek (BiLSTM)

Çift Yönlü LSTM veya BiLSTM modeli, Schuster & Paliwal tarafından geliştirilmiştir (Schuster ve Paliwal, 1997). LSTM algoritmasının geliştirilmiş bir versiyonudur. BiLSTM, iki ayrı LSTM katmanından oluşur. LSTM mimarisi tek yönlüyken BiLSTM çift yönlüdür. Bi-LSTM'de giriş, ileri (sağ yön) ve geri (sol yön) olmak üzere iki yönde akar (Xia ve ark., 2020). Hem geçmiş hem de gelecek bilgileri değerlendirebilir ve elde edilen bilgiler model çıkışını üretmek için kullanır. BiLSTM, metin sınıflandırma, konuşma tanıma, duygu sınıflandırma, tahmin vb. problemlerin çözümünde kullanılmıştır (Graves ve ark., 2005). Şekil 3.2, BiLSTM mimarisini göstermektedir. BiLSTM'de giriş veri dizisi (x_{t-1}, x_t, x_{t+1}) hem ileri LSTM hem de geri LSTM'de giriş olarak kullanılır. Her t adımında aynı anda $\vec{h}_t$ (ileri LSTM gizli durumu) ve $\overleftarrow{h}_t$ (geri LSTM gizli durumu) hesaplanır. BiLSTM'in nihai çıktısı y_t , hem $\vec{h}_t$ hem de $\overleftarrow{h}_t$ bilgileri kullanılarak elde edilir. Denklem 3.7, 3.8 ve 3.9 kullanılarak $\vec{h}_t$, $\overleftarrow{h}_t$, y_t çıkışları hesaplanmaktadır.



Şekil 3.2. BiLSTM mimarisi (Xia ve ark., 2020)

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}) \quad (3.7)$$

$$\tilde{h}_t = LSTM(x_t, \tilde{h}_{t+1}) \quad (3.8)$$

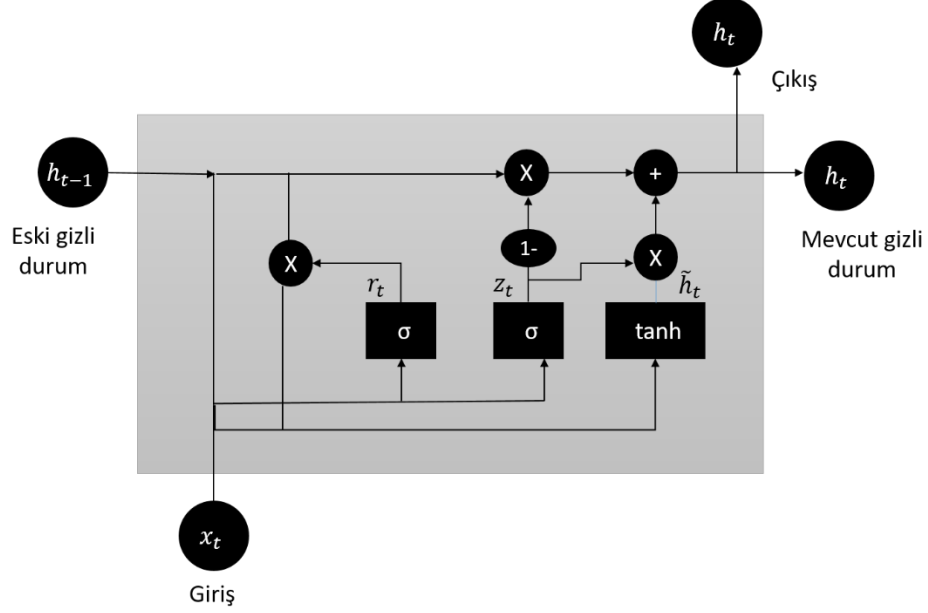
$$y_t = W_{\vec{h}_y} \vec{h}_t + W_{\tilde{h}_y} \tilde{h}_t + b_y \quad (3.9)$$

LSTM ($\cdot$) gösterimi, Denklem 3.1–3.6'da tanımlanan LSTM işlemlerini, $W_{\vec{h}_y}$ ve $W_{\tilde{h}_y}$ sırasıyla ileri LSTM ağırlığını ve geri LSTM ağırlığını, b_y çıktı katmanındaki bias değerini, $\vec{h}_t$ ve $\tilde{h}_t$ ileri ve geri gizli durumları, y_t çıkışı ifade etmektedir.

3.2.3. Geçitli tekrarlayan birim (GRU)

LSTM mimarisinin bir varyasyonu, Cho ve ark. (2014) tarafından önerilen GRU modelidir (Cho ve ark., 2014). GRU, LSTM ağına benzer. LSTM, GRU'nun genelleştirilmiş halidir. Kaybolan gradyan problemini çözmek için kullanılırlar. GRU'yu hesaplamak ve uygulamak LSTM'den daha kolay olabilmektedir. GRU, LSTM ile karşılaştırıldığında daha basit bir mimariye sahiptir ve LSTM'den daha hızlı eğitilebilir. GRU ayrılmış bellek hücresine sahip değildir, bilgileri gizli durumu kullanarak aktarır.

GRU, sıfırlama kapısı ve güncelleme kapısı olmak üzere iki kapıya sahiptir (Rana, 2016). GRU'nun mimarisi Şekil 3.3'te gösterilmiştir. GRU modeline ait denklemler, Denklem 3.10, 3.11, 3.12 ve 3.13'de verilmiştir.



Şekil 3.3. GRU mimarisi (Olah, 2015)

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (3.10)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (3.11)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad (3.12)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (3.13)$$

3.10, 3.11, 3.12, 3.13 denklemlerinde z_t , güncelleme kapısı ve r_t sıfırlama kapısı olarak ifade edilir. h_t, h_{t-1} sırasıyla mevcut ve önceki durumun gizli durum çıkışını, $\tilde{h}_t$ aday gizli durumu, W_z, W_r sırasıyla güncelleme ve sıfırlama kapılarının ağırlık değerlerini gösterir. σ , Sigmoid aktivasyon fonksiyonunu ve $\tanh$, hiperbolik tanjant aktivasyon fonksiyonunu ifade etmektedir. Güncelleme kapısı, LSTM'nin giriş ve unutma kapısına benzer şekilde hareket eden giriş ve unutma kapısından oluşur. Güncelleme kapısı, ne kadar geçmiş bilginin aktarılacağına, sıfırlama kapısı ise geçmiş bilgilerin ne kadarının ihmal edilmesi gerektiğine karar vermek için kullanılmaktadır.

3.3. Zaman Serisi Analiz Tekniği:

Bu bölümde, Covid-19 salgını tahmin etmek için ARIMA modeli tartışılmıştır.

3.3.1. Otoresif entegre hareketli ortalama (ARIMA):

ARIMA modeli Box ve Jenkin tarafından 1976 yılında önerilmiştir (Box ve Jenkins, 1976). Durağan olmayan zaman serileri için ARIMA modeli sıklıkla kullanılmaktadır. Kısa vadeli tahminde, en iyi bilinen ve yaygın olarak kullanılan yöntemlerden birisidir. Zaman serisinin mevsim unsuru içerip içermemesine göre kullanılan farklı modelleri vardır. ARIMA modeli, otoregresif (Autoregressive- AR), entegre (Integrated- I) ve hareketli ortalama (Moving Average- MA) olmak üzere üç bileşen içerir.

ARIMA modellerinin genel gösterimi, ARIMA (p, d, q) şeklinde tanımlanmıştır. p, AR modelinin derecesini, q, hareketli ortalama modelinin derecesini ve d, zaman serisini durağan hale getirmek için gereken fark alma işleminin sayısını ifade etmektedir.

Aşağıdaki denklemler p. derece AR modeli (Denklem 3.14) ve q. derece MA modelini (Denklem 3.15) göstermektedir.

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (3.14)$$

$$y_t = C + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (3.15)$$

ARIMA modeli, AR modeli, entegrasyon (I) ve MA modelinin birleşimi ile oluşturulabilir. Entegrasyon (I), tahmini oluşturmak için fark alma işleminin ters sürecidir. Genelleştirilmiş modelinin matematiksel ifadesi Denklem 3.16'da gösterilmiştir.

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3.16)$$

C, sabit terimi, ϕ_i ($i = 1, 2, \dots, p$) otoregresif model parametrelerini, θ_i ($i = 1, 2, \dots, q$) hareketli ortalama model parametrelerini, y_t güncel zaman serisi değerini,

$y_{t-1}, y_{t-2}, y_{t-p}$ geçmiş değerleri ve ε_t , t. gün için hata terimini simgeler ve aşağıdaki denklemde verilmiştir (ArunKumar ve ark., 2021):

$$\varepsilon_t = y_t - y_{t-1} \quad (3.17)$$

3.4. Performans Değerlendirmesi

Kullanılan yöntemlerin performansını değerlendirmek için RMSE, MAPE metrikleri kullanılmıştır.

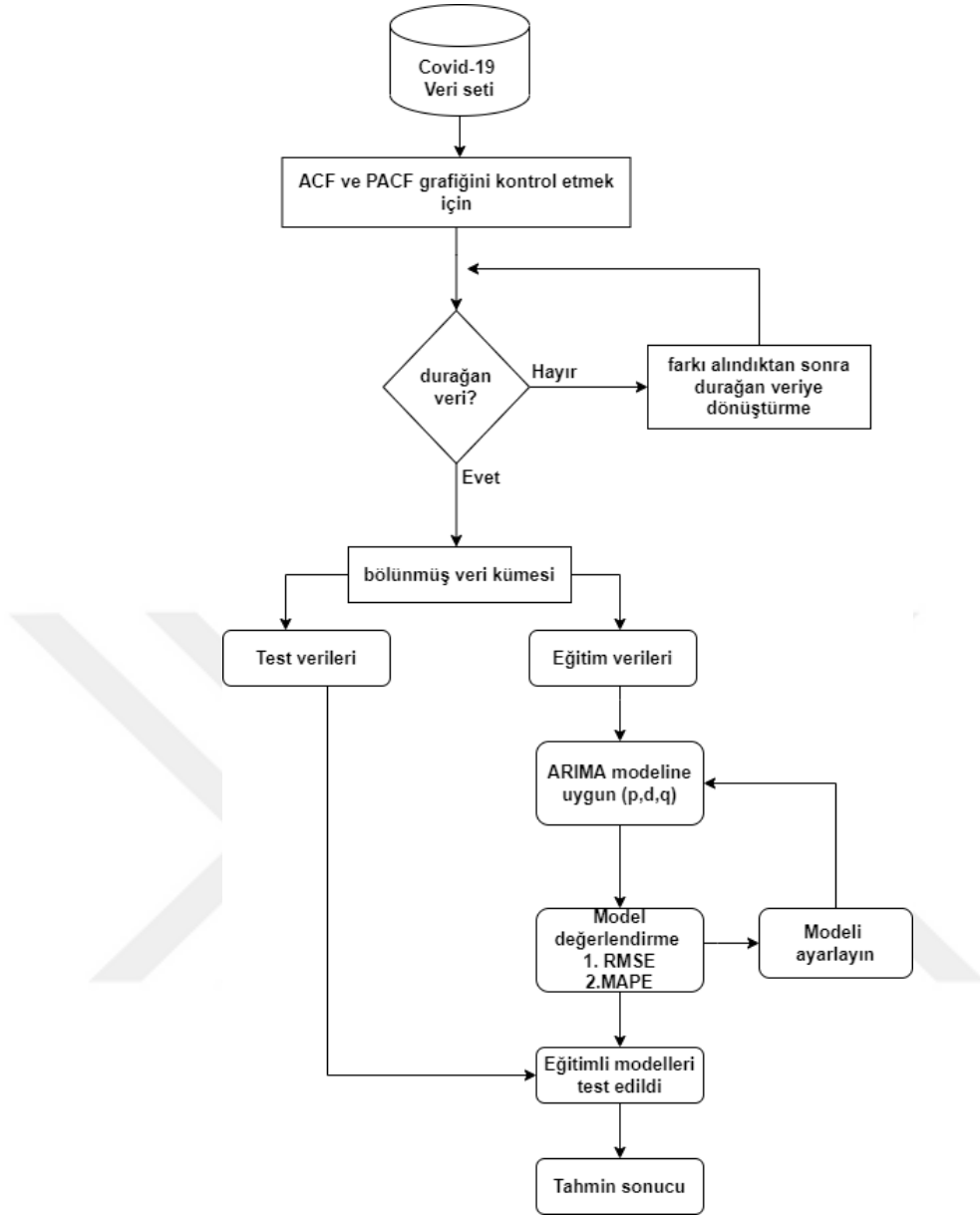
$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y - \hat{y})^2}{N}} \quad (3.18)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y - \hat{y}}{y} \right| \quad (3.19)$$

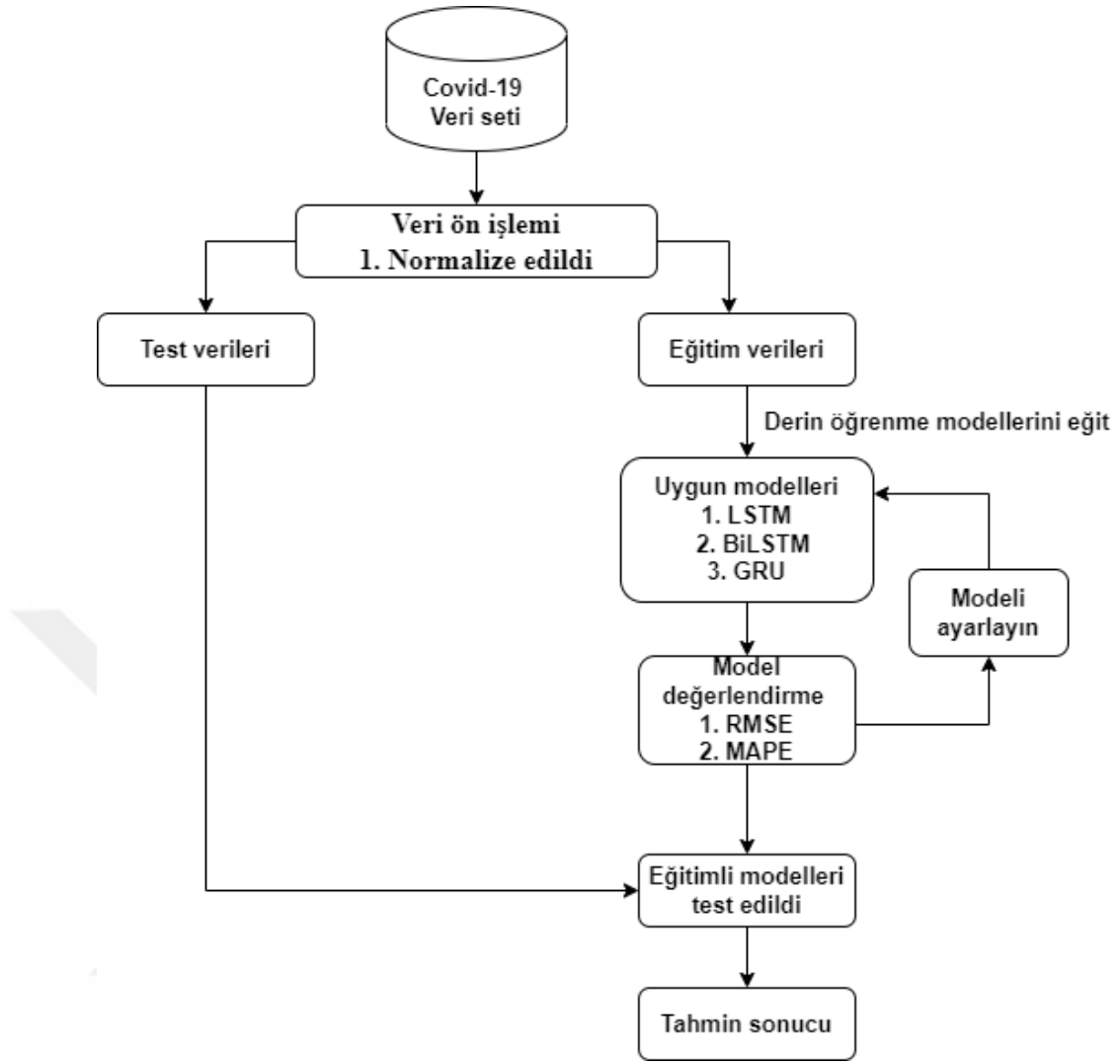
Denklem 3.18 ve 3.19’da, N gerçek test örneği sayısını, y gerçek sonucu, $\hat{y}$ ise tahmin sonucunu göstermektedir.

