

Pandemi Süreçlerinde S-I-R Gruplarının Tahmini: Farklı Regresyon ve Diferansiyel Modellerin Performans Analizi

Hamit ARMAĞAN^{*1}, Ulaş YAMANCI²

¹ Süleyman Demirel Üniversitesi, Enformatik Bölümü, 32200, Isparta, Türkiye

² Süleyman Demirel Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, İstatistik Bölümü, 32200, Isparta, Türkiye

(Alınış / Received: 24.04.2025, Kabul / Accepted: 21.08.2025, Online Yayınlanma / Published Online: 25.08.2025)

Anahtar Kelimeler

Covid-19,
SIR Yayılım Modeli,
Gauss Regresyonu,
Fourier Regresyonu,
Sinüs Regresyonu,
Diferansiyel Denklem
Tabanlı Regresyon

Öz: Bu çalışmanın amacı, pandemi süreçlerinde hastalık şüphesi taşıyan sağlıklı bireyler (S- susceptible), hastalığa yakalanan bireyler (I- infected) ve iyileşen ya da vefat eden bireylerden (R- recovered) oluşan S-I-R gruplarının sayısını tahmin etmek için farklı regresyon modellerinin performansını değerlendirmektir. Türkiye'deki COVID-19 verileri Türkiye Cumhuriyeti Sağlık Bakanlığı'nın resmi internet sitesinden alınmıştır. Çalışmada kullanılan veri seti, pandeminin başlangıcından itibaren seçilen zaman dilimini kapsamakta ve günlük S-I-R grup sayılarını içermektedir. Polinom regresyonu, Fourier regresyonu, Gauss regresyonu ve sin toplamı regresyonu gibi farklı regresyon modelleri incelenmiş ve tahmin performansları analiz edilmiştir. Ayrıca, Chebyshev polinomlarına dayalı doğrusal olmayan bir denklem sistemi olan SIR modelinin diferansiyel çözümleri ile karşılaştırılmıştır. Performans ölçütleri olarak R^2 (RSquare- Determination Coefficient) ve Ortalama Hatanın Karekökü (RMSE- Root Mean Squared Error) kullanılmıştır. Sonuçlar, farklı regresyon modellerinin S-I-R gruplarının sayısını tahmin etmedeki performansını değerlendirerek pandemi süreçlerinin analizi için en uygun modelin belirlenmesine katkı sağlamaktadır.

Prediction of S-I-R Groups in Pandemic Processes: A Performance Analysis of Various Regression and Differential Models

Keywords

Covid-19,
SIR Spread Model,
Gaussian Regression,
Fourier Regression,
Sine Regression,
Differential Equation Based
Regression

Abstract: The aim of this study is to evaluate the performance of various regression models in predicting the number of individuals in the S-I-R groups during pandemic processes: susceptible individuals (S), infected individuals (I), and those who have recovered or passed away (R). The COVID-19 data used in the study were obtained from the official website of the Ministry of Health of the Republic of Türkiye. The dataset covers a selected time period starting from the onset of the pandemic and includes daily counts of the S-I-R groups. Various regression models, such as polynomial regression, Fourier regression, Gaussian regression, and sine sum regression, were analyzed and their prediction performances were evaluated. Additionally, these models were compared with the differential solutions of the SIR model based on Chebyshev polynomials, which constitutes a nonlinear system of equations. The performance metrics used in the evaluation were the coefficient of determination (R^2) and the root mean squared error (RMSE). The results contribute to identifying the most appropriate model for analysing pandemic dynamics by assessing the effectiveness of different regression approaches in predicting S-I-R group counts.

1. Giriş

Salgın hastalıkların tanınması ve kontrol altına alınması, insan sağlığının korunmasında önemli bir unsurdur. Bu sebeple SIR modeli epidemiyolojideki en önemli araçlardan biri olarak öne çıkmaktadır. 1927 yılında Kermack ve McKendrick [3] tarafından ortaya atılan SIR modeli, salgın hastalıkların nasıl yayıldığını ve yönetilebileceğini anlamada temel bir model olmuştur. Basit olmasına rağmen SIR modeli, kısa ve uzun vadeli salgın dinamiklerini tanımak için güçlü bir model sunar. Nüfusu üç ana gruba ayırır: Duyarlı bireyler (S), enfekte bireyler (I) ve iyileşmiş ya da vefat etmiş bireyler (R). Bu basit sınıflandırma, hastalığın nasıl yayıldığını ve popülasyon üzerindeki etkisini modellerken matematiksel analiz ve tahminler için alt yapı oluşturur.

SIR modeli, çok sayıda araştırmacının çeşitli salgın hastalıkları incelemesine ve modellemesine imkân sağlamıştır. Bu modelin uygun parametre çözümleri, hastalık dinamiklerinin daha iyi anlaşılmasına ve etkili müdahale stratejilerinin geliştirilmesine olanak sağlamaktadır. Bhattacharya ve arkadaşları [1] hastalığın yayılma oranını ve pandeminin potansiyel boyutunu değerlendirmek için SIR modelini ve varyasyonlarını incelemiştir. COVID-19 pandemisi süresince, SIR modeli ve türevleri, virüsün yayılma dinamiklerini anlamak ve kontrol stratejileri geliştirmek açısından kritik bir rol oynamıştır. Örneğin, Uddin ve Algehyne [13] COVID-19'un yayılmasını modellemek için diferansiyel denklemler kullanmış ve zaman içinde koruma, maruz kalma ve ölüm oranlarındaki değişiklikleri incelemiştir. Zhu ve Shen [14] gibi diğer araştırmacılar, pandeminin dinamiklerini değerlendirmek ve müdahale stratejilerine rehberlik etmek için SIR modeline dayalı varyasyon modeller geliştirmiştir. Ayrıca SIR modelinin yaklaşık çözümleri Hermite serileri kullanılarak elde edilmiştir [10].

Geniş çapta kabul gören SIR modeli, hastalık yayılımını anlamak ve kontrol stratejileri geliştirmek için matematiksel biyoloji ve epidemiyoloji alanlarında temel bir araç olmaya devam etmektedir. Bu model, salgınların kontrolündeki zorlukları aşmada ve gelecekteki olası salgınlara karşı daha etkin hazırlık yapmada önemli katkılar sunmaktadır.

