



Türkiye'de COVID-19 Bulaşısının ARIMA Modeli ve LSTM Ağı Kullanılarak Zaman Serisi Tahmini

Abdullah Ammar Karcıoğlu^{1*}, Sezercan Tanışman², Hasan Bulut³

^{1*} Ege Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye, (ORCID: 0000-0002-0907-751X), abdullah.ammar.karcioglu@ege.edu.tr

² Ege Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye (ORCID: 0000-0002-8094-708X), sezercan.tanisman@ege.edu.tr

³ Ege Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye (ORCID: 0000-0002-4872-5698), hasan.bulut@ege.edu.tr

(International Conference on Design, Research and Development (RDCONF) 2021 – 15-18 December 2021)

(DOI: 10.31590/ejosat.1039394)

ATIF/REFERENCE: Karcıoğlu, A. A., Tanışman, S., & Bulut, H. (2021). Türkiye'de COVID-19 Bulaşısının ARIMA Modeli ve LSTM Ağı Kullanılarak Zaman Serisi Tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, (32), 288-297.

Öz

11 Mart 2020'de Dünya Sağlık Örgütü (WHO), 2019 yılı yeni tip korona virüsünü küresel salgın olarak ilan etmiştir. COVID-19 olarak da bilinen yeni tip korona virüsü, ilk olarak Aralık 2019'da Çin'in Wuhan şehrinde ortaya çıkmış ve birkaç hafta içinde tüm dünyaya yayılmıştır. Tüm dünyada 5 milyondan fazla insan, Türkiye'de ise 70 bine yakın insan bu hastalıktan dolayı vefat etmiştir. Küresel çapta insan sağlığını tehdit eden ve ekonomik krizlere neden olan bu salgınla mücadelede önceden önlem almak hayati önem taşımaktadır. COVID-19 salgınının yayılmasının tahmin edilmesi bu hastalıkla ilgili ne gibi önlemler alınacağı hakkında fikir verir. Bu çalışmada, ARIMA zaman serisi modeli ve LSTM ağı kullanılarak Türkiye'deki COVID-19 salgınında günlük vaka sayısı, günlük iyileşen sayısı ve günlük vefat sayısı tahmini gerçekleştirilmiştir. Modellerin doğruluğunu test etmek için bilinen ve bilinmeyen verilerin tahmini yapılarak, tahmini yapılan verilerin hata yüzdeleri karşılaştırılmıştır. Türkiye'de COVID-19 bulaşı seyrinin gelecek 15 günlük tahmini için yapılan deneysel çalışmalar sonucu günlük vaka sayılarında ve günlük vefat sayılarında ARIMA modeliyle yüksek doğrulukta tahminleme yapılırken, günlük iyileşen sayılarında ise LSTM modeliyle yüksek doğrulukta tahminleme yapıldığı gözlemlenmiştir. ARIMA ve LSTM modellerinde günlük vaka ve vefat sayılarında azalma seyri gözlemlenmiştir. Ancak günlük iyileşme sayılarında ARIMA modelinde azalma, LSTM modelinde artış gözlemlenmiştir.

Anahtar Kelimeler: Zaman Serisi Tahmini, Makine Öğrenmesi, COVID-19.

Time Series Forecasting of COVID-19 Transmission in Turkey Using ARIMA Model and LSTM Network

Abstract

On March 11, 2020, the World Health Organization (WHO) declared the new coronavirus 2019 as a global epidemic. The new type of coronavirus, also known as COVID-19, first appeared in Wuhan city of China in December 2019 and spread all over the world within a few weeks. More than 5 million people all over the world and nearly 70,000 people in Turkey have died due to this disease. It is vital to take precautionary measures in the fight against this epidemic, which threatens human health on a globally and causes economic crises. The forecasting of the spread of the COVID-19 epidemic gives an idea about what measures to take regarding this disease. In this study, daily number of cases, daily number of recoveries and daily number of deaths were estimated in the COVID-19 outbreak in Turkey using the ARIMA time series model and LSTM network. To test the accuracy of the models, known and unknown data were estimated and the error percentages of the estimated data were compared. As a result of the experimental studies carried out to predict the course of COVID-19 transmission in Turkey for the next 15-days, it has been observed that the daily number of cases and daily number of death sare estimated with high accuracy in the ARIMA model, while the daily number of recoveries are estimated with high accuracy in the LSTM model. In the ARIMA and LSTM models, a decrease was observed in the number of daily cases and deaths. However, a decrease in the daily number of recoveries in the ARIMA model and an increase in the LSTM model were observed.

Keywords: Time Series Forecasting, Machine Learning, COVID-19.

* Sorumlu Yazar: abdullah.ammar.karcioglu@ege.edu.tr

1. Giriş

COVID-19, 11 Mart 2020 tarihinden bu yana Dünya Sağlık Örgütü (WHO) tarafından küresel salgın ilan edildiğinden beri, şiddetli akut solunum sendromu salgına dönüşerek dünya çapında etkisi büyük bir pandemi haline gelmiştir. Küresel çapta ülkeleri olumsuz yönde etkileyen bu pandemi, ekonomik krizlere, işsizliğe, eğitim-öğretim programlarının aksamasına neden olmuştur [1]. Dünyada birçok sektörde olumsuzluklara neden olan COVID-19 salgını, hükümet ve araştırmacıların, akademik kurumların ve endüstri kuruluşlarının, salgını önlemeye yönelik olarak ortak hedefler etrafında birleşmesine neden olmuştur. Sağlık hizmetlerinin kaynaklarının planlanması, salgını önlemek için aşı geliştirme çalışmalarının başlatılması, gündelik hayatı normalleştirecek adımların atılması, yeni eğitim yöntemlerinin geliştirilmesi gibi salgınla alakalı birçok çıktının elde edilmesini sağlamıştır.[2].

Türkiye'de 10 Mart 2020'de ilk COVID-19 hastasıyla vakalar bildirilmeye başlanmıştır. Virüsün yayılma şiddeti ortaya çıkınca, pandemiye kontrol altına almak için Türkiye hükümeti tarafından kısmi sokağa çıkma yasağı uygulanmıştır. İlk izolasyon duyurulmuş ve kademeli olarak 10 Nisan'dan Haziran'a kadar uzatılmıştır [3]. 29 Nisan- 17 Mayıs 2021 tarihleri arası iki haftalık sokağa çıkma yasağı ilan edilmiştir. Yapılan bu karantina önlemlerinde vaka sayılarında ciddi bir düşüş elde edilmiştir. 1 Temmuz itibarıyla de yurt genelinde yasaklar kaldırılmış ve vaka sayılarında artış gözlemlenmiştir. Hükümet tarafından birçok maske, mesafe ve sokağa çıkma gibi önlemler alınmasına rağmen artan vaka sayılarının tahmin edilebilir olması, sağlık altyapısının daha verimli kullanılmasını ve ekonomik istikrarı yeniden sağlama yönünde olumlu izlenimler kazandıracaktır.

Hükümetler çok sayıda kısıtlama ve sosyal mesafe önlemi uygulasa da sağlık sistemlerine olan ihtiyaç önemli ölçüde artmıştır ve enfekte hastaların etkin yönetimi hastaneler için bir sorun haline gelmiştir. Bu nedenle, mevcut kaynakları optimize etmek ve hastalıkların ilerlemesini durdurmak veya yavaşlatmak için vaka sayılarının kısa vadeli tahmini önem arz etmektedir [4]. Ülke genelinde meydana gelen hastalığın yayılma hızı, sayısı ve yönü önceden tahmin edilebilirse hükümetlerin alacağı kararlar, sağlık hizmetlerinin hastalığı önlemede uygulayacağı algoritmalar öngörülebilir. Böylelikle, hastalığa karşı zaman serisi modellerine dayalı algoritmaları kullanarak önceden elde ettiğimiz tahminler sayesinde sağlık hizmetlerinde uygulanacak programların daha etkili oluşturulması sağlanacaktır.

Bu çalışmada, Türkiye'de meydana gelen COVID-19 salgınında günlük vaka sayısı, günlük iyileşen sayısı ve günlük vefat sayısı ARIMA zaman serisi modeli ve LSTM sinir ağı kullanılarak salgının zaman serisi tahmini gerçekleştirilmiştir. Yapılan çalışmada 15 günlük gelecek tahmini yapılarak farklı modellerle elde edilen sonuçlar karşılaştırılmıştır. Çalışmanın geri kalanı şu şekilde yapılandırılmıştır: İkinci bölümde, ilgili çalışmalar tanıtılmaktadır. Çalışmanın üçüncü bölümünde, materyal ve yöntemler, dördüncü bölümde deneysel sonuçlar, son bölümde ise çalışmanın genel sonuçlarına ve gelecekte yapılabilecek çalışmalara değinilmektedir.

