



Turizm Sektöründe Doluluk Oranları Tahminlerinin Makine Öğrenmesi Ve Zaman Serisi Analizleri ile İncelenmesi

Ayşe YİĞİTTEKİN^{1*}, Tuba KOÇ², Pelin AKIN³

¹Çankırı Karatekin Üniversitesi, Fen Bilimleri Fakültesi, İstatistik Bölümü, Çankırı

²Çankırı Karatekin Üniversitesi, Fen Bilimleri Fakültesi, İstatistik Bölümü, Çankırı

³Çankırı Karatekin Üniversitesi, Fen Bilimleri Fakültesi, İstatistik Bölümü, Çankırı

Özet

Bu çalışmada, Türkiye'deki konaklama işletmelerinin doluluk oranlarını tahmin etmek amacıyla zaman serisi analiz yöntemleri ile makine öğrenmesi tekniklerinden Aşırı Öğrenme Makinesi (ELM) ve Çok Katmanlı Algılayıcı (MLP) yöntemleri kullanılmıştır. Çalışmada 2000-2024 yılları arasındaki konaklama işletmelerine ait veriler analiz edilmiş olup, farklı eğitim-test veri oranları (%80-%20, %70-%30 ve %60-%40) ile modellerin performansları değerlendirilmiştir. Performans ölçütleri olarak Ortalama Mutlak Hata (MAE) ve Karekök Ortalama Hata (RMSE) kullanılmıştır. Sonuç olarak, veri ayırma oranına bağlı olarak yerli turist doluluk oranlarında ELM, yabancı turist doluluk oranlarında MLP ve toplam doluluk oranlarında SARIMA modeli en iyi tahmin performansını göstermiştir. Bu bulgular, konaklama sektöründe daha doğru kapasite planlamasına ve stratejik karar alma süreçlerinin iyileştirilmesine önemli katkılar sağlayabilir.

Makale Bilgisi

Başvuru:
19/03/2025
Kabul:
12/06/2025

Anahtar Kelimeler: Turizm Doluluk Oranı, ELM, MLP, Zaman Serisi Analizi, SARIMA, Turizm Talebi

Analysis of Occupancy Rate Forecasts in the Tourism Sector Using Machine Learning and Time Series Methods

In this study, Extreme Learning Machine (ELM) and Multilayer Perceptron (MLP) methods from time series analysis methods and machine learning techniques are used to predict the occupancy rates of accommodation establishments in Turkey. In the study, the data of accommodation establishments between 2000-2024 were analysed and the performances of the models were evaluated with different training-test data ratios (80%-20%, 70%-30% and 60%-40%). Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were used as performance measures. As a result, depending on the data separation rate, ELM model showed the best forecasting performance for domestic tourist occupancy rates, MLP model for foreign tourist occupancy rates and SARIMA model for total occupancy rates. These findings can make significant contributions to more accurate capacity planning and improvement of strategic decision-making processes in the hospitality industry.

Keywords: Tourism Occupancy Rate, ELM, MLP, Time Series Analysis, SARIMA, Tourism Demand

* İletişim e-posta: ayseyigittekin06@gmail.com

1 Giriş

Günümüzde en önemli sektörlerden biri olan turizmin, gelecek yüzyıllarda da etkisini sürdürmesi beklenmektedir. Turizm, insanların ticari/mesleki veya kişisel ihtiyaçları amaçları doğrultusunda çevre ülkelere veya yerlere gitmesini gerektiren sosyal, kültürel ve ekonomik bir olgudur [1]. Turizm, ulusal ekonomiye döviz girdisi ve istihdam sağlayan dinamik bir sektör olmanın yanı sıra, yerel düzeyde ekonomik kalkınmayı destekleyerek sosyal eşitsizlikleri azaltmada da önemli bir rol oynar. Hem gelişmiş hem de gelişmekte olan ülkeler için oldukça kritik bir sektördür [2]. Öyle ki; Türkiye ekonomisinde turizm sektörü, 2022 yılında GSYH'nin %4,4'ünü, 2023 yılında ise %5'ini oluşturmuştur. Sektör, 2022'de işgücünün %2,9'unu temsil ederken (885,000 kişi), yurt dışından gelen turist sayısı 51,4 milyona, turizm geliri ise 49,8 milyar ABD dolarına ulaşmıştır. 2023 yılında, turist sayısı 56,7 milyona, turizm geliri ise 55,9 milyar ABD dolarına yükselmiştir. Bu dönemde turizm geliri %12, yabancı ziyaretçi sayısı ise %9,32 oranında artış göstermiştir [3]. Bu nedenle, Türkiye'de turizm olanaklarının yaygınlığı, sektörün gelişimi ve yatırımlar açısından büyük önem taşımaktadır. Bu kadar dinamik bir yapıya sahip olan turizm sektöründen maksimum fayda sağlamak için kritik noktalara dikkat edilmelidir. Bunlardan biri, öngörülen tahminlerin doğru bir şekilde gerçekleşmesidir. Gelecekte meydana gelebilecek gelişmelerin önceden belirlenmesi, konaklama işletmelerinin kapasite planlamalarını taleplere göre yapması, doluluk oranlarını hesaplaması, yeni yatırımları hayata geçirmesi ve bu yatırımların risk yönetimini sağlaması, turizm sektörünün stratejik açıdan en önemli unsurlarından biridir [4]. Taleplerin esnek olması ve turizm sektöründe hizmetlerin stoklanamaması gibi faktörler, talep tahminleme modellerinin önemini artırmakta ve yöneticilerin doğru kararlar alarak kaynaklarını verimli kullanmasına yardımcı olmaktadır [2]. Konaklama işletmelerinin talepleri zaman içerisinde değişmektedir. Talepler günlük, haftalık, aylık ve mevsimlik değişim gösterir. Bu talepler aynı zamanda sosyal, siyasi ve ekonomik değişkenlerle de değişkenlik göstermektedir [5]. Konaklama işletmelerinde doluluk oranı tahmini, planlama ve karar alma süreçlerinde kritik bir rol oynar. Talebin doğru öngörülmesi, yöneticilere etkili planlamalar yapma fırsatı sunar. Bu süreçler, geliri en üst seviyeye çıkarmak ve maliyetleri en aza indirmek amacıyla gerçekleştirilmektedir [6].