3.5. Önerilen Yöntem

Önerilen yöntemlerin akış şeması Şekil 3.4 ve Şekil 3.5'te gösterilmiştir.



Şekil 3.4. ARIMA modelinin akış şeması



Şekil 3.5. Derin öğrenme modellerinin akış şeması

ARIMA modeli ve Derin öğrenme modellerinin akış şeması Şekil 3.4 ve 3.5'te gösterilmiştir. İlk olarak veri seti alınmış ve eksik olan değerler kontrol edilmiştir. Veri setinde eksik değer bulunmamaktadır. LSTM, BiLSTM ve GRU gibi üç derin öğrenme modeli çalışmada kullanılmıştır. Bu üç modeli uygulamadan önce veri ön işleme ve özellik ölçekleme teknikleri veri setine uygulanmıştır. Tüm özellikler kayan noktalı sayı olarak dönüştürülmüş ve ardından 0 ile 1 arasına normalize edilmiştir. ARIMA modeli için önce verilerin durağan olup olmadığı kontrol edilmiştir. Veri durağan değilse, veri durağan hale getirilmiştir. Veriyi durağan hale getirdikten sonra zaman serisi tekniğini olarak ARIMA modeli uygulanmıştır. Ön işleme, özellik ölçekleme ve veriyi durağan hale getirmeden sonra, veri seti 7:3'e oranına göre eğitim ve test setine bölünmüştür. Modeller eğitilip, doğrulandıktan sonra, LSTM, BiLSTM, GRU ve ARIMA modelleri, 1 Haziran 2021 ila 30 Haziran 2021 (30 gün) tarihleri arasında Covid-19 salgınının

toplam vaka sayıları ve toplam vefat sayıları tahmin etmek için kullanılmıştır. Model performanslarını değerlendirmek için RMSE ve MAPE metrikleri kullanılmıştır.



4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

Bu tez çalışması, Türkiye için 2021 yılı Haziran ayına ait (30 gün) günlük Covid-19 kaynaklı toplam vefat ve teyit edilmiş vaka sayılarını tahmin etmeyi amaçlamaktadır. Tahmin amacıyla LSTM, BiLSTM ve GRU olarak adlandırılan derin öğrenme yöntemleri ile ARIMA modeli kullanılmıştır. Simülasyonlar ve deneyler Google Colab'da yapılmış ve python ile kodlanmıştır. Pandas, numpy, scikit-learn, matplotlib, tensorflow, keras ve statsmodels gibi gerekli python kütüphanelerinden yararlanılmıştır. Öncelikle pandas dataframe kullanarak veri seti yüklemiştir. Ardından numpy dizisini çıkarılmış ve tamsayı değerleri kayan noktalı değerlere dönüştürülmüştür. Scikit-learn kitaplığından MinMaxScaler() kullanarak veriler 0 ile 1 arasına normalize edilmiştir. Zaman serisi veri analizinde veri dizisi sıraya göre eğitim (%70) ve test (%30) olarak bölünmüştür. Eğitim için 11 Mart 2020 ile 17 Ocak 2021 (312 gün) arasında yer alan veriler kullanılmıştır. 31 Mayıs 2021'e kadar kalan veriler modelleri test etmek için kullanılmıştır. Modellerin performansını değerlendirmek için farklı sayıda zaman adımı belirlenmiştir (1, 5, 7, 10, 20, 30); bu, bir sonraki zamanı tahmin etmek için giriş değişkenleri olarak kullanılacak önceki zaman adımının sayısı anlamına gelmektedir. numpy.reshape() işlevini kullanılarak giriş verileri [num_samples, timestep, num_features] 3D olacak şekilde yeniden şekillendirilmiştir. Son olarak, LSTM, BiLSTM ve GRU modelleri tasarlanmış ve uygun hale getirilmiştir. LSTM, BiLSTM ve GRU, keras kitaplığı ile oluşturulmuş ve eğitilmiştir. Keras açık kaynaklıdır ve python ile yazılmıştır ve arka uç olarak tensorflow kullanmaktadır. Derin öğrenme yöntemleri ile doğru bir sonuç elde etmek için nöron (neuron) sayısı, gizli katman (hidden layer) sayısı, yoğun katman (dense layer), optimize edici, epoch, toplu iş boyutu (batch size) vb. için uygun değerler belirlenmeye çalışılmıştır. Modeller 20 ila 200 epoch için eğitilmiş ve her bir gizli katman için nöron sayıları 5 ila 100 arasında değerlendirilmiştir. Üç gizli katman kullanılmıştır. Üç modelde de optimize edici olarak "Adam" metodu, aktivasyon fonksiyonu olarak sigmoid fonksiyonu seçilmiştir. RMSE ve MAPE değeri kullanılarak yöntemlerin performansı değerlendirilmiştir.

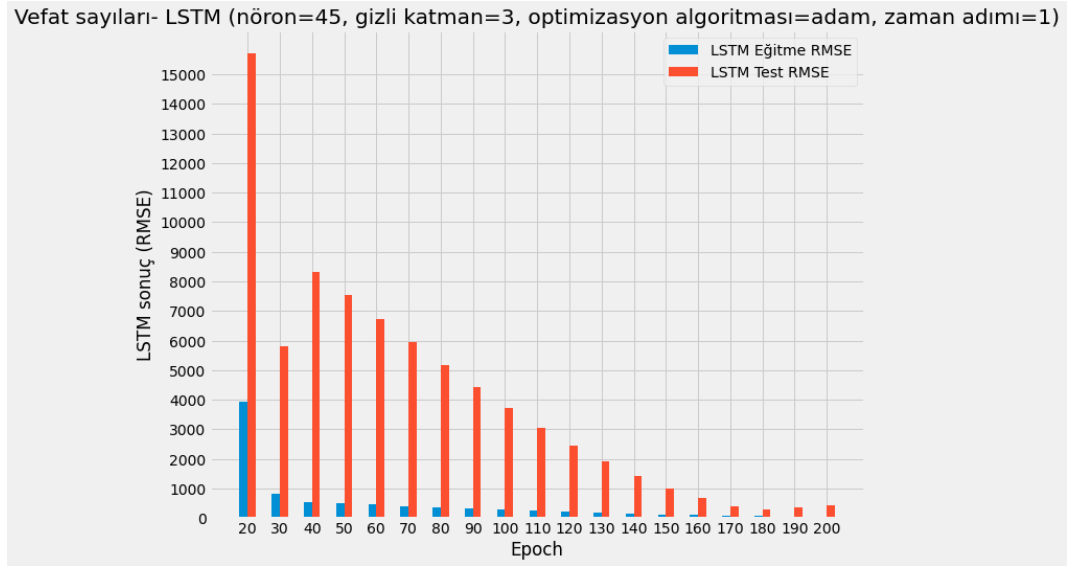
4.1. Derin Öğrenme Yöntemleri ile Vefat Sayılarının Tahmini

Vefat sayılarının tahmininde kullanılan LSTM, BiLSTM ve GRU yöntemlerinin uygun parametre değerlerini belirleyebilmek için deneysel çalışmalar gerçekleştirilmiştir.

LSTM modeli için uygun epoch sayısını belirlemek amacıyla, nöron sayısı 45, zaman adımı (timestep) 1 seçilmiş ve epoch sayısı 20 den 200'e kadar değiştirilmiştir. Her durumda eğitim ve test RMSE değerleri hesaplanmıştır. Elde edilen sonuçlar Çizelge 4.1'de gösterilmiştir. En iyi tahmin sonuçları 170 epoch'da elde edilmiştir. 180, 190 ve 200 epoch için toplam Covid-19 vefat sayıları azalma göstermiştir. Şekil 4.1'de epoch sayısına göre test RMSE değerleri grafiği gösterilmiştir.

Çizelge 4.1. LSTM için farklı epoch sayılarında RMSE değerleri

Epoch	Eğitim RMSE	Test RMSE
20	3939.8317	15711.4088
30	807.6641	5804.9227
40	530.5369	8315.5118
50	494.3507	7525.5777
60	455.9375	6729.1667
70	416.2272	5936.8291
80	375.9109	5162.8513
90	335.6366	4418.8987
100	296.0266	3714.4895
110	257.6817	3057.3910
120	221.1770	2453.8838
130	187.0494	1908.9695
140	155.7791	1426.6921
150	127.7705	1010.9298
160	103.3367	667.9362
170	82.6866	415.4750
180	65.9197	303.2855
		(tahmin sonucu azalıyor)
190	53.0120	348.3102
		(tahmin sonucu azalıyor)
200	43.7736	449.8723
		(tahmin sonucu azalıyor)

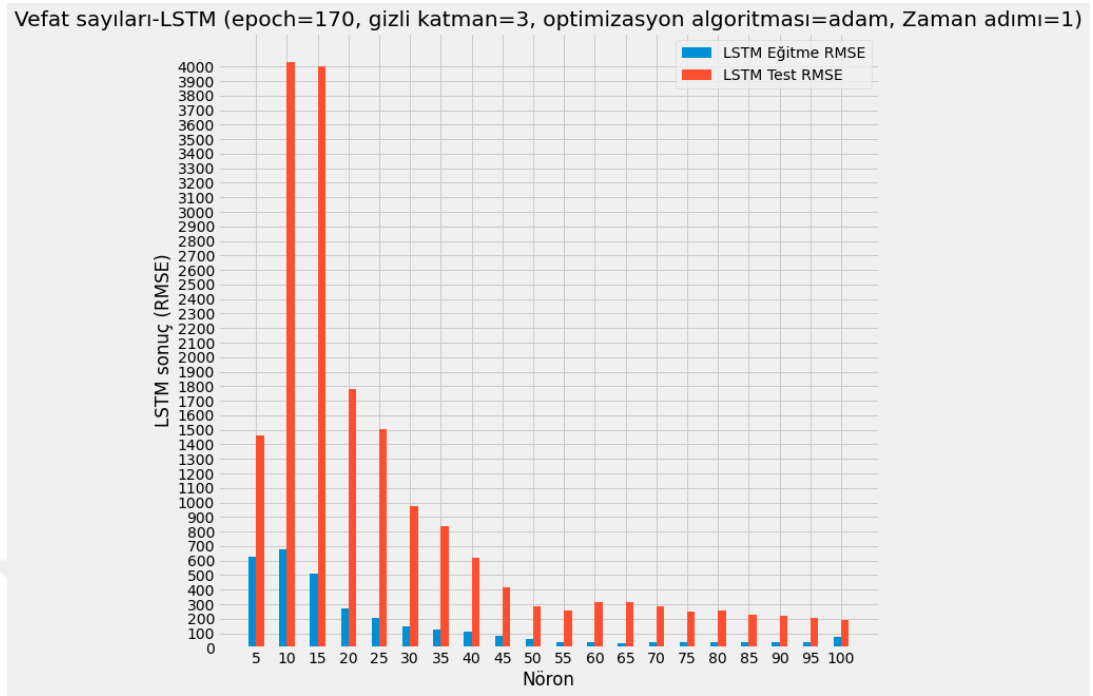


Şekil 4.1. LSTM için farklı epoch sayılarında RMSE değişim grafiği

LSTM için epoch sayısı belirlendikten sonra nöron sayıları araştırılmıştır. Epoch sayısı 170, timestep 1 olduğu durumda nöron sayıları 5’den 100’e kadar 5’er arttırılarak deneyler tekrar edilmiştir. En iyi tahmin sonuçları 45 nöron kullanıldığı durumda elde edilmiştir. 50 ve daha fazla nöron kullanıldığı durumlarda daha küçük RMSE değerleri elde edilmesine karşın aşırı uyum sorunu olduğu gözlemlenmiştir. Çizelge 4.2’de nöron sayısına göre eğitim ve test RMSE değişimleri, Şekil 4.2’de ise değişim grafiği gösterilmiştir.

Çizelge 4.2. LSTM için farklı nöron sayılarında RMSE değerleri

Nöron	Eğitim RMSE	Test RMSE
5	623.8870	1463.1512
10	676.2610	4031.4342
15	509.6927	4001.5070
20	271.1675	1779.7560
25	205.0271	1507.8011
30	150.3919	976.9456
35	123.9221	833.3429
40	109.1870	617.5146
45	82.6866	415.4750
50	59.0191	283.2667
55	41.3398	257.0123
60	36.5107	317.5868
65	34.6336	315.6593
70	35.1365	284.3684
75	35.4101	250.6520
80	35.2273	256.6608
85	35.8615	224.7294
90	35.7298	222.4589
95	36.4646	206.3484
100	77.1346	191.3513

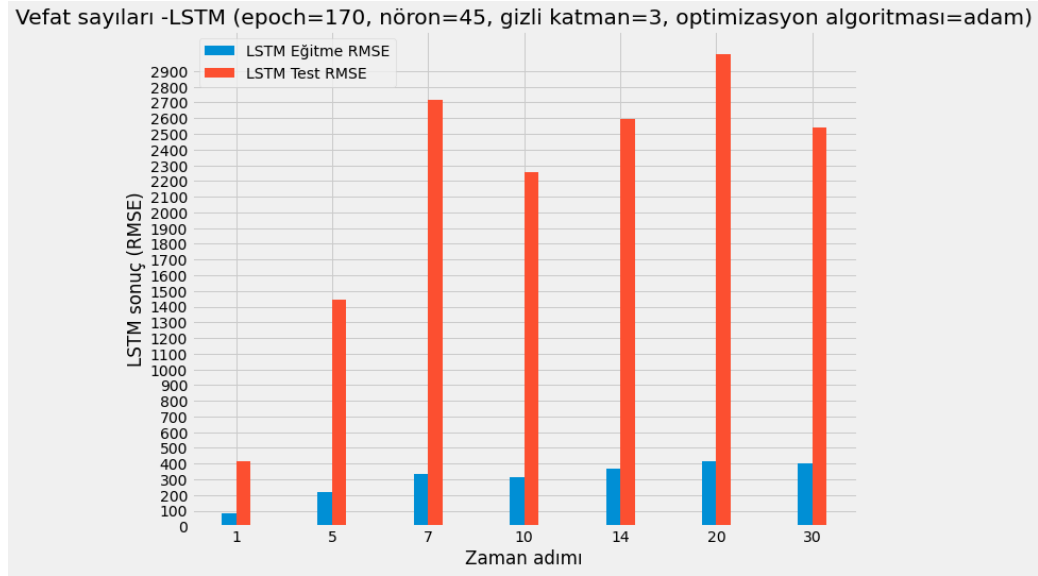


Şekil 4.2. LSTM için farklı nöron sayısında RMSE değişim grafiği

LSTM için epoch sayısı ve nöron sayıları belirlendikten sonra timestep değerlerinin etkisi araştırılmıştır. Epoch sayısı 170, nöron sayısı 45 olduğu durumda timestep değerleri 5, 7, 10, 14, 20 ve 30 olacak şekilde değiştirilerek deneyler tekrar edilmiştir. En iyi tahmin sonuçları timestep 1 olduğu durumda elde edilmiştir. Çizelge 4.3’de timestep sayısına göre eğitim ve test RMSE değişimleri, Şekil 4.3’de ise değişim grafiği gösterilmiştir.