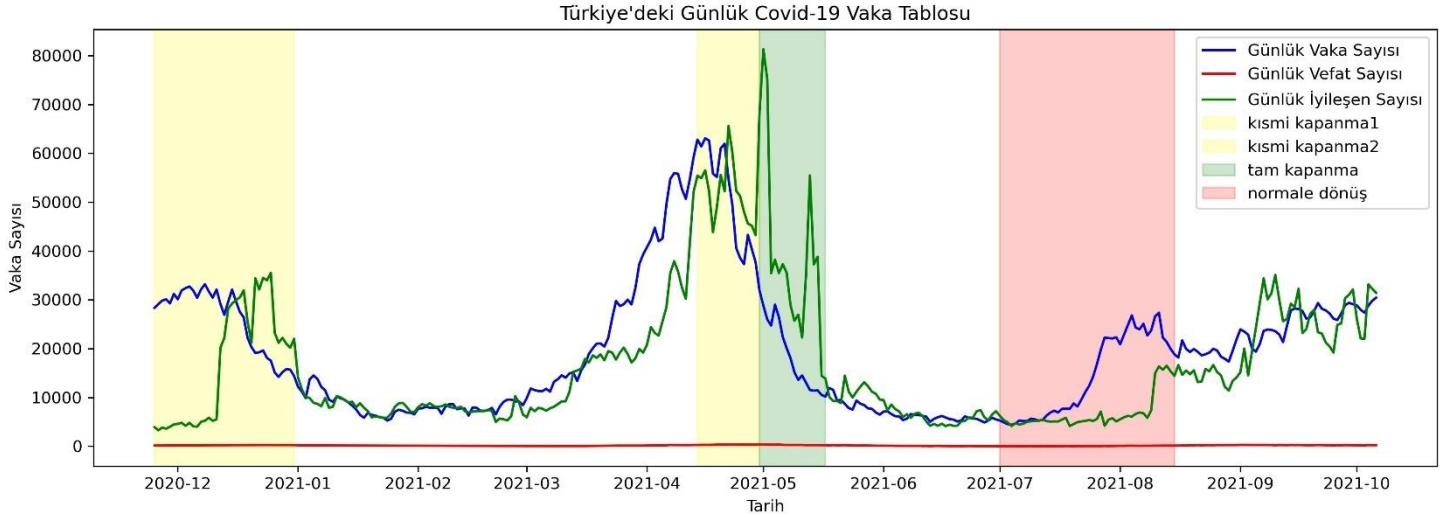
2. İlgili Çalışmalar

Zaman serisi tahmini önceden elde edilen bilgilerden yola çıkarak gelecek hakkında incelenen olayın zaman içindeki seyrinin tahmin edilmesine dayanmaktadır. Literatürde zaman

serisi modellerinden ARIMA modeli ve derin öğrenme modellerinden uzun kısa süreli bellek hafızasını kullanarak zaman serisi tahmini gerçekleştiren LSTM ağları birçok alanda tahminleme yapmak üzere kullanılmıştır. COVID-19 vakalarının zaman serileri analizine yönelik olarak gerçekleştirilmiş birçok bilimsel çalışma bulunmakla birlikte, bu bölümün geri kalanında ilgili alandaki temel çalışmalara değinilmektedir.

Baramn [5], ARIMA modelini ve birkaç LSTM mimarisini kullanarak 4 ülkedeki (ABD, İtalya, İspanya ve Almanya) COVID-19 vaka sayılarının zaman serisi analizini ve tahminini sunmuştur. Anne [6], katlanarak büyüyen COVID-19 salgınının zaman serisinin aktarımının kısa vadeli tahminini yapmak için ARIMA modelini kullanmıştır ve bir zaman aralığındaki toplam vaka sayısı üzerinde bunu gerçekleştirmiştir. Ding vd. [7] İtalya'da 24 Şubat-30 Mart 2020 arasındaki salgın verilerini istatistiksel olarak analiz etmiştir ve ARIMA modeline dayalı basit bir zaman serisi analiz modelini önermişlerdir. Shai vd. [8] salgının yayılmasını tahmin etmek için COVID-19'dan etkilenen ilk beş ülkenin ARIMA zaman serisi modeline dayalı zaman serisi analizini gerçekleştirmişlerdir. Maleki vd. [9] önerdikleri zaman serisi modellerini kullanarak COVID-19'un dünyadaki yayılma ve ölüm oranını tahmin etmişlerdir. Papastefanopoulos vd. [10], 4 Mayıs 2020 itibarıyla en fazla teyit edilmiş vaka sayısına sahip on farklı ülkede koronavirüs salgını tespiti için çeşitli zaman serisi modelleme yaklaşımlarının doğruluğunu araştırmışlardır. Chimmula vd. [11] Kanada ve dünyadaki COVID-19 salgınının eğilimlerini ve olası durma süresini tahmin etmek için temel özellikleri değerlendirmişlerdir ve gelecekteki COVID-19 vaka tahmini için derin öğrenme yaklaşımı olan LSTM ağlarını sunmuşlardır. Maleki vd. [12] doğrulanmış ve iyileşen COVID-19 vakalarını tahmin etmek için iki parça ölçekli normal dağılıma sahip zaman serisi modellemesine dayalı otoregresif modelleri önermişlerdir.

Türkiye'deki COVID-19 vakalarının zaman serisi modelleri kullanılarak tahmin edilmesine yönelik literatürde birçok çalışma gerçekleştirilmiştir. Akay ve Akay [13] 11 Mart 2020'den 24 Ağustos 2020'ye kadar olan toplam vaka sayısını tahmin etmek için ARIMA modelini kullanmışlardır. Tahmin gerçekleştirmek için 25 Ağustos 2020'den 7 Eylül 2020'ye kadar ki 14 günlük vakaların tahmini ve analizi yapılmıştır. Ankaralı [14] zaman serisi modellerini kullanarak Türkiye'de COVID-19 salgın sürecinde ihtiyaç duyulacak yoğun bakım yatak ve solunum cihazı sayılarının direkt tahminini gerçekleştirmiştir. Koçak [15] Türkiye'deki COVID-19 vaka sayılarının tahmini için farklı zaman serisi modellerini kullanarak analiz sonuçlarını karşılaştırmıştır. Ergül vd. [16] büyüme eğrilerini ve zaman serisi analizlerini kullanarak Türkiye'deki vaka sayılarının tahminini gerçekleştirmişlerdir. Bayar ve Varışlı [17] en küçük kareler yöntemi ve zaman serisi yöntemlerini kullanarak Türkiye'deki COVID-19 salgının tüketici kredileri hacmi üzerindeki etkisinin tahminini gerçekleştirmişlerdir. Demir [18] bir sekteli zaman serisi analizini kullanarak Türkiye'deki salgının seyrini ve bu salgının ekonomiye olan etkilerini tahmin etmiştir. Karasu ve Doğu [19] Johns Hopkins Üniversitesi resmi web sitesinde gerçek zamanlı derlenen verileri kullanarak Türkiye'deki COVID-19 salgının gerçek zamanlı seyrine olanak sağlayan bir web arayüz tasarlamışlardır. Er vd. [20] Türkiye'nin de içinde olduğu birçok ülkedeki COVID-19 vakalarının tahminini gerçekleştirmek için frekans güncellemesine dayalı ARIMA modelini kullanmışlardır. Ergül vd. [21] dünyada ve Türkiye'de nisan ayı itibarıyla COVID-19 salgın verilerinin istatistiksel değerlendirilmesi için ARIMA zaman serisi modelini kullanmışlardır.



Şekil 1. Türkiye'deki Günlük COVID-19 Grafiği

3. Materyal ve Yöntemler

3.1. Kullanılan Veri Seti

Çalışma kapsamında kullanılan veri seti Türkiye'deki günlük COVID-19 bulaşısının zaman serisi modelleri ile tahmin etmek için sağlık bakanlığının halka açık şekilde duyurduğu <https://covid19.saglik.gov.tr/TR-66935/genel-koronavirus-tablosu.html> internet sayfasından elde edilmiştir ve Şekil 1'de gösterilmiştir. 25 Kasım 2020 – 6 Ekim 2021 tarihleri arasındaki Türkiye'de meydana gelen günlük vaka sayısı, günlük vefat sayısı ve günlük iyileşen sayısı Şekil 2'de gösterildiği gibi veri seti olarak toplanmıştır. Başlangıç tarihi olarak 25 Kasım 2020 alınmasının nedeni, sağlık bakanlığının günlük vaka sayılarını bu tarihten itibaren açıklamaya başlamasıdır.

