



Acil servise başvuru sayısının zaman serisi analiz ve makine öğrenmesi yöntemleri ile tahmin edilmesine yönelik bir uygulama

An application for forecasting the number of applications to the emergency department with time series analysis and machine learning methods

Sema ÇİFTÇİ^{1*} , Gül Didem BATUR SİR² 

¹Endüstri Mühendisliği Bölümü, Mühendislik Fakültesi, Ostim Teknik Üniversitesi, Ankara, Türkiye.

semactc@gmail.com

²Endüstri Mühendisliği Bölümü, Mühendislik Fakültesi, Gazi Üniversitesi, Ankara, Türkiye.

dbatur@gazi.edu.tr

Geliş Tarihi/Received: 28.07.2022

Düzeltilme Tarihi/Revision: 12.12.2022

doi: 10.5505/pajes.2022.18488

Kabul Tarihi/Accepted: 29.12.2022

Araştırma Makalesi/Research Article

Öz

Günümüzde, acil sağlık servislerine yönelik talepler, salgın, deprem vb. doğal afetler ile patlamalar gibi durumlarda olağanüstü artış göstermektedir. Söz konusu talebin doğru bir şekilde tahmin edilmesi, acil servislere başvuracak kişi sayısının belirlenmesi ve ilgili kaynak planlamalarının etkin şekilde gerçekleştirilmesini sağlayacağından olağanüstü durumlar için kriz yönetim sürecinde kolaylık sağlayacaktır. Bu çalışmada, bir acil servise başvuru sayısının tahmini amaçlanmaktadır. Ele alınan, mevsimsel özelliklere sahip verilere yönelik olarak, zaman serisi analizi yöntemlerinden SARIMA, Holt-Winters ve ayrıştırma; makine öğrenmesi yöntemlerinden rastgele ağaç ve rastgele orman teknikleri kullanılmıştır. Tahmin çalışması için Ankara'da yer alan bir hastanenin 396 günlük "başvuran hasta sayısı" verisi kullanılmıştır. Her bir yöntemdeki tahminler, yedi, on beş ve otuz günlük olarak yapılmıştır. Talep tahmini yöntemlerinden en başarılı yöntemin belirlenebilmesi için korelasyon, düzeltilmiş RKARE ve ortalama mutlak yüzde hatası değerlerinden faydalanılmıştır. Yapılan analizlerde SARIMA yönteminin, acil servise yapılacak başvuru sayısının tahmin edilmesinde diğer yöntemlere göre daha etkili sonuçlar verdiği görülmüştür. Ayrıca, acil servislere yapılan başvuruların sürekli değişen, dinamik bir yapıya sahip olmasının bir sonucu olarak, tahmin edilen gün sayısındaki değişimin tahmin değerleri üzerinde önemli etkisi olduğu da anlaşılmıştır.

Anahtar kelimeler: Talep tahmini, Acil servis, Zaman serisi analizi, SARIMA, Holt-Winters, Makine öğrenmesi.

Abstract

Today, the demands for emergency health services show an extraordinary increase in cases such as epidemics, earthquakes, natural disasters, and explosions. The accurate estimation of the demand in question will facilitate the crisis management process for extraordinary situations, as it will enable the determination of the number of people who will apply to the emergency services and the effective realization of the relevant resource planning. In this study, it is aimed to estimate the number of applications to an emergency department. For the seasonal data, SARIMA, Holt-Winters and decomposition, which are among the time series analysis methods; Random tree and random forest techniques from machine learning methods are used. For this forecasting study, 396-day "number of patients admitted" data of a hospital located in Ankara is used. Forecasts in each method are performed for seven, fifteen, and thirty days. Correlation corrected square root and average absolute percentage error values are used to determine the most successful one among demand forecasting methods. In the analyzes made, it is observed that SARIMA method gives more effective results than others in forecasting the number of applications to the emergency department. In addition, because of the constantly changing and dynamic nature of the applications made to the emergency services, it is understood that the change in the forecasted number of days has a significant effect on the resulting forecast values.

Keywords: Demand forecast, Emergency service, Time series analysis, SARIMA, Holt-Winters, Machine learning.

1 Giriş

Hastanelerde yer alan acil servisler, sağlık sistemlerinin temel yapısını oluşturmaktadır. Acil servisler, hastanelere başvurulduğunda genellikle ilk gidilen yerler olmakla birlikte, karar vericiler (hastane yöneticileri) açısından planlaması zor sistemlerdir. Bu noktada en önemli etken, söz konusu sistemlerin doğası gereği sürekli değişen ve ani dalgalanmaların yaşandığı dinamik bir yapıya sahip olmasıdır. Bu aşamada, karar vericilerin insan (doktorlar, hemşireler vb.) ve malzeme (yataklar, kutular vb.) kaynak tahsisini optimize etmesine yardımcı olacak hasta akışını tahmin etmesi gerekmektedir [1]. Bu nedenle acil servislere başvuru sayısının tahmin edilmesi, hastane yönetim sisteminin işleyişini etkileyecek önemli bir problemdir.

Günümüzde insanların kitlesel olarak etkilendiği; sel, yangın, deprem vb. doğal afetler ile salgın, kaza, zehirlenme gibi tıbbi müdahalenin zorunlu olduğu durumlarda acil sağlık sistemlerine yönelik taleplerde ciddi artış gözlemlenmektedir [2]. Bu tür olağanüstü hallerde sağlık hizmetlerinin etkin ve kesintisiz sürdürülebilirliği açısından, acil servis birimlerine başvuran kişi sayısının doğru tahmin edilmesi; erken müdahale, kriz yönetimi ve acil servis kapasite planlaması açısından hayati önem taşımaktadır. Ayrıca, doğru tahmin çalışmalarının, hastanelerin iktisadi planlamalarına (kaynak oluşturma, malzeme temini, yıllık bilanço, ödemeler vb.) doğrudan bir etkisinin olduğu da bilinmektedir. Belirtilen olağanüstü durumlar halinde, acil servis yoğunluğunu belirleyen bir diğer kriter ise mevsimsel farklılıklardır [2]. Bu çalışmada, zaman serisi analizi ve makine öğrenmesi yöntemleri birlikte kullanarak mevsimsel döngüye sahip veri

*Yazışılan yazar/Corresponding author

setiyle bir acil servise başvuru sayısının tahmini amaçlanmaktadır.

Son yıllarda hastalık tipine bağlı olarak başvuru sayısının tahminine yönelik çalışmalar [3],[4], covid-19 salgınına yönelik yapılan tahmin çalışmaları [5],[6], malzeme ve ekipman sayısının tahmini [7], günlük dolu yatak sayısının tahmini [8], [9] hastaneden taburcu edilecek hasta sayısının tahmin edilmesi [10] gibi sağlık sistemlerine yönelik tahmin çalışmaları büyük ölçüde artış göstermiştir. Konunun daha detaylı bir şekilde analiz edilebilmesi amacıyla bu çalışmada, ilgili literatür, kullanılan çözüm yöntemi, tahmin edilen toplam süre, veri seti aralığı, mevsimsellik etkisi ve değerlendirme ölçütü açılarından incelenerek analiz edilmiştir. Analiz sonuçları Tablo 1’de verilmiştir.

Tablo 1 incelendiğinde; zaman serisi analizi yöntemleri [11], [18] ve yapay zekâ yöntemleri [2], [21] sıklıkla kullanıldığı görülmektedir. Ancak bu yöntemlerin veri seti tiplerine göre tahmin performansı değişkenlik göstermektedir ve acil sağlık servislerine başvuracak hasta sayısı probleminde kullanılacak doğru tahmin modelinin belirlenmesi oldukça zordur. Bu çalışmada, makine öğrenmesi yöntemlerinin tahmin çalışmalarında, literatürde sıklıkla kullanılması ve performans ölçüm değerlerinin yüksek olması nedeniyle; trend ve mevsimsel özelliklere sahip veri tiplerinde kullanılan zaman serisi analizi yöntemlerinden SARIMA, Holt-winters ve ayrıştırma yöntemleri birlikte kullanılmıştır. Zaman serisi analizi ve makine öğrenmesi yöntemlerinin birlikte kullanıldığı çalışma, acil servislere yapılan başvuru sayısının gerçeğe en yakın şekilde tahmin edilmesi açısından faydalıdır.

Tablo 1’deki çalışmalar “tahmin periyodu” açısından analiz edildiğinde, günlük [17], aylık [12] ve yıllık [11], [21] şeklinde değişkenlik gösterdiği görülmüştür. Sağlık sistemlerinin yapısı itibarıyla tahmin yapılan süre için en uygun değerlerin belirlenmesi oldukça zordur. Bu nedenle bu çalışmada tahmin periyodu; gün, ay ve yıl olarak ayrı ayrı değerlendirilmiştir.

Bir diğer parametre olarak, yapılan çalışmalar “veri seti aralığı” açısından değerlendirildiğinde, aylık [12] ve yıllık aralıklar [21] şeklinde verilerin ele alındığı görülmüştür. Bu çalışmada üzerinde çalışılan veri seti 01.31.2021- 31.03.2022 tarih aralığından oluşan 369 günlük acil servise başvuran günlük hasta sayısıdır.

Yapılan analizler ışığında, literature genel olarak bakıldığında, Türkiye’ de yer alan bir hastanenin acil servisindeki salgın dönemi başvuru sayısı verisini kullanarak yapılan bu uygulama çalışması, veri setinin trend ve mevsimsel özelliklere sahip yapısı itibarıyla kullanılacak yöntem ve yöntemlerde yer alan parametrelerin uygun bir şekilde belirlenmesi açısından zor bir problemdir. Bu amaçla bu çalışma:

- ✓ Mevsimler etkileri göz önünde bulundurularak, mevsimsellik özelliğinin periyodik olarak değerlendirilmesi,
- ✓ Zaman serisi analizi ve makine öğrenmesi yöntemlerini kullanarak karşılaştırılmalı bir analiz gerçekleştirilmesi,
- ✓ Zaman serisi analizleri yöntemlerinden; SARIMA, Holt-winters ve ayrıştırma yöntemlerini birlikte kullanılarak, yöntemlerin kıyaslamalı analiz edilmesi,
- ✓ Zaman serisi ve makine öğrenmesi yöntemlerinde tahmin edilen gün sayısını değişken olarak kullanarak tahmin gücünün analiz edilmesi,

açılarından literatüre katkı sağlamaktadır.

Çalışmanın devamı şu şekilde organize edilmiştir: Bölüm 2’ de çalışma kapsamında kullanılan zaman serisi analizi ve makine öğrenmesi yöntemini içeren materyal ve yöntem, Bölüm 3’ te çalışmanın zaman serisi analizi ve makine öğrenmesi yöntemleri kullanılarak acil servise yapılan günlük hasta sayısının tahmin edilmesi uygulaması, Bölüm 4 ve Bölüm 5’te sonuçlar ve tartışma verilerek çalışma sonlandırılmıştır.