Çizelge 4.3. LSTM için farklı zaman adımı sayılarında RMSE değerleri

Zaman adımı	Eğitim RMSE	Test RMSE
1	82.6866	415.4750
5	215.8990	1444.1449
7	330.7028	2716.4568
10	314.4414	2253.6115
14	366.5564	2593.4956
20	417.5132	3008.1912
30	403.3665	2541.4671

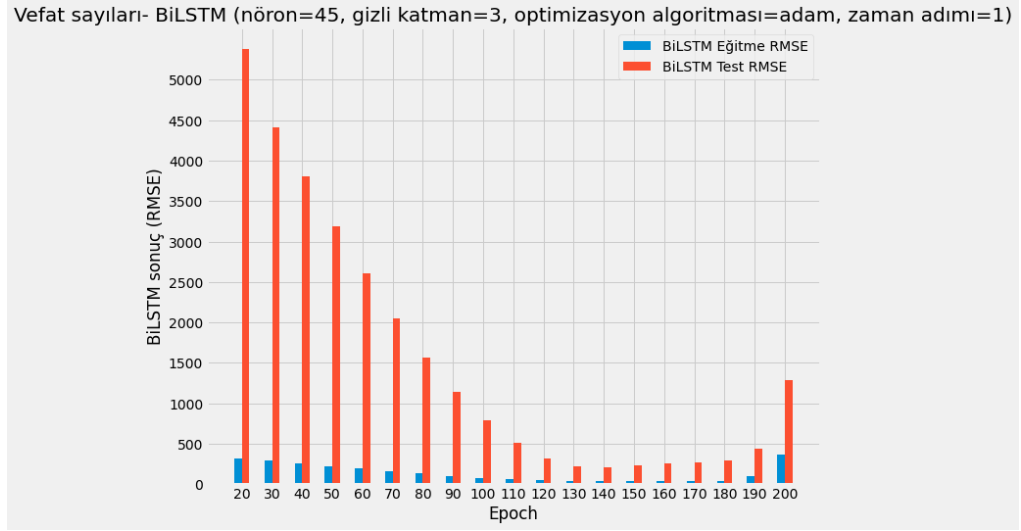


Şekil 4.3. LSTM için farklı zaman adımı sayısında RMSE değişim grafiği

BİLSTM modeli için LSTM modelinde yapılan deneyler tekrarlanmıştır. Çizelge 4.4’de farklı epoch sayıları için RMSE değişimleri, Çizelge 4.5’de farklı nöron sayıları için RMSE değişimleri, Çizelge 4.6’da farklı timestep değerleri için RMSE değişimleri sunulmuştur. Şekil 4.4, Şekil 4.5 ve Şekil 4.6’da sırasıyla epoch, nöron ve timestep değişimlerine karşın RMSE değerlerindeki değişimlerin grafikleri gösterilmiştir. En iyi değerler, 130 epoch, 45 nöron ve timestep 1 değerlerinde elde edilmiştir.

Çizelge 4.4. BİLSTM için farklı epoch sayılarında RMSE değerleri

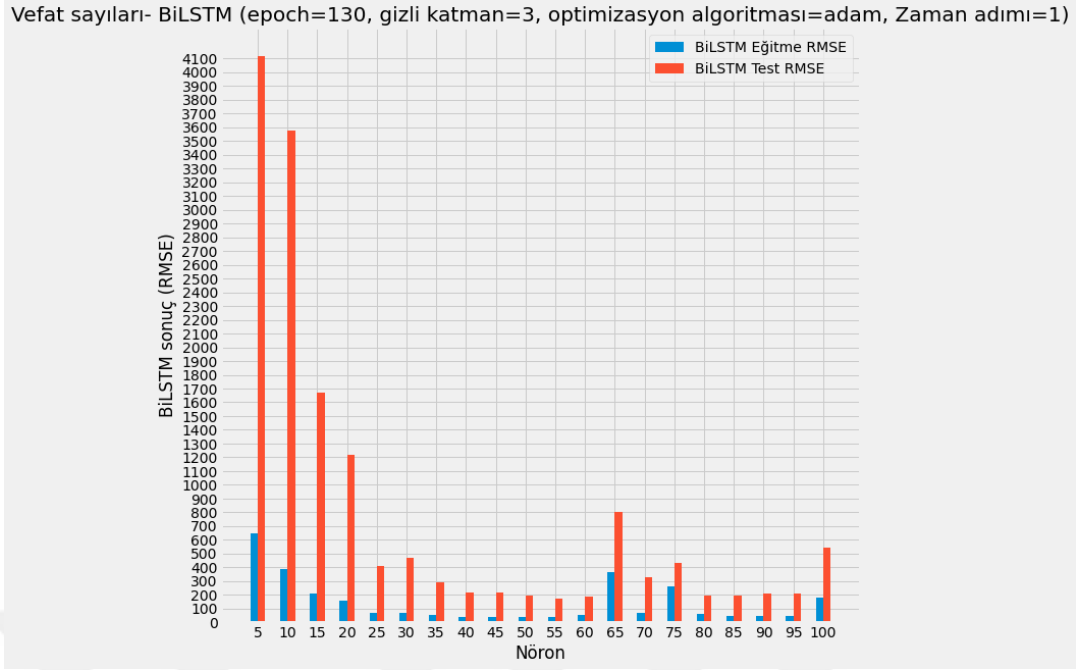
Epoch	Eğitme RMSE	Test RMSE
10	3829.5869	15736.7992
20	315.1156	5380.7652
30	291.6590	4407.1322
40	260.2735	3805.1298
50	227.1763	3194.2669
60	193.8495	2603.4188
70	161.6101	2053.9505
80	131.6043	1561.4255
90	104.8015	1136.3216
100	81.9828	784.8619
110	63.7277	510.5901
120	50.3606	318.1735
130	41.7927	218.3816
140	37.2889	206.0982
		(tahmin sonucu azalıyor)
150	35.5518	231.3602
160	35.2685	256.8431
170	35.3480	273.9190
180	35.5828	292.5428
190	105.8953	442.2149
200	372.5208	1290.0533



Şekil 4.4. BiLSTM için farklı epoch sayısında RMSE değişim grafiği

Çizelge 4.5. BiLSTM için farklı nöron sayılarında RMSE değerleri

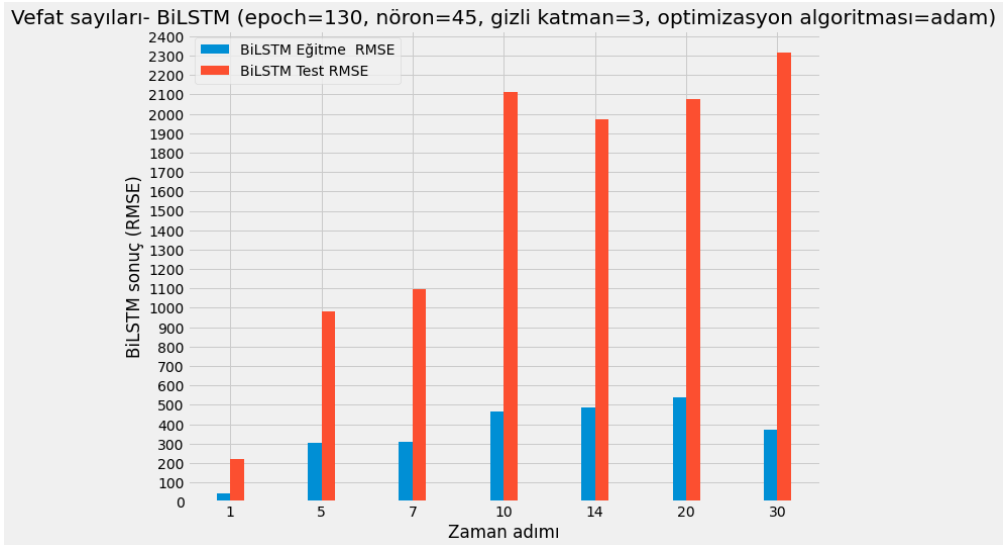
Nöron	Eğitme RMSE	Test RMSE
5	644.0347	4117.8225
10	386.8393	3576.7798
15	208.1333	1670.2150
20	156.2268	1217.8300
25	70.9393	408.9101
30	70.7601	472.7785
35	55.6723	294.3728
40	41.9014	216.4764
45	41.7927	218.3816
50	37.5134	191.7795
55	36.1683	176.1182
		(tahmin sonucu azalıyor)
60	53.8441	185.6376
65	366.9460	799.5757
70	68.6658	327.4097
75	264.4070	430.7611
80	58.2459	197.2544
85	43.8869	191.4720
90	46.4049	207.5559
95	46.4314	211.3734
100	182.1737	541.9249



Şekil 4.5. BiLSTM için farklı nöron sayısında RMSE değişim grafiği

Çizelge 4.6. BiLSTM için farklı zaman adımı sayılarında RMSE değerleri

Zaman adımı	Eğitme RMSE	Test RMSE
1	41.7927	218.3816
5	303.3298	980.8232
7	309.9942	1096.2251
10	467.8726	2113.4785
14	487.2636	1969.6037
20	538.7626	2075.2051
30	370.9996	2315.9051

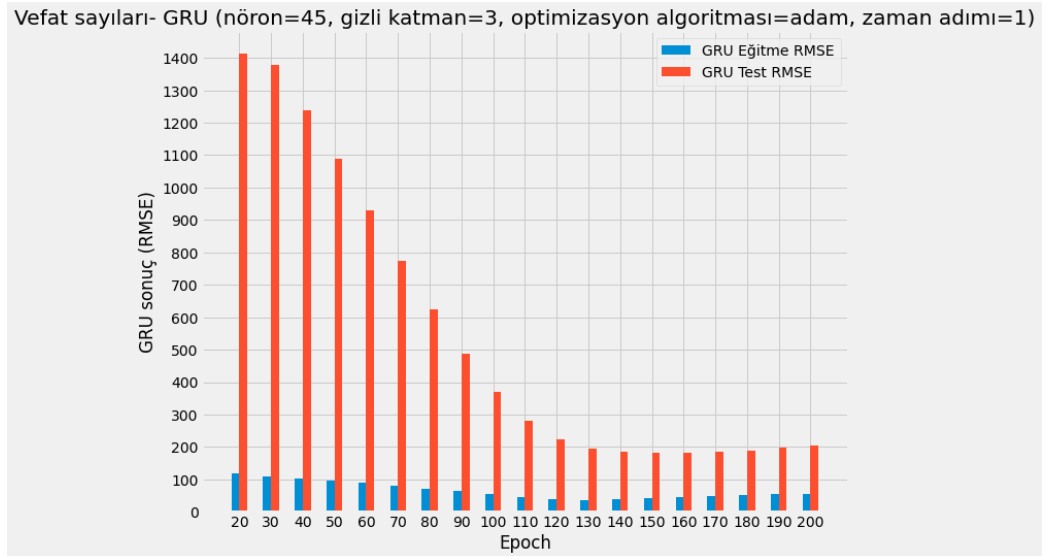


Şekil 4.6. BiLSTM için farklı zaman adımı sayısında RMSE değişim grafiği

GRU modeli için LSTM ve BİLSTM modellerinde yapılan deneyler tekrarlanmıştır. Çizelge 4.7’de farklı epoch sayıları için RMSE değişimleri, Çizelge 4.8’de farklı nöron sayıları için RMSE değişimleri, Çizelge 4.9’da farklı timestep değerleri için RMSE değişimleri sunulmuştur. Şekil 4.7, Şekil 4.8 ve Şekil 4.9’da sırasıyla epoch, nöron ve timestep değişimlerine karşın RMSE değerlerindeki değişimlerin grafikleri gösterilmiştir. En iyi değerler, 140 epoch, 45 nöron ve timestep 1 değerlerinde elde edilmiştir.

Çizelge 4.7. GRU için farklı epoch sayılarında RMSE değerleri

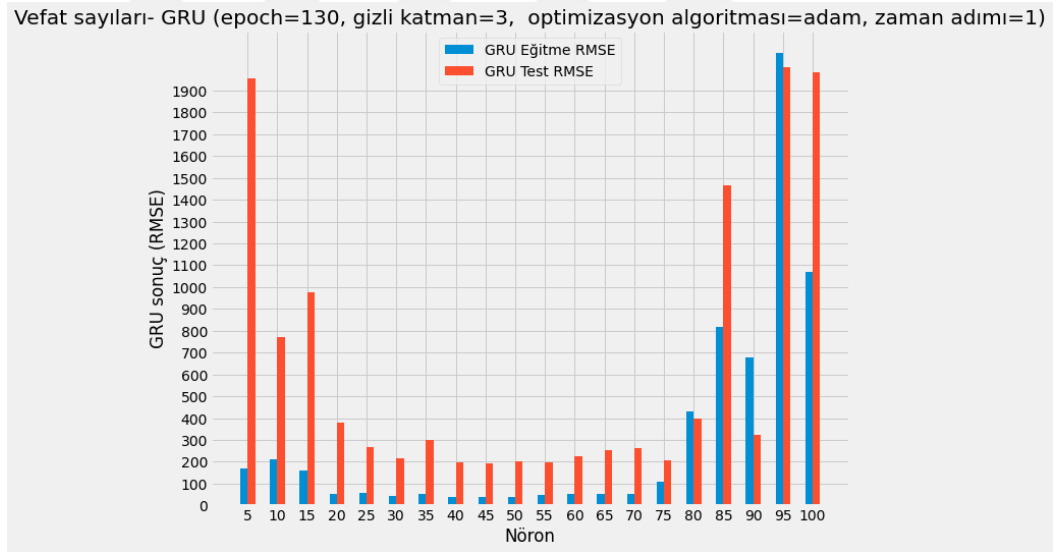
Epoch	Eğitme RMSE	Test RMSE
20	120.0519	1413.2149
30	109.6398	1378.7898
40	103.7024	1239.7014
50	96.8191	1087.959
60	89.1622	930.1865
70	80.8707	773.1173
80	72.1122	623.3136
90	63.1361	487.2414
100	54.3491	371.3170
110	46.3885	281.8349
120	40.2773	223.0659
130	37.2587	194.0024
140	37.7570	184.6765
150	40.9727	182.7554
160	45.2897	183.0989
170	49.2929	184.9241
180	52.0913	189.4369
190	53.4802	196.5877
200	53.8382	204.7852