Konaklama işletmelerinde doluluk oranları iki şekilde hesaplanır: oda doluluk oranı, yatak doluluk oranı. Yatak doluluk oranı, satılan yatak sayısının satışa hazır yatak sayısına bölünmesiyle; oda doluluk oranı ise, aynı yatak doluluk oranı gibi satılan oda sayısının satışa hazır oda sayısına bölünmesiyle bulunmaktadır. Her iki oran da yüzde olarak ifade edilmektedir. [5].

Turizm talebi tahmini ve doluluk oranları üzerine literatürde çeşitli çalışmalar bulunmaktadır. Burger vd. [7], Güney Afrika Cumhuriyeti'nin Durban kentine gelen Amerikalı turist sayısını tahmin etmek için hareketli ortalamalar, üstel düzleştirme, ARIMA, genetik regresyon ve YSA yöntemlerini uygulamışlardır. Elde edilen sonuçlar gerçek değerlerle karşılaştırıldığında, en iyi performansı YSA modelinin gösterdiği görülmüştür. Fernandes vd. [8], Portekiz'in kuzey ve güney bölgelerindeki konaklama işletmelerinin doluluk oranlarını tahmin etmek için zaman serisi yöntemleri ve YSA modelini kullanmışlardır. Box-Jenkins yöntemiyle elde edilen tahminlerle karşılaştırıldığında, YSA modelinin daha uygun tahminler ürettiği sonucuna varmışlardır. Çuhadar vd. [9], Antalya iline yönelik dış turizm talebini tahmin etmek için Üstel Düzleştirme, Box-Jenkins ve YSA modellerini kullanmış ve bu yöntemlerin tahminlerini karşılaştırmışlardır. Sonuçlar, YSA modelinin diğer yöntemlere kıyasla daha yüksek doğruluk sağladığını göstermiştir. Önder ve Hasgül [10], Türkiye genelinde gelen yabancı turist sayısını tahmin etmek için Üstel Düzleştirme, Box-Jenkins ve YSA yöntemlerini kullanmışlardır. Sonuçlar, YSA modelinin geleneksel yöntemlere alternatif olarak kullanılabileceğini göstermiştir. Çuhadar [11], Türkiye'ye yönelik aylık dış turizm talebini farklı yapay sinir ağı modelleriyle tahmin etmiş ve en yüksek doğruluğu sağlayan MLP modeli ile 2013 yılı tahminlerini üretmiştir. Caicedo-Torres ve Payares [6], dört farklı makine öğrenmesi tekniği kullanarak konaklama sektöründeki doluluk oranlarını tahmin etmeye çalışmıştır. Sonuçlar, bu tekniklerden biri olan Ridge Regresyon modelinin diğer modellere kıyasla daha iyi performans gösterdiğini ortaya koymuştur. Zorlutuna ve Bircan [12], turizm talebi tahmininde zaman serisi analiz yöntemleri (ARMA, ARIMA) ile yapay sinir ağlarından MLP yöntemini karşılaştırmış ve MLP yönteminin daha yüksek öngörü doğruluğu sağladığını belirlemiştir. Güler [13], Türkiye'de turizm gelirini etkileyen faktörleri makine öğrenmesi yöntemleriyle belirlemiş ve bu yöntemler arasında en yüksek öngörü doğruluğunu

MLP yapay sinir ağlarının sağladığını göstermişlerdir.

Bu çalışmanın amacı, Türkiye’de bulunan konaklama işletmelerinin doluluk oranlarını tahmin etmek için zaman serisi analiz yöntemleri ile birlikte makine öğrenmesi tekniklerinden Aşırı Öğrenme Makinesi (ELM) ve Çok Katmanlı Algılayıcı (MLP) yöntemlerini kullanmak ve bu belirtilen yöntemlerin tahmin performanslarını karşılaştırmalı olarak değerlendirmektir. Zaman serisine dayalı olan geleneksel yöntemlerle (ARIMA, SARIMA) modern makine öğrenmesi yaklaşımlarının (ELM, MLP) tahmin gücü farklı eğitim- test oranlarında karşılaştırılmış; böylece yöntemlerin üstün ve zayıf olan yönlerini ortaya koymayı başarmıştır. Bu karşılaştırma, turizm sektöründe stratejik karar alma süreçlerine daha etkili modeller önererek öne çıkarmayı hedeflemektedir.

Bu çalışma dört ana bölümden oluşmaktadır: Bölüm 2’de MLP, ELM ve ARIMA modelleri detaylı olarak açıklanmıştır. Bölüm 3’te algoritmalarından elde edilen sonuçlar sunulmuştur. Son olarak, Bölüm 4’te çalışmanın genel değerlendirmesi ve sonuçları tartışılmıştır.

2 Materyal ve metod

2.1 ARIMA (Otoregresif bütünleşik hareketli ortalama)

Durağan olmayan zaman serilerinin, analiz veya modelleme için kullanılmadan önce durağan hale getirilmesi gerekmektedir. Eğer bir seri eğer durağan değilse, uygun derecede fark alma işlemi uygulanarak o seri durağan hale getirilir. Fark alma işlemiyle durağan hale getirilen serilere ARIMA modelleri uygulanmaktadır. Bu modeller, otoregresif (AR) ve hareketli ortalama (MA) süreçlerinin kombinasyonu olup fark işlemini de (I) içermektedir. Otoregresif derecesini p ile, hareketli ortalama derecesini q ile ve fark alma işleminin derecesini d ile ifade eder ve ARIMA(p,d,q) formatında gösterilir [14]. Genel ARIMA(p,d,q) modeli aşağıdaki gibidir:

$$y_t = \psi_1 y_{t-1} + \psi_2 y_{t-2} + \dots + \psi_m y_{t-m} + \kappa + \alpha_t - \lambda_1 \alpha_{t-1} - \lambda_2 \alpha_{t-2} - \dots - \lambda_n \alpha_{t-n} \quad (1)$$

Bu modelde, $y_t, y_{t-1}, \dots, y_{t-m}$, geçmiş gözlem değerlerini temsil etmektedir. $\psi_1, \psi_2, \dots, \psi_m$ bu gecikmeli gözlemlere ait otoregresif katysayıları

belitirken, κ sabit değeri temsil etmektedir. $\alpha_t, \dots, \alpha_{t-n}$ hata terimini temsil ederken, $\lambda_1, \dots, \lambda_n$ bu hata terimlerine ait hareketli ortalama katsayılarını göstermektedir [15]. Burada d, zaman serisi üzerinde uygulanan fark alma (differencing) derecesini ifade etmektedir.