Tarih	BugunkuVakaSayisi	BugunkuVefatSayisi	Bugunkuyilesen Sayisi
0 2020-11-25	28351	168	3911
1 2020-11-26	29132	174	3291
2 2020-11-27	29845	177	3845
3 2020-11-28	30103	182	3611
4 2020-11-29	29281	185	4015
...
311 2021-10-02	27973	203	22110
312 2021-10-03	27351	194	22004
313 2021-10-04	28810	248	33152
314 2021-10-05	29802	228	32269
315 2021-10-06	30438	236	31413

Şekil 2. Veri Seti

3.2. Zaman Serisi Tahmin Modelleri

3.2.1. ARIMA Zaman Serisi Modeli

Otoregresif hareketli ortalama modeli ARIMA (p, d, q) bir zaman serisi analizi ve tahmin yöntemidir. Zaman serisinin durağan olmadığı durumlarda kullanılır. ARIMA modelinin üç bileşeni vardır:

- **Otoregresyon (AR):** Otoregresif bir model, kendi önceki (veya gecikmeli) değerlerine dayalı olarak değişen bir değişken arasındaki regresyon ilişkisini türetir.
- **Entegrasyon (I):** Entegrasyon, ham gözlemler arasındaki farkları modellemeyi ifade eder, böylece zaman serileri

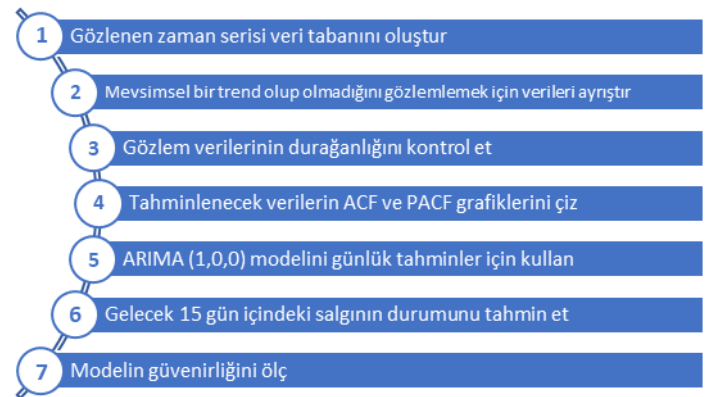
durağan kabul edilebilir. Başka bir deyişle, ham değerler, ham değerler ile önceki değerler arasındaki farklarla değiştirilir.

- **Hareketli Ortalama (MA):** Modeldeki hareketli ortalama bileşeni, bir gözlem ile önceki (veya gecikmeli) gözlemlere uygulanan hareketli ortalama modelinden kalan bir hata arasındaki ilişkiyi içerir. ARIMA modeli 3 parametreye sahiptir.
1. p , otoregresif (AR) modelin derecesini temsil eder.
 2. d , serinin durağanlaştırılabilmesi için kaçınıcı dereceden farkının alındığını temsil eder.
 3. q , hareketli ortalama penceresinin boyutunu gösterir.

ARIMA modelinin (1,0,0) modeli bu çalışmada kullanılmıştır ve tahmin edilecek gün sayısı $t = 15$ gün olarak belirlenmiştir. Birinci mertebeden otoregresif model olan ARIMA (1,0,0) modeli, seri durağan ise ve otokorelasyonlu ise, kendi önceki değerinin bir katı artı bir sabit olarak tahmin edilebilir. ARIMA modelinde gelecekteki doğrulanmış COVID-19 vakalarını tahmin etme modeli Denklem 1'de gösterilmiştir.

$$\text{ARIMA}(p, d, q): X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} + Z_t \text{ ve } Z_t = X_t - X_{t-1} \quad (1)$$

Burada X_t , t . günde tahmini doğrulanmış COVID-19 vakası sayısıdır, $\alpha_1, \alpha_2, \beta_1$ ve β_2 parametrelerdir, Z_t ise t . gün için kalan terimdir. Çalışmada kullanılan ARIMA zaman serisi modelinin genel çerçevesi Şekil 3'te gösterilmiştir.



Şekil 3. ARIMA Modeli Genel Çerçevesi

3.2.2. LSTM Sinir Ağı

LSTM sinir ağı, bir önceki adımın çıktısını mevcut adımın girdisi olarak kullanan ve kısa bir süre belleğinde tutan RNN ağları ile çok benzer olan ancak RNN ağlarının karşılaştığı sorunları çözmek için geliştirilmiş yapay sinir ağıdır. Böylelikle, doğal dil işleme ve zaman serileri gibi problemlerin çözülmesinde kullanılabilirler. RNN ağları bilgiyi uzun süreli tutmakta başarısız olduğu için önemli bazı bilgiler ağına başında kaybolabilir. Aynı zamanda RNN ağları, geri yayılım süresince gradyan yok olması gibi problemlere sahiptir. Genel işleyişi Şekil 4'te gösterilen LSTM ağı bahsedilen bu kısıtlamalara karşı geliştirilmiş bir modeldir. LSTM hücreleri kısa süreli bellek sorununu çözmek için bilgi akışını düzenleyen ve kontrol eden kapılar oluştururlar. Bir hücre girdi kapısı, hücre durumu, unutmaya kapısı ve çıktı kapısı içerir ve bu kapılar sigmoid veya tanh aktivasyon fonksiyonu kullanır [22]. Bu kapılar ve görevleri aşağıda verilmiştir:

1. *Girdi kapısı*: Hücre durumunu güncellemek için bir önceki gizli durumu ve şu anki girdiyi alarak sigmoid fonksiyonundan geçirir.
2. *Hücre durumu*: Diğer kapıların yardımıyla bilgiyi tutmayı veya unutmayı sağlar.
3. *Çıktı kapısı*: Hücrenin çıktısını oluşturur.
4. *Unutmaya kapısı*: Bilginin unutulacak ve tutulacak parçalarına karar verir.

Girdi kapısı, unutmaya kapısı, çıktı kapısı hücre durumu ve gizli durum değerleri sırasıyla i_t , f_t , o_t , c_t ve h_t temsil edilmektedir. x_t , t zamanındaki girdi vektörü, σ sigmoid aktivasyon fonksiyonu ve W , b parametre matrisi ve vektörü olmak üzere LSTM hücresi Denklem 2-6'da gösterildiği gibi hesaplanmaktadır [23,24].

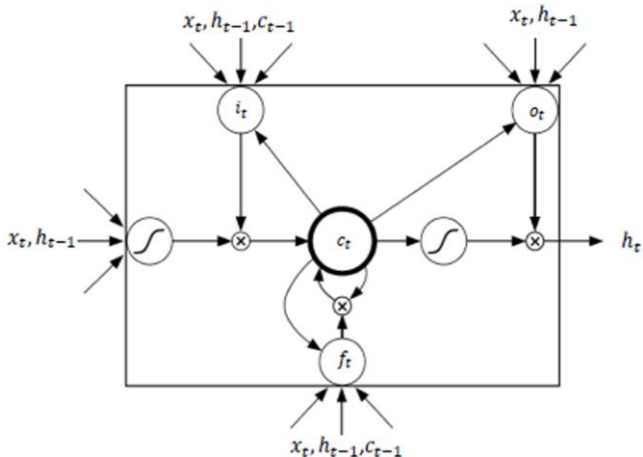
$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (4)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$



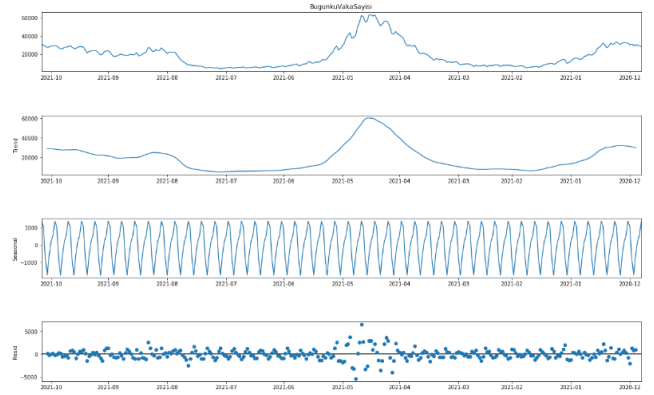
Şekil 4. LSTM Ağı Modeli [23]

4. Deneysel Sonuçlar

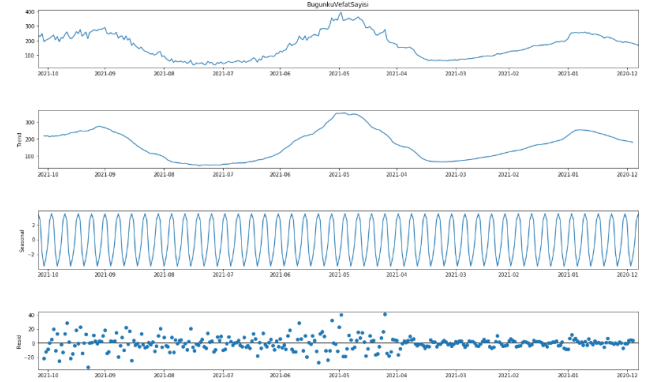
Deneysel çalışmaları gerçekleştirmek için Python programlama dilinin 3.6.15 versiyonu kullanılmıştır. Kullanılan makinenin teknik özellikleri ise Intel(R) Core (TM) 64-bit, i7-9750H CPU, 2.60 GHz, 32 GB RAM, Nvidia GeForce RTX 2060.