2 Materyal ve yöntem

2.1 Zaman serisi analizi yöntemleri

2.1.1 ARIMA modeli

Box ve Jenkins olarak da bilinen ARIMA modeli, gelecek tahmini için kullanılan istatistiksel yöntemlerden biridir. Box-Jenkins yöntemi, tek değişkenli zaman serilerinin geleceğini tahmin etmek için kullanılır. Eşit zaman aralıklarında elde edilen gözlem değerlerinden oluşan kesikli ve durağan zaman serilerinin gelecek tahmin modellerinin oluşturulmasında sistematik bir yaklaşım gösterir.

Tablo 1. Acil servise başvuru sayısı için literatür araştırması.

Table 1. Literature search for the number of admissions to the emergency department.

Çalışmalar	Çözüm Yöntemi	Tahmin Edilen Toplam Süresi	Veri Seti Aralığı	Mevsimsellik Etkisi	Değerlendirme Ölçütü
[11]	SARIMA, YSA	1-7-14-21-20 günlük	2005-2007	✓	MAPE
[12]	Çoklu Zaman Serisi	24 saatlik	2006	✓	MAE
[13]	İstatistiksel Analiz	Aylık	2009-2010	×	Medyan- Ortalama
[14]	Çoklu Düzleme Yöntemi	Aylık	2005-2011	×	RMAE
[15]	Doğrusal Regresyon	2 Aylık	2010	×	R ² - MAPE
[16]	Lojistik Regresyon	30 günlük	2005-2008	×	Güven Aralığı
[1]	ARMA	×	2013-2014	×	MAE- RMAE- RMAP
[17]	ARIMA- SARIMA	14 gün	76 günlük	✓	MAPE
[18]	SARIMA	1 Yıllık	2012-2013	✓	MAPE
[19]	İstatistiksel Analiz	Yıllık	2001-2009	×	×
[20]	Lojistik Regresyon- Sinir Ağı	Aylık	2012	×	MAPE
[21]	Rassal Orman	4 Yıllık	2005-2019	×	MAPE-MAD-MSE
[2]	Derin Öğrenme	3-6-9-12-15-18-21 günlük	2011-2012	×	RKARE- MAE
[22]	Veri Madenciliği	Yıllık	2014	×	X ² -F değeri
[23]	İstatistiksel Analiz	Yıllık	2006-2010	×	X ² - Varyans Analizi
[24]	Lojistik Regresyon	14 günlük	2020	×	Güven Aralığı
Bu çalışma	Zaman serisi analizi yöntemleri- Makine Öğrenmesi	7-15-30 günlük	2021-2022	✓	R- RKARE- MAPE

MAE: Ortalama mutlak hata. RMAE: Göreceli ortalama mutlak hata. RMAP: Göreceli ortalama mutlak performans. YSA: Yapay sinir ağları.

Eşit zaman aralıklarında elde edilen gözlem değerlerinden oluşan seriler, kesikli ve durağan olan Box-Jenkins yönteminin önemli varsayımlarıdır [25]. Aynı zamanda zaman serisi verilerini tahmin etmek için çok etkili bir araçtır. ARIMA yöntemi, verileri analiz etmek için AR (otoregresif) ve MA'yı (hareketli ortalamalar) birleştirir. ARIMA modelleri durağan zaman serileri için kullanılmaktadır. Verilerin stabilizasyonu d (Integrated- d) işlemindeki fark alınarak yapılır. Otopregresyon parametresinin derecesi p, fark derecesi parametresi d ve hareketli ortalama parametresinin derecesi q ise bu modele derece cinsinden Otopregresif Entegre Hareketli Ortalama modeli denmektedir [26].

Bir ARIMA'nın genel ifadesi ARIMA (p, d, q) Denklem 1'deki gibidir:

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \dots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (1)$$

Durağan bir seri elde etmek için serilerin 1 derece farkının alınması sonucunda Denklem 2' de verilen formül elde edilir. Bu işlem ile verilerin d derece farkı alınarak durağan bir veri elde edilmektedir.

$$\nabla x_t = w_t = x_t - x_{t-1} \quad (2)$$

∇ = Fark operatörü
 d = Fark derecesi
{ w_t } = Farklı seriler

ARIMA modelinin gerçekleştirilmesinde örneğin birinci derece için kullanılan formül Denklem 3 ve 4'te gösterilmektedir [27].

$$y_t = \beta_0 + \beta_1 y_{(t-1)} + u_t \quad (3)$$

$$u_t = \phi u_{(t-1)} + \varepsilon_t \quad (4)$$

Burada birinci derece otopregresif istatistiksel modelinde, -1 ile 1 arasında değer alan bilinmeyen otokorelasyon parametresi, sıfır ortalama ve sabit bir varyansa sahip bağımsız hata terimidir. İstatistiksel model yapısı ARIMA zaman serisi modeli olarak tanımlanmaktadır [27].

Daha genel ifade ile p' inci dereceden ARIMA modelinin çalışma sürecini ifade eden formül Denklem 5 ve Denklem 6'da gösterilen istatistiksel model ile tanımlanmaktadır [27].

$$y_t = \beta_0 + \beta_1 y_{(t-1)} + \dots + \beta_p y_{(t-p)} + u_t \quad (5)$$

$$u_t = \phi u_{(t-1)} + \phi_2 u_{(t-2)} + \dots + \phi_p u_{(t-p)} + \varepsilon_t \quad (6)$$

2.1.2 SARIMA modeli

ARIMA modeline benzeyen ancak daha güçlü olan SARIMA modeli genel olarak SARIMA (p, d, q) (P, D, Q)_m olarak ifade edilir. ARIMA modeli ile karşılaştırıldığında, zaman serisinin mevsimsel bileşeninin özelliğini açıklamak için SARIMA modeline dört yeni parametre eklenmiştir. p, q ve d trend bileşeninin otopregresif düzeni, fark alma sırası ve hareketli ortalama sırası; P, D ve Q mevsimsel bileşeninin mertebeleridir. Diğer bir parametre olan m, tek bir mevsimsel dönemin zaman adımlarını göstermektedir. SARIMA modelinin ifadesi Denklem 7'de verilmiştir [28]:

$$\phi_p(B^m)\varphi(B)\nabla_s^D\nabla^d x_t = O_Q(B^m)\theta(B)w_t \quad (7)$$

X_t: Durağan olmayan zaman serisi, W_t: Gauss beyaz gürültü sürecidir, $\phi_p(B^m)$ ve $\theta_Q(B^m)$ mevsimsel bileşeninin otopregresif

ve hareketli ortalamasını temsil eder. ∇_s^D ve ∇^d mevsimsel bileşen ve trend bileşeninin farklılaşan bileşenlerinin gösterir. B geri kaydırma operatörüdür. $\varphi(B)$ ve $\theta(B)$ trend bileşeninin otopregresif ve hareketli ortalama süreci polinomlarla temsil edilmektedir.

2.1.3 Holt-Winters üstel düzleme yöntemi

Winters [29] ve Holt [30] tarafından geliştirilen Holt-Winters modeli olarak da bilinen üçlü üstel yumuşatma yöntemi, düzeltilmiş zaman serileri üretmek için trend ve mevsimsellik bileşenlerini yakalamak için kullanışlıdır. Holt-Winters modelinin toplamalı ve çarpmalı yöntemler olmak üzere iki çeşidi bulunmaktadır. Mevsimsel değişimlerin kabaca sabit olduğu durumlarda toplama yöntemi tercih edilirken, mevsimsel değişimlerin serinin seviyesine orantılı olarak değiştiği durumlarda çarpma yöntemi tercih edilir [31].

Holt-Winters toplamsal modeli Denklem 8, 9, 10 ve 11 kullanılarak hesaplanır [31]:

$$trend = L_t = \alpha(y_t - s_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (8)$$

$$mevsimsel = b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (9)$$

$$seviye = s_t = \gamma(Y_t - L_t) + (1 - \gamma)s_{t-s} \quad (10)$$

$$tahmin = F_{t-m} = L_t + b_t m + S_{t-s+m} \quad (11)$$

α, β ve γ her biri [0,1] aralığında bulunan düzleme parametresi olmak üzere;

α : ortalama düzeyi

β : eğim parametresi

γ : mevsimsellik parametresi

m: tahminin kaçınıcı ileriki döneme ait olduğunu belirten değer

s: mevsimselliğin uzunluğunu (1 yıldaki ayların ya da çeyrek dönemlerin sayısı)

Y_t : değişkenin t anındaki gözlem değeri

L_t : Serinin t dönemindeki genel seviyesi

bt: Trend bileşeni

St: Mevsimsel bileşen

F_{t+m} : m ileri dönem için tahmin değerini göstermektedir.

Çarpımsal Holt Winters Yöntemi Denklem 12, 13, 14 ve 15 kullanılarak hesaplanır [32]:

$$trend = L_t = \frac{Y_t}{S_{t-s}}\alpha + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (12)$$

$$mevsimsel = b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (13)$$

$$seviye = s_t = \gamma \frac{Y_t}{L_t} + (1 - \gamma)s_{t-s} \quad (14)$$

$$tahmin = F_{t-m} = (L_t + b_t m)S_{t-s+m} \quad (15)$$

2.1.4 Ayırıştırma yöntemi

Ayırıştırma yöntemi; bir seriyi trend, mevsimsellik ve hata bileşenlerinin bir kombinasyonu olarak modelleyen bir zaman serisi analizi yöntemidir. Toplamsal ve çarpımsal şeklinde iki farklı türü bulunmaktadır. Toplamsal ve çarpımsal ayırıştırma yöntemi formülleri Denklem 16 ve 17' de verilmiştir [33]:

$$y_t = w_{(x_{t-1})} + R_{(x_{t-1})}\varepsilon_t \quad (16)$$

$$X_t = F_{(x_{t-1})} + G_{(x_{t-1})}\varepsilon_t \quad (17)$$

y_t , t zamanındaki gözlemi ve x_t gözlemlenmeyen bileşenleri ifade etmektedir.

W, R, F ve G durum vektörünün fonksiyonlarıdır. ε_t ise hata serisini ifade etmektedir.

Bu çalışmada, veri seti ilk dönemlerde çarpımsal özellikler gösterirken, sonlara doğru sabit özelliklere sahip olduğu için, uygun modelin belirlenmesi amacıyla Holt-Winters ve ayrıştırma yöntemlerinde toplamsal ve çarpımsal yöntemler birlikte kullanılarak MAPE, Korelasyon ve RKARE değerlerine göre model seçimi gerçekleştirilmiştir.