2.2 SARIMA

SARIMA modeli, zaman serilerinin hem mevsimsel hem de mevsimsel olmayan bileşenlerini ele alarak genişletilmiş bir ARIMA modelidir [16]. SARIMA(P,D,Q)(p,d,q)[s] şeklinde gösterilmektedir ve bu ifade ARIMA modelinde kullanılan p,d,q bileşenleri haricinde şu bileşenlerden oluşmaktadır:

P: Mevsimsel AR derecesi

D: Mevsimsel fark alma işleminin derecesi

Q: Mevsimsel MA derecesi

s: Mevsimsel dönem

Modelin mevsimsel kısmını içeren terimler, ARIMA modelinin denklemindeki gibi mevsimsel olmayan bileşenlerine benzemektedir ancak SARIMA modeli mevsimsel dönemin geriye kaymasını içermektedir [17]. Matematiksel ifade şu şekildedir:

$$\phi_p(B^s)(1 - B^s)^D(1 - B)^d y_t = \theta_q(B^s)\theta_q(B)\varepsilon_t \quad (2)$$

Bu modelde, B geri kaydırma operatörünü, ϕ mevsimsel AR terimlerini, θ mevsimsel MA terimlerini ifade etmektedir.

2.3 ELM (Aşırı öğrenme makinesi)

ELM algoritması, klasik YSA algoritmasının sınırlamalarına bir çözüm olarak Huang vd. (2006) tarafından tanıtılmıştır. Bu algoritma, öğrenme sürecini basitleştirmek amacıyla bir eğitim yaklaşımını kullanmaktadır ve bir tür ileri beslemeli sinir ağı olarak tasarlanmıştır [18]. ELM algoritmasında, geleneksel yöntemlerden farklı olarak giriş katmanı değerleri ve eşik değerleri rastgele atanırken, çıktı katmanı ağırlıkları ise analitik bir yaklaşım ile hesaplanmaktadır. Bu özellik sayesinde SVM ve YSA gibi yöntemlere kıyasla çok daha yüksek bir performans göstermekte ve çok daha hızlı bir şekilde eğitilebilmektedir. ELM modeli şu şekildedir:

$$\sum_{i=1}^M \beta_i g(W_i X_k + b_i) = Y_k, \quad k = 1, 2, \dots, N \quad (3)$$

Bu modelde, β_i çıkış ağırlığını, W_i giriş ağırlığını ifade eden matrislerdir. X_k giriş, b_i i. nörondaki eşik değeri ve Y_k ise ağırlık çıktısıdır [15].

2.4 MLP (Çok katmanlı algılayıcı)

YSA modellerinden biri olan MLP, en yaygın kullanılan modeller arasında yer almaktadır [19]. MLP ağlarının eğitiminde, anlaşılabilirliği ve matematiksel olarak ispatlanabilirliği nedeniyle geri yayılım algoritması tercih edilmektedir [20]. MLP, en az üç temel katmandan oluşur: bir girdi katmanı, bir çıktı katmanı ve bir veya daha fazla gizli katman. Girdi katmanı, modelin öğrenme süreci için gerekli bilgilerin alındığı ancak herhangi bir bilgi işleme yapılmadığı katmandır. Gizli katman ise girdi katmandan gelen bilgileri işleyen bölümlerdir ve problemin gereksinimlerine göre sayısı artırılabilir. Çıktı katmanı, gizli katmandan gelen bilgileri işleyerek girdi katmanına verilen veri için modelin tahmin ettiği çıktıyı üretir ve iletir [21]. MLP algoritması, zaman serisi gibi dinamik yapıya sahip problemlerle de başarıyla kullanılabilir. Zaman serisi tahmininde, çıktı ($\hat{z}_t$) ile serinin geçmiş gecikmeli değerlerinden ($u_{t-1}, u_{t-2}, \dots, u_{t-N}$) oluşan girdiler arasındaki ilişki şu şekilde ifade edilebilir:

$$\hat{z}_t = c_0 + \sum_{k=1}^p \alpha_k h\left(c_k \sum_{i=1}^N \beta_{i,k} u_{t-i}\right) + \varepsilon_t \quad (4)$$

Bu ifadede α_k ve $\beta_{i,k}$ terimleri, nöronlar arasındaki ağırlık değerlerini temsil ederken, p gizli nöronlarına sayısını, h ise gizli katmandaki doğrusal olmayan aktivasyon fonksiyonunu tanımlamaktadır [15].

2.5 Performans ölçütleri

Tahmin etme çalışmaları iki temel aşamadan oluşmaktadır. İlk aşama veriyi tahmin etmek için uygun hale getirmek için gerekli işlemlerin gerçekleştirilmesini kapsarken, ikinci aşama ise farklı tahmin edici modelleri performans açısından karşılaştırılmasını içermektedir. Modelleri karşılaştırmak için kullanılan ölçütlerin hız, doğruluk, sağlamlık, ölçeklenebilirlik ve yorumlanabilirlik gibi temel özellikleri taşıması gerekmektedir. Makine öğrenmesi yöntemlerinde model performansını değerlendirmek amacıyla kullanılan temel performans ölçütleri arasında R-kare (R^2), Ortalama Kare Hatası (MSE), Kök Ortalama Kare Hatası (RMSE), Ortalama Mutlak Kare (MAE) yer almaktadır [22]. Bu çalışmada performans ölçütlerinden MAE ve RMSE kullanılmış olup matematiksel gösterimleri şu şekildedir:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

Burada y_i gerçek değeri, $\hat{y}_i$ ise tahmin edilen değeri ifade etmektedir [23].

3 Uygulama

T.C. Kültür ve Turizm Bakanlığı'nın Turizm-Konaklama istatistiklerinden elde edilen veri seti, 2000-2024 yılları arasındaki toplam, yerli ve yabancı konaklama istatistiklerini içermektedir. Veri setinde bulunan eksik gözlemler, R programında "mice" paketi yardımı ile doldurulmuştur ve doldurulan veri seti üzerinde farklı eğitim - test veri oranlarının model performansına etkisi incelenmiştir. Eğitim ve test veri setleri sırasıyla %80-%20, %70-%30 ve %60-%40 ayrılarak üç farklı senaryoda ele alınacaktır. Modellerin performansı, MAE (Ortalama Mutlak Hata) ve RMSE (Karekök Ortalama Hata) ölçütleri kullanılarak değerlendirilmiştir.

a) Yerli Doluluk Oranları Verisi Analiz Sonuçları

Bu çalışmada ilk olarak yerli turist doluluk oranlarının analizi gerçekleştirilmiştir. Tablo 1' de, yerli turist verilerini kullanarak elde edilen modellerin %80-%20, %70-%30 ve %60-%40 eğitim-test ayırma oranlarına göre hesaplanan MAE ve RMSE değerleri gösterilmiştir.