4.1. ARIMA Zaman Serisi Modeli Sonuçları

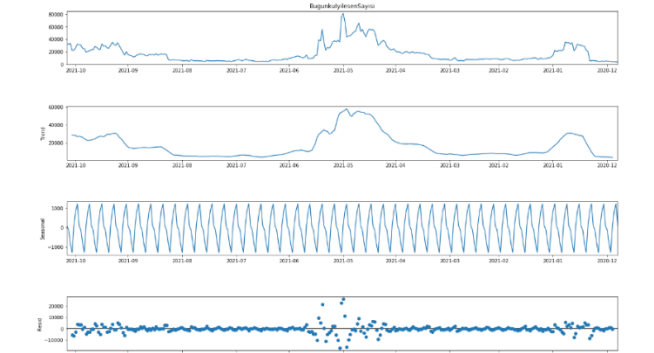
Çalışmada zaman serisi veri tabanı oluşturulduktan sonra mevsimsel bir trend olup olmadığını gözlemlemek için verilerin ayrıştırma grafikleri çizilmiştir. Günlük vaka sayısının, günlük vefat sayısının ve günlük iyileşen sayısının trend eğilimleri sırasıyla Şekil 5-6-7'de gösterildiği gibidir. 2021 yılı Nisan-Mayıs aylarında hastalığın ciddi bir peak (tepe) yaptığı görülmektedir. Mevsimsel olarak bu aylarda yükseliş beklense de hastalığın verilerini toplamının üzerinden birkaç sene geçmediğinden bu mevsimsel trendi bu tarih aralıklarına dayandırmak şimdilik modelin doğruluğu için doğru değildir.



Şekil 5. Günlük Vaka Sayısının Mevsimsel Trend Grafiği



Şekil 6. Günlük Vefat Sayısının Mevsimsel Trend Grafiği

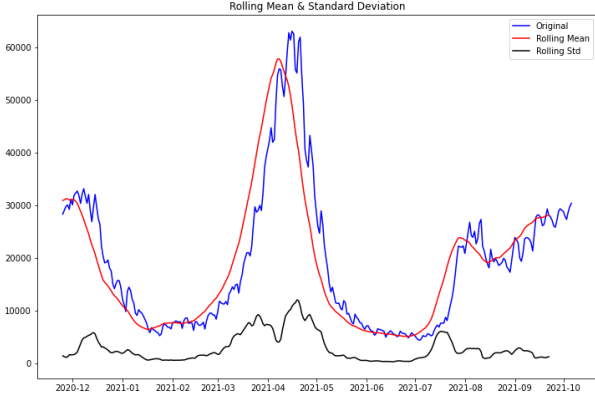


Şekil 7. Günlük İyileşen Sayısının Mevsimsel Trend Grafiği

ARIMA modeli daha sonra normallik ve durağanlıktaki varyans için test edilmiştir. Tahmin yapmadan önce gözlem verilerinin durağanlığı kontrol edilmiştir. Ayrıca günlük tahmini yapılan her bir verinin durağanlığını sağlamak için yapılan istatistiksel bir test olan Dickey-Fuller testleri yapılmıştır. Şekil 8 günlük vaka sayısının durağanlığını göstermektedir. Yapılan analiz sonucu günlük vaka sayılarının durağanlık gösterdiği gözlemlenmiştir.

```
Results of Dickey-Fuller Test:
p-value = 0.0043. The series is likely stationary.
Test Statistic      -3.684941
p-value            0.004330
#Lags Used         14.000000
Number of Observations Used  301.000000
Critical Value (1%)  -3.452263
Critical Value (5%)  -2.871191
Critical Value (10%) -2.571912
dtype: float64
```

<Figure size 864x288 with 0 Axes>



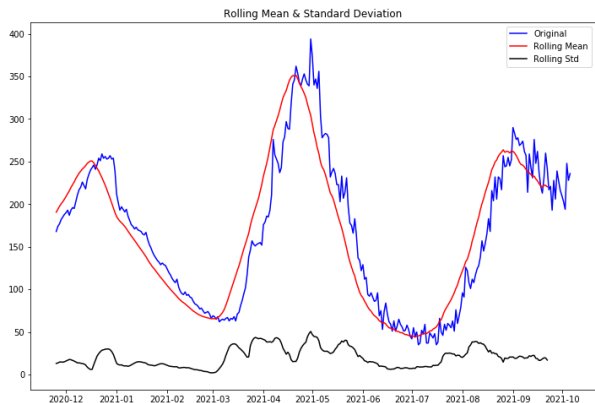
Şekil 8. Günlük Vaka Sayısının Durağanlık ve DF Testi

Şekil 9 günlük vefat sayısının durağanlığını göstermektedir. Yapılan analiz sonucu günlük vefat sayılarının durağanlık gösterdiği gözlemlenmiştir. Şekil 10 günlük iyileşen sayısının durağanlığını göstermektedir. Yapılan analiz sonucu günlük iyileşen sayılarının durağanlık göstermediği gözlemlenmiştir.

ACF grafiği, bir zaman serisi ile gecikmeli değerler arasındaki korelasyon katsayılarının bir çubuk grafiğidir. Basitçe ifade edecek olursak: ACF, belirli bir zaman serisinin şimdiki değerinin geçmiş değerleriyle nasıl ilişkili olduğunu açıklar. ACF grafiğinde, x eksenini korelasyon katsayısını, y eksenini ise gecikme sayısını ifade eder. PACF, seriler ve kendi gecikmeleri arasındaki kısmi korelasyonu açıklayan kısmi otokorelasyon fonksiyonudur.

```
Results of Dickey-Fuller Test:
p-value = 0.0025. The series is likely stationary.
Test Statistic      -3.838941
p-value            0.002533
#Lags Used         14.000000
Number of Observations Used  301.000000
Critical Value (1%)  -3.452263
Critical Value (5%)  -2.871191
Critical Value (10%) -2.571912
dtype: float64
```

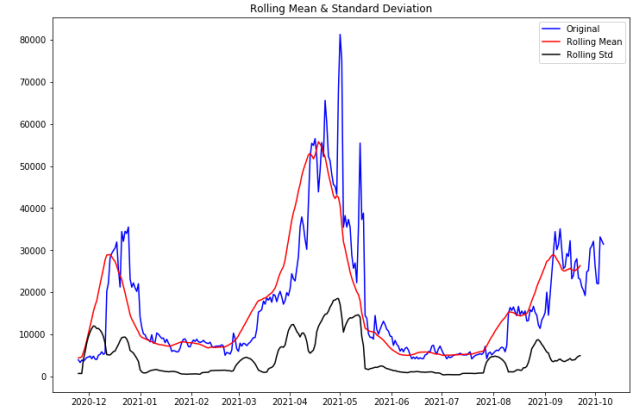
<Figure size 864x288 with 0 Axes>



Şekil 9. Günlük Vefat Sayısının Durağanlık ve DF Testi

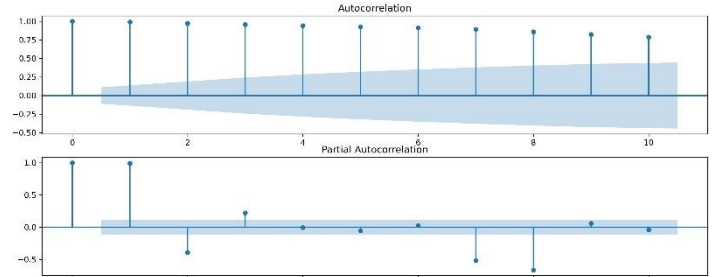
```
Results of Dickey-Fuller Test:
p-value = 0.0567. The series is likely non-stationary.
Test Statistic      -2.811028
p-value            0.056716
#Lags Used         15.000000
Number of Observations Used  300.000000
Critical Value (1%)  -3.452337
Critical Value (5%)  -2.871223
Critical Value (10%) -2.571929
dtype: float64
```