2.2 Makine öğrenmesi yöntemleri

Makine öğrenimi, bilgisayar biliminin bir alt alanıdır ve bir yapay zekâ yöntemidir. Bu yöntemin avantajı, bir modelin, açık algoritmalarla temsil edilmesi imkânsız olan sorunları çözebilme yeteneğidir [34].

Makine öğrenmesi teknikleri doğal dil işleme, veri madenciliği, görüntü işleme gibi çok kompleks problemlerin yanı sıra tahmin ve sınıflandırma problemlerinde kullanılmaktadır [35]. Kullanılan yöntemler, matematiksel ve istatistiksel işlemler ile veriler üzerinden çıkarımlar yaparak tahminler yapmaktadır. Literatürde tahmin çalışmalarında rastgele ağaç ve rastgele orman yöntemleri sıklıkla tercih edildiği görülmektedir [36], [37],[38]. Burada tercihin en belirgin sebebi, bu yöntemlerin başarı oranlarının (R ve RKARE) diğer yapay zekâ tabanlı tahmin yöntemlerine göre daha yüksek olmasıdır.

Bu çalışmada da yine rastgele ağaç ve rastgele orman yöntemleri tahmin amaçlı olarak kullanılmıştır.

2.2.1 Rastgele ağaç yöntemi

Rastgele Ağaç Yöntemi, başlangıç noktasından hedef noktasına giderken, çevresinde rastgele noktalar seçerek ve bu noktalara sürekli dallanmalar gerçekleştirerek rastgele yollar belirlemektedir. Süreç, başlangıçtan başlayarak çevresinde belirlenen sınır değerinde rastgele noktalar atayarak ilerler. Sonrasında, nokta seçimlerini en yakın olan düğümü başlangıç noktası düşünerek yine rastgele yani önceki düğümleri dikkate almadan herhangi bir hesap yapmadan belirlenir. Fakat yeni noktalardan, önceki belirlenen noktalara en yakın olanlar birbirine bağlanmaktadır. Düğümler birbirlerine bağlanma sürecinde, bir engele çarpıyorsa o dallanma iptal edilir. Böylece işlemler bulunan en yakın noktadan, hedef noktasına varana kadar sürekli tekrarlanmaktadır. Bağlantılardan belirlenen harita içerisinde çeşitli dallanmalar gerçekleştirilerek yollar meydana getirilmekte ve süreç sonlandırılmaktadır [39].

2.2.2 Rastgele orman yöntemi

Karar ağaçları tıp, mühendislik, eğitim gibi birçok alanda yaygın olarak kullanılmaktadır. Ağaç yapısı halinde karar vermek için oluşturulmuş kuralların, insanların okuyabileceği ve kullanabileceği formatta olması insan karar verme sistemine olan benzerliği ve yöntemin kullanım kolaylığı özellikle tercih nedeni olmaktadır [40].

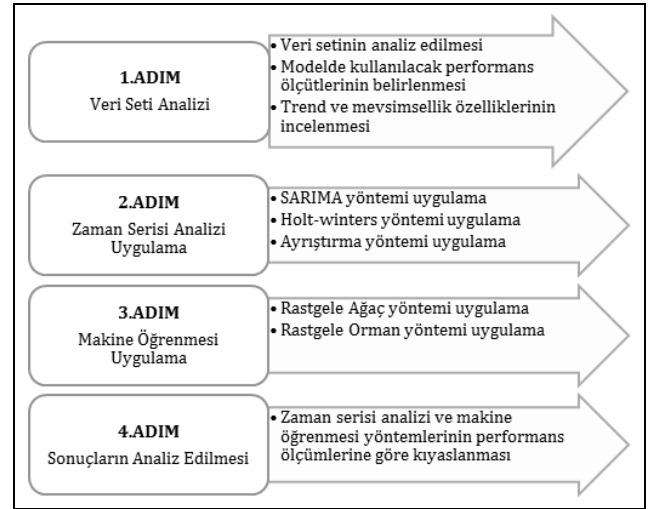
Klasik karar ağaçlarının kavramsal olarak oluşturulması kolay olsa da yüksek varyansa karşı gösterdikleri hassasiyet, sınıflandırmada başarısını olumsuz yönde etkilemektedir. Bu sorunun çözümü için rastgele tabanlı çalışan, her seferinde eğitim setinin farklı bir alt kümesini seçerek karar ağacının

farklı varyantlarını üreten Rastgele Orman algoritması oluşturulmuştur [41].

Rastgele orman algoritması, ağacı oluşturma sırasında kendini test edebildiğinden, yüksek boyutlu parametre uzayında diğer makine öğrenme algoritmalarına göre daha başarılı olduğu kabul edilmektedir [42].

3 Uygulama

Acil servise başvuru sayısının tahmin edilmesinde zaman serisi analizi yöntemleri ve makine öğrenmesi yöntemlerinin kullanıldığı çalışmada, Şekil 1'de verilen uygulama adımları takip edilmiştir.

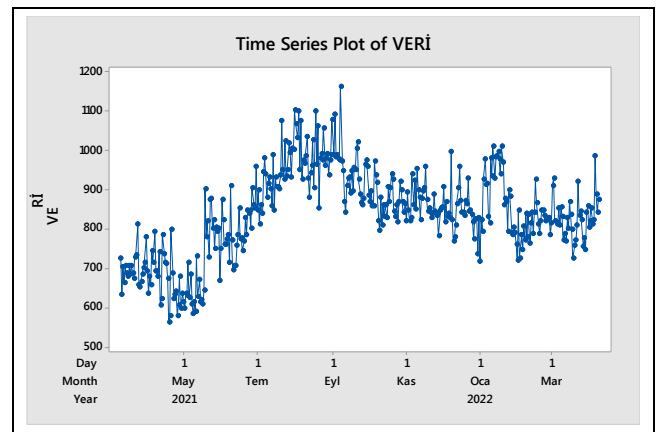


Şekil 1. Uygulama adımları.

Figure 1. Application steps.

3.1 Veri seti analizi

Çalışmada Ankara'da bulunan bir hastanenin 01.03.2021-31.03.2022 tarihleri arasındaki 396 günlük, Acil Servis'e başvuran günlük hasta sayısı veri seti kullanılmıştır. Veri setine uygun tahmin modellerinin belirlenebilmesi için verinin grafiksel dağılımı incelenmiştir. Dağılım sonuçları Şekil 2'de verilmiştir.



Şekil 2. Acil servise başvuru sayısı grafiği.

Figure 2. Graph of the number of applications to the emergency department.

Şekil 2 incelendiğinde, bir yıllık acil servis başvuru sayısında dönemlere göre ani artış ve azalışlar şeklinde dalgalanmalar gerçekleştiği görülmektedir.

Modellerin tahmin başarı ölçümlerinin sayısal ifade edilmesi için korelasyon katsayısı (R), düzeltilmiş R-KARE ve ortalama mutlak yüzde hatası (MAPE) değerleri kullanılmıştır. Kullanılan formüller Denklem 18, 19 ve 20’ de verilmiştir.

$$R = \frac{\sum(Y - \bar{Y}) * (\hat{Y} - \bar{Y})}{\sqrt{\sum(Y - \bar{Y}) * (\hat{Y} - \bar{Y})}} \quad (18)$$

$$R^2 = \left\{ \frac{1}{n} \times \sum[(Y - \bar{Y}) * (\hat{Y} - \bar{Y}) \times (\hat{Y} - \bar{Y})] / \sigma_Y \times \sigma_Y \right\}^2 \quad (19)$$

$$MAPE = \frac{100}{n} \sum \frac{|e_j|}{|A_j|} \quad (20)$$

A_j : j. tahminin gerçek değeri,

e_j : j. verinin gerçek değeri ile tahmin edilen değeri arasındaki mutlak fark değeri.

Literatürde, R ve RKARE değerlerinin, iyi bir model statüsü için 1’ e yakın olması beklenmektedir. MAPE değeri için ise %10’dan düşük olan modeller “çok iyi” model statüsündeyken; %10- %20 arasında olan modeller “iyi”, %20- %50 arasında olan modeller ise “kabul edilebilir”, ve %50’nin üzerinde olan modeller ise “yanlış ve hatalı” olarak kabul edilmektedir [43].

3.2 Zaman serisi analizi yöntemleri uygulama

3.2.1 SARIMA yöntemi uygulama

Zaman serisi analizi yöntemlerinden Box- Jenkins modellerinin kullanabilmek için, öncelikle uygun modelin belirlenmesi gerekmektedir. Uygun model belirlendikten sonra, serinin trend etkisinden ve mevsimsel dalgalanmadan kurtulması gerekmektedir. Mevsimlere göre acil servise başvuru sayısı dalgalanmalar gösterdiği için, bu çalışmada SARIMA yöntemi tercih edilmiştir.

SARIMA yöntemi kullanılarak yapılan tahminde, aşamalar aşağıdaki gibidir:

1. Verilerin durağanlığının analiz edilmesi,
2. Serilerin durağanlaştırılması,
3. SARIMA parametre değerlerinin belirlenmesi,
4. Model performansının, tahmin sonuçlarının ölçülmesi.

Veri setine uygun SARIMA model parametrelerini tespit etmeden önce veri setinin durağan olup, olmadığını tespit edilmesi için birim kök testinden Augmented Dickley Fuller ADF ve ACF ve PACF grafiklerinden yararlanılmıştır. Birim kök testi sonuçları Tablo 2, ACF ve PACF grafiği Şekil 3 ve Şekil 4’te verilmiştir.

Birim kök testi için sıfır hipotezi şu şekilde ifade edilmektedir:

H_0 : Veri seti birim köke sahiptir.

Sıfır hipotezinde belirtilen birim kök, verinin durağanlığına zarar vermektedir ve test sonucunda sıfır hipotezini reddedebilecek değerlere (test istatistiği, p-değeri) ulaşamıyorsa, veri durağan değildir ve entegre edilmelidir sonucuna varılır. Test sonuçları Tablo 2’de verilmiştir.

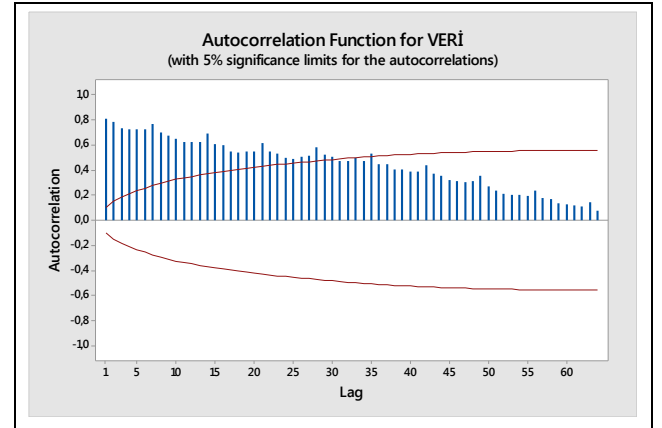
Tablo 2’de verilen birim kök testi sonuçları incelendiğinde, ham verinin t istatistik değerinin -0.7013 ve p değeri 0.4126’ dır. P değeri 0.05’ den büyük olması nedeniyle durağan bir veri yapısına sahip olmadığı görülmüştür. Bu nedenle verinin durağan bir hale dönüştürülmesi gerekmektedir.