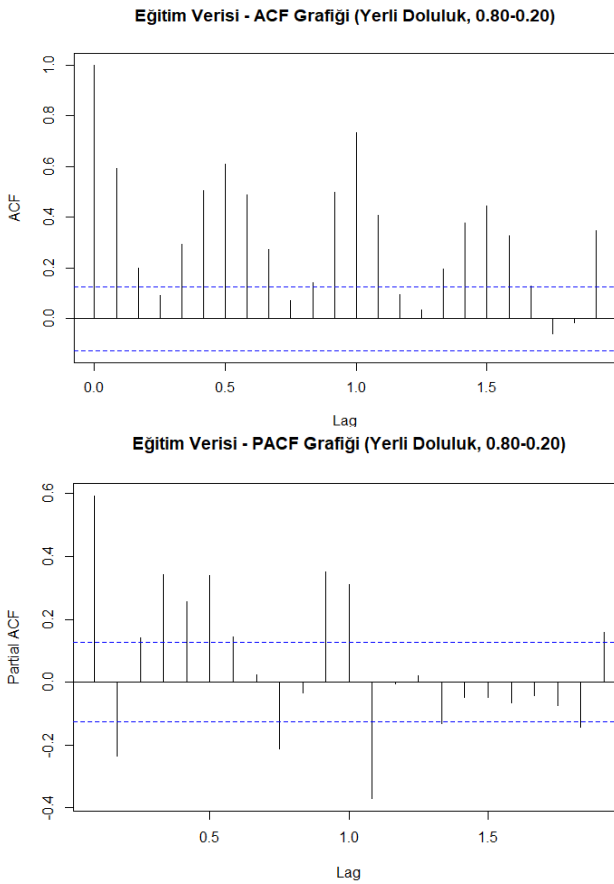
Tablo 1. Yerli Turist Verisi analiz sonuçları için Modellerin MAE ve RMSE değerleri.

Yerli Veri Hata Türü		%80-%20 Eğitim-Test	%70-%30 Eğitim-Test	%60-%40 Eğitim-Test
ELM	MAE	3.168	5.791	3.091
	RMSE	4.417	6.997	4.132
MLP	MAE	3.804	4.830	3.501
	RMSE	4.617	6.055	4.592
SARIMA	MAE	3.573	6.359	4.218
	RMSE	4.193	7.570	5.118

ELM modeli için MAE ve RMSE değerleri incelendiğinde, en düşük hata değerleri sırasıyla

3.091 ve 4.132 olup, bu değerler %60 eğitim – %40 test oranında elde edilmiştir. Ancak, %70 eğitim – %30 test oranında MAE ve RMSE değerleri en yüksek seviyeye ulaşmış ve modelin bu koşullarda düşük performans sergilediği gözlemlenmiştir.

MLP modeli için MAE değeri incelendiğinde, en düşük hata değeri 3.501 olup, bu değer %60 eğitim – %40 test oranında elde edilmiştir. RMSE değeri ise en düşük hata değeri 4.167 olup, %80 eğitim – %20 test oranında gözlemlenmiştir. Genel olarak, MLP modelinin ELM modeline kıyasla biraz daha yüksek hata değerleri ürettiği görülmektedir.



Şekil 1. Yerli Turist Verisi Doluluk oranı için %80-%20 oranında ACF-PACF grafikleri.

ARIMA modeli için MAE ve RMSE değerlerine bakıldığında, %70 eğitim – %30 test oranında en kötü performansı sergilediği ve hata değerlerinin diğer modellere kıyasla oldukça yüksek olduğu söylenebilir. Ancak, %80 eğitim – %20 test oranında ELM ve MLP modellerine yakın bir performans gösterdiği görülmüştür. Öte yandan, %60 eğitim – %40 test oranında ise MAE ve RMSE

değerleri diğer modellere göre daha yüksek bulunmuştur.

Hata katsayısının en düşük olduğu zaman serisi ayırma oranı %80-%20 olup, bu oran detaylı olarak incelenmiştir. Zaman serisi modellemesi öncesinde verilerin yapısal özelliklerini anlamak için otokorelasyon (ACF) ve kısmi otokorelasyon (PACF) grafiklerinden yararlanılmıştır. Bu grafikler serinin durağanlığını ve otoregresif (AR) veya hareketli ortalama (MA) yapısı taşıyıp taşımadığını belirlemede yardımcı olmaktadır.

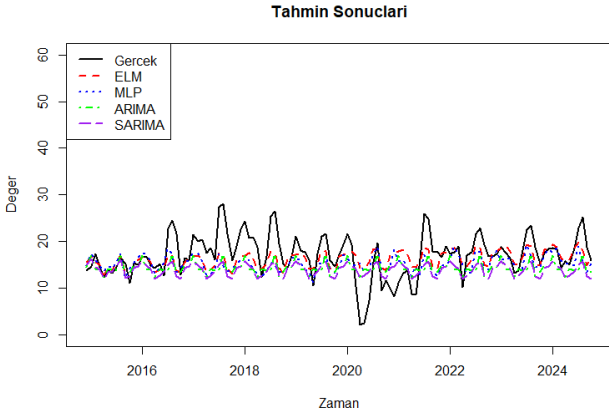
ACF Grafikleri; yerli doluluk oranı verisinin kendi geçmiş değerleri ile olan ilişkisinin zaman içerisindeki değişimini ifade etmektedir. PACF Grafikleri; her gecikmenin seriye olan bağımsız etkisini ortaya koyan grafiklerdir.

ACF grafiğine bakıldığında; 1-3 gecikmelerin anlamlı olduğu görülmekte olup, 12. Gecikmede kuvvetli bir mevsimsel sıçrama olduğu belli olmaktadır. Bu durum, verinin hem kısa vadeli otokorelasyona sahip olduğunu hem de yıllık döngüye sahip olduğunu göstermektedir. PACF grafiğinde, 1-2 gecikmelerde anlamlı kısmi otokorelasyonlar olduğu görülmektedir. AR(2) yapısının uygun olduğu söylenebilir.