<Figure size 864x288 with 0 Axes>

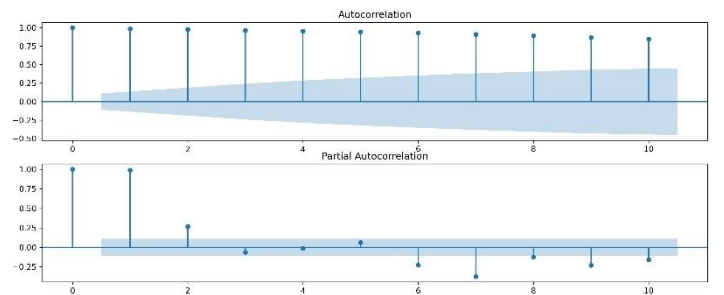


Şekil 10. Günlük İyileşen Sayısının Durağanlık ve DF Testi

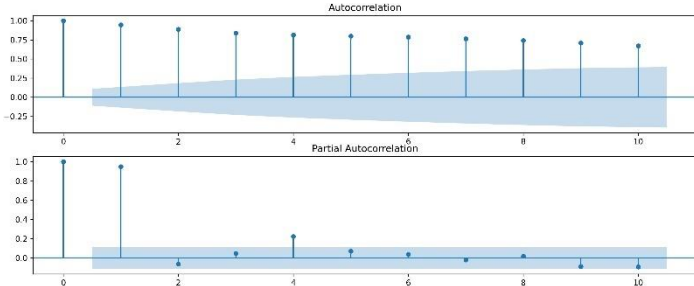
ACF ve PACF grafikleri, süreci tanımlamak için birlikte düşünülmelidir. AR süreci için, ACF grafiğinin kademeli olarak azalmasını ve aynı anda PACF'nin p önemli gecikmelerden sonra keskin bir düşüşe sahip olmasını bekliyoruz. Bir MA sürecini tanımlamak için, ACF ve PACF grafiklerinden tam tersini bekleriz, bu şu anlama gelir: ACF belirli bir q gecikme sayısından sonra keskin bir düşüş göstermeli, PACF ise geometrik veya kademeli bir düşüş eğilimi göstermelidir. Öte yandan hem ACF hem de PACF grafikleri kademeli bir azalan model gösteriyorsa, modelleme için ARIMA süreci dikkate alınmalıdır [25]. ARIMA modellerinin başlangıç sayısını bulmak için otokorelasyon fonksiyonu (ACF) grafiği ve kısmi otokorelasyon (PACF) grafiği kullanılmıştır. Günlük vaka sayısının, günlük vefat sayısının ve günlük iyileşen sayısının trend eğilimleri sırasıyla Şekil 11-12-13'te gösterildiği gibidir. Otokorelasyon fonksiyonu, olası tüm gecikmelerde otokorelasyonu verir. Hem ACF hem de PACF grafikleri kademeli bir azalan model gösterdiği için modellemede ARIMA süreci ele alınmıştır.



Şekil 11. Günlük Vaka Sayısının Oto&Kısmi Korelasyon Grafiği



Şekil 12. Günlük Vefat Sayısının Oto&Kısmi Korelasyon Grafiği



Şekil 13. Günlük İyileşen Sayısının Oto&Kısmi Korelasyon Grafiği

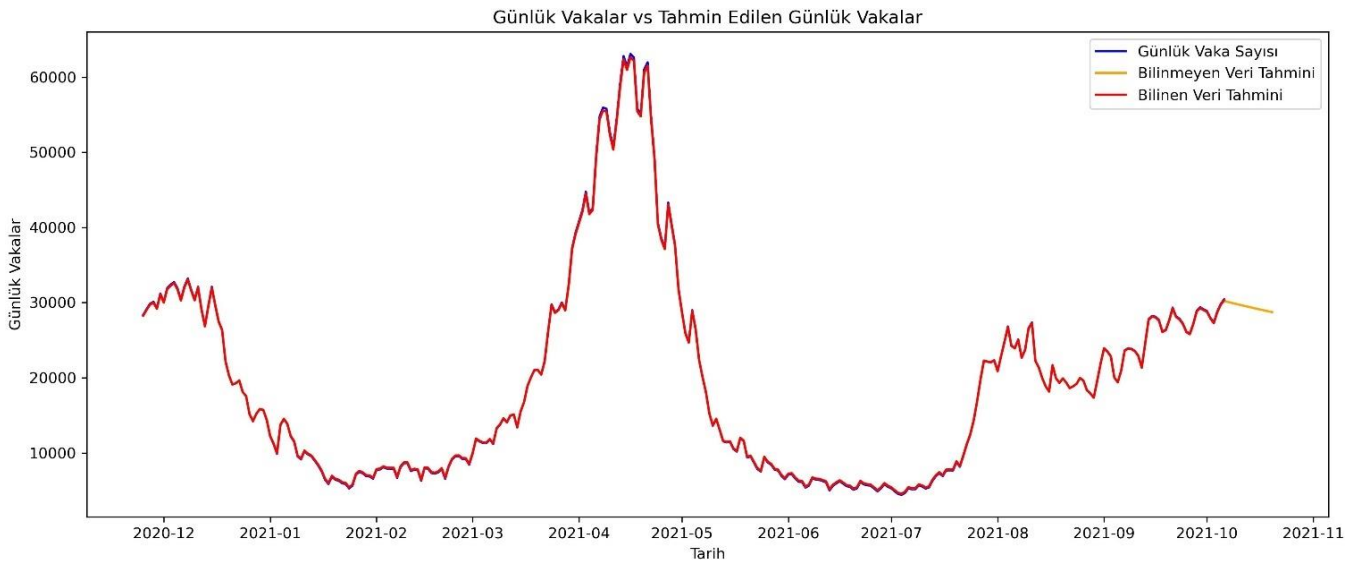
ARIMA zaman serisi modelini kullanarak Türkiye'deki COVID-19 bulaşısının tahmin edilen 15 günlük vaka sayısında, vefat sayısında ve iyileşen sayısında azalmaya yönelik bir regresyon eğrisi Şekil 14-15-16'da gösterildiği gibi elde edilmiştir. Grafikler üzerinde var olan veri, bilinen verinin tahmini ve bilinmeyen 15 günlük tahmin gösterilmiştir. Bilinen

verinin tahminine bakıldığında yüksek doğruluk yüzdesiyle grafiğin çizdirildiği gözlemlenmektedir.

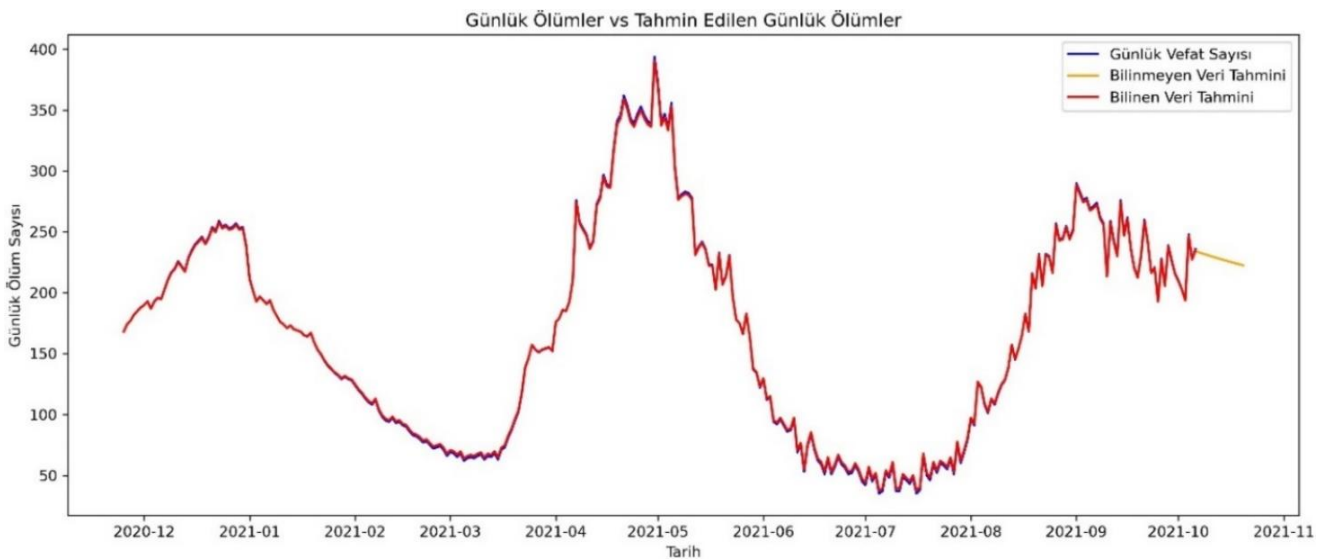
Tahmin modelinin doğruluğunu göstermek için performans değerlendirme metriklerinden Denklem 7'de formülü verilmiş olan ortalama mutlak yüzde hatası (mean absolute percentage error-MAPE) kullanılmıştır. MAPE, gerçek ve tahmin edilen değerler arasındaki farkı hesaplar. MAPE yüzdesi ne kadar düşükse modelin doğruluğu o kadar yüksektir. Burada G gerçek değeri, T tahmin edilen değeri ve n ise tahmin edilen toplam veri sayısını temsil etmektedir.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{G_t - T_t}{G_t} \right| \quad (7)$$

Tablo 1'de gösterildiği gibi, tahmini yapılan gelecek 15 günlük sağlık bakanlığının ilan ettiği verilerle karşılaştırıldığında, ARIMA modelinde Türkiye'deki COVID-19 bulaşısının gelecek 15 günlük vaka sayısı %5,71, vefat sayısı %11,87 ve iyileşen sayısı %9,87 hata oranları ile elde edilmiştir.