Şekil 4.7. GRU için farklı epoch sayısında RMSE değişim grafiği

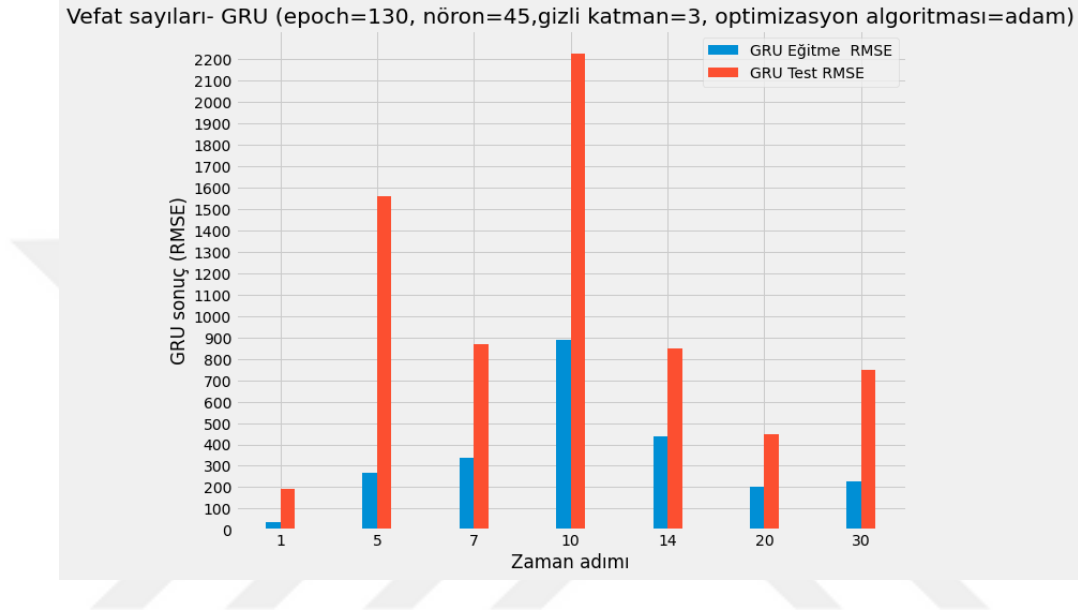
Çizelge 4.8. GRU için farklı nöron sayılarında RMSE değerleri

Nöron	Eğitme RMSE	Test RMSE
5	170.6590	1955.9175
10	213.5186	769.2636
15	159.8600	976.3126
20	50.8502	380.8727
25	56.3800	267.0657
30	42.6128	217.4926
35	52.7737	300.9714
40	37.6637	199.1339
45	37.2587	194.0024
50	38.4112	202.4234
55	45.7592	195.6999
60	54.9787	225.1272
65	53.3832	253.4544
70	50.9739	264.3742
75	108.2217	206.2428
80	431.0272	395.8891
85	819.1485	1465.4150
90	676.5695	324.0652
95	2072.8237	2008.4454
100	1069.9550	1984.1859

**Şekil 4.8.** GRU için farklı nöron sayısında RMSE değişim grafiği

Çizelge 4.9. GRU için farklı zaman adımı sayılarında RMSE değerleri

Zaman adımı	Eğitme RMSE	Test RMSE
1	37.2587	194.0024
5	265.5751	1562.4078
7	336.1697	869.1873
10	889.0438	2227.3275
14	437.2346	849.3323
20	203.3180	447.9569
30	228.8760	750.5305

**Şekil 4.9.** GRU için farklı zaman adımı sayısında RMSE değişim grafiği

Çizelge 4.10’da LSTM, BiLSTM ve GRU modeli kullanılarak 01/06/2021 ile 30/06/2021 arasında tahmin edilen vefat sayıları ile gerçek vefat sayılarının karşılaştırılması gösterilmiştir. LSTM modeli, gerçek vefat sayılarından daha yüksek tahmin değerleri üretirken, BiLSTM ve GRU modelleri daha düşük vefat sayıları tahmin etmiştir. En hızlı sonucu LSTM modeli üretmektedir.

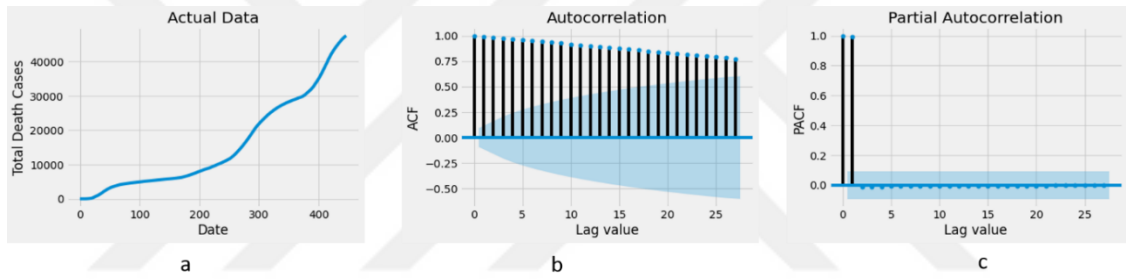
Çizelge 4.10. Derin Öğrenme modelleri tahmin değerleri ve gerçek değerleri

Tarih	LSTM Nöron =45 katman =3, Epoch=170 Optimizer= adam Timestep=1	BiLSTM Nöron =45 katman =3 Epoch=130 Optimizer= adam Timestep=1	GRU Nöron =45 katman =3 Epoch=130 Optimizer= adam Timestep=1	Gerçek vefat sayıları
01/06/2021	47809	47607	47615	47656
02/06/2021	48070	47683	47699	47768
03/06/2021	48310	47755	47779	47882
04/06/2021	48530	47823	47856	47976
05/06/2021	48731	47887	47929	48068
06/06/2021	48915	47948	47999	48164
07/06/2021	49083	48006	48065	48255
08/06/2021	49236	48060	48129	48341
09/06/2021	49375	48111	48189	48428
10/06/2021	49501	48159	48247	48524
11/06/2021	49616	48205	48302	48593
12/06/2021	49720	48248	48354	48668
13/06/2021	49814	48289	48404	48721
14/06/2021	49899	48327	48451	48795
15/06/2021	49976	48363	48496	48879
16/06/2021	50045	48397	48539	48950
17/06/2021	50108	48429	48580	49012
18/06/2021	50165	48459	48619	49071
19/06/2021	50216	48488	48656	49122
20/06/2021	50262	48514	48691	49185
21/06/2021	50303	48540	48724	49236
22/06/2021	50340	48563	48756	49293
23/06/2021	50374	48586	48786	49358
24/06/2021	50404	48607	48815	49417
25/06/2021	50431	48627	48842	49473
26/06/2021	50456	48645	48868	49524
27/06/2021	50478	48663	48893	49576
28/06/2021	50497	48679	48916	49634
29/06/2021	50515	48695	48939	49687
30/06/2021	50531	48709	48960	49732
RMSE	923.400	603.006	449.771	
MAPE	0.01785	0.01095	0.00815	
MAPE(%)	1.785	1.095	0.815	
Zaman(eğitime)	23s	32s	34s	

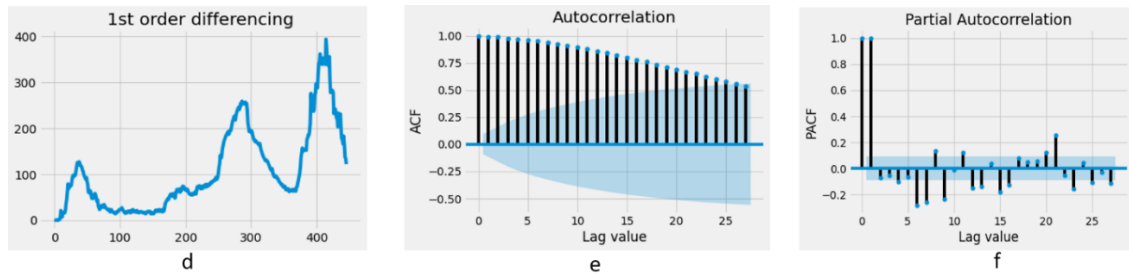
4.2. ARIMA Modeli ile Vefat Sayılarının Tahmini

İlk olarak ACF ve PACF grafikleri kullanılarak p,d,q değerleri manuel olarak seçilmiştir. ACF ve PACF, otokorelasyon fonksiyonunu ve kısmi otokorelasyon fonksiyonunu dikkate almaktadır. PACF ve ACF sırasıyla AR (p), MA (q) sırasını belirtmiştir. ACF, zaman serisi mevcut verilerinin önceki zaman serisi verileriyle korelasyonunu tanımlar. Ancak PACF, zaman serileri belirli zaman birimleriyle ayrılmış kendi gecikmeli (time step/ lag value) değerleriyle ilişkilendirir.

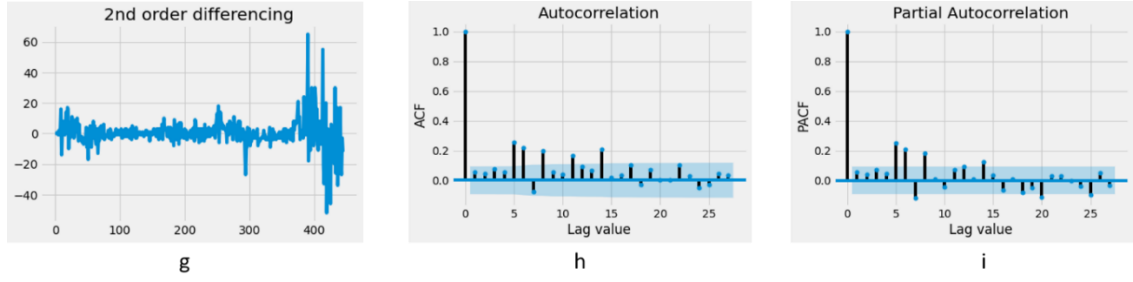
Şekil 4.10. (a), 4.10. (b), 4.10. (c)'de vefat sayılarının gerçek değerleri, gerçek verilerin ACF grafiği ve gerçek verilerin PACF grafiği gösterilmiştir. Bu şekil veri setinin durağan olmadığını göstermektedir. Veri kümesini durağan hale getirmek için bir fark alma yöntemi uygulanmıştır. Şekil 4.11. (d), 4.11. (e), 4.11. (f) farkın alınmasından sonra gerçek verileri, ACF çizim verisini, PACF çizimini göstermektedir. Şekil 4.11. (e) ACF grafiği gecikme değeri azalan düzendedir ve şekil 4.11. (f) Verilerin PACF grafiği, gecikme değeri 1'den sonra bir kesme gösterir. Şekil 4.12. (g), 4.12. (h), 4.12. (i) iki fark alındıktan sonra gerçek veriyi, ACF çizim verisini, PACF çizim verisini gösterir. Şekil 4.12. (g), zaman serisi verilerinin durağan olduğunu göstermektedir. Şekil 4.12. (h) ACF grafiği gecikme 0'dan sonra salınıyor ve şekil 4.12. (i) PACF çizim verileri gecikme 0'dan sonra keskin bir kesme göstermektedir. Şekil 4.12.'ye göre $(p,d,q) = (0,2,0)$ olarak tespit edilebilir.



Şekil 4.10. (a) gerçek veriyi, (b) gerçek verinin ACF grafiğini, (c) gerçek verinin PACF grafiğini gösterir



Şekil 4.11. (d) bir farktan sonraki gerçek verileri, (e) bir farktan sonraki ACF çizim verilerini, (f) bir farktan sonraki PACF çizim verilerini gösterir.



Şekil 4.12. g iki farktan sonraki gerçek verileri, h iki farktan sonra ACF çizim verilerini, i iki farktan sonra PACF çizim verilerini gösterir

GridSearch() kullanarak da p, d, q değerleri bulunabilir. Toplam araştırma süresi 1 saat 30 dk sürmektedir. p değerleri aralığı [0-9], q değerleri aralığı [0-2] ve d değerleri aralığı [0-2] olarak belirlenmiştir. Çizelge 4.11.'de ARIMA modeline ait bazı RMSE değerleri gösterilmiştir.

Çizelge 4.11. ARIMA modelin RMSE değerleri

ARIMA(p,d,q)	RMSE
ARIMA(0, 0, 0)	22605.088
ARIMA(0, 0, 1)	125.817
ARIMA(0, 1, 1)	69.026
ARIMA(0, 2, 0)	14.280
ARIMA(0, 2, 1)	14.454
ARIMA(0, 2, 2)	14.701
ARIMA(1, 0, 0)	199.514
ARIMA(1, 1, 0)	14.234
ARIMA(1, 2, 0)	14.560
ARIMA(2, 1, 0)	14.521
ARIMA(2, 2, 0)	14.776
ARIMA(2, 2, 1)	14.298
ARIMA(2, 2, 2)	14.521
ARIMA(3, 1, 0)	14.171
ARIMA(3, 2, 0)	14.795
ARIMA(3, 2, 1)	14.320
ARIMA(3, 2, 2)	14.419
ARIMA(4, 1, 0)	14.772
ARIMA(4, 2, 0)	14.895
ARIMA(4, 2, 1)	14.337
ARIMA(4, 2, 2)	14.452
ARIMA(5, 1, 0)	14.854
ARIMA(5, 2, 0)	14.381
ARIMA(5, 2, 1)	14.214
ARIMA(5, 2, 2)	14.563
ARIMA(6, 1, 0)	14.363
ARIMA(6, 2, 0)	14.310
ARIMA(6, 2, 1)	14.237
ARIMA(7, 1, 0)	14.251
ARIMA(7, 2, 0)	14.342
ARIMA(7, 2, 1)	14.271
ARIMA(8, 1, 0)	14.321
ARIMA(8, 2, 0)	14.112
ARIMA(8, 2, 1)	14.151
ARIMA(9, 1, 0)	14.059
ARIMA(9, 2, 0)	14.191
ARIMA(9, 2, 1)	14.245

En iyi ARIMA parametre deęerleri yapılan arařtırma sonucunda gridsearch ile (9, 1, 0) ve manuel (0, 2, 0) olarak bulunmuřtur.

Çizelge 4.12’de en iyi ARIMA modelleri için vefat tahmin sayıları ve gerçek vefat sayıları gösterilmiřtir.

Çizelge 4.12. ARIMA modeli tahmin deęeri ve gerçek deęer

Tarih	ARIMA (0,2,0) manuel olarak	ARIMA (9,1,0) GridSearch olarak	Gerçek vefat sayıları
01/06/2021	47539	47655	47656
02/06/2021	47673	47777	47768
03/06/2021	47808	47884	47882
04/06/2021	47944	47990	47976
05/06/2021	48079	48091	48068
06/06/2021	48215	48186	48164
07/06/2021	48351	48278	48255
08/06/2021	48487	48365	48341
09/06/2021	48624	48448	48428
10/06/2021	48761	48531	48524
11/06/2021	48898	48609	48593
12/06/2021	49036	48684	48668
13/06/2021	49174	48758	48721
14/06/2021	49312	48828	48795
15/06/2021	49451	48898	48879
16/06/2021	49590	48966	48950
17/06/2021	49729	49031	49012
18/06/2021	49868	49096	49071
19/06/2021	50008	49160	49122
20/06/2021	50148	49223	49185
21/06/2021	50288	49285	49236
22/06/2021	50429	49347	49293
23/06/2021	50570	49408	49358
24/06/2021	50711	49468	49417
25/06/2021	50853	49528	49473
26/06/2021	50994	49589	49524
27/06/2021	51137	49649	49576
28/06/2021	51279	49709	49634
29/06/2021	51422	49769	49687
30/06/2021	51565	49829	49732
RMSE	923.312	42.842	
MAPE	0.01460	0.00071	
MAPE (%)	1.4	0.7	
Zaman	1s	5 dk	

ARIMA modelleri ile tahmin edilen vefat sayıları derin öğrenme yöntemlerine göre gerçek vefat sayılarına daha yakındır. ARIMA modeli (9,1,0) katsayıları ile elde edilen hatalar, derin öğrenme yöntemleri ile elde edilen hatalardan daha azdır.