Tablo 2. Ham veri birim kök testi sonuçları.

Table 2. Raw data unit root test results.

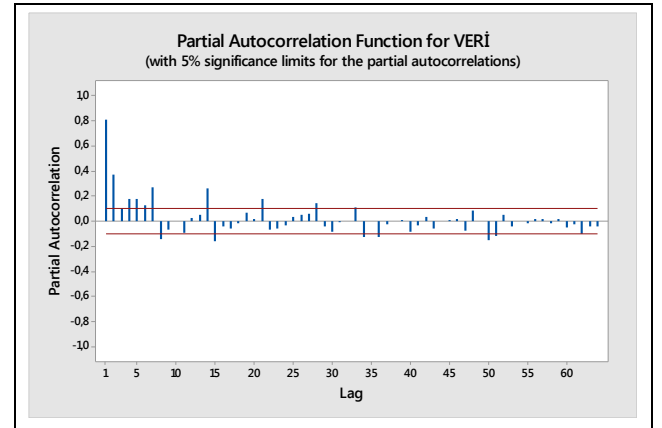
Ham Veri	t- istatistik değeri	p-değeri
ADF Test İstatistiği:	-0.701312	0.4126
Kritik test değerleri:	1% seviye -2.570.830	
	5% seviye -1.941.628	
	10% seviye -1.616.161	

Şekil 3 ve Şekil 4’te verilen ACF ve PACF grafikleri incelendiğinde serinin trende sahip olduğu görülmüştür. Trend etkisinden kurtulmak için 1. Dereceden (d=1) mevsimsel olmayan fark alma işlemi uygulanmıştır. 1 derece fark alınmış serinin ACF ve PACF grafikleri Şekil 5 ve Şekil 6’da verilmiştir.



Şekil 3. Acil servise başvuran sayısı için ACF grafiği.

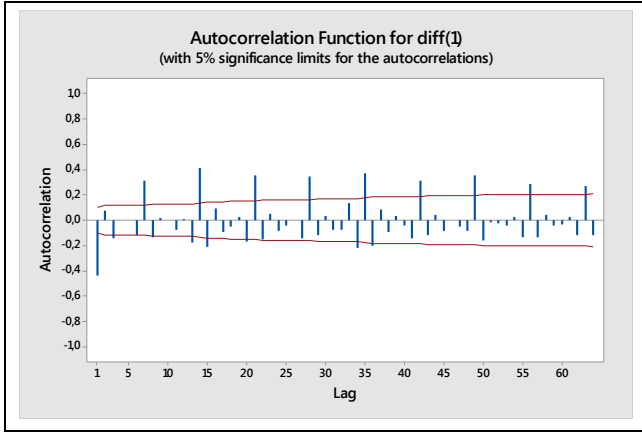
Figure 3. ACF graph for the number of emergency department admissions.



Şekil 4. Acil servise başvuran sayısı için PACF grafiği.

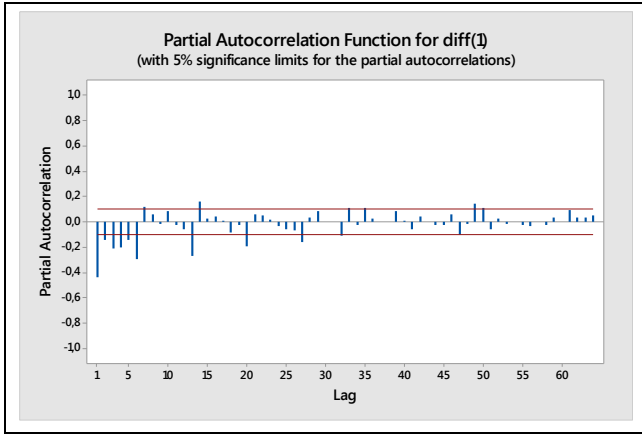
Figure 4. PACF graph for the number of emergency department admissions.

Şekil 5 ve Şekil 6 incelendiğinde, serinin durağan hale geldiği, ancak dalgalanmalar olduğu yani mevsimsellik etkisinin olduğu görülmektedir. Bu nedenle verinin 1 derece (D=1) mevsimsel farkı alınmıştır. Mevsimsel fark alma işleminde ACF grafiğinden yararlanılmış ve periyod 14 (m=14) olarak belirlenmiştir. Fark alma sonucunda, verinin ACF ve PACF grafikleri Şekil 7 ve Şekil 8’de verilmiştir.



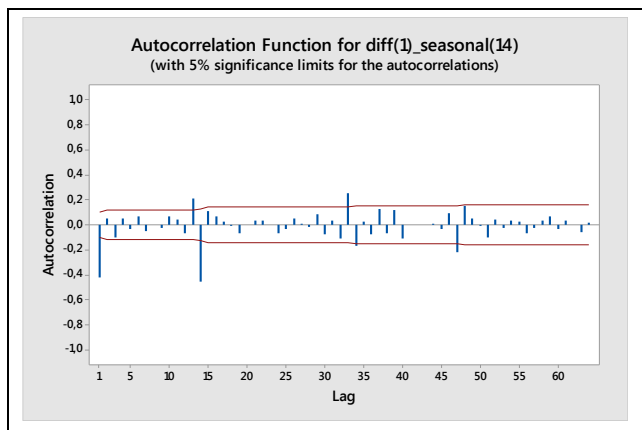
Şekil 5. Acil servise başvuran sayısı için 1 derece farkı alınmış ACF grafiği.

Figure 5. ACF graph with 1 degree difference for the number of emergency department admissions.



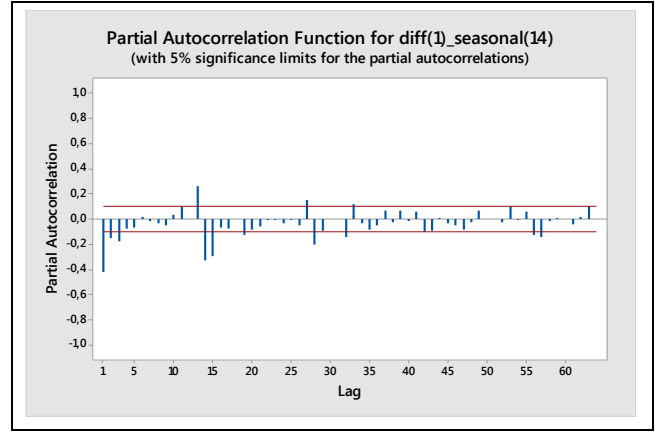
Şekil 6. Acil servise başvuran sayısı için 1 derece farkı alınmış PACF grafiği.

Figure 6. PACF graph with 1 degree difference for the number of emergency department admissions.



Şekil 7. Acil servise başvuran sayısı için 1 derece mevsimsel farkı alınmış ACF grafiği

Figure 7. ACF graph with 1-degree seasonal difference for the number of emergency department admissions.



Şekil 8. Acil servise başvuran sayısı için 1 derece mevsimsel farkı alınmış PACF grafiği.

Figure 8. PACF graph with 1-degree seasonal difference for the number of emergency department admission.

1 derece trend ($d=1$) ve mevsimsel farkı ($D=1$) alınmış veri setinin, ACF (Şekil 7) ve PACF (Şekil 8) grafiklerine göre, trend ve mevsimsel etkilerden kurtularak durağan bir hale geldiği görülmektedir. Verinin durağanlığının sayısal olarak da ispat edilmesi amacıyla, $d=1$ ve $D=1$ farkı alınmış veri setine ADF testi uygulanmıştır. Test sonuçları Tablo 3'te verilmiştir.

Tablo 3. $d=1$ ve $D=1$ farkı alınmış verinin birim kök testi sonuçları.

Table 3. Unit root test results of $d=1$ and $D=1$ differenced data.

$d=1$ ve $D=1$ Veri	t-istatistik değeri	p-değeri
	-5.830.713	0.0000
ADF Test İstatistiği		
Kritik test değerleri	1% seviye	-3.982.393
	5% seviye	-3.421.694
	10% seviye	-3.133.645

Tablo 3'te verilen birim kök testi sonuçları incelendiğinde, $d=1$ ve $D=1$ farkı alınmış verinin t istatistik değerinin -5.830 ve p değeri 0.00'dır. P değeri 0.05'ten küçük olması nedeniyle verinin durağan hale geldiği görülmüştür.

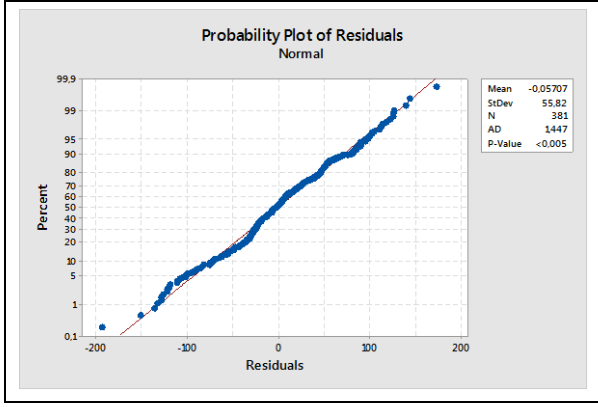
Durağan hale gelen veri setinin diğer parametre değerlerinin (p , q , P , Q) belirlenebilmesi için ön çalışmalar gerçekleştirilmiştir. En iyi tahmin modelinin belirlenebilmesi için, R, RKARE ve MAPE değerleri incelenmiştir. Farklı SARIMA(p , 1, q)(P , 1, Q)₁₄ model parametrelerine göre değerlendirme ölçütü sonuçları Tablo 4'te verilmiştir.

Tablo 4 incelendiğinde, SARIMA (2,1,3)*(1,0,0)₁₄, SARIMA (0,1,1)*(0,1,1)₁₄, SARIMA (0,1,1)*(1,1,0)₁₄, SARIMA (1,1,0)*(0,1,1)₁₄ ve SARIMA (1,1,0)*(1,1,0)₁₄ modelinin R, RKARE ve MAPE değerlerine göre en iyi tahmini yaptığı görülmektedir. Modeller arasında en iyi tahmin modelinin seçilebilmesi için hataların normal dağılım olasılık grafiği (probability plot of residual) incelenmiştir. Her bir model için sırasıyla Şekil 9, Şekil 10, Şekil 11, Şekil 12 ve Şekil 13'te normal dağılım grafikleri verilmiştir. Grafikler, tahmin edilen/uygulanan değerler ile regresyondan gözlemlenen deneysel değerler arasındaki farkı temsil etmektedir. Hataların normal dağılım grafiği, bir veri setinin yaklaşık olarak normal dağılıp dağılmadığını değerlendirmek için kullanılan grafiksel bir araçtır [44].