Yapay zeka algoritmalarında, veriler arasındaki ilişkilerin modellenmesinde polinom çekirdekli regresyon algoritmaları yaygın biçimde tercih edilmektedir. Ancak birçok durumda, yalnızca tek bir girdi değişkeni kullanan doğrusal regresyon yöntemleri, özellikle günlük yaşamdan elde edilen verilerle oluşturulan modellerde yeterli tahmin doğruluğu sağlayamamaktadır [8]. Bu durum, araştırmacıları birden fazla girdi değişkenini aynı anda değerlendiren doğrusal olmayan regresyon tekniklerine yönlendirmiştir [6,9]. Bununla birlikte, bu tür modellerin doğrudan veya optimizasyon süreci

olmadan uygulanması, regresyon analizlerinde istenilen başarı düzeyine ulaşılmasını engelleyebilmektedir. Özellikle bazı veri kümelerinde, modelin performansını artırmak amacıyla polinom derecesinin yükseltilmesi ya da algoritmada kullanılan parametrelerin optimize edilmesi gerekmektedir. Ancak, polinom derecesinin artırılması modelin yapısını daha karmaşık hale getirebilmekte ve aşırı uyum (overfitting) riskini beraberinde getirmektedir [11,12]. Bu durum, yazılım algoritmalarının verimliliğini azaltmakta ve işlem süresini önemli ölçüde artırmaktadır. Buna karşın, parametre optimizasyonu ile desteklenen parametrik matematiksel modeller kullanan algoritmalar, daha yüksek başarı oranı ve işlem hızıyla öne çıkmaktadır [15,16].

Regresyon modelleri, girdi ve çıktı değişkenleri arasındaki matematiksel fonksiyonlar aracılığıyla sürekli değişkenlerin yanıtlarını tahmin eder [6,8]. Regresyon modellemesi, bağımlı değişken ile bağımsız değişken arasındaki ilişkinin bir fonksiyon ile temsil edilmesidir [6,9]. Bu fonksiyonun temel amacı, modellenen sistemle ilgili tanım kümesindeki değerlere dayanarak tahminleri hesaplamaktır [6]. Literatürde birçok regresyon modeli olup, bu modellerden polinom regresyon, Fourier regresyon, Gaussian regresyon, sin regresyon ve diferansiyel denklem tabanlı regresyon bu çalışmada tercih edilmiştir [9,12].

Bu çalışmanın amacı, pandemi süreçlerinde S-I-R gruplarındaki bireylerin sayısını tahmin etmek için farklı regresyon modellerinin performansını değerlendirmektir. Bu amaçla polinom, Fourier, Gaussian, Sinus ve diferansiyel denklem bazlı çeşitli regresyon modelleri Türkiye'deki COVID-19 verileri kullanılarak incelenmiş ve tahmin performansları karşılaştırılmıştır. Bu çalışma, pandemi süreçlerinin analizinde en uygun tahmin modelinin belirlenmesini amaçlamaktadır.

2. Materyal ve Metot

Bu çalışmada kullanılan COVID-19 verileri, Türkiye Cumhuriyeti Sağlık Bakanlığı'nın resmi internet sitesinden elde edilmiştir [2]. Sağlık Bakanlığı, pandeminin başlangıcından bu yana Türkiye'deki günlük vaka, iyileşme, test ve ölüm sayıları gibi çeşitli istatistikleri yayınlamıştır. Bu veriler, Türkiye'deki pandeminin seyrini anlamak için çok önemlidir. Analizde kullanılan veri seti, pandeminin başlangıcından itibaren seçilen zaman dilimini kapsamakta ve üç ana gruba ayrılan nüfus bilgilerini içermektedir: hastalıktan şüphelenilen sağlıklı bireyler (S), hastalığa yakalanan bireyler (I) ve iyileşen ya da vefat eden bireyler (R).

Bu çalışmada, Türkiye'de COVID-19 pandemisi sırasında üç ana grubun (S, I, R) sayısını tahmin etmek için farklı regresyon modelleri incelenmiştir. Bu

modeller arasında polinom regresyonu, Fourier regresyonu, Gauss regresyonu, sinus regresyonu ve diferansiyel bazlı regresyon yer almaktadır. Her model belirli matematiksel formüllere göre özelleştirilmiş ve veri setine uygulanmıştır.

Modellerin performansı R^2 (Determination Coefficient), RMSE (Root Mean Squared Error) ve MAPE (Mean Absolute Percentage Error) gibi istatistiksel ölçütler kullanılarak değerlendirilmektedir. R^2 değeri modelin bağımsız değişkenlerin hedef değişkeni ne kadar iyi açıkladığını ölçen bir istatistiktir. RMSE değeri ise modelin tahminlerinin gerçek değerlerden ne kadar sapıp saptığını bir göstergesidir. MAPE (Mean Absolute Percentage Error) yüzdelik hata miktarını göstermekte olup tahmin hatalarının gerçek değerlere göre yüzdesel büyüklüğünü ifade eder. Bu metrikler modeller arasında karşılaştırma yapılmasını ve en uygun modelin seçilmesini sağlamıştır.

2.1. Matematiksel modelleme

Bu çalışmada polinom regresyon (1), gauss regresyon (2), sigmoidal lojistik regresyon (3), fourier regresyon (4) ve sinus regresyon (5) modelleri kullanılmıştır. Her bir model, hastalığın yayılma dinamiklerinin farklı yönlerini ve karakteristiklerini yansıtmada özgün yeteneklere sahiptir. Örneğin, Fourier serisi mevsimsel değişimleri ve periyodik yapıları etkili biçimde modelleyebilirken; Gauss modeli, veri setindeki tepe noktalarını ve yoğunluk dağılımını temsil etmede güçlüdür. Sinüs toplamı modeli ise, periyodik olmayan dalgalanmalara odaklanarak daha karmaşık dalga biçimlerini yakalayabilme kapasitesine sahiptir.

$$y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (1)$$

$$f(x) = \sum_{i=1}^3 a_i \exp \left(-\frac{(x - b_i)^2}{c_i^2} \right) \quad (2)$$

$$f(x) = d + \frac{a - d}{1 + \left(\frac{x}{c}\right)^b} \quad (3)$$

$$f(x) = y = a_0 + \sum_{i=1}^n a_i \cos(iwx) + b_i \sin(iwx) \quad (4)$$

$$f(x) = \sum_{i=1}^5 a_i \sin(b_i x + c_i) \quad (5)$$

2.1.1. Polinom regresyon

Polinom regresyon, bağımlı bir değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi ifade etmek için polinom fonksiyonları kullanarak bir regresyon modeli oluşturma yöntemidir. Genel olarak, n. dereceden bir polinom regresyon modeli aşağıdaki gibidir:

$$y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n. \quad (6)$$

Burada, $a_0, a_1, a_2, \dots, a_n$ katsayıları modelin öğrenilen parametreleridir ve değerleri regresyon analizi ve doğrusal denklem sistemleri ile hesaplanır. Polinom regresyonu uygularken modelin derecesini seçmek önemlidir. Bunun nedeni, düşük dereceli bir modelin veri setinin karmaşıklığını yakalayamaması ve benzer şekilde yüksek dereceli bir modelin aşırı uyuma yol açabilmesidir. Bu model, değişkenler arasındaki karmaşık veya doğrusal olmayan ilişkileri modellemek için kullanılır.