Verilen ARIMA(2,0,2)(1,1,0)[12], zaman serisinde hem kısa dönemli ($ar1 = -0.2117$, $ar2 = 0.6141$; $ma1 = 0.6497$, $ma2 = -0.1419$) hem de yıllık mevsimsel yapıyı ($sar1 = -0.2048$) yakalamayı hedeflemektedir. Özellikle $ar2$, $ma1$ ve $sar1$ katsayılarının görece anlamlı (standart hataları sırasıyla 0.1214, 0.1844 ve 0.0740) olması, seride geçmiş değerlere ve mevsimsel bileşene bağımlılık olduğunu göstermektedir. Modelin hata varyansı 3.647, log likelihood değeri -466.38, AIC değeri 946.75, AICc değeri 947.26 ve BIC değeri 970.73'tür.

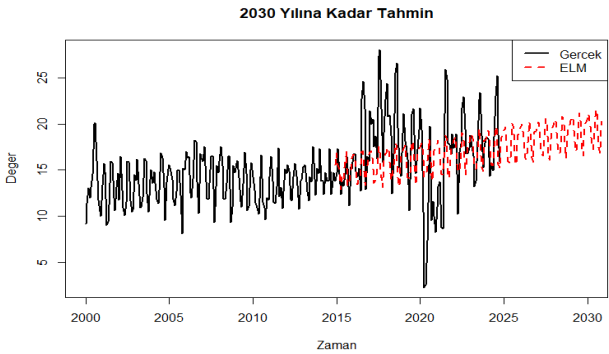
Yerli Doluluk Oranları verisi üzerinde gerçekleştirilen analizler sonucunda, %60 eğitim %40 test oranının en düşük hata değerlerine sahip olduğu tespit edilmiştir. Özellikle ELM modeli, MAE = 3.091 ve RMSE = 4.132 değerleriyle bu veri setinde en başarılı tahmin performansını göstermiştir. Bu durum, modelin test verisi üzerindeki genelleme yeteneğinin yüksek olduğunu ve aşırı öğrenme problemini minimize ettiğini göstermektedir. Bu bulgular doğrultusunda, yerli turist doluluk oranlarının tahmini için %60 eğitim – %40 test oranı en uygun strateji olarak

önerilmektedir. Şekil 2’de bu orana göre elde edilen tahmin grafiği sunulmuştur.



Şekil 2. Yerli Turist Verisi için %60-%40 oranlarının grafiği.

Şekil 2’de görüldüğü üzere, gerçek değerlere en yakın tahminleri ELM modeli üretmiştir. MLP modelinin tahminleri ELM modeline oldukça yakın olmakla birlikte, karşılaştırıldığında daha düşük hata değerlerine sahip olan ELM modeli öne çıkmaktadır. Ayrıca, ELM modeli veriyi %60 eğitim ve %40 test olarak ayırarak, 2030 yılına kadar yerli turist sayısına ilişkin tahminlerini Şekil 3’da sunmaktadır.



Şekil 3. Yerli Turist Verisi için %60-%40 oranında ELM modelinin 2030 yılına kadar tahmin grafiği.

Şekil 3’te ise ELM modelinin 2000–2025 yılları arasındaki gerçek veriler üzerinden eğitilerek, 2025–2030 dönemi için gerçekleştirdiği tahminler gösterilmektedir. Grafik incelendiğinde, modelin geleceğe yönelik tahminlerinde dengeli ve artış eğilimli bir seyir izlendiği ve yerli doluluk oranlarının %15–%20 aralığında devam edeceği öngörülmektedir. Bu sonuçlar, ilgili kamu kurumları ve turizm planlaması yapan paydaşlar

açısından ileriye dönük stratejiler belirlemede önemli bir karar destek mekanizması sunmaktadır.

b) Yabancı Doluluk Oranları Verisi Analiz Sonuçları

Bu çalışmada ikinci aşamasında, yabancı turist doluluk oranlarının ilişkin analizler gerçekleştirilmiştir. Tablo 2’de, %80-%20, %70-%30 ve %60-%40 eğitim-test oranlarına göre oluşturulan ELM, MLP, ARIMA ve SARIMA modellerine ait MAE ve RMSE hata değerleri sunulmuştur.

Tablo 2. Yabancı Turist Verisi analiz sonuçları için Modellerin MAE ve RMSE değerleri.

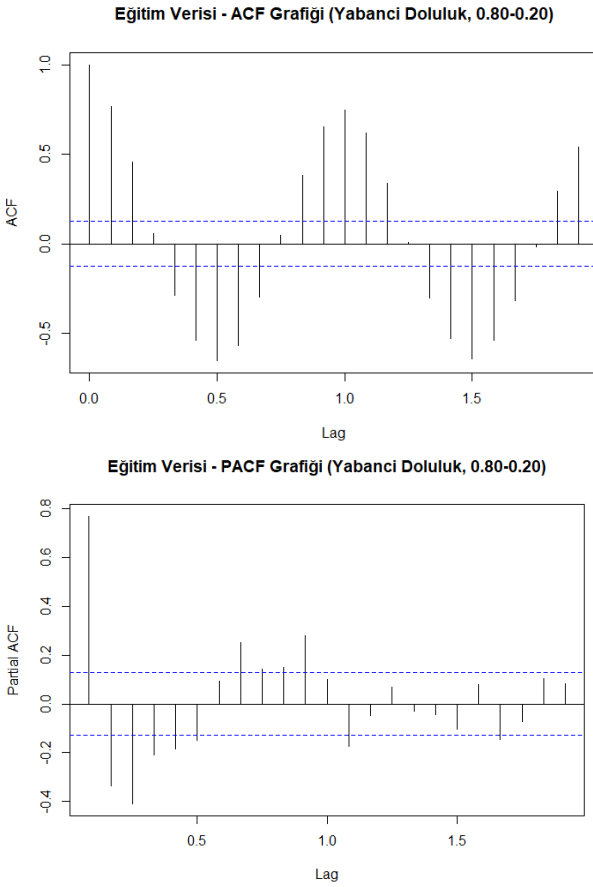
Yabancı Veri		%80-%20	%70-%30	%60-%40
Hata Türü		Eğitim-Test	Eğitim-Test	Eğitim-Test
ELM	MAE	21.524	11.189	11.029
	RMSE	23.954	13.739	13.679
MLP	MAE	15.002	10.576	21.343
	RMSE	18.920	13.006	27.101
ARIMA	MAE	8.841	13.965	13.973
	RMSE	13.527	16.063	17.246
SARIMA	MAE	8.837	13.388	13.913
	RMSE	13.538	15.504	17.201

ELM modeli için; MAE ve RMSE değerleri incelendiğinde en düşük hata değerinin %60 eğitim - %40 test oranında sırasıyla (11.029), (13.679) olduğu görülmektedir. Ancak, %80 eğitim - %20 test oranında ELM modelinin MAE ve RMSE değerleri en yüksek seviyeye ulaşmakta olup, bu koşullarda düşük performans sergilediği gözlemlenmiştir.