Tablo 4. SARIMA yöntemi en iyi tahmin modelleri ve parametreleri.

Table 4. SARIMA method best estimation models and parameters.

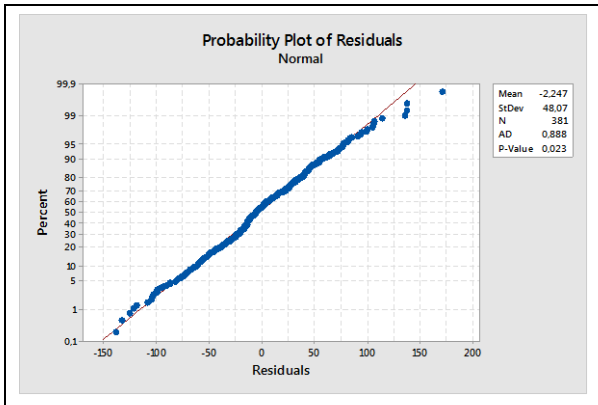
Model	Tahmin Edilen Gün Sayısı	R	RKARE	MAPE
SARIMA (2,1,3)*(1,1,0)₁₄	7	%88.51	%78.34	%4.62
SARIMA (2,1,3)*(1,1,0) ₁₄	15	%53.26	%28.36	%4.98
SARIMA (2,1,3)*(1,1,0) ₁₄	30	%38.92	%15.15	%6.60
SARIMA (0,1,1)*(0,1,1)₁₄	7	%83.93	%70.45	%0.05
SARIMA (0,1,1)*(0,1,1) ₁₄	15	%53.84	%28.99	%4.83
SARIMA (0,1,1)*(0,1,1) ₁₄	30	%66.86	%44.70	%4.43
SARIMA (0,1,1)*(1,1,0)₁₄	7	%86.31	%74.50	%5.01
SARIMA (0,1,1)*(1,1,0) ₁₄	15	%56.07	%31.44	%4.98
SARIMA (0,1,1)*(1,1,0) ₁₄	30	%39.71	%15.77	%6.37
SARIMA (1,1,0)*(0,1,1)₁₄	7	%81.55	%66.51	%4.18
SARIMA (1,1,0)*(0,1,1) ₁₄	15	%53.57	%28.70	%4.99
SARIMA (1,1,0)*(0,1,1) ₁₄	30	%66.60	%44.36	%4.30
SARIMA (1,1,0)*(1,1,0)₁₄	7	%82.92	%68.75	%3.63
SARIMA (1,1,0)*(1,1,0) ₁₄	15	%56.39	%31.80	%5.28
SARIMA (1,1,0)*(1,1,0) ₁₄	30	%40.04	%16.03	%6.53



Şekil 9. SARIMA (2,1,3)*(1,1,0)₁₄ hataların normal dağılım grafiği.

Figure 9. SARIMA (2,1,3)*(1,1,0)₁₄ normal distribution graph of residuals.

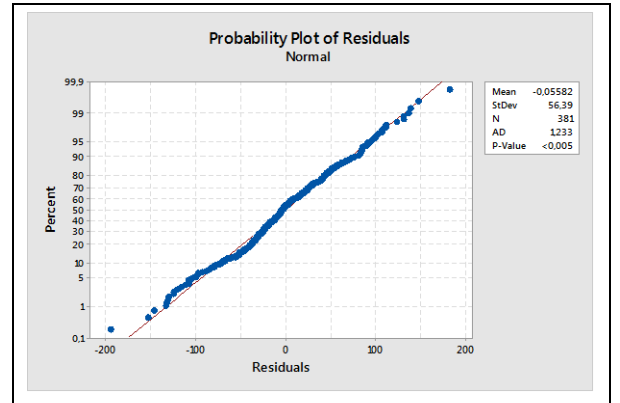
Şekil 9 incelendiğinde, SARIMA (2,1,3)*(1,1,0)₁₄ modelinde hataların normal dağılım grafiğinin, normal dağılıma uymadığı görülmektedir. Bu nedenle ilgili parametre değerlerinin tahmin için uygun olmadığı söylenebilir.



Şekil 10. SARIMA (0,1,1)*(0,1,1)₁₄ hataların normal dağılım grafiği.

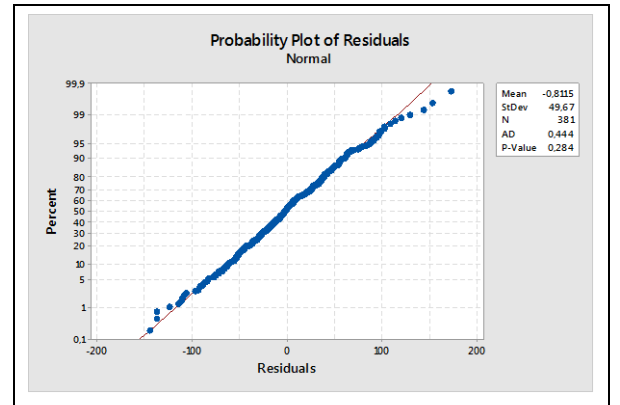
Figure 10. SARIMA (0,1,1)*(0,1,1)₁₄ normal distribution graph of residuals.

Şekil 10 incelendiğinde, SARIMA (0,1,1)*(0,1,1)₁₄ modelinde hataların normal dağılım grafiğinin normal dağılıma uyduğu görülmektedir. Böylece, ilgili parametre değerlerinin tahmin için kullanılması mümkündür.



Şekil 11. SARIMA (0,1,1)*(1,1,0)₁₄ hataların normal dağılım grafiği.

Figure 11. SARIMA (0,1,1)*(1,1,0)₁₄ normal distribution graph of residuals.

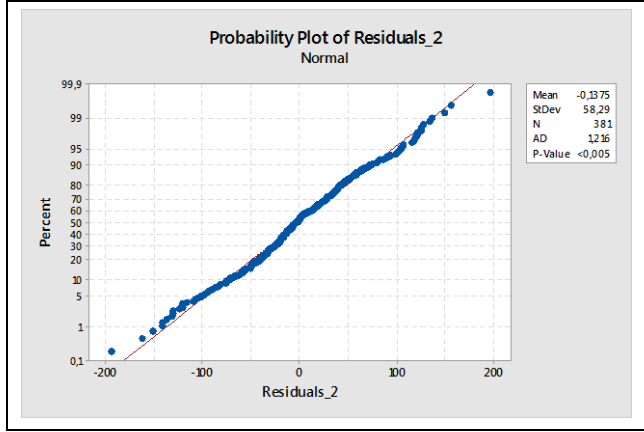


Şekil 12. SARIMA (1,1,0)*(0,1,1)₁₄ hataların normal dağılım grafiği.

Figure 12. SARIMA (1,1,0)*(0,1,1)₁₄ normal distribution graph of residuals.

Şekil 11 incelendiğinde, SARIMA (0,1,1)*(1,1,0)₁₄ modelinde hataların normal dağılım grafiğinin, normal dağılıma uymadığı görülmektedir. Bu nedenle ilgili parametre değerleri tahmin için uygun değildir.

Şekil 12 incelendiğinde, SARIMA (1,1,0)*(0,1,1)₁₄ modelinde hataların normal dağılım grafiğinin, p<0,005 olduğu için normal dağılıma uyduğu görülmektedir. Buradaki ilgili parametre değerleri de tahmin için kullanılabilir.



Şekil 13. SARIMA (1,1,0)*(1,1,0)₁₄ hataların normal dağılım grafiği.

Figure 13. SARIMA (1,1,0)*(1,1,0)₁₄ normal distribution graph of residuals.

Şekil 13 incelendiğinde, SARIMA (1,1,0)*(1,1,0)₁₄ modelinin hataların normal dağılım grafiğinin, p<0,005 olduğu için normal dağılıma uymadığından, söz konusu parametre değerleri tahmin için uygun değildir.

Hataların normal dağılım grafikleri incelendiğinde; SARIMA(0,1,1)((0,1,1)₁₄ ve SARIMA(1,1,0)(0,1,1)₁₄ modellerinin tahmin yapmak için uygun olduğu görülmüştür. İki model arasında R, RKARE ve MAPE değerleri değerlendirildiğinde; sırasıyla modellerin R, RKARE ve MAPE değerleri; (%83.33, %70.45, %0.05), (%81.55, %66.51, %4.18) şeklindedir. SARIMA(0,1,1)((0,1,1)₁₄ modelinin tahmin performansı açısından daha başarılı sonuçlar ürettiği gözlemlenmiştir.

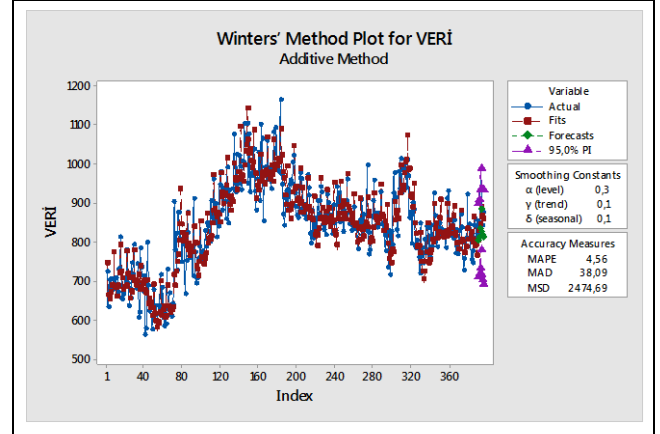
3.2.2 Holt-Winters yöntemi uygulama

Çalışmada, veri seti ilk dönemlerde çarpımsal özellikler gösterirken (Şekil 2), sonlara doğru sabit özelliklere sahip olduğu için, uygun modelin belirlenmesi amacıyla Holt-Winters yönteminde toplamsal ve çarpımsal yöntemler birlikte kullanılarak MAPE, Korelasyon ve RKARE değerlerine göre modeller arasında seçim gerçekleştirilmiştir.

Toplamsal ve çarpımsal modellerde yer alan (α , β , γ) parametrelerinin optimal değerleri hata terimlerini en küçükleyecek şekilde belirlenmiştir. Toplamsal ve çarpımsal modellerde kullanılan parametre değerleri sırasıyla (0.3, 0.1, 0.1) ve (0.7, 0.1, 0.1) şeklindedir. Toplamsal ve çarpımsal modele ait tahmin grafikleri sırasıyla Şekil 14 ve Şekil 15'te verilmiştir.