2.1.2. Gauss regresyonu

Gauss regresyon modeli kullanılarak, veri setindeki çeşitli eğilimleri ve düzensizlikleri modellemek için Gauss fonksiyonlarının kombinasyonu kullanılmıştır.

$$f(x) = \sum_{i=1}^3 a_i \exp \left(-\frac{(x - b_i)^2}{c_i^2} \right) \quad (7)$$

Bu formüldeki a_1, a_2 ve a_3 katsayıları her bir Gauss fonksiyonunun yüksekliğini belirlerken, b_1, b_2 ve b_3 katsayıları fonksiyonun merkezini, yani x eksenindeki tepe noktasının konumunu ayarlar. c_1, c_2 ve c_3 katsayıları fonksiyonun genişliğini kontrol eder.

Gauss regresyon modeli, özellikle zaman serisi verilerindeki karmaşık eğilimleri ve dalgalanmaları yakalama kabiliyeti nedeniyle pandemi gibi dinamik ve hızlı değişen durumları modellemek için en uygun seçimlerden biridir. Bu modelin yüksek doğruluğu ve düşük hata oranı, sağlıklı bireylerin sayısını tahmin etmede büyük başarı sağlamış ve pandemi yönetiminde kullanılan önemli modellerden biri olmuştur. Dolayısıyla, sağlık politikaları ve müdahale stratejileri geliştirilirken bu matematiksel modellerin kullanılması daha doğru ve etkili kararlar alınmasına yardımcı olmaktadır.

2.1.3. 4-Parametrelili sigmoidal lojistik regresyon

4-P Sigmoidal lojistik regresyon modelinin matematiksel ifadesi aşağıda verilmiştir.

$$f(x) = d + \frac{a - d}{1 + \left(\frac{x}{c}\right)^b} \quad (8)$$

Bu modelde a parametresi fonksiyonun üst sınırını, d parametresi ise fonksiyonun alt sınırını temsil etmektedir. c parametresi eğimin derecesini belirler. b parametresi x eksenini boyunca kayma miktarını temsil eder ve eğrinin orta noktasının yerini belirler. Bu model, belirli bir duruma veya koşula duyarlı nüfus oranının belirli bir zaman aralığında nasıl değiştiğini modellemek için kullanılabilir. a, b, c ve d parametreleri, modelin hastalık yayılımı üzerindeki etkilerini yansıtmak şeklinde optimize edilir. 4-P sigmoidal lojistik model, bir hastalığın yayılma hızının başlangıçta yavaş olduğu, orta derecede hızlandığı ve

sonunda bir doygunluğa ulaştığı durumları modellemek için özellikle uygundur.

2.2.4. Fourier regresyonu

Fourier regresyonu, periyodik bir fonksiyonu sinüs ve kosinüs terimlerinin toplamı olarak ifade eden bir regresyon yöntemidir. Genel olarak Fourier regresyon modeli denklem 9'da verilmiştir:

$$f(x) = y = a_0 + \sum_{i=1}^n a_i \cos(iwx) + b_i \sin(iwx) \quad (9)$$

Model, mevsimsel etkileri, yayılma dönemlerindeki artış ve azalışları ve hastalık verilerinde gözlemlenen diğer periyodik özellikleri yakalamayı amaçlamaktadır. Her bir terim, farklı frekanslarda hastalığın yayılma dinamiklerindeki dalgalanmaları temsil etmektedir. Bu, modelin hastalık vakalarının zaman içinde nasıl değişebileceğini daha kesin bir şekilde tahmin etmesini sağlar.

a_0 terimi veri setinin genel ortalamasını ifade eder ve sabit bir değerdir. Cosinus ve Sinus terimleri farklı frekanslardaki dalgalanmaları modellemek için kullanılır ve $a_1, b_1, a_2, b_2, a_3, b_3$ katsayıları dalgalanmaların genliklerini (yükseklikleri ve derinlikleri) belirler. w dalgaının açısal frekansını ifade eder.

2.2.5 Sinüs toplamı modeli

Sinüs toplam modelleri, bir dizi sinüs fonksiyonunun toplamı olarak temsil edilir. Bu model Fourier serisine benzer, ancak genellikle daha basit bir formdadır ve belirli sayıda terimle sınırlıdır. Bu model, hastalık verilerinin periyodik olmayan ancak dalgalanan karakterini modellemek için kullanılabilir. Sinüs toplam modeli, hastalığın zamana göre değişen yayılma oranını yakalayabilir.

$$f(x) = \sum_{i=1}^5 a_i \sin(b_i x + c_i) \quad (10)$$

Modelin her bir bileşeni, farklı frekanslarda hastalığın yayılmasındaki dalgalanmaları ifade eder. Özellikle, hastalık vakalarının zaman içinde nasıl değişebileceğine dair daha ayrıntılı bir analiz sağlar.

a_i ($i=1,2,\dots,5$) katsayıları dalgalanmaların genliklerini göstermektedir. b_i ($i=1,2,\dots,5$) katsayıları dalgalanmaların frekanslarını belirler. Bu, hastalığın yayılma hızındaki periyodik değişikliklerin ne sıklıkla meydana geldiğini gösterir. c_i ($i=1,2,\dots,5$) sabitleri dalgalanmaların zaman eksenindeki konumunu, yani her bir dalgalanmanın başlangıç noktasını belirler.