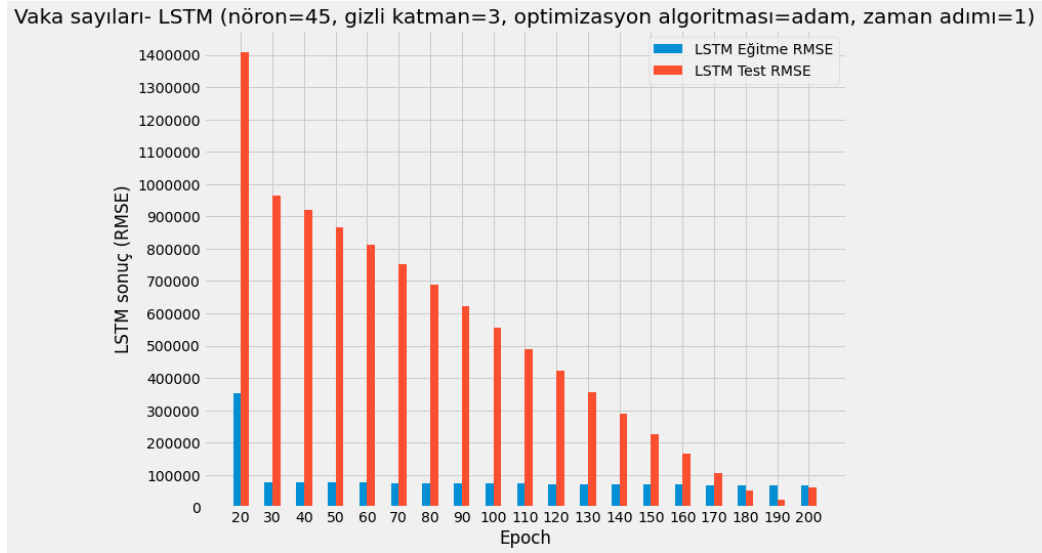
4.3. Derin Öğrenme Yöntemleri ile Vaka Sayılarının Tahmini

Vaka sayılarının tahmininde kullanılan LSTM, BiLSTM ve GRU yöntemlerinin uygun parametre değerlerini belirleyebilmek için deneysel çalışmalar gerçekleştirilmiştir.

LSTM modeli için uygun epoch sayısını belirlemek amacıyla, nöron sayısı 45, zaman adımı (timestep) 1 seçilmiş ve epoch sayısı 20 den 200'e kadar değiştirilmiştir. Her durumda eğitim ve test RMSE değerleri hesaplanmıştır. Elde edilen sonuçlar Çizelge 4.13'te gösterilmiştir. En iyi tahmin sonuçları 180 epoch'da elde edilmiştir. 190 ve 200 epochta Covid-19 vaka sayıları azalma göstermiştir. Şekil 4.13'te epoch sayısına göre test RMSE değerleri grafiği gösterilmiştir.

Çizelge 4.13. LSTM için farklı epoch sayılarında RMSE değerleri

Epoch	Eğitme RMSE	Test RMSE
20	351874.63076	1408144.7489
30	77177.63821	962922.2467
40	76711.66920	918708.8161
50	76203.15080	867102.5728
60	75649.72587	810770.2278
70	75059.97517	750765.9498
80	74442.66820	687926.3514
90	73806.25014	623013.0681
100	73158.97779	556730.3754
110	72509.06967	489735.5608
120	71864.63364	422651.0769
130	71233.70708	356067.5531
140	70624.20662	290554.2484
150	70043.72760	226676.1104
160	69499.52103	165028.6631
170	68998.19267	106383.8400
180	68545.59102	52673.6136
190	68146.49256	24341.2526
200	67804.39821	60375.4398

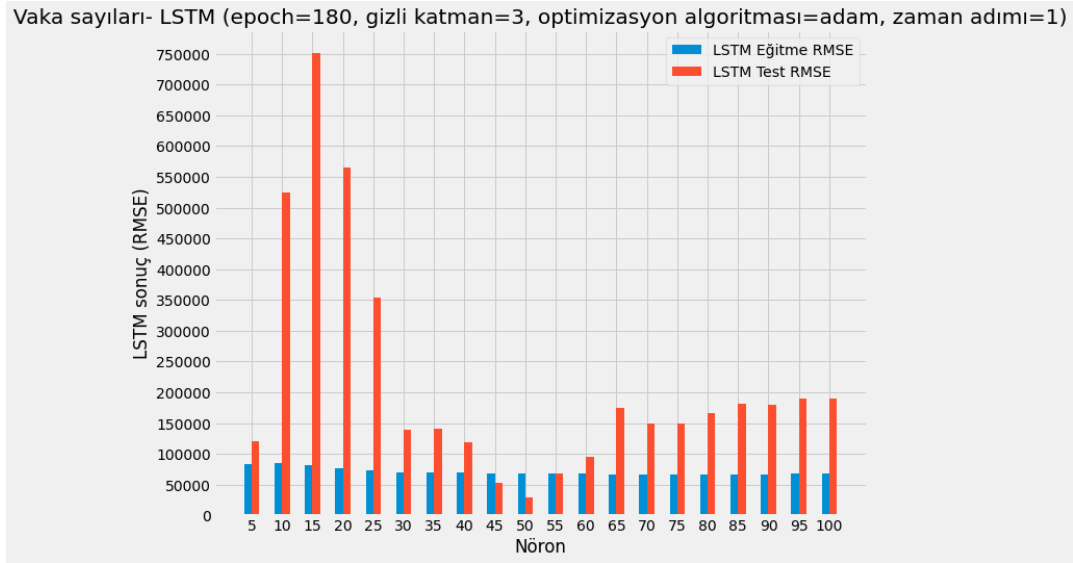


Şekil 4.13. LSTM için farklı epoch sayılarında RMSE değişim grafiği

LSTM için epoch sayısı belirlendikten sonra nöron sayıları araştırılmıştır. Epoch sayısı 180, timestep 1 olduğu durumda nöron sayıları 5’den 100’e kadar 5’er arttırılarak deneyler tekrar edilmiştir. En iyi tahmin sonuçları 45 nöron kullanıldığı durumda elde edilmiştir. Nöron sayısı 50 olduğunda, elde edilen vaka tahmin değerleri azalma göstermiştir. Çizelge 4.14’te nöron sayısına göre eğitim ve test RMSE değişimleri, Şekil 4.14’te ise değişim grafiği gösterilmiştir.

Çizelge 4.14. LSTM için farklı nöron sayılarında RMSE değerleri

Nöron	Eğitim RMSE	Test RMSE
5	84085.44530	121389.1604
10	85227.03947	525000.7031
15	82058.07114	751403.6539
20	76456.32102	565024.7675
25	72379.81597	354706.3380
30	69821.88448	138773.4079
35	69201.35213	141078.7931
40	69248.47320	118647.7494
45	68545.59102	52673.6136
50	68216.31101	29672.0097
55	67515.05574	68642.5062
60	67305.80623	95836.2978
65	67027.26371	174672.0931
70	67047.72080	148823.2034
75	67069.01520	149656.3060
80	66967.77016	166926.3797
85	67217.87572	181447.9011
90	66919.09671	179958.8049
95	67508.95834	189773.6090
100	67662.42041	190310.5511

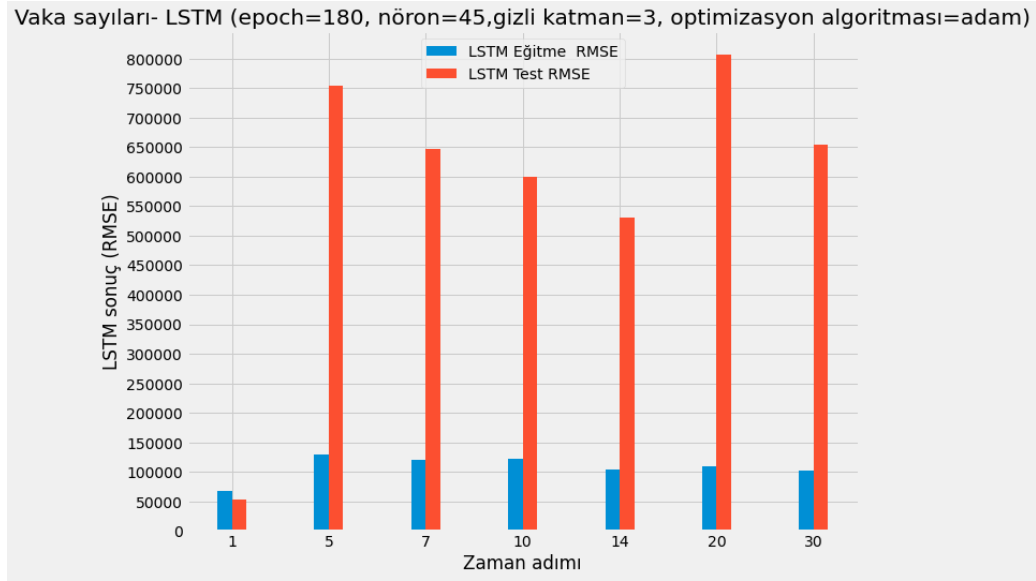


Şekil 4.14. LSTM için farklı nöron sayısında RMSE değişim grafiği

LSTM için epoch sayısı ve nöron sayıları belirlendikten sonra timestep değerlerinin etkisi araştırılmıştır. Epoch sayısı 180, nöron sayısı 45 olduğu durumda timestep değerleri 5, 7, 10, 14, 20 ve 30 olacak şekilde değiştirilerek deneyler tekrar edilmiştir. En iyi tahmin sonuçları timestep 1 olduğu durumda elde edilmiştir. Çizelge 4.15'te timestep sayısına göre eğitim ve test RMSE değişimleri, Şekil 4.15'te ise değişim grafiği gösterilmiştir.

Çizelge 4.15. LSTM için farklı zaman adımı sayılarında RMSE değerleri

Zaman adımı	Train RMSE	Test RMSE
1	68545.59102	52673.6136
5	130120.58604	753910.8726
7	119574.37050	646709.9898
10	122743.46164	600040.1633
14	104169.30957	530979.5167
20	109573.35034	806785.8175
30	102123.02542	653610.0887

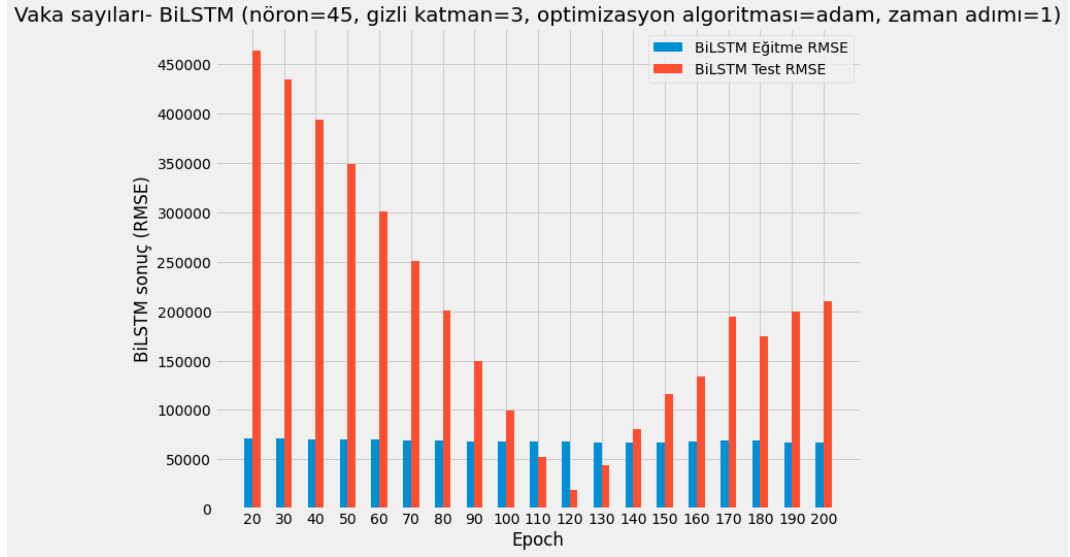


Şekil 4.15. LSTM için farklı zaman adımı sayısında RMSE değişim grafiği

BİLSTM modeli için LSTM modelinde yapılan deneyler tekrarlanmıştır. Çizelge 4.16’de farklı epoch sayıları için RMSE değişimleri, Çizelge 4.17’de farklı nöron sayıları için RMSE değişimleri, Çizelge 4.18’de farklı timestep değerleri için RMSE değişimleri sunulmuştur. Şekil 4.16, Şekil 4.17 ve Şekil 4.18’de sırasıyla epoch, nöron ve timestep değişimlerine karşın RMSE değerlerindeki değişimlerin grafikleri gösterilmiştir. En iyi değerler, 110 epoch, 45 nöron ve timestep 1 değerlerinde elde edilmiştir. 120, 130 epoch için toplam Covid-19 vaka sayıları azalma göstermiştir.

Çizelge 4.16. BİLSTM için farklı epoch sayılarında RMSE değerleri

Epoch	Eğitme RMSE	Test RMSE
20	71186.598	463716.318
30	70875.153	434517.111
40	70547.315	394190.937
50	70187.286	349300.520
60	69806.282	301258.898
70	69415.129	251228.142
80	69024.085	200276.770
90	68642.980	149426.288
100	68281.424	99744.231
110	67948.762	52836.678
120	67654.713	19032.588
130	67407.911	43732.610
140	67219.464	80705.488
150	67078.042	115979.968
160	67938.641	133970.842
170	68683.432	194486.502
180	68759.270	174154.884
190	67309.610	199507.659
200	67504.017	210390.594

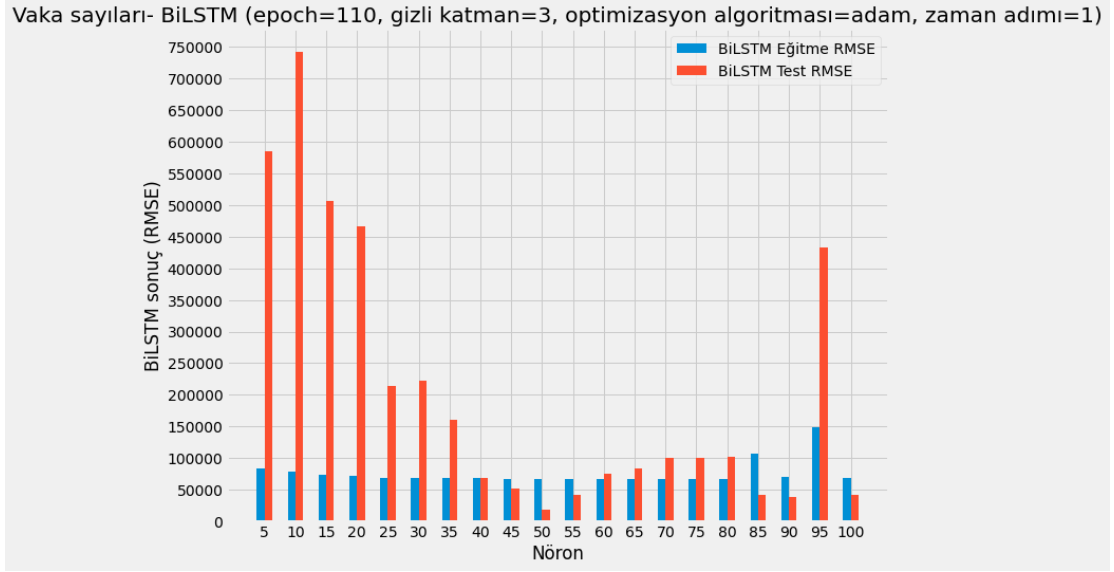


Şekil 4.16. BiLSTM için farklı epoch sayısında RMSE değişim grafiği

Nöron sayısının 50, 55 olduğu durumlarda yeterince öğrenme olmamış ve toplam Covid-19 vaka sayılarında azalma gözlemlenmiştir.