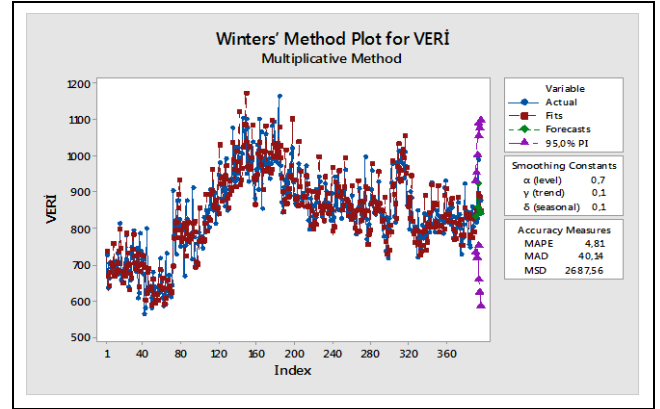
Şekil 14 ve Şekil 15 incelendiğinde; toplamsal ve çarpımsal Holt-Winters tahmin yöntemlerinin MAPE değerlerinin düşük olduğu görülmektedir. İki model arasında en iyi tahmin sonucunu veren modelin belirlenmesi amacıyla; her iki yöntem kullanılarak 7, 15 ve 30 günlük tahminler yapılmıştır. Tahmin

sonuçları R, RKARE ve MAPE değerlerine göre analiz edilmiş ve değerlendirme sonuçları Tablo 5'te verilmiştir.



Şekil 14. Holt-Winters toplamsal model tahmin grafiği.

Figure 14. Holt-Winters additive model prediction plot.



Şekil 15. Holt-Winters çarpımsal model tahmin grafiği.

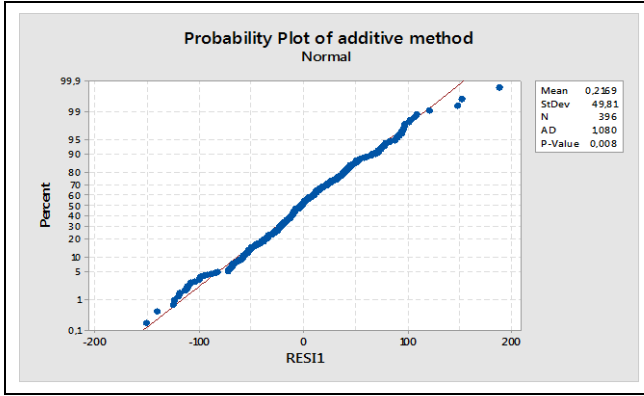
Figure 15. Holt-Winters multiplicative model prediction plot.

Tablo 5. Holt-Winters tahmin yöntemi en iyi tahmin yöntemi ve parametreleri.

Table 5. Holt-Winters estimation method best estimation method and parameters.

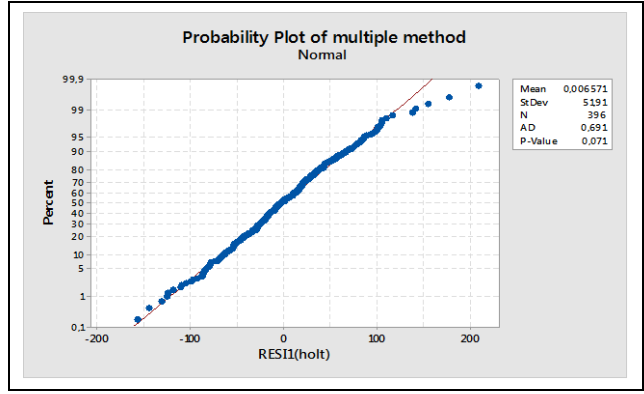
Model	Tahmin Edilen Gün Sayısı	R	RKARE	MAPE
Toplamsal	7	%82.51	%68.07	
	15	%65.66	%43.11	%4.56
	30	%49.11	%24.12	
Çarpımsal	7	%86.66	%75.09	
	15	%75.60	%57.16	%4.81
	30	%51.51	%57.16	

Tablo 5 incelendiğinde, toplamsal ve çarpımsal modelin 7 günlük tahminde daha başarılı olduğu görülmektedir. Modellerin 7 günlük tahmin sonuçları incelendiğinde ise, çarpımsal modelin R ve RKARE değerleri, toplamsal modele göre daha yüksek çıkmıştır. Toplamsal modelin ise MAPE değeri daha düşüktür. Uygun modelin seçilebilmesi için, yöntemlerin hatalarının normal dağılım grafikleri incelenmiştir. Normal dağılım grafikleri Şekil 16 ve Şekil 17'de verilmiştir.



Şekil 16. Holt -Winters toplamsal modeli normal dağılım grafiği.

Figure 16. Holt-Winters additive model normal distribution plot.



Şekil 17. Holt -Winters çarpımsal modeli normal dağılım grafiği.

Figure 17. Holt-Winters multiple model normal distribution plot.

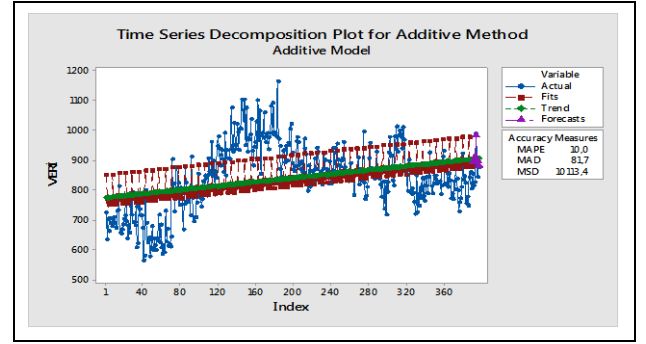
Şekil 16 ve Şekil 17'deki normal dağılım grafikleri incelendiğinde, toplamsal ve çarpımsal Holt-Winters yöntemlerinin p değeri 0.05'ten büyük olduğu için normal dağılıma uymaktadır. Ancak, çarpımsal modelin p değerinin daha yüksek olması ve R ve RKARE değerlerinin toplamsal modele göre daha büyük olması nedeniyle Holt-Winters yöntemlerinden çarpımsal model daha başarılı bir tahmin modeli olarak tercih edilmiştir.

3.2.3 Ayrıştırma yöntemi uygulama

Çalışmada, veri seti ilk dönemlerde çarpımsal özellikler gösterirken (Şekil 2), sonlara doğru sabit özelliklere sahip olduğu için, uygun modelin belirlenmesi amacıyla Holt-Winters yönteminde kullanıldığı gibi, toplamsal ve çarpımsal yöntemler birlikte kullanılarak MAPE, Korelasyon ve RKARE değerlerine göre modeller arasında seçim yapılmıştır.

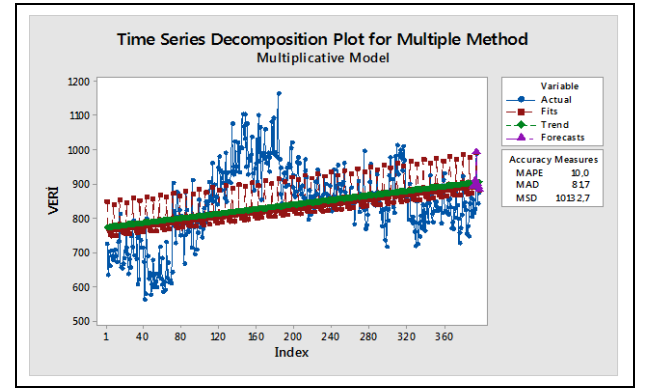
Ayrıştırma modelinde yer alan toplamsal ve çarpımsal modele ait tahmin grafikleri sırasıyla Şekil 18 ve Şekil 19'da verilmiştir.

Şekil 18 ve Şekil 19 incelendiğinde her iki modelin de MAPE değerlerinin yüksek olduğu görülmektedir. İki model arasında hangisinin daha iyi tahmin sonucu verdiğinin belirlenmesi amacıyla; 7, 15 ve 30 günlük tahminler yapılmıştır. Tahmin sonuçları R, RKARE ve MAPE değerlerine göre analiz edilmiş ve değerlendirme sonuçları Tablo 6'da verilmiştir.



Şekil 18. Ayrıştırma toplamsal modeli tahmin grafiği.

Figure 18. Decomposition additive model prediction plot.



Şekil 19. Ayrıştırma çarpımsal modeli tahmin grafiği.

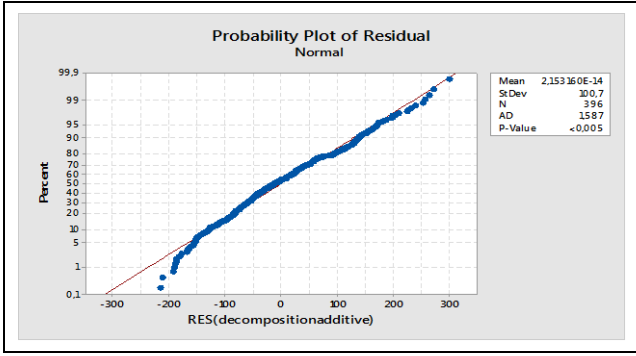
Figure 19. Decomposition multiplicative model prediction plot.

Tablo 6. Ayrıştırma tahmin yöntemi en iyi tahmin yöntemi ve parametreleri.

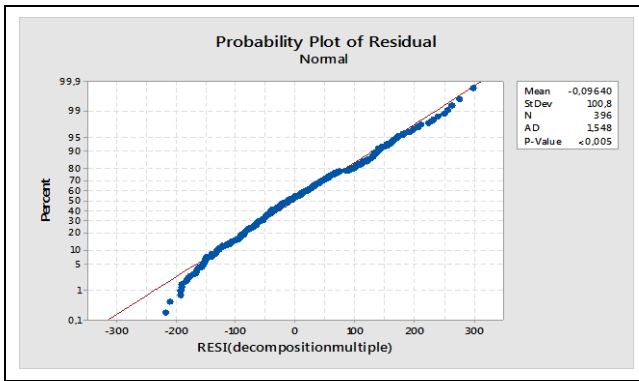
Table 6. Decomposition estimation method best estimation method and parameters.

Model	Tahmin Edilen Gün Sayısı	R	RKARE	MAPE
Toplamsal	7	%82.13	%67.46	
	15	%57.04	%32.53	%10.00
	30	%66.49	%44.21	
Çarpımsal	7	%82.84	%68.62	
	15	%58.37	%34.07	%10.00
	30	%67.67	%45.80	

Tablo 6'dan görüleceği üzere, her iki model de 7 günlük tahminde daha başarılı sonuçlar üretmiştir. Ancak çarpımsal modelin, toplamsal modele göre R ve RKARE değerleri daha yüksektir. Modellerin MAPE değerleri yüksek sonuçlar ürettiği için uygun modelin seçilebilmesi amacıyla, yöntemlerin hatalarının normal dağılım grafikleri incelenmiştir. Normal dağılım grafikleri Şekil 20 ve Şekil 21'de verilmiştir.