2.2.6 SIR modeli

SIR salgın modeli matematiksel olarak doğrusal olmayan adi diferansiyel denklemler kümesi aşağıdaki formüller ile temsil edilebilir

$$\begin{aligned} \frac{dS(t)}{dt} &= -\frac{\alpha}{P} S(t)I(t), \\ \frac{dI(t)}{dt} &= \frac{\alpha}{P} S(t)I(t) - \mu I(t), \\ \frac{dR(t)}{dt} &= \mu I(t). \end{aligned} \quad (11)$$

Burada $S(0) = S_0, I(0) = I_0, R(0) = R_0$ başlangıç koşulları ve $P = S(t) + I(t) + R(t)$ dir.

[2]'de bildirilen ilk tarih 4 Nisan 2020 olduğundan, S_0, I_0 ve R_0 'nin başlangıç değerleri olarak 4 Nisan 2020'de Türkiye'deki şüpheli kişi sayısı ile enfekte ve iyileşen/vefat eden kişi sayısı alınmıştır. Popülasyon büyüklüğünün 84.000.000, enfekte kişi sayısının 3013 ve iyileşen/vefat eden kişi sayısının 378 (günde 76 ölüm ve günde 302 iyileşme) olduğu varsayıldığında, duyarlı birey sayısı 83.996.609'dur. Bireylerin COVID-19'a maruz kaldığından şüpheleniliyorsa önerilen karantina süresi 14 gün olduğundan dolayı bulaşma oranı $\alpha=1/14 \approx 0,0714$ olarak alınmaktadır. 4 Nisan 2020'de enfekte olan toplam kişi sayısı 23934 ve iyileşen/vefat eden toplam kişi sayısı 1287 (toplam 501 ölüm ve toplam 786 iyileşen hasta) olduğundan, iyileşme oranı $\mu=1287/23934 \approx 0.0538$ 'dir [7].

3. Bulgular

3.1. Duyarlı birey sayısı (S) ve zaman modeli

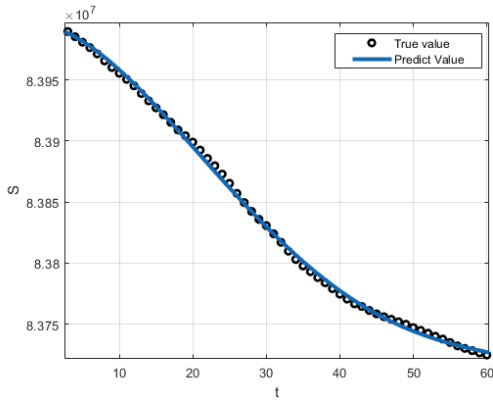
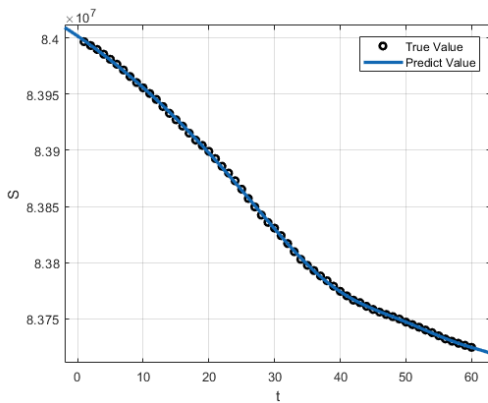
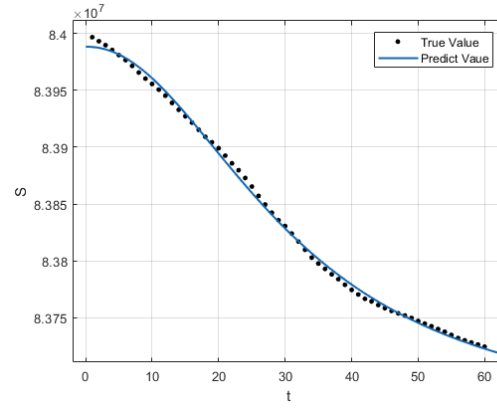
Hastalık şüphesi taşıyan sağlıklı bireylerin sayısını tahmin etmek için farklı regresyon modelleri geliştirilmiş ve bu modellerin performansları karşılaştırılmıştır. Polinom, Gauss ve sigmoidal lojistik regresyon modelleri seçilen zaman dilimi için Türkiye örneğinde geliştirilmiş ve her modelin R^2 , RMSE, MAE ve MAPE gibi hata ölçütleri ayrıntılı olarak analiz edilmiştir. Polinom regresyon modeli 0.99926'lık yüksek R^2 değeri ve 2519.7'lik kök ortalama kare hatası (RMSE) ile karakterize edilirken, gauss regresyon modeli 0.99993'lük en yüksek R-kare değeri ve 813.97'lik en düşük RMSE değeri ile daha iyi performans göstermektedir. Sigmoidal lojistik regresyon modeli 0.99842 R^2 ve 3658.6 RMSE değerine sahiptir.

Öte yandan, polinom ve sigmoidal lojistik regresyon modelleri de yeterli bir performans sergilemiş, ancak Gauss regresyon modeli daha yüksek tahmin doğruluğu ve daha düşük hata değerleri ile öne çıkmıştır. Bu sonuçlar, pandemi sürecinde kullanılan matematiksel modellerin seçiminde Gauss regresyonunun tercih edilebileceğini ve daha etkili tahminler sağlayabileceğini göstermektedir.

Tablo 2. S-t regresyonu hata ve yaklaşım değerleri

Regresyon Modeli	R ²	MAE	MAPE	RMSE
polinom (derece=4)	0.99926	2057.33	0.0025	2412.4
gaussian (terim sayısı=3)	0.99993	597.84	0.00071	750.44
4 parametrelili lojistik (sigmoidal)	0.99842	3051.4	0.0036	3534.6

Uzman görüşleri, pandemi sürecinde S parametresindeki azalmanın devam edeceğinin altını çizmektedir. Bu doğrultuda, seçilen zaman aralığında S parametresindeki azalma oranına bağlı olarak Gauss regresyon fonksiyonu en uygun model olarak bulunmuştur. Farklı regresyon modellerinin etkin bir şekilde kullanılabileceği ve uzman görüşlerinin bu süreçte hayati bir rol oynadığı da belirlenmiştir. SIR modelinin Chebysev polinomundan yararlanılarak elde edilen yaklaşık değerlerin gerçek değerleri sağlamadığı da gözlemlenmiştir [7].