MLP modeli için; MAE açısından en düşük hata değeri %70 eğitim- %30 test oranında (10.576) hesaplanırken, RMSE açısından da en düşük (13.006) olarak bulunmuştur. Ancak, %60 eğitim - %40 test oranında en yüksek hata değerleri (MAE: 21.343, RMSE: 27.101) gözlemlenmiştir, dolayısıyla bu modelin bu koşullarda kötü performans sergilediği gözlemlenebilir.

ARIMA modeli için; MAE ve RMSE açısından %80 eğitim- %20 test oranında en düşük hata değerlerine sahip olduğu, ancak %70 ve %60 eğitim

oranlarında hata değerlerinin önemli ölçüde arttığı görülmüştür. En düşük hata değerine sahip ayırma oranın sonuçlarını detaylı inceleyelim.



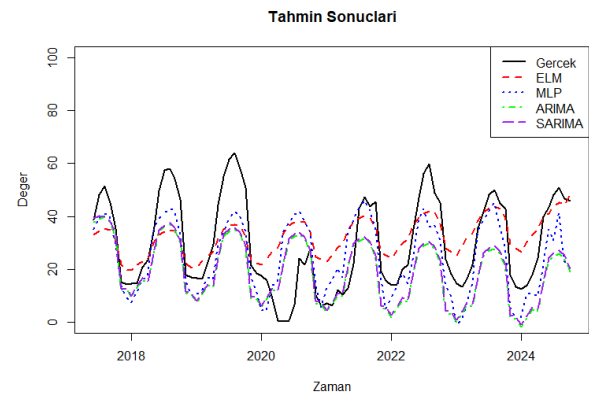
Şekil 4. Yabancı Turist Verisi Doluluk oranı için %80-%20 oranında ACF-PACF grafikleri.

Şekil 4'da sunulan ACF ve PACF grafiklerine göre; ACF grafiğinde 1. ve 2. gecikmelerde belirgin otokorelasyon gözlenmiştir. Mevsimsel yapı ise 12. gecikmede daha zayıf bir görünüm sergilemiştir. PACF grafiğinde 1-3. gecikmeler arasında anlamlılık dikkat çekmekte, özellikle 2 ve 3. gecikmeler sınıra yakın çıkmaktadır. Bu sonuçlar, serinin kısa vadeli etkilerinin ön planda olduğunu ve AR(1) ya da AR(2) gibi daha sade modellerin uygun olabileceğini göstermektedir.

Durağanlık analizine göre, $d=1$ alındığında ACF grafiği durağan olmayan bir yapı sergilemiştir. $p=1$ olduğunda PACF grafiği anlamlılık göstermekte, $q=2$ için ACF dalgalı yapı sergileyerek MA etkisinin zayıfladığını göstermektedir.

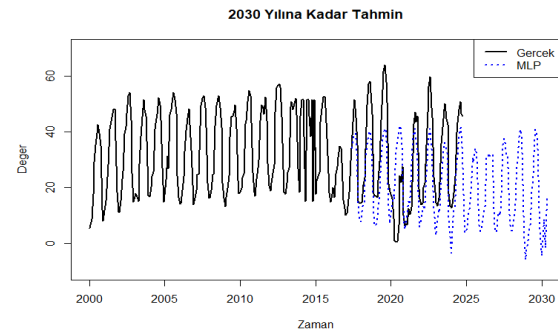
Auto ARIMA tarafından önerilen model ARIMA(3,0,1)(0,1,1)[12] olup; özellikle sma1 (-0.8838) mevsimsel etkilerin güçlü biçimde yakalandığını göstermektedir. Ayrıca ar2 (0.5981) ve ma1 (0.7684) katsayıları da dikkat çekmektedir. Ancak, hata varyansının yüksekliği ($\sigma^2 = 14.46$) ve AIC değerinin 1282.27 olması, modelin belirsizlik içerdiğini göstermektedir.

MLP modeli %70 eğitim oranında en iyi performansı sergilemektedir. Bu durumda Yabancı Doluluk Oranları veri seti için %70-%30 oranlarının tercih edilmesi önerilmektedir. Şekil 11'de %70 eğitim - %30 test oranlarının tahmin grafiği görülmektedir:



Şekil 5. Yabancı Turist Verisi için %70-%30 oranlarının tahmin grafiği.

Şekil 5de görüldüğü üzere, gerçek değerlere en yakın tahminleri MLP modeli üretmiştir. Dolayısıyla, MLP modeli seçildiğinde en iyi sonucu vererek gerçeğe daha yakın tahminler yapmaktadır. Ayrıca, MLP modeli veriyi %70 eğitim ve %30 test olarak ayırarak, 2030 yılına kadar yabancı turist sayısına ilişkin tahminlerini Şekil 12'de sunmaktadır.



Şekil 6. Yabancı Turist Verisi için %70-%30 oranında MLP modelinin 2030 yılına kadar tahmin grafiği.