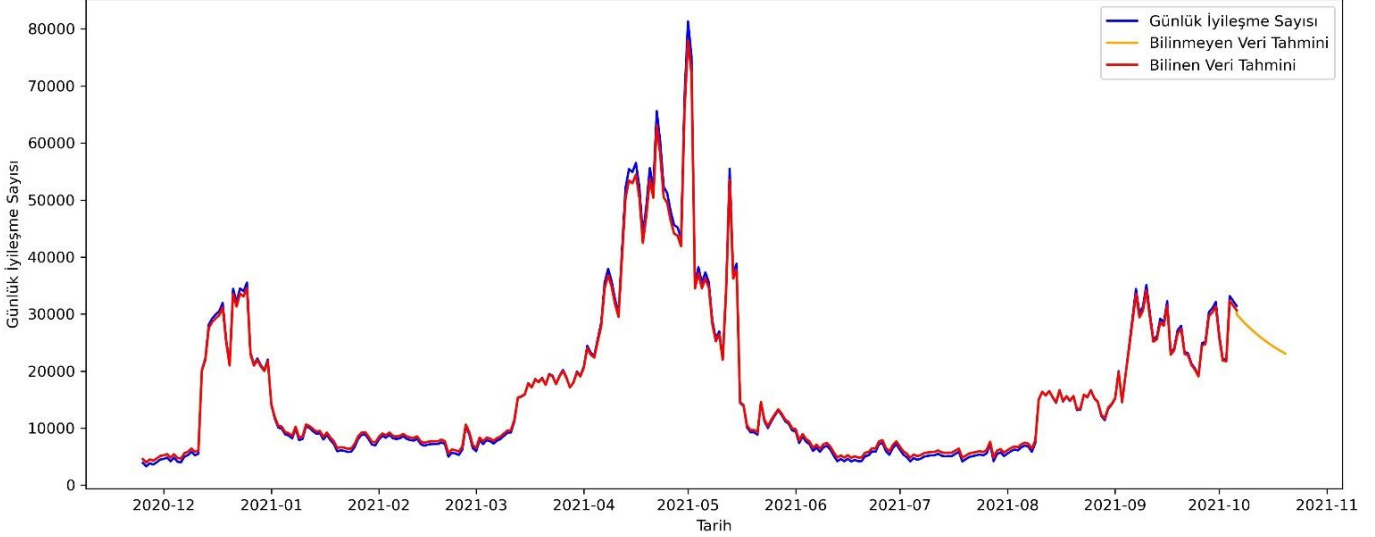


Şekil 14. ARIMA Modelinde Günlük Vaka Sayıları vs Tahmin Edilen Günlük Vaka Sayıları



Şekil 15. ARIMA Modelinde Günlük Vefat Sayıları vs Tahmin Edilen Günlük Vefat Sayıları

Günlük İyileşmeler vs Tahmin Edilen Günlük İyileşmeler



Şekil 16. ARIMA Modelinde Günlük İyileşen Sayıları vs Tahmin Edilen Günlük İyileşen Sayıları

Tablo 1. ARIMA Modelinde Tahmin Edilen COVID-19 Bulaşının MAPE Sonuçları

Tarih	Günlük Vaka			Günlük Vefat			Günlük İyileşen		
	Gerçek Vaka Sayısı	Tahmin Edilen Vaka Sayısı	Hata Yüzdesi	Gerçek Vefat Sayısı	Tahmin Edilen Vefat Sayısı	Hata Yüzdesi	Gerçek İyileşen Sayısı	Tahmin Edilen İyileşen Sayısı	Hata Yüzdesi
7 Ekim 2021	30019	30208	%0,63	217	234	%7,89	32240	29946	%7,11
8 Ekim 2021	30201	30096	%0,34	188	233	%24,04	28167	29269	%3,91
9 Ekim 2021	28645	29984	%4,67	206	232	%12,76	24217	28626	%18,21
10 Ekim 2021	28370	29873	%5,30	196	231	%18,06	25772	28017	%8,71
11 Ekim 2021	30563	29764	%2,61	188	230	%22,62	33861	27439	%18,96
12 Ekim 2021	33860	29656	%12,41	237	229	%3,08	30110	26890	%10,69
13 Ekim 2021	31248	29549	%5,43	236	228	%3,03	30331	26370	%13,05
14 Ekim 2021	30709	29443	%4,12	203	228	%12,31	26461	25877	%2,20
15 Ekim 2021	30694	29338	%4,41	181	227	%25,51	25513	25408	%0,41
16 Ekim 2021	28537	29234	%2,44	212	226	%6,78	25611	24964	%2,52
17 Ekim 2021	24114	29131	%20,81	186	225	%21,28	24403	24543	%0,57
18 Ekim 2021	29240	29029	%0,71	214	224	%5,04	26539	24143	%9,02
19 Ekim 2021	30862	28928	%6,26	223	224	%0,46	27352	23764	%13,11
20 Ekim 2021	29760	28829	%3,12	214	223	%4,33	28803	23405	%18,74
21 Ekim 2021	28465	28730	%0,93	198	222	%12,39	29160	23064	%20,90
			MAPE			MAPE			MAPE
			%4,95			%11,87			%9,87

4.2. LSTM Sinir Ağı Sonuçları

Tahminleme için tek değişkenli zaman serisi modeli geliştirilmiştir. Tahmin modeli 3 LSTM katmanı içermektedir. Aktivasyon fonksiyonu olarak tanh seçilmiştir. Her bir katmandaki nöron sayısı hiperparametre optimizasyonu kullanılarak elde edilmiştir ve optimizasyon olarak adam optimizasyonu kullanılmıştır ve öğrenme katsayısı $\beta = 0.01$ seçilmiştir. Geliştirilen model tensorflow keras ile gerçekleştirilmiştir. Veri seti %80 eğitim, %20 test olarak ayrılmıştır. Eğitim verisi 14 günlük zaman dilimi olarak biçimlendirilmiştir. Model 100 epoch ve 4 batch boyutu kullanılarak eğitilmiştir. Model eğitim verisi ile doğrulanmış, test verisi ile test edilmiştir. Ortalama mutlak hata yüzdesi (MAPE) ölçüğü ile model başarısı ölçülmüştür. Model başarısı değerleri Tablo 2’de ARIMA ve LSTM modellerindeki

doğrulama ve test MAPE değerleri gösterilmiştir. Doğrulama MAPE değerleri, bilinen verinin tahmin edilmesi olup yüzde kaç hata ile doğru bilindiğini göstermektedir. ARIMA modelinde test MAPE değerleri, bilinen verinin son %20’lik kısmının test verisi kabul edilip tahmin edilmesidir. Burada yüzde kaç hata ile doğru bilindiğini göstermektedir. LSTM modelinde test MAPE değerleri, %20’lik test verisinin yüzde kaç hata ile doğru tahminlendiğini göstermektedir. Tablo 2’de gösterildiği gibi, günlük vaka, vefat ve iyileşen sayısında doğrulama ve test MAPE değerlerinde ARIMA modelinde daha az hata ile modelin tahminlendiği gözlemlenmiştir. LSTM ağı modelinde bir sonraki gün tahmini gerçekleştirildiğinden hata yüzdesinde daha yüksek oranlar elde edildiği anlaşılmaktadır.

LSTM ağı modeli kullanılarak vaka, vefat ve iyileşen sayısının gelecek 15 günlük tahmini Şekil 17-18-19'da gösterildiği gibidir. LSTM ağı kullanılarak Türkiye'deki COVID-19 bulaşı seyri deneysel sonuçları ve gelecek 15 günlük tahmin grafiklerine bakıldığında günlük vaka ve günlük vefat sayısında azalma, günlük iyileşen sayısında ise artma seyri gösterdiği gözlemlenmiştir.