**Şekil 1.** S-t Polinom regresyonu için gerçek ve tahmin değerleri**Şekil 2.** S-t Gauss regresyonu için gerçek ve tahmin değerleri**Şekil 3.** S-t Sigmoidal regresyonu için gerçek ve tahmin değerleri

3.2. Enfekte birey sayısı (I) ve zaman modeli

Türkiye'deki vaka çalışması, hastalık tahmininde farklı zaman serisi regresyon modellerinin performansını değerlendirmeyi amaçlamıştır. Belirli bir zaman aralığında hastalığa yakalanan bireylerin sayısını tahmin etmek için Fourier, Gauss ve Sinüs regresyon modelleri kullanılmış ve her bir modelin performansı hata oranları ve yaklaşım değerleri açısından değerlendirilmiştir.

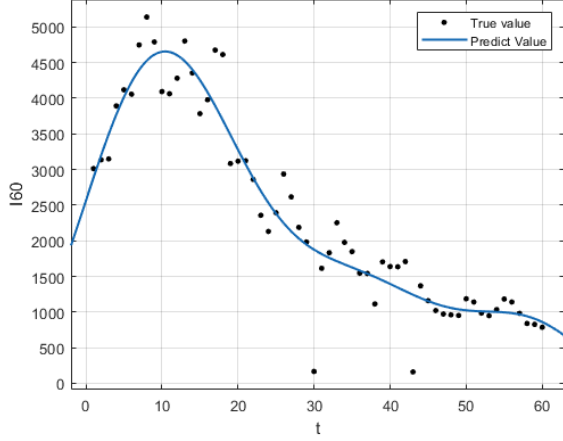
Çalışmada kullanılan Fourier regresyon modeli, zaman serisi verilerindeki periyodik değişiklikleri yakalamak için yeterli bir yöntemdir. Hastalık yayılımının tepe noktalarını ve dağılımını modellemek için Gauss regresyon modeli kullanılırken, daha karmaşık periyodik yapıları modellemek için Sinüs regresyon modeli tercih edilmiştir. Modeller arasındaki performans farklarını göstermek için her bir modelin hata ve yaklaşım değerleri Tablo 3.'de detaylı olarak gösterilmiştir. Bu değerlendirmeler sayesinde hastalık yayılma senaryoları için en uygun modelleme tekniklerinin belirlenmesi ve tahminlerin doğruluğunun artırılması amaçlanmaktadır. Böyle bir analiz, sağlık sektöründe daha bilinçli planlama ve kaynak tahsisi için oldukça önemlidir.

Tablo 3. I-t regresyonu hata ve yaklaşım değerleri

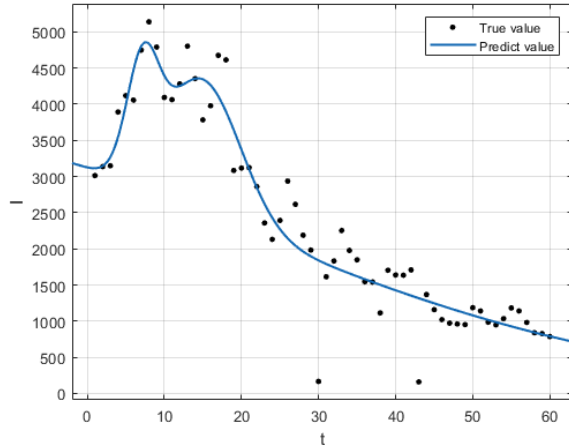
Regresyon Modeli	R ²	MAE	MAPE	RMSE
fourier (terim sayısı=3)	0.9134	270.4982	13.1367	403.0680
gaussian (terim sayısı=3)	0.9200	259.0115	13.1867	387.3738
sinus (terim sayısı=5)	0.9392	262.1306	17.8995	337.8367

Uzman görüşleri, pandemi sürecinde "I" parametresinde artan/azalan/kararlı durumlar arasında bir geçiş olabileceğini vurgulamaktadır. Bu kapsamda seçilen zaman aralığında "I" parametresinin azalma hızına bağlı olarak Sinüs regresyon

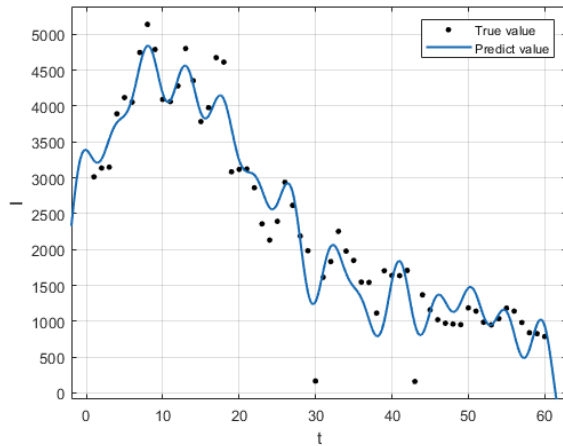
fonksiyonunun toplamının en uygun model olduğu tespit edilmiştir. Farklı regresyon modellerinin etkin bir şekilde kullanılabileceği ve uzman görüşlerinin bu süreçte kritik rol oynadığı da görülmüştür. SIR modelinin Chebyshev polinomu kullanılarak elde edilen yaklaşık değerlerin gerçek değerleri sağlamadığı da gözlemlenmiştir [7].



Şekil 4. I-t Fourier regresyonu için gerçek ve tahmin değerleri



Şekil 5. I-t Gaussian regresyonu için gerçek ve tahmin değerleri



Şekil 6. I-t Sinüs regresyonu için gerçek ve tahmin değerleri

Pandemi süreci, özellikle sağlık verilerinin anlamlandırılmasında birçok bilimsel ve istatistiksel analiz yönteminin test edildiği ve karmaşık modellerin ön plana çıktığı bir dönemdir. Bu dönemde "I" ve "S" parametreleri gibi kritik değişkenlerin dinamiklerini anlamak, pandeminin yayılımını ve etkilerini tahmin etmek için hayati önem taşımaktadır. Uzmanlar, pandemi süresince "I" parametresinin zaman içinde artma, azalma ya da sabit kalma gibi farklı durumlar arasında geçiş yapabildiğini ve bu geçişlerin pandeminin seyrini anlamada önemli bilgiler verdiğinin altını çizmektedir. Bu kapsamda, "S" parametresinin belirli bir zaman aralığındaki azalma oranının, Sinüs regresyon fonksiyonu kullanılarak modellenmesinin en uygun yaklaşım olduğu tespit edilmiştir. Bu fonksiyon, "S" parametresinin azalma hızındaki değişimleri dikkate alarak pandemi sürecindeki eğilimleri yakalayabilen esnek bir modelleme kapasitesine sahiptir.