Çizelge 4.17. BiLSTM için farklı nöron sayılarında RMSE değerleri

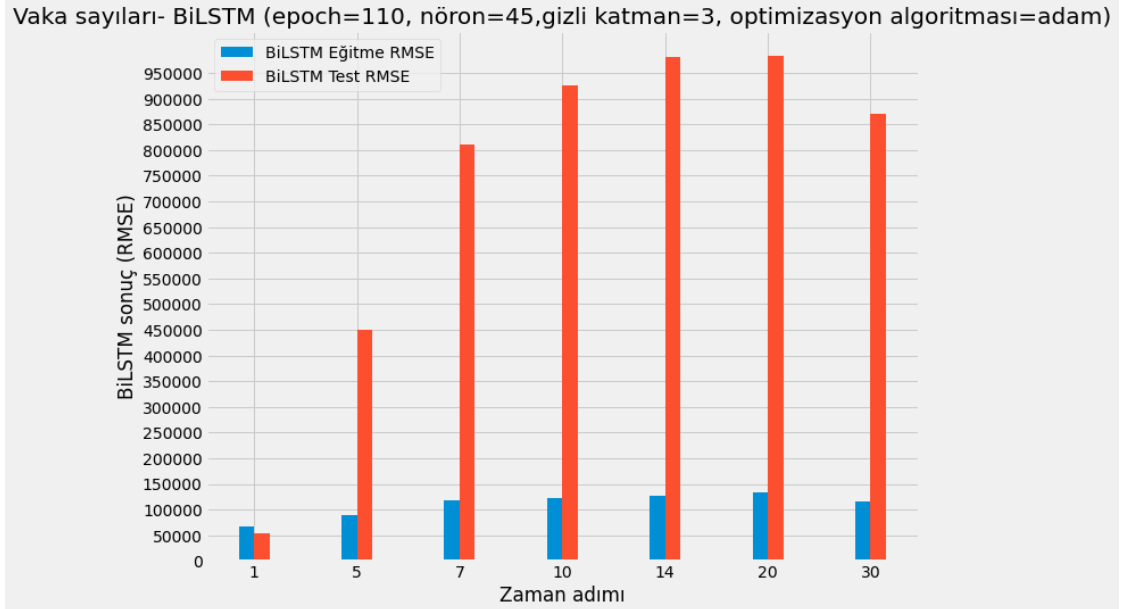
Nöron	Eğitme RMSE	Test RMSE
5	84322.553	584551.602
10	78277.203	741494.002
15	73766.518	506611.824
20	72664.010	465471.028
25	69639.033	214147.309
30	69537.152	222790.858
35	68948.380	160346.396
40	68148.266	69030.290
45	67948.762	52836.678
50	67538.068	18695.381
55	67337.463	42016.778
60	67165.648	75303.714
65	67125.399	84765.297
70	67105.564	100229.563
75	67110.245	100738.739
80	67209.062	101713.605
85	106487.278	42593.746
90	70989.106	38371.326
95	148187.512	432478.000
100	68978.292	42142.068



Şekil 4.17. BiLSTM için farklı nöron sayısında RMSE değişim grafiği

Çizelge 4.18. BiLSTM için farklı zaman adımı sayılarında RMSE değerleri

Zaman adımı	Eğitme RMSE	Test RMSE
1	67948.762	52836.678
5	89576.734	449860.093
7	118305.900	811494.359
10	121849.197	926762.571
14	125840.780	981432.653
20	132415.033	983503.424
30	114964.061	871056.957

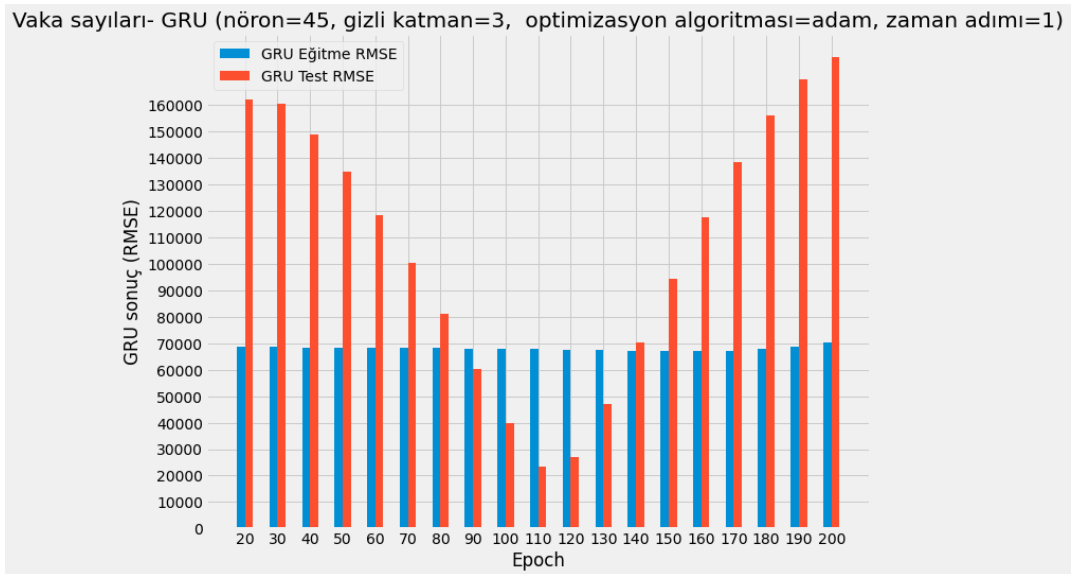


Şekil 4.18. BiLSTM için farklı zaman adımı sayısında RMSE değişim grafiği

GRU modeli için LSTM ve BİLSTM modellerinde yapılan deneyler tekrarlanmıştır. Çizelge 4.19’de farklı epoch sayıları için RMSE değişimleri, Çizelge 4.20’de farklı nöron sayıları için RMSE değişimleri, Çizelge 4.21’de farklı timestep değerleri için RMSE değişimleri sunulmuştur. Şekil 4.19, Şekil 4.20 ve Şekil 4.21’de sırasıyla epoch, nöron ve timestep değişimlerine karşın RMSE değerlerindeki değişimlerin grafikleri gösterilmiştir. En iyi değerler, 110 epoch, 45 nöron ve timestep 1 değerlerinde elde edilmiştir.

Çizelge 4.19. GRU için farklı epoch sayılarında RMSE değerleri

Epoch	Eğitme RMSE	Test RMSE
20	68615.69733	162126.3191
30	68542.17365	160712.2706
40	68482.44108	148905.1156
50	68413.66881	134756.8976
60	68335.78459	118570.8562
70	68248.43465	100571.5456
80	68150.53455	81032.5209
90	68040.46984	60396.2722
100	67916.81312	39724.1464
110	67778.54953	23507.3531
120	67622.40197	26972.4743
130	67457.29031	46881.8785
140	67290.59798	70511.0470
150	67152.03240	94552.3219
160	67100.27648	117634.2677
170	67244.44625	138610.3336
180	67747.80150	156313.1970
190	68750.41934	169688.2246
200	70130.98054	178228.3544

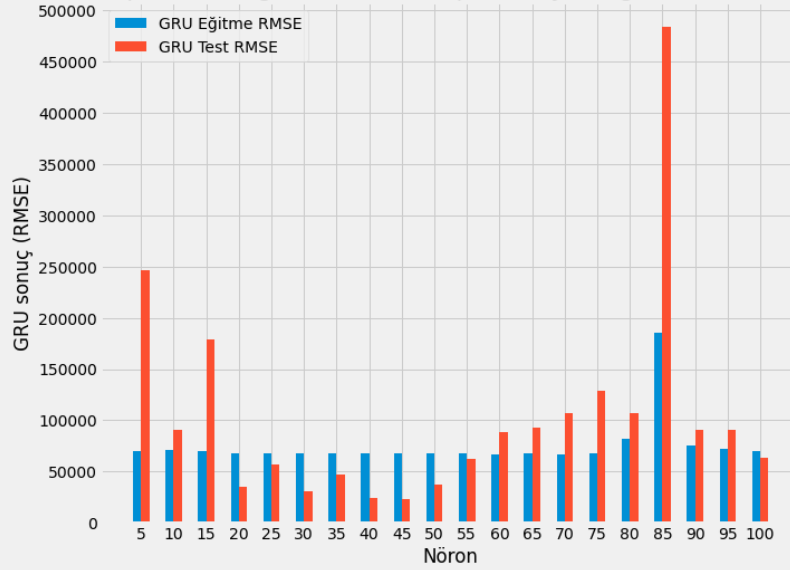


Şekil 4.19. GRU için farklı epoch sayısında RMSE değişim grafiği

Çizelge 4.20. GRU için farklı nöron sayılarında RMSE değerleri

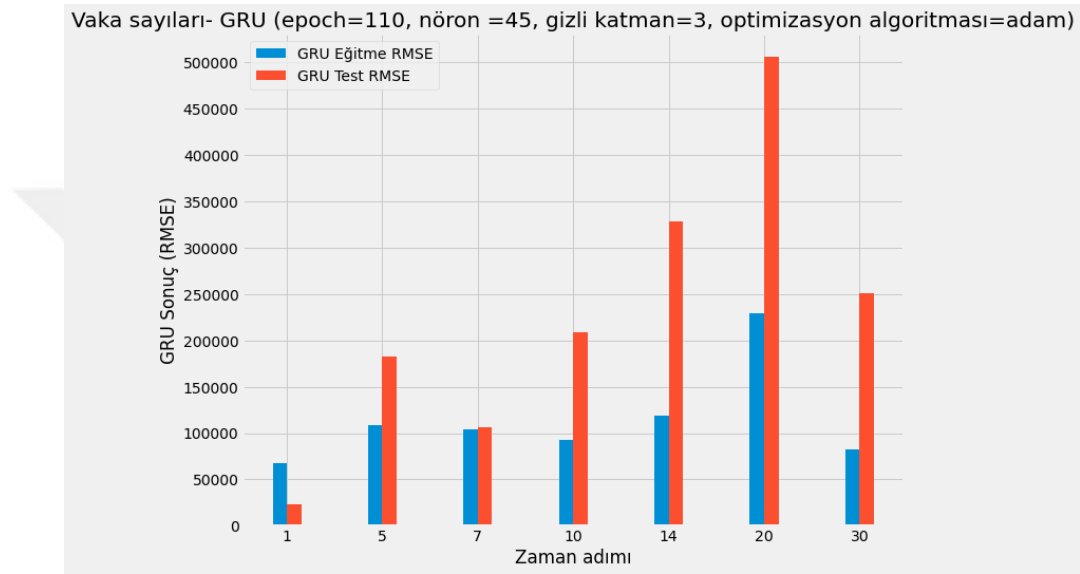
Nöron	Eğitme RMSE	Test RMSE
5	69970.06903	245858.9551
10	70847.51581	90897.4603
15	70216.22194	178624.4873
20	68302.66009	35081.3801
25	68206.16652	57185.9429
30	67868.35634	30449.5501
35	68004.24646	46868.7140
40	67810.14437	24095.2404
45	67778.54953	23507.3531
50	67630.43102	37496.0947
55	67353.76313	62045.5533
60	67273.17159	88426.4749
65	67280.27586	92375.0990
70	67195.26460	107338.7234
75	67381.86818	128315.1480
80	81699.88867	107110.9553
85	185014.02118	483709.7475
90	74985.49150	90755.6501
95	71957.23886	90668.0599
100	70028.49614	63774.4365

Vaka sayıları- GRU (epoch=110, gizli katman=3, optimizasyon algoritması=adam, zaman adımı=1)

**Şekil 4.20.** GRU için farklı nöron sayısında RMSE değişim grafiği

Çizelge 4.21. GRU için farklı zaman adımı sayılarında RMSE değerleri

Zaman adımı	Eğitme RMSE	Test RMSE
1	67778.549	23507.3531
5	108171.516	182436.1340
7	103751.1755	105953.2274
10	92202.4485	208657.9919
14	118397.34051	328911.3116
20	229788.16540	506121.6145
30	82580.56178	250714.1983

**Şekil 4.21.** GRU için farklı zaman adımı sayısında RMSE değişim grafiği

Çizelge 4.22’de LSTM, BiLSTM ve GRU modeli kullanılarak 01/06/2021 ile 30/06/2021 arasında tahmin edilen vaka sayıları ile gerçek vaka sayılarının karşılaştırılması gösterilmiştir.

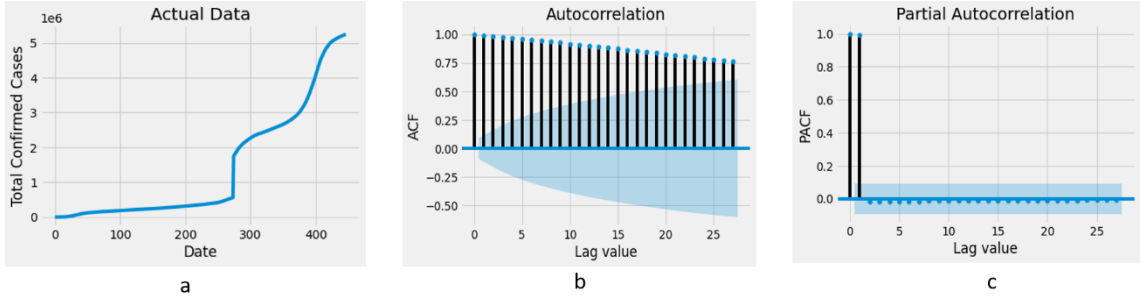
Çizelge 4.22. Derin Öğrenme modelleri tahmin değerleri ve gerçek değerleri

Tarih	LSTM Nöron =45 katman =3, Epoch=180 Optimizer= adam Timestep=1	BiLSTM Nöron =45 katman =3 Epoch=110 Optimizer= adam Timestep=1	GRU Nöron =45 katman =3 Epoch=110 Optimizer= adam Timestep=1	Gerçek vaka sayıları
01/06/2021	5333093	5353051	5274652	5256516
02/06/2021	5412086	5456023	5298909	5263697
03/06/2021	5486187	5557964	5322194	5270299
04/06/2021	5555290	5658516	5344528	5276468
05/06/2021	5619369	5757331	5365934	5282594
06/06/2021	5678470	5854068	5386431	5287980
07/06/2021	5732705	5948413	5406046	5293627
08/06/2021	5782245	6040068	5424804	5300236
09/06/2021	5827296	6128769	5442728	5306690
10/06/2021	5868102	6214279	5459844	5313098
11/06/2021	5904924	6296403	5476180	5319359
12/06/2021	5938043	6374982	5491761	5325435
13/06/2021	5967737	6449893	5506613	5330447
14/06/2021	5994288	6521055	5520762	5336073
15/06/2021	6017968	6588422	5534234	5342028
16/06/2021	6039041	6651982	5547056	5348249
17/06/2021	6057757	6711763	5559252	5354153
18/06/2021	6074348	6767818	5570848	5359728
19/06/2021	6089034	6820228	5581869	5365208
20/06/2021	6102013	6869094	5592338	5370299
21/06/2021	6113471	6914539	5602279	5375593
22/06/2021	6123574	6956698	5611716	5381736
23/06/2021	6132474	6995720	5620671	5387545
24/06/2021	6140307	7031763	5629165	5393248
25/06/2021	6147195	7064986	5637220	5398878
26/06/2021	6153250	7095556	5644855	5404144
27/06/2021	6158567	7123635	5652091	5409027
28/06/2021	6163236	7149384	5658947	5414310
29/06/2021	6167332	7172964	5665438	5420156
30/06/2021	6170925	7194526	5671585	5425652
RMSE	620332.697	1256305.770	186107.675	
MAPE	0.09786	0.17084	0.03110	
MAPE(%)	9.786	17.084	3.11	
Zaman	25 s	33s	18s	