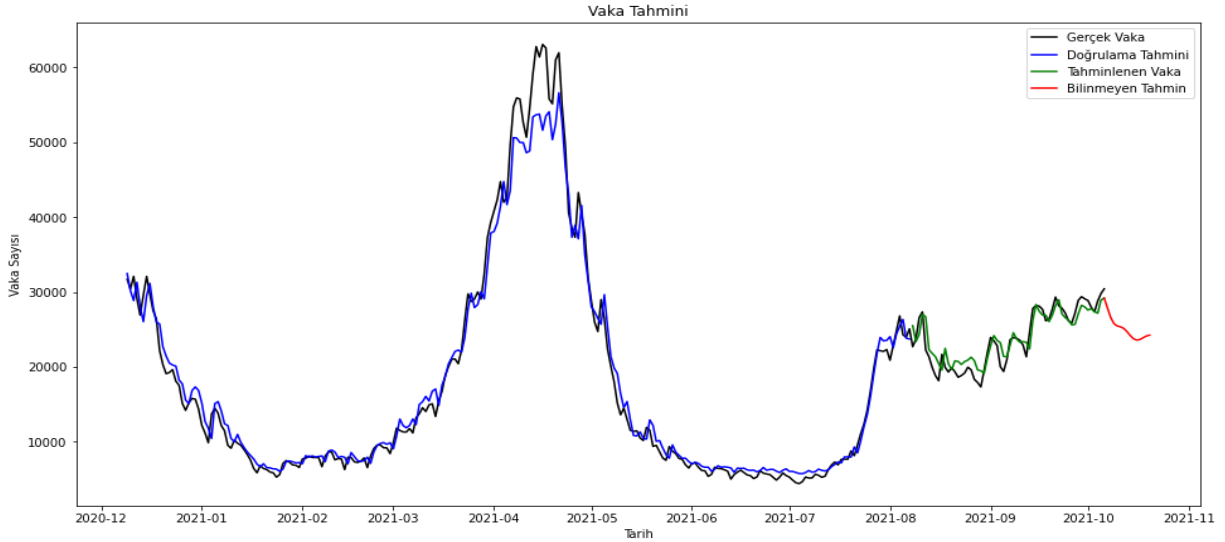
Tablo 3'te gösterildiği gibi, tahmini yapılan gelecek 15 günlük sağlık bakanlığının ilan ettiği verilerle karşılaştırıldığında, LSTM ağı modelinde Türkiye'deki COVID-19 bulaşısının gelecek 15 günlük vaka sayısı %14,85, vefat sayısı %8,60 ve iyileşen sayısı %20,98 hata oranları ile elde edilmiştir.

Tablo 1 ve Tablo 3'te gösterildiği gibi, Türkiye'de bulaşı seyrinde gelecek tahmini ARIMA ve LSTM modelleriyle gerçekleştirildiğinde, bilinmeyen verilerin tahmininde ARIMA modelinin daha yüksek doğrulukla tahmin ettiği gözlemlenmektedir. Günlük vaka sayısında LSTM ağına elde edilen hata yüzdesi %61,54 oranında, günlük iyileşen sayısında %52,95 oranında azalma ARIMA modelinde elde edilmiştir.

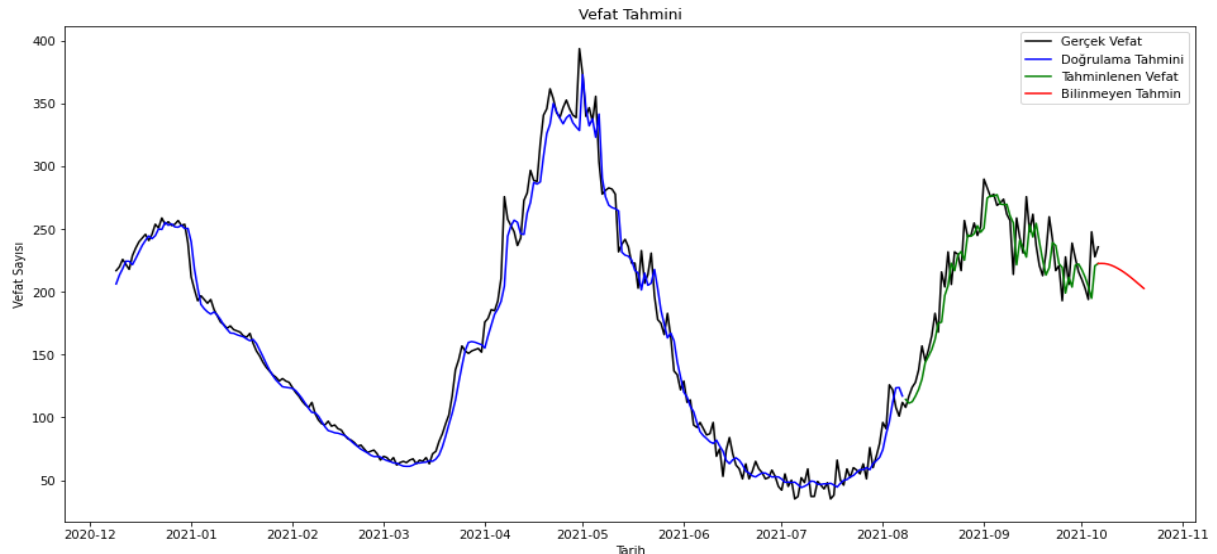
Günlük vefat sayısında ise ARIMA modelinde elde edilen hata yüzdesi %27,54 oranında azalma LSTM ağı modelinde elde edilmiştir.

Tablo 2. Doğrulama ve Test MAPE Yüzdeleri

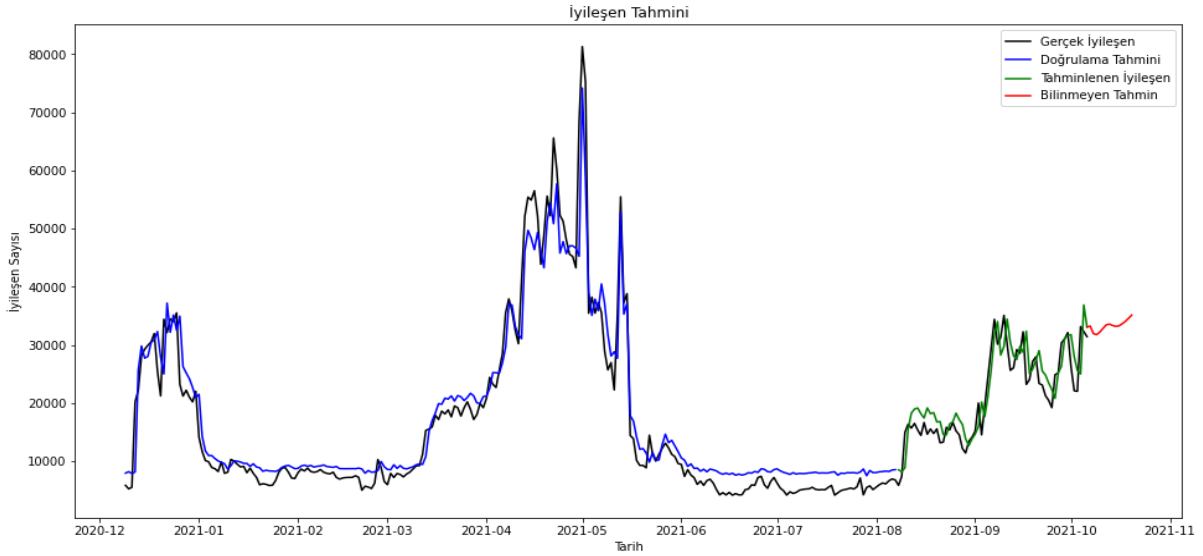
	ARIMA		LSTM	
	Doğrulama MAPE	Test MAPE	Doğrulama MAPE	Test MAPE
Günlük Vaka	%0,91	%0,19	%9,10	%5,61
Günlük Vefat	%1,21	%0,43	%6,76	%6,64
Günlük İyileşen	%5,41	%2,01	%25,21	%15,02



Şekil 17. LSTM Ağına Günlük Vaka Sayıları vs Tahmin Edilen Günlük Vaka Sayıları



Şekil 18. LSTM Ağına Günlük Vefat Sayıları vs Tahmin Edilen Günlük Vefat Sayıları



Şekil 19. LSTM Ağında Günlük İyileşen Sayıları vs Tahmin Edilen Günlük İyileşen Sayıları