Öte yandan, farklı regresyon modellerinin pandemi verileri üzerinde etkin bir şekilde kullanılması, bu süreçte veri analizi stratejilerinin çeşitliliğini göstermektedir. Doğrusal regresyon, lojistik regresyon ve makine öğrenmesi tabanlı regresyon modelleri gibi çeşitli yaklaşımlar pandemi dinamiklerinin karmaşıklığını ve değişkenliğini anlamak için önemli araçlar olarak ortaya çıkmıştır. Bu modeller özellikle "I" ve "S" parametreleri gibi kritik değişkenlerin zaman içindeki davranışlarını analiz etme ve tahmin etme açısından değerlendirmektedir. Uzman görüşü, model seçimi ve uygulamasında kritik bir rol oynar ve pandemi sürecinde bilimsel analizlerin değerlendirilmesinde önemli rol oynar. Bu durum, pandemi gibi küresel sağlık krizlerinin yönetiminde multidisipliner yaklaşımların ve bilimsel işbirliğinin önemini bir kez daha göstermektedir.

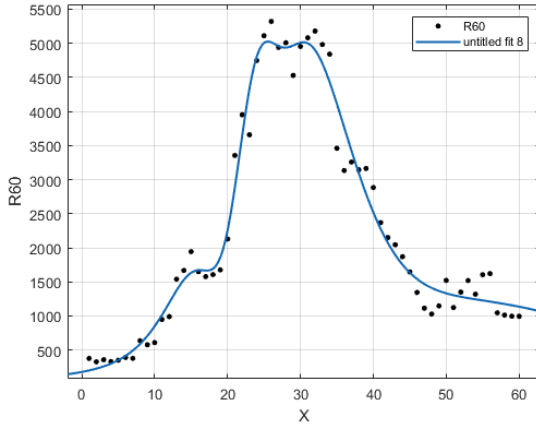
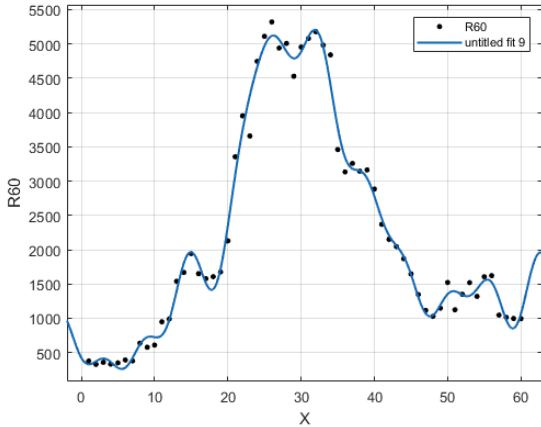
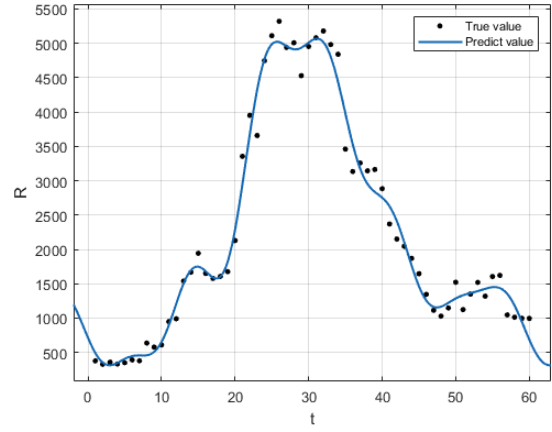
3.3. İyileşen/Vefat birey sayısı (R) ve zaman modeli

Türkiye'de pandemi boyunca hastalıktan kurtulan veya iyileşen bireylerin sayısını tahmin etmek için regresyon modellerinin iyileştirilmesi ve bu modellerin performansını karşılaştıran veriler Tablo 4.'de verilmiştir. Fourier, Gaussian ve Sinüs regresyon modelleri kullanılarak yapılan analizler, bu modellerin hastalığın yayılmasını ve etkisini modellemedeki etkinliğini değerlendirmektedir. Fourier modeli $0.9825 R^2$ ve $243.2374 RMSE$ ile yüksek bir uyum gösterirken, bunu 0.9739 'luk biraz daha düşük R^2 ve $303.9395 RMSE$ ile Gauss modeli takip etmektedir. Sinüs modeli bu özel durumda en iyi performansı göstermiş, $0.9898 R^2$ değeri ve $197.4230 RMSE$ değeri ile analiz edilen modeller arasında en yüksek uyum ve en düşük hata oranını sergilemiştir. Bu sonuçlar, farklı regresyon modellerinin pandemi verileri üzerindeki etkisinin kapsamlı bir şekilde anlaşılması için önemli bir temel oluşturmaktadır.

Tablo 4. R-t regresyonu hata ve yaklaşım değerleri

Regresyon Modeli	R ²	MAE	MAPE	RMSE
fourier (terim sayısı=7)	0.9825	156.8949	9.4161	208.2961
gaussian (terim sayısı=6)	0.9739	198.2659	13.0551	254.2940
sin toplamı (terim sayısı=7)	0.9898	116.7696	7.7674	159.1675

Tablo 4.'de sunulan hata ve yaklaşım değerleri, modellerin hastalık verileri üzerindeki tahmin yeteneklerini somut bir şekilde göstermektedir. Bu karşılaştırmalı analiz, pandemi sırasında hastalığın yayılma dinamiklerini ve nüfus üzerindeki etkisini anlamada matematiksel ve istatistiksel modellerin önemini göstermektedir. Özellikle Sinüs modeli en düşük RMSE ve en yüksek R² değerine ulaşarak zaman serisi verilerindeki periyodik dalgalanmaları yakalamadaki üstünlüğünü ortaya koymaktadır. Dolayısıyla, bu modeller pandemi müdahale stratejilerinin planlanması ve uygulanmasında önemli bir rol oynamaktadır. Ayrıca SIR modelinin Chebysev polinomu kullanılarak elde edilen yaklaşık değerlerin gerçek değerleri sağlamadığı görülmüştür [7].