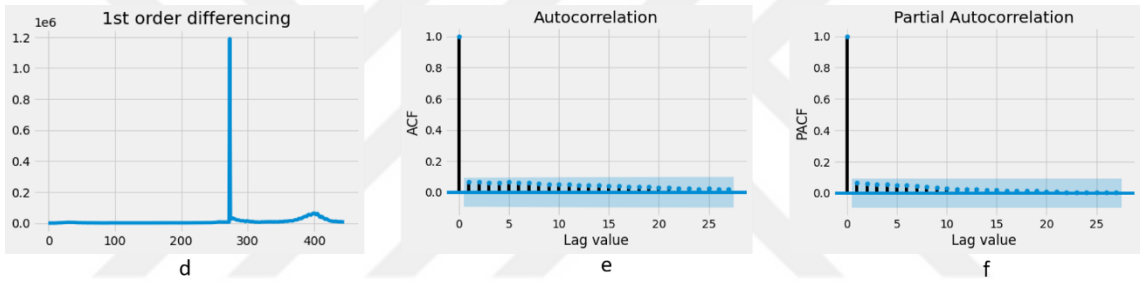
4.4. ARIMA Yöntemi ile Vaka Sayılarının Tahmini

Şekil 4.22. (a), 4.22. (b), 4.22. (c)'de vaka sayılarının gerçek değerleri, gerçek verilerin ACF grafiği ve gerçek verilerin PACF grafiği gösterilmiştir. Bu şekil veri setinin durağan olmadığını göstermektedir. Veri kümesini durağan hale getirmek için bir fark alma yöntemi uygulanmıştır. Şekil 4.23. (d), 4.23. (e), 4.23. (f) farkın alınmasından sonra gerçek verileri, ACF çizim verisini, PACF çizimini göstermektedir. Şekil 4.23. (d), zaman serisi verilerinin durağan olduğunu göstermektedir. Şekil 4.23.

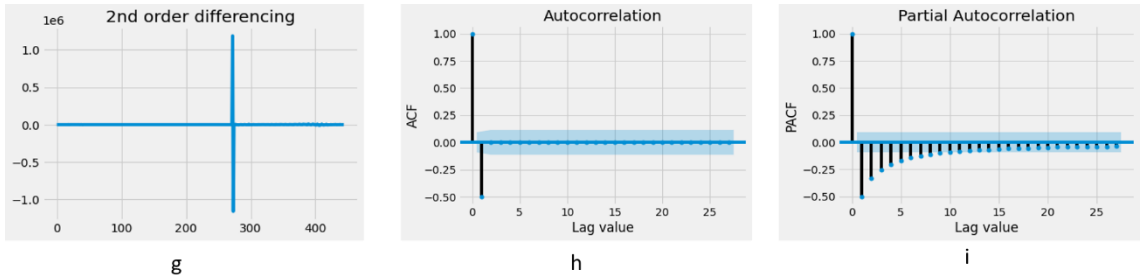
(e) ACF grafiği gecikme değeri 0'dan sonra bir kesme gösterir ve şekil 4.23. (f) Verilerin PACF grafiği, gecikme değeri 0'dan sonra bir kesme gösterir. Şekil 4.23.'e göre $(p,d,q) = (0,1,0)$ olarak tespit edilebilir. Şekil 4.24. (g), 4.24. (h), 4.24. (i) iki fark alındıktan sonra gerçek veriyi, ACF çizim verisini, PACF çizim verisini gösterir.



Şekil 4.22. a gerçek veriyi, b gerçek verinin ACF grafiğini, c gerçek verinin PACF grafiğini gösterir



Şekil 4.23. d bir farktan sonraki gerçek verileri, e, bir farktan sonraki ACF çizim verilerini, f, bir farktan sonraki PACF çizim verilerini gösterir.



Şekil 4.24. g iki farktan sonraki gerçek verileri, h iki farktan sonra ACF çizim verilerini, i iki farktan sonra PACF çizim verilerini gösterir

GridSearch() kullanarak da ARIMA modelinin p, d, q değerleri tespit edilmiştir. p değerleri $[0-9]$ aralığında, q değerleri $[0-2]$ aralığında ve d değerleri $[0-2]$ aralığında araştırılmıştır. İşlem süresi toplam 1 saat 39 dk sürmüştür. Araştırma işlemine ilişkin bazı veriler aşağıda sunulmuştur. Çizelge 4.23.'te ARIMA modelin RMSE değerleri gösterilmiştir.

Çizelge 4.23. ARIMA modelin RMSE degerli

ARIMA (p,d,q)	RMSE
ARIMA(0, 0, 0)	2799254.168
ARIMA(0, 0, 1)	1431674.951
ARIMA(0, 1, 0)	20828.249
ARIMA(0, 1, 1)	19957.846
ARIMA(0, 1, 2)	19280.862
ARIMA(0, 2, 0)	2317.981
ARIMA(0, 2, 1)	17010.873
ARIMA(1, 0, 0)	28488.054
ARIMA(1, 1, 0)	19839.789
ARIMA(1, 1, 1)	14565.334
ARIMA(1, 2, 0)	2947.232
ARIMA(2, 0, 0)	26677.720
ARIMA(2, 1, 0)	19015.025
ARIMA(2, 1, 1)	14546.026
ARIMA(2, 2, 0)	3433.785
ARIMA(3, 0, 0)	25884.848
ARIMA(3, 1, 0)	18321.310
ARIMA(3, 1, 1)	14521.822
ARIMA(3, 2, 0)	3821.913
ARIMA(4, 0, 0)	24343.245
ARIMA(4, 1, 0)	17669.830
ARIMA(4, 1, 1)	14492.197
ARIMA(4, 2, 0)	4170.740
ARIMA(5, 0, 0)	22586.506
ARIMA(5, 1, 0)	17057.990
ARIMA(5, 1, 1)	14472.867
ARIMA(5, 2, 0)	4498.357
ARIMA(6, 0, 0)	21367.445
ARIMA(6, 1, 0)	16535.811
ARIMA(6, 1, 1)	14466.371
ARIMA(6, 2, 0)	4858.105
ARIMA(7, 0, 0)	20443.804
ARIMA(7, 1, 0)	16105.759
ARIMA(7, 1, 1)	14462.725
ARIMA(7, 2, 0)	5293.272
ARIMA(8, 0, 0)	19534.080
ARIMA(8, 1, 0)	15743.745
ARIMA(8, 1, 1)	14459.018
ARIMA(8, 2, 0)	5753.734
ARIMA(9, 0, 0)	18989.793
ARIMA(9, 1, 0)	15479.880
ARIMA(9, 1, 1)	14458.008
ARIMA(9, 2, 0)	6204.144

En iyi ARIMA parametre değerleri yapılan araştırma sonucunda gridsearch ile (0, 2, 0) ve manuel olarak (0, 1, 0) bulunmuştur. Çizelge 4.24.'te en iyi ARIMA modelleri için vaka tahmin sayıları ve gerçek vaka sayıları gösterilmiştir.

Çizelge 4.24. ARIMA modeli vaka tahmin sayısı ve gerçek vaka sayısı

Tarih	ARIMA manual (0,1,0)	ARIMA Auto (0,2,0)	Gerçek vaka sayıları
01/06/2021	5254692	5256823	5256516
02/06/2021	5266474	5263803	5263697
03/06/2021	5278256	5270799	5270299
04/06/2021	5290038	5277810	5276468
05/06/2021	5301820	5284836	5282594
06/06/2021	5313601	5291879	5287980
07/06/2021	5325383	5298937	5293627
08/06/2021	5337165	5306010	5300236
09/06/2021	5348947	5313099	5306690
10/06/2021	5360729	5320204	5313098
11/06/2021	5372511	5327324	5319359
12/06/2021	5384292	5334460	5325435
13/06/2021	5396074	5341612	5330447
14/06/2021	5407856	5348779	5336073
15/06/2021	5419638	5355962	5342028
16/06/2021	5431420	5363161	5348249
17/06/2021	5443201	5370375	5354153
18/06/2021	5454983	5377604	5359728
19/06/2021	5466765	5384850	5365208
20/06/2021	5478547	5392111	5370299
21/06/2021	5490329	5399387	5375593
22/06/2021	5502111	5406679	5381736
23/06/2021	5513892	5413987	5387545
24/06/2021	5525674	5421310	5393248
25/06/2021	5537456	5428649	5398878
26/06/2021	5549238	5436004	5404144
27/06/2021	5561020	5443374	5409027
28/06/2021	5572801	5450760	5414310
29/06/2021	5584583	5458161	5420156
30/06/2021	5596365	5465578	5425652
RMSE	96824.002	20486.0945	
MAPE	0.01524	0.00305	
MAPE(%)	1.524	0.3	

ARIMA modeli ile elde edilen vaka sayısı tahmini, derin öğrenme modellerine ait vaka sayısı tahminleri ile karşılaştırıldığında, ARIMA modelinin gerçek vaka sayısına daha yakın sonuçlar ürettiği gözlemlenmiştir. ARIMA modeli (0, 2, 0) katsayıları ile elde edilen hatalar, derin öğrenme yöntemleri ile elde edilen hatalardan daha azdır.

Daha önceki çalışmaların ve tez çalışmalarının özeti Çizelge 4.25'te verilmiştir.

Çizelge 4.25. Literatür karşılaştırması

Referans	vakalar	Ülke	modeli	Sonuçlar
(Chintalapudi ve ark., 2020)	Kayıtlı Vakalar, Kurtarılan Vakalar	İtalya	ARIMA	RMSE Kayıtlı Vakalar = 514,74 RMSE Kurtarılan Vakalar = 186,85
(Alzahrani ve ark., 2020)	Onaylanmış vakalar	Suudi Arabistan	AR, MA, ARMA, ARIMA	AR RMSE = 180,29 R2 Puan = 0,69 MA RMSE = 241,05 R2 Puan = 0,46 ARMA RMSE = 117,22 R2 Puan = 0,87 ARIMA RMSE = 21,17 R2 Puan = 0,99
(Chakraborty ve Ghosh, 2020)	Onaylanmış vakalar	Kanada, Fransa, Hindistan, Güney Kore, İngiltere.	ARIMA, WBF, Hybrid ARIMA-WBF model	Kanada RMSE ARIMA=150,05 WBF=202,64 ARIMA-WBF=149,60 Fransa RMSE ARIMA=710,46 WBF=740,06 ARIMA-WBF=631,91 Hindistan RMSE ARIMA= 50,83 WBF= 68,38 ARIMA-WBF= 55, 25 Güney Kore RMSE ARIMA=81,81 WBF=82,78 ARIMA-WBF=90,29 İngiltere RMSE ARIMA=209,36 WBF=405,87 ARIMA-WBF=180,66
(Ardabili ve ark., 2020)	Onaylanmış vakalar	İtalya, Almanya, İran, ABD, İngiltere Çin	MLP, ANFIS	Italy RMSE MLP= 191,27 ANFIS= 1418,1 China RMSE MLP=2318,22 ANFIS=2960,81 Iran RMSE MLP=391,8 ANFIS=391,39 Germany RMSE MLP=55,52 ANFIS=623,31 USA RMSE MLP=22,1 ANFIS=323,05
(Al-Qaness ve ark., 2020)	Günlük vakalar	Çin	ANFIS FPASSA	ANFIS: RMSE=7375 R ² score=0,90 FPASSA: RMSE=5779 R ² score=0,96
(Zeroual ve ark., 2020)	Onaylanmış vakalar, Kurtarılan vakalar	İtalya, İspanya, Fransa, Çin, Avustralya, ABD	RNN, LSTM, BiLSTM, GRU, VAE	Onaylanmış vakalar RMSE İtalya: RNN=1.070.474 LSTM=1.054.089 Kurtarılan vakalar RMSE İtalya: RNN= 4.006.731 LSTM= 3.887.027

BiLSTM=1.041.374	BiLSTM=3.886.682
GRU=113.775	GRU= 3.892.321
VAE= 1.386.225	VAE= 2.273.616
İspanya:	İspanya:
RNN= 1.683.011	RNN= 9.460.326
LSTM=	LSTM= 368.623
1.254.449	BiLSTM=
BiLSTM=	1.850.373
1.194.711	GRU= 2.221.248
GRU=1.795.678	VAE= 4.022.143
VAE= 5.315.748	Fransa:
Fransa:	RNN= 100.203
RNN= 1.287.786	LSTM=
LSTM=	7.880.727
1.085.008	BiLSTM=
BiLSTM=	785.049
1.168.893	GRU= 8.467.965
GRU= 1.204.139	VAE= 3.946.933
VAE= 3.688.083	Çin:
Çin:	RNN= 856.465
RNN= 1.252.034	LSTM= 784.579
LSTM= 101.482	BiLSTM=
BiLSTM=	1.031.264
1.205.955	GRU= 957.498
GRU= 1.085.698	VAE= 415.008
VAE= 11.103	Avustralya:
Avustralya:	RNN= 630.847
RNN= 39.928	LSTM= 35.956
LSTM= 327.123	BiLSTM=
BiLSTM=	429.146
335.033	GRU= 48.319
GRU= 295.978	VAE= 972.236
VAE= 18.732	ABD:
ABD:	RNN= 2.301.648
RNN= 5.227.287	LSTM=
LSTM=	2.227.272
1.129.183	BiLSTM=
BiLSTM=	2.227.271
4.330.228	GRU= 2.252.458
GRU= 4.369.108	VAE= 1.194.272
VAE= 4.079.244	

(Elsheikh ve ark., 2021)	Onaylanmış vakalar, vefat sayılar	Brezilya, Hindistan, Suudi Arabistan, Güney Afrika, İspanya, Amerika	LSTM	Onaylanmış vakalar (Brezilya): RMSE=18609,7 R ² =0,979 (Hindistan): RMSE=34625,08 R ² =0,996 (Suudi Arabistan): RMSE=160,60 R ² =0,99 (Güney Afrika): RMSE=5346,5 R ² =0,97 (İspanya): RMSE=8257,56 R ² =0,97 (Amerika): RMSE=9205,8 R ² =0,996	vefat sayılar (Brezilya): RMSE=388,2 R ² =0,983 (Hindistan): RMSE=735,183 R ² =0,998 (Suudi Arabistan): RMSE=100,03 R ² =0,998 (Güney Afrika): RMSE=143,23 R ² =0,97 (İspanya): RMSE=112,78 R ² =0,969 (Amerika): RMSE=709,7 R ² =0,971
(Shastri ve ark., 2020)	Onaylanmış vakalar, Ölüm	Hindistan, Amerika Birleşik	Stacked LSTM, BiLSTM,	Onaylanmış vakalar (Hindistan):	Ölüm Vakaları (Hindistan):

Vakaları	Devletleri	ConVLSTM	Stacked LSTM MAPE =4,00 BiLSTM MAPE =3,33 ConVLSTM MAPE =2,17 (ABD): Stacked LSTM MAPE =10,00 Bi-LSTM MAPE =6,66 ConVLSTM MAPE =2,00	Stacked LSTM MAPE =6,66 BiLSTM MAPE =4,00 ConVLSTM MAPE =3,33 (ABD): Stacked LSTM MAPE =8,00 BiLSTM MAPE =6,00 ConVLSTM MAPE =2,50
Tez çalışmamız	Onaylanmış vakalar, Ölüm Vakaları	Türkiye	LSTM, BiLSTM GRU, ARIMA	Onaylanmış vakalar RMSE LSTM =52673,6136 BiLSTM=52836,0 4 GRU=23507,35 ARIMA = 2317,98
				Ölüm Vakaları RMSE LSTM=415,47 BiLSTM=218,38 GRU=194,00 ARIMA=14,059

Tez çalışması kapsamında vaka sayılarının tahmini ve vefat sayılarının tahmini amacıyla kullanılan yöntemlerden elde edilen sonuçlar literatürdeki sonuçlar ile karşılaştırıldığında,

- ARIMA modeli ile elde edilen vefat sayılarına ait hata sonuçlarının literatüre göre iyi performans gösterdiği
 - Derin öğrenme yöntemleri ile elde edilen vefat sayılarına ait hata sonuçlarının literatüre benzer kabul edilebilir performans gösterdiği
 - ARIMA modeli ile elde edilen vaka sayılarına ait hata sonuçlarının literatüre benzer kabul edilebilir performans gösterdiği
 - Derin öğrenme yöntemleri ile elde edilen vaka sayılarına ait hata sonuçlarının literatüre göre daha düşük performans gösterdiği
- sonuçlarına ulaşılmıştır.