Şekil 20. Ayırıştırma toplamsal modeli normal dağılım grafiği.
Figure 20. Decomposition additive model normal distribution plot.



Şekil 21. Ayırıştırma toplamsal modeli normal dağılım grafiği.
Figure 21. Decomposition additive model normal distribution plot.

Şekil 20 ve Şekil 21'deki normal dağılım grafikleri incelendiğinde, toplamsal ve çarpımsal ayırıştırma yöntemlerinin hatalarının normal dağılıma uymadığı görülmektedir. Çünkü her iki modelde de p değeri 0.05'ten oldukça küçüktür. Bu nedenle çalışmada ayırıştırma yöntemleri tercih edilmemiştir.

3.3 Makine Öğrenmesi Yöntemleri Uygulama

Önceki bölümlerde sunulan değerlendirmeler ışığında, acil servise günlük başvuru sayısının makine öğrenmesi yöntemleri ile tahmin edilmesinde rastgele orman ve rastgele ağaç yöntemlerinin kullanılmasına karar verilmiştir. Her bir yöntemle 7, 15 ve 30 günlük tahminler yapılmış; tahmin sonuçlarının analiz edilmesi için, R, RKARE ve MAPE değerleri incelenmiştir. Analiz sonuçları Tablo 7'de verilmiştir.

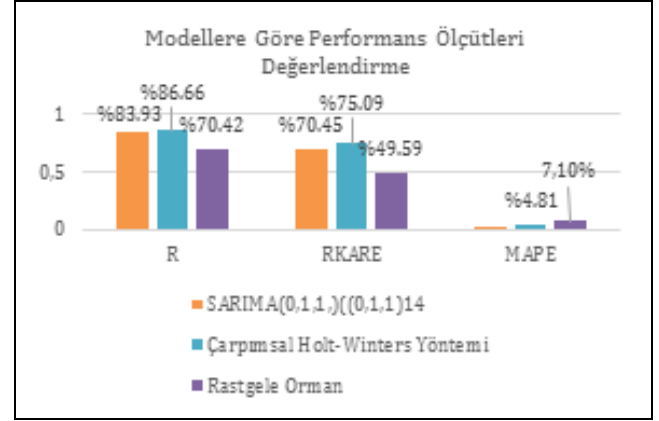
Tablo 7. Makine öğrenmesi yöntemleri tahmin sonuçları.

Table 7. Machine learning methods prediction results.

Yöntemler	Tahmin Edilen Gün	R	RKARE	MAPE
	Sayı			
Rastgele Orman	7	%70.42	%49.59	%7.1
	15	%56.82	%32.2	%4.74
	30	%70.18	%49.26	%4.71
Rastgele Ağaç	7	%56.98	%32.47	%7.83
	15	%38.76	%15.02	%5.06
	30	%15.60	%2.43	%5.37

Tablo 7 incelendiğinde, rastgele orman yöntemi, rastgele ağaç yöntemine göre R, RKARE ve MAPE değerlerine göre daha başarılı olduğu görülmektedir.

Yapılan analizler ile, zaman serisi analizi ve makine öğrenmesi yöntemlerinden başarılı model olarak seçilen yöntemlere yönelik performans ölçütü sonuçları Şekil 22'de verilmiştir.



Şekil 22. SARIMA ve Lineer Regresyon performans ölçütü sonuçları.

Figure 22. SARIMA and Linear Regression performance criteria results.

Şekil 22'de verilen, SARIMA (0,1,1) * (0,1,1)₁₄ modeli, Holt-Winters çarpımsal ve rastgele orman yöntemleri karşılaştırıldığında, yedi günlük tahmin sonucunda bütün modeller, MAPE değerleri %10'un altında olduğu için, çok iyi model statüsündedir. Ancak SARIMA (0,1,1) * (0,1,1)₁₄ yönteminin MAPE değeri, yaklaşık olarak 0'a (Tablo 4) eşittir. Bu durum yöntemin tahmin performansı açısından oldukça başarılı olduğunu göstermektedir.

Yöntemler, R ve RKARE değerleri açısından değerlendirildiğinde ise; çarpımsal Holt-Winters yönteminin diğer yöntemlere göre bire daha yakın sonuçlar ürettiği görülmektedir. Ancak, model tahmin sonuçlarında MAPE değerinin sıfıra yakın çıkması ve model hata performansının oldukça düşük olması nedeniyle SARIMA yöntemi en başarılı yöntem olarak tercih edilmiştir.

4 Sonuçlar ve tartışma

Acil servise yapılan başvuru sayısının, zaman serisi analizi ve makine öğrenmesi yöntemleri ile tahmin edilmesi, hastane kaynak planlamasının etkili bir şekilde gerçekleştirilmesi açısından önem taşımaktadır. Sunulan bu çalışma kapsamında, zaman serisi analizi yöntemlerinden SARIMA, Holt-Winters ve ayırıştırma; makine öğrenmesi yöntemlerinden rastgele ağaç ve rastgele orman yöntemleri kullanılarak acil servise yapılacak günlük başvuru sayısı tahmin edilmiştir. Tahmin sonuçları model değerlendirme kriterlerine göre incelenmiş ve SARIMA yönteminin söz konusu diğer yöntemlere göre daha yüksek bir başarı oranı elde ettiği görülmüştür.

Acil servise başvuru sayısının tahmin edilmesine yönelik literatür araştırması yapıldığında bu çalışmanın, tahmin edilen gün sayısındaki değişkenlik, salgın dönemi veri setiyle çalışılması ve zaman serisi analizi yöntemlerinden SARIMA, Holt-Winters ve ayırıştırma yöntemlerini birlikte kullanarak kapsamlı bir analiz gerçekleştirilmesi nedeniyle, konuyla ilgili sınırlı sayıda çalışma arasında yer aldığı görülmektedir. Bu durum, acil servise yapılan başvuru sayısının tahmin

edilmesinde farklı yöntemlerin kullanılması ve gerçeğe yakın değerler elde edilmesinde açısından kritik önem taşımaktadır. Uygulama sonuçlarına göre zaman serisi analiz yöntemlerinin tahminde etkili bir araç olacağı düşünülmektedir. Aynı zamanda yapılan literatür araştırmasına göre bu çalışma; COVID-19 salgın dönemi acil servise günlük başvuru sayısı veri setini kullanması itibarıyla de az sayıdaki çalışma arasında yer alarak literatüre katkı sağlamaktadır. Çünkü salgın dönemi veri seti, yapısı itibarıyla tahmin edilmesi zor bir veri setidir ve bu etki, tahmin performansını olumsuz yönde etkilemektedir. Buna rağmen, SARIMA yöntemi ile sırasıyla korelasyon, RKARE ve MAPE değerleri olarak %83.93, %70.45 ve %0.005 değerine ulaşılmıştır. R ve RKARE değerlerinin bire yakın değerlere ulaşması ve MAPE değerinin yaklaşık sıfır olması model performansının başarısını açıkça göstermektedir.

İlerleyen çalışmalarda, acil servise başvuru sayısının tahmin edilmesinde birden fazla yılın verisini içeren veri setleri ile çalışarak SARIMA yönteminin performans değerlerinde artış sağlanabileceği öngörülmektedir. Bununla birlikte, derin öğrenme gibi farklı çözüm yöntemlerini kullanarak analiz çıktılarının verimliliğinin artırılması mümkündür. Ayrıca, uzun vadeli tahminler yapılarak hastanelerin, periyodik kapasite planlamalarına katkı sağlanabileceği gibi, başvuru yapacak hastaların hastalık tipinin de göze alındığı tahmin modellerinin geliştirilmesi de kapsamlı bir kapasite planlaması sağlayacaktır.

Yine gelecekte yapılacak çalışmalarda, Holt-Winters yönteminde yer alan parametre değerlerinin optimum şekilde belirlenmesi için farklı algoritmalar (genetik algoritma vb.) kullanılması da mümkündür. Böylelikle model tahmin performans ölçüm değerlerinde iyileşmeler olabileceği öngörülmektedir. Bunlara ek olarak, çalışmada kullanılan makine öğrenmesi yöntemleri zaman serisi analizi yöntemlerine göre düşük başarı göstermiştir. Ancak makine öğrenmesi yöntemlerinin büyük veri setleri üzerindeki performansları yüksektir. Bu nedenle bir yıllık veri seti yerine daha büyük veri setleri kullanıldığında, makine öğrenmesi yöntemlerinin performans ölçütlerinin, zaman serisi analizi yöntemlerine göre yeniden değerlendirilmesi, yeni bir çalışma konusu olacaktır.

5 Conclusion and discussion

Estimating the number of applications made to the emergency service using time series analysis and machine learning methods is important for effective hospital resource planning. Within the scope of this study, the number of daily admissions to the emergency department was estimated using SARIMA, Holt-Winters and decomposition methods from time series analysis methods, random tree, and random forest methods from machine learning methods. The prediction results have been examined according to the model evaluation criteria and it has been seen that the SARIMA method achieved a higher success rate compared to the other methods in question.

A literature search has been conducted to estimate the number of admissions to the emergency department. According to the results of the research, it is seen that this study is among a limited number of studies on the subject due to the variability in the estimated number of days, working with the epidemic period data set, and performing a comprehensive analysis by using the time series analysis methods SARIMA, Holt-Winters, and decomposition methods. This situation is of critical importance in terms of using different methods to estimate the number of applications to the emergency department and

obtaining values close to reality. According to the results of the application, it is thought that time series analysis methods will be an effective tool in forecasting. According to the literature survey conducted, this study is among the few studies that use the data set of the number of daily admissions to the emergency department during the COVID-19 pandemic period. Because the epidemic period data set is a dataset that is difficult to predict due to its structure. This effect negatively affects the forecast performance. Despite this, the correlation, RCARE and MAPE values are 83.93%, 70.45% and 0.005%, respectively, by the SARIMA method. The fact that the R and RKARE values reach values close to one and the MAPE value is approximately zero clearly shows the success of the model performance.

In future studies, it is expected that an increase in the performance values of the SARIMA method can be achieved by working with data sets containing data from more than one year in estimating the number of applications to the emergency department. However, it is possible to increase the efficiency of analysis outputs by using different solution methods such as deep learning. In addition, by making long-term forecasts, hospitals can contribute to periodic capacity planning, as well as the development of forecasting models that consider the type of disease of patients who will apply, will also provide comprehensive capacity planning.