**Şekil 7.** R-t Fourier regresyonu için gerçek ve tahmin değerleri**Şekil 8.** R-t Gaussian regresyonu için gerçek ve tahmin değerleri**Şekil 9.** R-t Sinus regresyonu için gerçek ve tahmin değerleri

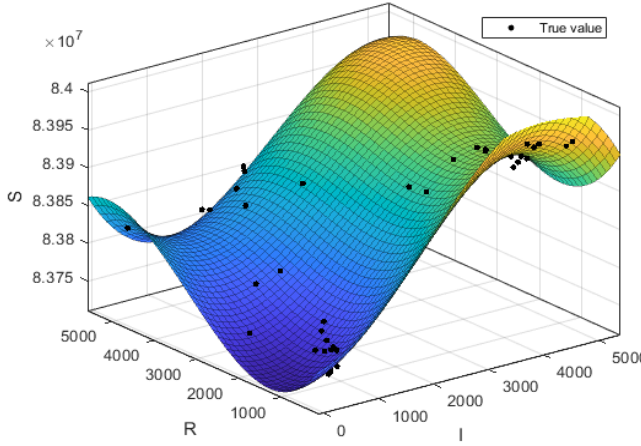
3.4. Duyarlı birey sayısı – enfekte birey sayısı – iyileşen/vefat birey sayısı ve zaman modeli

Bu analizde, Türkiye'de pandemi süresince hastalığa duyarlı sağlıklı bireyler (S), hastalığa yakalanan bireyler (I) ve iyileşen/vefat bireyler (R) gibi temel grupların sayısını tahmin etmek için polinom regresyon modeli kullanılmıştır. Polinom regresyon modeli, pandemi sırasındaki veri setlerinin karmaşıklığı ve bu verilerde gözlemlenen doğrusal olmayan ilişkiler nedeniyle seçilmiştir. Modelin R² değeri 0,9648 olarak hesaplanmış ve RMSE değeri 16667.2364 olarak belirlenmiştir. Bu değerler modelin veri setine ne kadar iyi uyduğunu ve tahminlerinin gerçek verilere ne kadar yakın olduğunu göstermektedir. Bu sonuçlar, polinom regresyon modelinin Türkiye örneğinde pandemi sürecindeki S-I-R gruplarının sayısını tahmin etmede önemli bir rol oynadığını göstermektedir.

Tablo 5. S-I-R regresyonu hata ve yaklaşım değerleri

Regresyon Modeli	R ²	MAE	MAPE	RMSE
polinom (derece x=3 ve y=3)	0.9648	13835.1047	0.0165	16667.2364

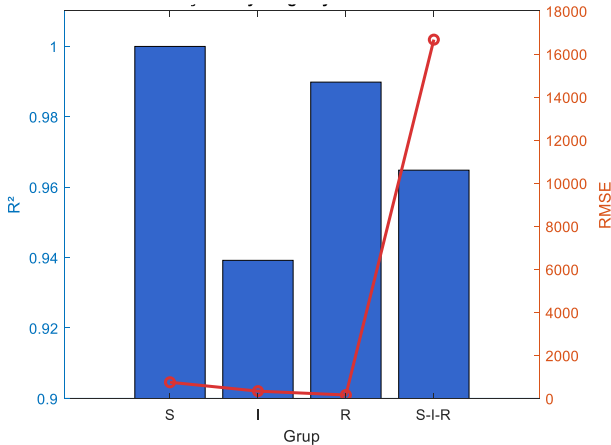
Tablo 5.'de ayrıntılı olarak verilen hata ve yaklaşım değerleri, polinom regresyon modelinin performansının daha iyi anlaşılmasını sağlamaktadır. Polinom regresyon, özellikle pandemi verileri gibi karmaşık ve zamanla değişen veri setlerinde doğrusal olmayan ilişkileri yakalama kapasitesi nedeniyle seçilmiştir. Modelin yüksek R² değeri, veri setinin büyük bir bölümünü başarıyla açıklayabildiğini ve genel eğilimleri yakalayabildiğini göstermektedir. Bununla birlikte, RMSE değerinin büyüklüğü, modelin bazı tahminlerde önemli hatalar yapabileceğine işaret edebilir ki bu, özellikle büyük nüfuslu ülkelerde pandemi yönetimi için dikkate alınması gereken bir faktördür. Bu tür modellerin geliştirilmesi ve değerlendirilmesi, pandemi müdahale stratejilerinin planlanması ve uygulanması için değerli bilgiler sağlar, böylece daha etkili sağlık politikaları ve müdahaleleri geliştirilebilir.



Şekil 10. S-I-R regresyonu için gerçek ve tahmin değerleri

3.5. Regresyon modellerinin S-I-R gruplarına göre genel karşılaştırması

Regresyon modellerinin S-I-R parametrelerine göre karşılaştırmalı analizi, her parametre için farklı modellerin öne çıktığını göstermektedir. Duyarlı birey sayısı (S) için Gauss regresyon modeli, en yüksek R^2 (0.99993) ve en düşük RMSE (750.44) ile en başarılı model olmuştur. Enfekte birey sayısı (I) için sinüs toplam regresyon modeli, en yüksek R^2 (0.9392) ve en düşük RMSE (337.8367) değerleriyle en iyi tahmini sağlamıştır. İyileşen/vefat eden birey sayısı (R) tahmininde ise sinüs toplam regresyon modeli en yüksek başarıyı göstermiştir ($R^2 = 0.9898$, RMSE = 159.1675). Genel S-I-R tahmininde kullanılan polinom regresyon modeli, yüksek R^2 (0.9648) sunmasına rağmen alt gruplara göre daha yüksek hata değerlerine sahiptir. Bu sonuçlar, her bir S-I-R bileşeni için ayrı ve uygun model seçiminin tahmin doğruluğunu artırdığını ortaya koymaktadır.



Şekil 11. S-I-R grupları için en uygun regresyon modellerinin R^2 ve RMSE değerleri

4. Tartışma ve Sonuç

Bu çalışmada SIR modeli ile farklı regresyon modelleri, aynı S-I-R zaman serisi verileri üzerinde karşılaştırılmıştır. SIR modeli diferansiyel denklemler aracılığıyla nüfus gruplarının değişim hızlarını

modelleyen sürekli yapı bir sistem iken; regresyon modelleri, doğrudan gözlemsel verilere dayalı zaman serisi tahminleri üretmektedir. Modellerin matematiksel temelleri farklı olsa da, her biri aynı tarihsel veri seti ve başlangıç koşullarıyla çalıştırılmış; böylece $S(t)$, $I(t)$ ve $R(t)$ değerlerine ilişkin tahminler, ortak hata ölçütleri (R^2 , RMSE vb.) ile nesnel biçimde değerlendirilmiştir. Bu yaklaşım, modellerin teorik yapılarından ziyade, tahmin doğruluklarını esas alarak uygulama düzeyinde karşılaştırılabilirliğini sağlamaktadır.