Şekil 6'de, 2000-2025 yılları arasındaki gerçek veriler siyah çizgi ile gösterilirken, 2025-2030 yılları için MLP modeli kullanılarak yapılan tahminler mavi kesikli çizgi ile verilmiştir. Gerçek verilerde belirgin mevsimsel dalgalanmalar ve zaman içinde artan oynaklık gözlemlenirken, MLP modeli gelecekte de bu dalgalanmaların devam edeceğini öngörmektedir.

c) Toplam Doluluk Oranları Verisi Analiz Sonuçları

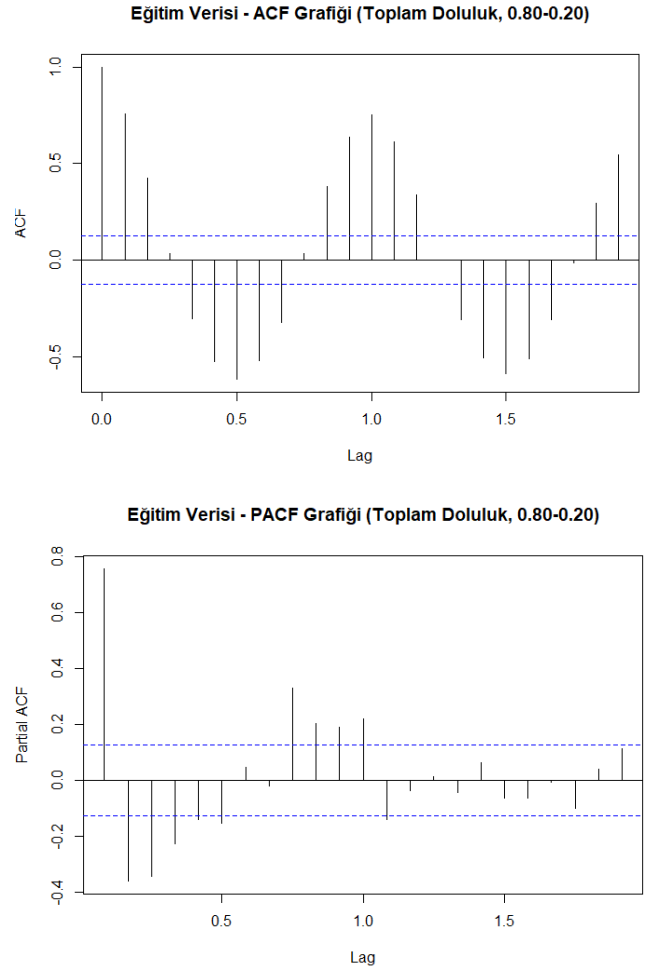
Bu çalışmada son olarak toplam turist doluluk oranlarının analizi gerçekleştirilmiştir. Tablo 3' te, toplam turist verilerini kullanarak elde edilen modellerin %80-%20, %70-%30 ve %60-%40 eğitim-test ayırma oranlarına göre hesaplanan MAE ve RMSE değerleri gösterilmiştir.

Tablo 3. Toplam Turist Verisi analiz sonuçları için Modellerin MAE ve RMSE değerleri.

Toplam Veri		%80-%20	%70-%30	%60-%40
Hata Türü		Eğitim-Test	Eğitim-Test	Eğitim-Test
ELM	MAE	19.824	13.066	32.184
	RMSE	24.459	16.933	36.065
MLP	MAE	17.897	12.604	93.953
	RMSE	22.626	16.781	103.179
SARIMA	MAE	9.807	11.675	15.856
	RMSE	15.989	14.673	19.854

Toplam Doluluk Oranları verisi üzerinde yapılan analizler sonucunda, ELM modeli için; MAE ve RMSE değerleri incelendiğinde en düşük hata değerinin %70 eğitim - %30 test oranında sırasıyla (13.066), (16.933) olduğu görülmektedir. Ancak, %60 eğitim - %40 test oranında ELM modelinin MAE (32.184) ve RMSE (36.065) değerleri en yüksek seviyeye ulaşmış olup, bu koşullarda düşük performans sergilediği gözlemlenmiştir.

MLP modeli için; MAE açısından en düşük hata değeri %70 eğitim - %30 test oranında (12.604) hesaplanırken, RMSE açısından da en düşük hata değeri %70 eğitim - %30 test oranında (16.781) olarak bulunmuştur. Ancak, %60 eğitim - %40 test oranında MLP modelinin MAE (93.953) ve RMSE (103.179) değerleriyle en kötü performansı sergilediği belirlenmiştir.



Şekil 7. Toplam Turist Verisi Doluluk oranı için %80-%20 oranında ACF-PACF grafikleri.

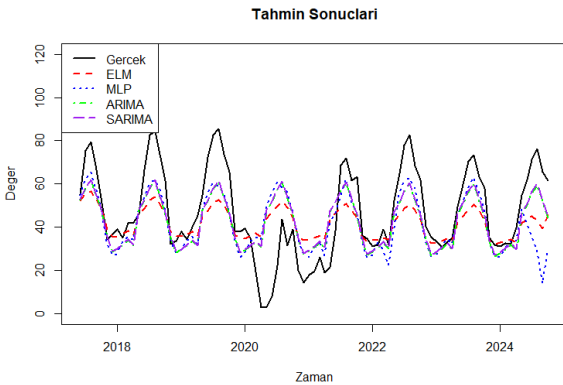
ACF grafiğine bakıldığında, ilk 3 gecikme boyunca anlamlı ve pozitif otokorelasyonlar görülmekte, ayrıca 12. gecikmede güçlü bir mevsimsel sıçrama gözlenmektedir. Bu durumda, hem kısa vadeli MA bileşeni hem de yıllık döngülerin varlığı görülmektedir. PACF grafiğinde, 1-3 arasındaki gecikmelerde anlamlı olduğu görülmektedir. Bu durumda, AR(3) yapısının varlığı söylenebilir.

ARIMA(3,0,2)(0,1,1)[12], hem kısa dönemli hem de yıllık mevsimsel bileşenleri içermektedir. Kısa dönemli yapıyı yansıtan $ar1 = -1.0363$, $ar2 = 0.2538$, $ar3 = 0.4779$, $ma1 = 1.3160$ ve $ma2 = 0.5302$ katsayılarının çoğu anlamlı olup, özellikle $ar1$, $ar3$ ve $ma1$ katsayılarının yüksek olması, seride güçlü otoregresif ve hareketli ortalama etkilerinin bulunduğunu göstermektedir. Mevsimsel hareketli ortalama bileşeni olan $sma1 = -0.8243$, yıllık döngüsel yapıların belirginliğini vurgulamaktadır. Ayrıca, $drift = 0.0439$ katsayısı pozitif bir trend

etkisine işaret etmektedir. Modelin hata varyansı $\sigma^2 = 36.39$, log likelihood değeri -737.12, AIC değeri 1490.25, AICc değeri 1490.90 ve BIC değeri 1517.68

ARIMA modeli için, en düşük MAE değeri %80 eğitim - %20 test oranında (9.862), en düşük RMSE değeri ise %80 eğitim - %20 test oranında (16.195) olarak hesaplanmıştır. Ancak, %60 eğitim - %40 test oranında MAE (16.004) ve RMSE (20.019) değerleri artarak daha kötü bir performans sergilemiştir. Genel olarak, %70 eğitim - %30 test oranı, en düşük hata değerlerine sahip olması nedeniyle en uygun oran olarak öne çıkmaktadır. ELM ve MLP modelleri bu oran altında en iyi performansı sergilerken, ARIMA modeli %80 eğitim - %20 test oranında daha düşük hata değerleri ile öne çıkmaktadır, fakat SARIMA modeli de unutulmamalıdır. SARIMA modeli %70 eğitim - %30 test oranında en düşük değerleri almıştır (11.675, 14.673).