5. SONUÇLAR VE ÖNERİLER

5.1 Sonuçlar

Günümüzde salgın hastalıkların sayısı hızla artmaktadır. Salgın hastalıklar, bireyler, toplumlar ve ülkeler için yıkıcı sonuçlar ortaya çıkarabilmektedir. Covid-19, 2019 yılında ortaya çıkan insandan insana bulaşma yeteneğine sahip bir solunum yolu hastalığıdır ve Dünya Sağlık Örgütü tarafından pandemi olarak tanımlanmıştır. Günümüzde farklı varyantları ile halen etkinliğini sürdürmektedir. Ekonomik, toplumsal, kültürel, siyasi vb. birçok alanda önemli sonuçları olmuştur.

Salgın hastalık durumlarında, yöneticiler, karar alıcılar için kısa, orta ve uzun vadede vaka sayıları, salgın kaynaklı vefat sayıları, iyileşebilecek hasta sayısı vb. tahmin edilebilmesi oldukça önemli olmaktadır.

Bu tez çalışması kapsamında Türkiye'deki Covid-19 kaynaklı vaka ve vefat sayılarının kısa dönemli tahmini için LSTM, BiLSTM, GRU derin öğrenme modelleri ve ARIMA yöntemine dayalı tahmin modelleri geliştirilmiştir. Sağlık Bakanlığı'nın web sayfasında yer alan günlük Covid-19 verileri kullanılarak önerilen modeller eğitilip test edilmiştir. Sonra gelecek 1 ay içerisinde oluşabilecek vaka sayıları ve vefat sayıları tahmin edilmiştir. Elde edilen sonuçlar aynı dönem için gerçekleşen gerçek vaka ve vefat sayıları ile karşılaştırılmıştır.

Hem vaka, hem de vefat sayısı açısından gerçek sonuçlara en yakın sonuçlar, ARIMA modeli ile elde edilmiştir. Derin öğrenme modelleri içerisinde en iyi sonuçlar GRU modeli ile elde edilmiştir. Vefat sayısı tahmin sonuçları, vaka sayısı tahmin sonuçlarından daha yüksek performansa sahiptir. Vaka ve vefat sayısı açısından ARIMA modeli ile elde edilen sonuçlar literatürde yer alan diğer ülkelere ait sonuçlarla karşılaştırılabilir düzeydedir. Derin öğrenme modelleri kullanılarak gerçekleştirilen vaka sayısı tahmin sonuçları literatüre göre düşük performansa sahiptir. LSTM modellerinin hatalı sonuçlar üretebildiği gözlemlenmiştir.

GRU, ARIMA tahmin modelleri ile elde edilen sonuçlar gerçekleşen sonuçlara oldukça yakındır. Bu sonuçlar önerilen modellerden bazılarının kısa dönemli tahmin işlemleri için kullanılabileceğini göstermektedir. Önerilen modeller, yöneticiler için kısa zamanda meydana gelecek salgın kaynaklı vaka ve vefat sayılarının değişimi hakkında bilgi sağlayabilecektir.

5.2 Öneriler

Bu çalışma kapsamında kullanılan veriler, salgına ilişkin verilerin bir kısmını içermektedir. Salgının başlangıcından günümüze kadar olan veriler kullanılarak modellerin performansı yeniden değerlendirilebilir.

Salgın kaynaklı vaka ve vefat sayılarının tahmini için farklı makine öğrenmesi modelleri ile çalışmalar gerçekleştirilebilir.

Açık kaynaklarda yayınlanan, farklı ülkelere ait, salgın kaynaklı vaka ve vefat sayıları kullanılarak, önerilen modellerin performansı değerlendirilebilir.

Vaka ve vefat sayılarının orta ve uzun dönemdeki değişimlerini tahmin edebilmek için yeni tahmin modelleri kullanılabilir.



KAYNAKLAR

- Abdulmajeed, K., Adeleke, M. ve Popoola, L., 2020, Online forecasting of COVID-19 cases in Nigeria using limited data, *Data in Brief*, 30, 105683.
- Al-Qaness, M. A., Ewees, A. A., Fan, H. ve Abd El Aziz, M., 2020, Optimization method for forecasting confirmed cases of COVID-19 in China, *Journal of Clinical Medicine*, 9 (3), 674.
- Alzahrani, S. I., Aljamaan, I. A. ve Al-Fakih, E. A., 2020, Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions, *Journal of Infection and Public Health*, 13 (7), 914-919.
- Ardabili, S. F., Mosavi, A., Ghamisi, P., Ferdinand, F., Varkonyi-Koczy, A. R., Reuter, U., Rabczuk, T. ve Atkinson, P. M., 2020, COVID-19 Outbreak Prediction with Machine Learning, *MedRxiv*, 2020.2004.2017.20070094.
- ArunKumar, K., Kalaga, D. V., Kumar, C. M. S., Chilkoor, G., Kawaji, M. ve Brenza, T. M., 2021, Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA), *Applied Soft Computing*, 107161.
- Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M. ve Kalhori, S. R. N., 2020, Predicting COVID-19 incidence through analysis of google trends data in iran: data mining and deep learning pilot study, *JMIR Public Health and Surveillance*, 6 (2), e18828.
- Balkhair, A. A., 2020, COVID-19 pandemic: a new chapter in the history of infectious diseases, *Oman Medical Journal*, 35 (2), e123.
- Benvenuto, D., Giovanetti, M., Vassallo, L., Angeletti, S. ve Ciccozzi, M., 2020, Application of the ARIMA model on the COVID-2019 epidemic dataset, *Data in Brief*, 29, 105340.
- Box, G. E. ve Jenkins, G. M., 1976, Time Series Analysis. Forecasting and Control (rev. ed.), San Francisco: Holden-Day.
- Bulut, C. ve Kato, Y., 2020, Epidemiology of COVID-19, *Turkish Journal of Medical Sciences*, 50 (SI-1), 563-570.
- Chakraborty, T. ve Ghosh, I., 2020, Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: A data-driven analysis, *Chaos, Solitons & Fractals*, 135, 109850.
- Chaudhry, R. M., Hanif, A., Chaudhary, M. ve Minhas, S., 2020, Coronavirus disease 2019 (COVID-19): forecast of an emerging urgency in Pakistan, *Cureus*, 12 (5).
- Chen, D.-G., Chen, X. ve Chen, J. K., 2020, Reconstructing and forecasting the COVID-19 epidemic in the United States using a 5-parameter logistic growth model, *Global Health Research and Policy*, 5 (1), 1-7.
- Chimmula, V. K. R. ve Zhang, L., 2020, Time series forecasting of COVID-19 transmission in Canada using LSTM networks, *Chaos, Solitons & Fractals*, 135, 109864.
- Chintalapudi, N., Battineni, G. ve Amenta, F., 2020, COVID-19 virus outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach, *Journal of Microbiology, Immunology and Infection*, 53 (3), 396-403.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. ve Bengio, Y., 2014, Learning phrase representations using RNN encoder-decoder for statistical machine translation, *ArXiv Preprint ArXiv:1406.1078*.

- De Wit, E., Van Doremalen, N., Falzarano, D. ve Munster, V. J., 2016, SARS and MERS: recent insights into emerging coronaviruses, *Nature Reviews Microbiology*, 14 (8), 523-534.
- Dehesh, T., Mardani-Fard, H. ve Dehesh, P., 2020, Forecasting of covid-19 confirmed cases in different countries with arima models, *MedRxiv*.
- Dil, S., Dil, N. ve Maken, Z. H., 2020, COVID-19 trends and forecast in the Eastern Mediterranean Region with a Particular Focus on Pakistan, *Cureus*, 12 (6).
- Elsheikh, A. H., Saba, A. I., Abd Elaziz, M., Lu, S., Shanmugan, S., Muthuramalingam, T., Kumar, R., Mosleh, A. O., Essa, F. ve Shehabeldeen, T. A., 2021, Deep learning-based forecasting model for COVID-19 outbreak in Saudi Arabia, *Process Safety and Environmental Protection*, 149, 223-233.
- European Centre for Disease Prevention and Control, 2021, European Centre for Disease Prevention and Control (ECDC), COVID-19 situation update worldwide, as of week 1 2021, <https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>.: [30 Haziran 2021].
- Gorbalenya, A. E., Baker, S. C., Baric, R. S., Groot, R. J. d., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L. M., Samborskiy, D. V., Sidorov, I. A., Sola, I. ve Ziebuhr, J., 2020, The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2, *Nature Microbiology*, 5 (4), 536-544.
- Graves, A., Fernández, S. ve Schmidhuber, J., 2005, Bidirectional LSTM networks for improved phoneme classification and recognition, *International Conference On Artificial Neural Networks*, 799-804.
- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R. ve Schmidhuber, J., 2016, LSTM: A search space odyssey, *IEEE Transactions On Neural Networks and Learning Systems*, 28 (10), 2222-2232.
- Gupta, R. ve Pal, S. K., 2020, Trend Analysis and Forecasting of COVID-19 outbreak in India, *MedRxiv*.
- Hochreiter, S. ve Schmidhuber, J., 1997, Long short-term memory, *Neural Computation*, 9 (8), 1735-1780.
- Maleki, M., Mahmoudi, M. R., Wraith, D. ve Pho, K.-H., 2020, Time series modelling to forecast the confirmed and recovered cases of COVID-19, *Travel Medicine and Infectious Disease*, 37, 101742.
- McCarty, M. F. ve DiNicolantonio, J. J., 2020, Nutraceuticals have potential for boosting the type 1 interferon response to RNA viruses including influenza and coronavirus, *Progress in Cardiovascular Diseases*.
- Moftakhar, L., Mozghan, S. ve Safe, M. S., 2020, Exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models, *Iranian Journal of Public Health*, 49 (Supple 1), 92-100.
- Naudé, W., 2020, Artificial Intelligence against COVID-19: An early review.
- Olah, C., 2015, Understanding LSTM Networks, <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> [15 Ocak 2021].
- Pang, L., Liu, S., Zhang, X., Tian, T. ve Zhao, Z., 2020, Transmission dynamics and control strategies of COVID-19 in Wuhan, China, *Journal of Biological Systems*, 28 (03), 543-560.
- Paules, C. I., Marston, H. D. ve Fauci, A. S., 2020, Coronavirus infections—more than just the common cold, *Jama*, 323 (8), 707-708.

- Rana, R., 2016, Gated recurrent unit (GRU) for emotion classification from noisy speech, *ArXiv Preprint ArXiv:1612.07778*.
- Roosa, K., Lee, Y., Luo, R., Kirpich, A., Rothenberg, R., Hyman, J., Yan, P. ve Chowell, G., 2020, Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020, *Infectious Disease Modelling*, 5, 256-263.
- Schuster, M. ve Paliwal, K. K., 1997, Bidirectional recurrent neural networks, *IEEE Transactions on Signal Processing*, 45 (11), 2673-2681.
- Shastri, S., Singh, K., Kumar, S., Kour, P. ve Mansotra, V., 2020, Time series forecasting of Covid-19 using deep learning models: India-USA comparative case study, *Chaos, Solitons & Fractals*, 140, 110227.
- Shim, E., Tariq, A., Choi, W., Lee, Y. ve Chowell, G., 2020, Transmission potential and severity of COVID-19 in South Korea, *International Journal of Infectious Diseases*, 93, 339-344.
- T.C. SAĞLIK BAKANLIĞI, 2020, COVID-19 BİLGİLENDİRME SAYFASI, Genel Koronavirüs Tablosu, <https://covid19.saglik.gov.tr/TR-66935/genel-koronavirus-tablosu.html> [31 Haziran 2021].
- Tamang, S., Singh, P. ve Datta, B., 2020, Forecasting of Covid-19 cases based on prediction using artificial neural network curve fitting technique, *Global Journal of Environmental Science and Management*, 6 (Special Issue (Covid-19)), 53-64.
- TR Ministry of Health, 2021, Covid-19 information page, <https://covid19.saglik.gov.tr/TR-66935/genel-koronavirus-tablosu.html> [10 Ocak, 2021].
- Wang, C., Horby, P. W., Hayden, F. G. ve Gao, G. F., 2020a, A novel coronavirus outbreak of global health concern, *The lancet*, 395 (10223), 470-473.
- Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z. ve Xiong, Y., 2020b, Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China, *Jama*, 323 (11), 1061-1069.
- World Health Organization, 2020, World Health Organization (WHO), WHO Director-General's opening remarks at the media briefing on COVID-19, <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>. [30 Kasım 2020].
- Wu, Z. ve McGoogan, J. M., 2020, Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention, *Jama*, 323 (13), 1239-1242.
- Xia, T., Song, Y., Zheng, Y., Pan, E. ve Xi, L., 2020, An ensemble framework based on convolutional bi-directional LSTM with multiple time windows for remaining useful life estimation, *Computers in Industry*, 115, 103182.
- Yu, Y., Si, X., Hu, C. ve Zhang, J., 2019, A review of recurrent neural networks: LSTM cells and network architectures, *Neural Computation*, 31 (7), 1235-1270.
- Zeroual, A., Harrou, F., Dairi, A. ve Sun, Y., 2020, Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study, *Chaos, Solitons & Fractals*, 140, 110121.
- Zontur, E. Ç., 2020, Turkey confirms first death from coronavirus, <https://www.aa.com.tr/en/health/turkey-confirms-first-death-from-coronavirus/1769849> [30 Kasım 2020].

Zorlu, F., 2020, Turkey confirms first case of coronavirus, <https://www.aa.com.tr/en/latest-on-coronavirus-outbreak/turkey-confirms-first-case-of-coronavirus/1761522>: [30 Kasım 2020].