Tablo 3. LSTM Ağında Tahmin Edilen COVID-19 Bulaşının MAPE Sonuçları

Tarih	Günlük Vaka			Günlük Vefat			Günlük İyileşen		
	Gerçek Vaka Sayısı	Tahmin Edilen Vaka Sayısı	Hata Yüzdesi	Gerçek Vefat Sayısı	Tahmin Edilen Vefat Sayısı	Hata Yüzdesi	Gerçek İyileşen Sayısı	Tahmin Edilen İyileşen Sayısı	Hata Yüzdesi
7 Ekim 2021	30019	29185	%2,77	217	222	%7,89	32240	33056	%7,11
8 Ekim 2021	30201	27880	%7,68	188	222	%24,04	28167	33239	%3,91
9 Ekim 2021	28645	26605	%7,12	206	222	%12,76	24217	31937	%18,21
10 Ekim 2021	28370	25792	%9,08	196	222	%18,06	25772	31750	%8,71
11 Ekim 2021	30563	25473	%16,65	188	221	%22,62	33861	32192	%18,96
12 Ekim 2021	33860	25361	%25,10	237	220	%3,08	30110	32818	%10,69
13 Ekim 2021	31248	25150	%19,51	236	218	%3,03	30331	33434	%13,05
14 Ekim 2021	30709	24725	%19,48	203	217	%12,31	26461	33579	%2,20
15 Ekim 2021	30694	24199	%21,16	181	215	%25,51	25513	33339	%0,41
16 Ekim 2021	28537	23772	%16,69	212	213	%6,78	25611	33182	%2,52
17 Ekim 2021	24114	23586	%2,18	186	211	%21,28	24403	33259	%0,57
18 Ekim 2021	29240	23663	%19,07	214	209	%5,04	26539	33620	%9,02
19 Ekim 2021	30862	23904	%22,54	223	207	%0,46	27352	34034	%13,11
20 Ekim 2021	29760	24134	%18,90	214	205	%4,33	28803	34568	%18,74
21 Ekim 2021	28465	24254	%14,79	198	202	%12,39	29160	35125	%20,91
			MAPE %14,85			MAPE %8,60			MAPE %20,98

5. Sonuçlar ve Gelecek Çalışmalar

COVID-19 küresel salgını tüm dünyada olduğu gibi ülkemizde de birçok alanda olumsuz etkilere neden olmuştur. Alınan tedbirlere rağmen vakaların hayatı olumsuz etkilediği kaçınılmaz bir gerçektir. COVID-19 bulaşısının seyrinin tahmin edilmesi alınacak tedbirler için hayati önem taşımaktadır. Bu çalışmada, Türkiye'deki COVID-19 bulaşısının günlük vaka sayısının, günlük vefat sayısının ve günlük iyileşen sayısının gelecek 15 günlük tahmini için ARIMA zaman serisi modeli ve LSTM sinir ağı modeli kullanılmıştır. Yapılan deneysel çalışmalar sonucu ARIMA ve LSTM modellerinde bulaşı seyrinin düşme eğiliminde olduğu gözlemlenmiştir.

ARIMA modelinde MAPE tahmin hatası vaka sayıları için %4,95, vefat sayıları için %11,87 ve iyileşen sayıları için %9,87 olarak elde edilirken, LSTM modelinde MAPE tahmin hatası vaka sayıları için %14,85, vefat sayıları için %8,60 ve iyileşen sayıları için %20,98 olarak elde edilmiştir. Modellerin doğruluğunu test etmek için bilinen ve bilinmeyen verilerin tahmini yapılarak tahmini yapılan verilerin hata yüzdeleri karşılaştırılmıştır. Türkiye'de COVID-19 bulaşı seyrinin gelecek 15 günlük tahmini için yapılan deneysel çalışmalar sonucu günlük vaka sayılarında ve günlük vefat sayılarında ARIMA modeliyle daha yüksek doğrulukta tahminleme yapıldığı, günlük iyileşen sayılarında ise LSTM modeliyle daha yüksek doğrulukta tahminleme yapıldığı gözlemlenmiştir. Yapılan tüm deneysel çalışmalar sonucu, ARIMA ve LSTM modellerinde günlük vaka ve vefat

sayılarında azalma seyri gözlemlenmiştir. Günlük iyileşme sayılarında ise ARIMA modelinde azalma gözlemlenirken LSTM ağ modelinde artış gözlemlenmiştir.

Gelecekte yapmayı planladığımız çalışmalar ise şu şekildedir;

- Vakaların, vefatların ve iyileşmelerin birbiriyle olan ilişkileri incelenebilir.
- Başka bir zaman serisi modeli olan SARIMAX modeliyle de karşılaştırma yapılabilir.
- Aşılama sürecinin salgının seyrine olan etkisi zaman serisi modelleriyle incelenebilir.

Kaynakça

1. Chawla, S., Mittal, M., Chawla, M., & Goyal, L. M. (2020). Corona virus-SARS-CoV-2: an insight to another way of natural disaster. *EAI Endorsed Transactions on Pervasive Health and Technology*, 6(22).
2. Wang, L. L., & Lo, K. (2021). Text mining approaches for dealing with the rapidly expanding literature on COVID19. *Briefings in Bioinformatics*, 22(2), 781-799.
3. Er, B., Emeç, M., & Özcanhan, M. H. (2020). Analysis Of Covid-19 Data Using Arima Time Series Model.
4. 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.
5. Barman, A. (2020). Time series analysis and forecasting of covid-19 cases using LSTM and ARIMA models. *arXiv preprint arXiv:2006.13852*.
6. Anne, R. (2020). ARIMA modelling of predicting COVID-19 infections. *medRxiv*.
7. Ding, G., Li, X., Shen, Y., & Fan, J. (2020). Brief Analysis of the ARIMA model on the COVID-19 in Italy. *medRxiv*.
8. Sahai, A. K., Rath, N., Sood, V., & Singh, M. P. (2020). ARIMA modelling & forecasting of COVID-19 in top five affected countries. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5), 1419-1427.
9. Maleki, M., Mahmoudi, M. R., Heydari, M. H., & Pho, K. H. (2020). Modeling and forecasting the spread and death rate of coronavirus (COVID-19) in the world using time series models. *Chaos, Solitons & Fractals*, 140, 110151.
10. Papastefanopoulos, V., Linardatos, P., & Kotsiantis, S. (2020). COVID-19: a comparison of time series methods to forecast percentage of active cases per population. *Applied sciences*, 10(11), 3880.
11. Chimmula, V. K. R., & Zhang, L. (2020). Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 135, 109864.
12. Maleki, M., Mahmoudi, M. R., Wraith, D., & Pho, K. H. (2020). Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel medicine and infectious disease*, 37, 101742.
13. Akay, S., & Akay, H. (2021). Time series model for forecasting the number of COVID-19 cases in Turkey. *Turkish Journal of Public Health*, 140-145.
14. ANKARALI, H. (2020). Türkiye’de COVID-19 Salgın Sürecinde İhtiyaç Duyulacak Yoğun Bakım Yatak ve Solunum Cihazı Sayılarının Direkt Tahmini. *Anatolian Clinic the Journal of Medical Sciences*, 25(Special Issue on COVID 19), 59-62.
15. Koçak, M. (2020). A comparison of time-series models in predicting COVID-19 cases. *Türkiye Klinikleri Biyoistatistik*, 12(1), 89-96.
16. Ergül, B., Altın Yavuz, A., Gündoğan Aşık, E., & Kalay, B. (2020). Dünya’da ve Türkiye’de nisan ayı itibariyle Covid-19 salgın verilerinin istatistiksel değerlendirilmesi. *Anadolu Kliniği Tıp Bilimleri Dergisi*, 25(1), 130-141.
17. Bayar, M., & VARIŞLI, N. (2020). Covid-19 Pandemisinin Türkiye’de Tüketici Kredileri Hacmi Üzerindeki Etkisi: 2013-2020 Dönemi Zaman Serisi Analizi. *Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 21(3), 85-99.
18. Demir, İ. (2020). COVID-19 Salgınının Seyri ve Türkiye Ekonomisi: Bir Sektörel Zaman Serisi Analizi (Szsa) Denemesi. *Disiplinlerarası Politika Vizyonu ve Stratejiler Dergisi*.
19. Karasoy, O., & EREN DOĞU, Z. F. (2020). COVID19Takip: Türkiye’de COVID-19 Salgınının Gerçek Zamanlı İzlenmesi için Web Arayüzü. *Türkiye Klinikleri Journal of Biostatistics*, 12(1).
20. Er, B., Emeç, M., & Özcanhan, M. H. ANALYSIS OF COVID-19 DATA USING ARIMA TIME SERIES MODEL.
21. Ergül, B., Altın Yavuz, A., Gündoğan Aşık, E., & Kalay, B. (2020). Dünya’da ve Türkiye’de nisan ayı itibariyle Covid-19 salgın verilerinin istatistiksel değerlendirilmesi. *Anadolu Kliniği Tıp Bilimleri Dergisi*, 25(1), 130-141.
22. Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2017, September). Stock price prediction using LSTM, RNN and CNN-sliding window model. In *2017 international conference on advances in computing, communications and informatics (icacci)* (pp. 1643-1647). IEEE.
23. Graves, A. (2013). Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*.
24. Karevan, Z., & Suykens, J. A. (2020). Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Networks*, 125, 1-9.
25. Ulyah, S. M., & Mardianto, M. F. F. (2019, December). Comparing the Performance of Seasonal ARIMAX Model and Nonparametric Regression Model in Predicting Claim Reserve of Education Insurance. In *Journal of Physics: Conference Series* (Vol. 1397, No. 1, p. 012074). IOP Publishing.