Again, in future studies, different algorithms (genetic algorithm, etc.) will be used to optimally determine the parameter values contained in the Holt-Winters method.) Thus, it is expected that there may be improvements in model prediction performance measurement values. In addition to these, the machine learning methods used in the study showed low success compared to the time series analysis methods. However, the performance of machine learning methods on large data sets is high. For this reason, when larger data sets are used instead of a one-year data set, re-evaluation of the performance criteria of machine learning methods according to time series analysis methods will be a new study.

6 Yazar katkı beyanı

Gerçekleştirilen çalışmada Sema ÇİFTÇİ, fikrin oluşturulması, literatür taraması, uygulama ve sonuçların analiz edilmesi, Gül Didem BATUR SİR, yazım ve eleştirel inceleme, içerik bakımından makalenin kontrol edilmesi konularında katkı sunmuştur.

7 Etik kurul onayı ve çıkar çatışması beyanı

Hazırlanan makalede etik kurul izni alınmasına gerek yoktur.

Hazırlanan makalede herhangi bir kişi/kurum ile çıkar çatışması bulunmamaktadır.

8 Kaynaklar

- [1] Afilal M, Yalaoui F, Dugardin F, Amodeo L, Laplanche D, Blua P. "Forecasting the emergency department patients flow". *Journal of Medical Systems*, 40(7), 1-18, 2016.
- [2] Harrou F, Dairi A, Kadri F, Sun Y. "Effective forecasting of key features in hospital emergency department: Hybrid deep learning-driven methods". *Machine Learning with Applications*, 7, 1-13, 2022.
- [3] Ramgopal S, Pelletier JH, Rakkar J, Horvat CM. "Forecast modeling to identify changes in pediatric emergency department utilization during the COVID-19 pandemic". *The American Journal of Emergency Medicine*, 49, 142-147, 2021.

- [4] Becerra M, Jerez A, Aballay B, Garcés HO, Fuentes A. "Forecasting emergency admissions due to respiratory diseases in high variability scenarios using time series: A case study in Chile". *Science of the Total Environment*, 706, 1-11, 2020.
- [5] Bekker R, Broek M, Koole G. "Modeling COVID-19 hospital admissions and occupancy in the Netherlands". *European Journal of Operational Research*, 304(1), 207-218, 2023.
- [6] Dam V, Zelis P, Kuijk S, Linkens A, Brüggeman R, Spaetgens B, Horst I, Stassen P. "Performance of prediction models for short-term outcome in COVID-19 patients in the emergency department: a retrospective study". *Annals of Medicine*, 53(1), 402-409, 2021.
- [7] Ozudogru AG, Gorener A. "Method selection for demand forecasting: Application in a private hospital". *International Journal of Decision Sciences & Applications*, 1(1), 13-22, 2020.
- [8] Jones SA, Joy MP, Pearson J. "Forecasting demand of emergency care". *Health Care Management Science* 2002, 5(4), 297-305, 2002.
- [9] Noel G, Bonte N, Persico N, Bar C, Luigi S, Roch, A, Viudesa G. "Real-time estimation of inpatient beds required in emergency departments". *European Journal of Emergency Medicine*, 26(6), 440-445, 2019.
- [10] McCoy TH, Pellegrini AM, Perlis RH. "Assessment of time-series machine learning methods for forecasting hospital discharge volume". *JAMA Network Open*, 1(7), 1-9, 2018.
- [11] Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. "Forecasting daily patient volumes in the emergency department". *Academic Emergency Medicine*, 15(2), 159-170, 2008.
- [12] Jones SS, Evans RS, Allen TL, Thomas A, Haug PJ, Welch SJ, Snow GL. "A multivariate time series approach to modeling and forecasting demand in the emergency department". *Journal of Biomedical Informatics*, 42(1), 123-139, 2009.
- [13] Luo W, Cao J, Gallagher M, Wiles J. "Estimating the intensity of ward admission and its effect on emergency department access block". *Statistics in medicine*, 32(15), 2681-2694, 2013.
- [14] Bergs J, Heerinckx P, Verelst S, "Knowing what to expect, forecasting monthly emergency department visits: a time-series analysis". *International Emergency Nursing*, 22(2), 112-115, 2014.
- [15] Peck JS, Benneyan JC, Nightingale DJ, Gaehde SA. "Predicting emergency department inpatient admissions to improve same-day patient flow". *Academic Emergency Medicine*, 19(9), 1045-1054, 2012.
- [16] Barrett TW, Martin AR, Storrow AB, Jenkins CA, Harrell Jr FE, Russ S, Darbar D. "A clinical prediction model to estimate risk for 30-day adverse events in emergency department patients with symptomatic atrial fibrillation". *Annals of Emergency Medicine*, 57(1), 1-12, 2011.
- [17] Sariyer G. "Acil servislerde talebin zaman serileri modelleri ile tahmin edilmesi". *Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi*, 10(1), 66-77, 2017.
- [18] Carvalho-Silva M, Monteiro MTT, de Sá-Soares F, Dória-Nóbrega S. "Assessment of forecasting models for patients arrival at Emergency Department". *Operations Research for Health Care*, 18, 112-118, 2018.
- [19] Singhal S, Allen MW, McAnnally JR, Smith KS, Donnelly JP, Wang HE. "National estimates of emergency department visits for pediatric severe sepsis in the United States". *PeerJ*, 1, 1-12, 2013.
- [20] Golmohammadi D. "Predicting hospital admissions to reduce emergency department boarding". *International Journal of Production Economics*, 182, 535-544, 2016.
- [21] Esen H, Kaya Ü. "Bir eğitim araştırma hastanesi acil servis birimine başvuran hasta sayısı tahmini". *Verimlilik Dergisi*, 3(3), 129-145, 2021.
- [22] Lucini FR, Fogliatto FS, da Silveira GJ, Neyeloff JL, Anzanello MJ, Kuchenbecker RS, Schaan, BD. "Text mining approach to predict hospital admissions using early medical records from the emergency department". *International journal of medical informatics*, 100, 1-8, 2017.
- [23] Singer AJ, Thode Jr HC, Peacock IV WF. "Admission rates for emergency department patients with venous thromboembolism and estimation of the proportion of low risk pulmonary embolism patients: a US perspective". *Clinical and experimental emergency medicine*, 3(3), 126-131, 2016.
- [24] Webb BJ, Levin NM, Grisel N, Brown SM, Peltan ID, Spivak ES, Bledsoe J. "Simple scoring tool to estimate risk of hospitalization and mortality in ambulatory and emergency department patients with COVID-19". *Plos one*, 17(3), 1-12, 2022.
- [25] Box G, Jenkins G, Reinsel G, Ljung G. *Thrid Edition Time Series Analysis Forecasting and Control*. New Jersey, Kanada, John Wiley & Sons, 1995.
- [26] Toğa G, Atalay B, Toksari MD. "COVID-19 prevalence forecasting using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN): Case of Turkey". *Journal Infect Public Health*, 14(7), 811-816, 2021.
- [27] Dagum EB, Giannerini S. "A critical investigation on detrending procedures for non-linear processes". *Journal of Macroeconomics*, 28(1), 175-191, 2006.
- [28] Liu J, Zhao Z, Zhong Y, Zhao C, Zhang G. "Prediction of the dissolved gas concentration in power transformer oil based on SARIMA model". *Energy Reports*, 8(5), 1360-1367, 2022.
- [29] Winters PR. "Forecasting Sales by Exponentially Weighted Moving Averages". *Management science*, 6(3), 324-342, 1960.
- [30] Holt CC. "Forecasting seasonals and trends by exponentially weighted moving averages". *International journal of forecasting*, 20(1), 5-10, 2004.
- [31] Omar MS, Kawamukai H. "Prediction of NDVI using the Holt-Winters model in high and low vegetation regions: A case study of East Africa". *Scientific African*, 14, 1-10, 2021.
- [32] Demir İ, Genç T, Karaboğa HA. "Türkiye Cumhuriyet Merkez Bankası altın rezervinin holt-winters üstel düzleme yöntemi ve yapay sinir ağları ile incelenmesi". *Uluslararası Ekonomi, İşletme ve Politika Dergisi*, 2(1), 131-146, 2018.
- [33] Yang D, Sharma V, Ye ZI, Lim L, Zhao L, Aryaputera AW. "Forecasting of global horizontal irradiance by exponential smoothing, using decompositions". *Energy*, 81, 111-119, 2015.
- [34] C. Voyant, Notton G, Kalogirou S, Nivet ML, Paoli C, Motte Fabrice, Fouilloy A. "Machine learning methods for solar radiation forecasting: A review". *Renewable Energy*, 105, 569-582, 2017.

- [35] Namlı E, Ünlü R, Gül E. "Fiyat tahminlemede makine öğrenmesi teknikleri ve doğrusal regresyon yöntemlerinin kıyaslanması; Türkiye'de satılan ikinci el araç fiyatlarının tahminlenmesine yönelik bir vaka çalışması". *Konya Mühendislik Bilimleri Dergisi*, 7(4), 806-821, 2019.
- [36] Januschowski T, Wang Y, Torkkola K, Erkkilä T, Hasson H, Gasthaus J. "Forecasting with trees". *International Journal of Forecasting*, 38(4), 1473-1481, 2022.
- [37] Gradojevic N, Kukolj D, Adcock R, Djakovic V. "Forecasting Bitcoin with technical analysis: A not-so-random forest?". *International Journal of Forecasting*, 39(1), 1-17, 2021.
- [38] Basher SA, Sadorsky P. "Forecasting Bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility?". *Machine Learning with Applications*, 9, 1-19, 2022.
- [39] Torun Y, Ergül Z, Aksöz A, "Optimum enerji verimliliğini hedefleyen rastgele ağaçlar ve yapay arı kolonisi yöntemi ile otonom robotlarda yol planlama algoritması". *Gazi University Journal of Science Part C: Design and Technology*, 7(4), 903-915, 2019.
- [40] Ali J, Khan R, Ahmad N, Maqsood I. "Random forests and decision trees". *International Journal of Computer Science Issues*, 9(5), 272-278, 2012.
- [41] Breiman L. "Random forests". *Machine Learning 2001*, 45(1), 5-32, 2001.
- [42] Caruana R, Niculescu-Mizil A. "An empirical comparison of supervised learning algorithms". *Proceedings of the 23rd International Conference on Machine Learning*, Pennsylvania, USA, 25 June 2006.
- [43] Lewis CD. "Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting". Boston, London, Butterworth Scientific, 1982.
- [44] Hasan SH, Srivastava P, Talat M. "Biosorption of Pb(II) from water using biomass of *Aeromonas hydrophila*: Central composite design for optimization of process variables". *Journal of Hazardous Materials*, 168(2-3), 1155-1162, 2009.