Bu çalışmanın sonuçları, pandemi süreci boyunca farklı regresyon modellerinin kullanılmasının S-I-R verilerinin tahmini için etkili yöntemler olduğunu göstermektedir. Özellikle Gauss regresyon modeli, "S" parametresinin tahmininde en yüksek doğruluk ve en düşük hata değerleri ile öne çıkmaktadır. Bu sonuçlar, pandemi yönetimi ve stratejileri için matematiksel modellerin seçiminde Gauss regresyonunun tercih edilebileceğini ve daha güvenilir tahminler üretebileceğini göstermektedir. Buna ek olarak, Sinüs regresyon modelinin "I" parametresinin tahmini için en uygun seçenek olduğu görülmüştür. Bu model, pandemi sırasında hastalığa yakalanan bireylerin sayısını daha iyi yakalayarak doğru bir tahmin kabiliyeti sağlamaktadır. Bu sonuçlar, pandeminin seyrini ve etkilerini anlamak için farklı regresyon modelleri kullanmanın daha doğru tahminler ürettiğini göstermiştir. Son olarak, polinom regresyon modeli, pandemi sırasında S-I-R verilerinin tahmin edilmesinde etkili bir seçenek olarak görünmektedir. Bu model, pandemi müdahale stratejilerinin planlanması ve yürütülmesinde önemli bir rol oynayabilir. Bununla birlikte, büyük nüfuslu ülkelerde tahminlerde bazı hatalar olabileceği göz önünde bulundurulmalıdır. Bu çalışma, pandemi müdahale stratejilerinin geliştirilmesi ve uygulanmasında veri odaklı yaklaşımların önemini arz etmektedir.

Etik Beyanı

Bu çalışmada, "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" kapsamında uyulması gerekli tüm kurallara uyulduğunu, bahsi geçen yönergenin "Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler" başlığı altında belirtilen eylemlerden hiçbirinin gerçekleştirilmediğini taahhüt ederiz.

Kaynakça

- [1] P. Bhattacharya, S. Paul, P. Biswas, Mathematical Modeling of Treatment SIR Model with Respect to Variable Contact Rate, International Proceedings of Economics Development and Research, 83(2015), 34-41.
- [2] "Genel Koronavirüs Tablosu", T.C. Sağlık Bakanlığı COVID 19 Bilgi Platformu, [Çevrimiçi]. Erişim: <https://covid19.saglik.gov.tr/TR-66935/genel-koronavirus-tablosu.html>.

- [3] W.O. Kermack ve A.G. McKendrick, A contribution to the mathematical theory of epidemics, Proceedings of the Royal Society of London A, 115(1927), s. 700-721.
- [4] B.M. Ndiaye, L. Tendeng ve D. Seck, "Makine öğrenimi, deterministik ve stokastik SIR modelleri ile COVID-19 pandemisi ile teyit edilmiş vakaların karşılaştırmalı tahmini," ArXiv:2004, Kantitatif Biyoloji: Populations ans Evolution, 2004(2020), s.13489.
- [5] C.N. Ngonghala, E. Iboi, S. Eikenberry, M. Scotch, C.R. MacIntyre, H.M. Bonds ve A.B. Gumel, İlaç dışı müdahalelerin 2019 yeni Koronavirüsünü azaltma üzerindeki etkisinin Matematiksel Değerlendirmesi, Matematiksel Biyobilimler, 325(2020), 108364.
- [6] S. Nusinovici ve diğerleri, Logistic regression was as good as machine learning for predicting major chronic diseases, J. Clin. Epidemiol., 122(2020), 56-69.
- [7] D. Özdek, COVID 19'un SIR Modeli için Chebyshev Polinom Çözümü, Gazi Mühendislik Bilimleri Dergisi, 9(4)(2023), 39-47.
- [8] P. Rivas-Perea, J. Cota-Ruiz, D. G. Chaparro, J. A. P. Venzor, A. Q. Carreón, and J. G. Rosiles, Support Vector Machines for Regression: A Succinct Review of Large-Scale and Linear Programming Formulations, Int. J. Intell. Sci., 3(1)(2013), 5-14.
- [9] M. Shahhosseini, G. Hu, and H. Pham, Optimizing ensemble weights and hyperparameters of machine learning models for regression problems, Mach. Learn. with Appl., 7(2022), 100251.
- [10] A. Secer, N. Ozdemir ve M.A. Bayram, Salgın hastalıkların SIR modelinin çözümü için Hermite polinom yaklaşımı, Matematik, 6(12)(2018), s. 305.
- [11] V. Strijov ve G. W. Weber, Hyperparameter optimization kullanarak doğrusal olmayan regresyon modeli oluşturma, Comput. Math. with Appl., 60(4)(2010), 981-988.
- [12] O. Torrealba-Rodríguez, R. A. Conde-Gutiérrez, A. L. Hernández-Javier, Matematiksel ve hesaplamalı modeller uygulayarak Meksika'da COVID-19'un modellenmesi ve tahmini. Kaos, Solitonlar ve Fraktallar, 138(2020), 109946.
- [13] R. Uddin ve E.A. Algehyne, Konveks İnsidans Oranlı SIR Modeli Kullanılarak COVID-19'un Matematiksel Analizi, Fizikte Sonuçlar, 23(2021), 103970.
- [14] W. Zhu, S. Shen, Çin'deki COVID-19 Salgın Dinamiklerini Tanımlayan Geliştirilmiş Bir SIR Modeli. Fizikte Sonuçlar, 25(2021), 104289.
- [15] P. Wang, X. Zheng, J. Li ve B. Zhu, COVID-19'daki salgın eğilimlerinin lojistik model ve makine öğrenimi teknikleri ile tahmini, Chaos, Solitons & Fractals, 139(2020), s. 110058.
- [16] H. Xue, Y. Bai, H. Hu, H. Liang, Çoklu Regresyon Modeli ve Yapay Sinir Ağına Dayalı İnfluenza Aktivite Gözetimi, IEEE Access, 6(2017), 563-575.