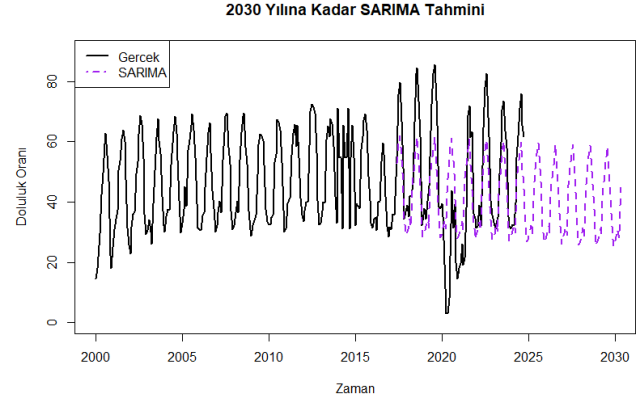
Genel olarak tüm modellerin performans ölçütlerine bakıldığında, %70 eğitim oranına sahip veri setinin diğer oranlara göre daha tutarlı ve daha düşük hata değerleri içerdiği görülmektedir. ELM, MLP ve ARIMA modellerinde bu veri seti üzerinde daha düşük hatalar gözlenmiştir. Bu durum ise diğer veri setlerine göre daha uygun olduğunu işaret etmektedir. Sonuç olarak, %70 eğitim - %30 test oranına sahip veri seti genel modeller için daha sabit ve daha düşük hatalarla sonuçlanmış, bu durumda tahmin ve modelleme açısından daha başarılı olduğunu göstermiştir. Şekil 17'de %70 eğitim - %30 test oranlarının tahmin grafiği görülmektedir:



Şekil 8. Toplam Turist Verisi için %70-%30 oranlarının grafiği.

Şekil 8'de görüldüğü üzere, ARIMA ve SARIMA modellerinin değerleri diğer model değerlerine

kıyasla gerçek değerlere çok daha yakın gözükmektedir. SARIMA değerlerini gözlemlendiğinde en uygun olacak modelin bu olduğuna kanaat getirilmiştir.



Şekil 9. Toplam Turist Verisi için %70-%30 oranında SARIMA modelinin 2030 yılına kadar tahmin grafiği.

Şekil 9'da, 2000-2025 yılları arasındaki gerçek veriler siyah çizgi ile gösterilirken, 2025-2030 yılları için SARIMA modeli kullanılarak yapılan tahminler mor kesik çizgi ile sunulmuştur. Gerçek verilerde düzenli mevsimsel dalgalanmalar ve zaman içinde artan oynaklık gözlemlenirken, SARIMA modeli gelecekte bu dalgalanmaların devam edeceğini ancak daha dengeli bir trend izleyeceğini öngörmektedir.

4 Tartışma ve Sonuç

Bu çalışma, 2000-2024 yılları arasındaki konaklama istatistiklerini kullanarak doluluk oranlarının tahmini üzerine yoğunlaşan bir analiz sunmaktadır. Zaman serisi analiz yöntemleri ve makine öğrenmesi modelleri ile gerçekleştirilen çalışmada, turizm talebi tahmini için farklı modelleme yaklaşımlarının performansları karşılaştırılmıştır. Analiz kapsamında, her bir modele farklı eğitim ve test oranları uygulanmış olup, bu oranlar %80-%20, %70-%30 ve %60-%40 olarak üç farklı durumda incelenmiştir.

Yerli doluluk oranında, %60 eğitim-%40 test oranı ile en başarılı tahminleri ELM modeli üretmiştir. 2000-2015 yılları arasında stabil dalgalanmalar gözlemlenirken, 2015 sonrasında gerçek verilerin volatilitesi belirgin şekilde artmıştır. Gerçek veriler, siyah çizgi ile gösterilmektedir. 2020 yılında yaşanan keskin düşüş, COVID-19 pandemisinin etkisiyle açıklanabilir; ancak bu düşüşün ardından tekrar artış eğilimi gözlenmiştir. 2024 yılı

sonrasında, kırmızı çizgi ile gösterilen MLP modeli tahminleri, dalgalanmalardaki sert değişimlerin azalacağını ve daha stabil bir yapı izleneceğini öngörmektedir. Tahminlerin belirli bir bant içinde hareket etmesi, gelecekteki değerlerin nispeten istikrarlı olacağını göstermektedir. Bununla birlikte, ELM modelinin mevsimsel desenleri yeterince güçlü yansıtmadığı ve daha düz, sabit bir eğilim sunduğu görülmektedir. Bu durum, turizm talebinde göreceli olarak daha dengeli bir toparlanmanın olabileceğini işaret etmektedir. İşletmelere, daha planlı ve kapasite odaklı stratejik çözümler geliştirmeleri önerilmektedir. Yabancı doluluk oranında, %70 eğitim-%30 test oranı ile en başarılı tahmini MLP modeli sunmuştur. Siyah çizgi, gerçek verileri temsil ederken, mavi çizgi MLP modelinin tahminlerini göstermektedir. Gerçek veriler, belirgin bir mevsimsel döngü ve dalgalı bir yapıya sahiptir. 2020 civarındaki keskin düşüş, COVID-19 pandemisinin etkisiyle açıklanabilir. MLP modelinin tahminleri, turizm sektöründe doluluk oranlarının yeniden mevsimsel hareketliliğini sürdüreceğini öngörmektedir. Model, ileriye yönelik olarak yıllık periyotlarla belirgin dalgalanmaların devam edeceğini göstermektedir. Bu tür tahminler, turizm işletmeleri ve stratejik karar vericiler için geleceğe yönelik önemli bir karar destek aracı olarak değerlendirilebilir.

Toplam doluluk oranında, %80 eğitim-%20 test oranı ile SARIMA(3,0,2)(0,1,1)[12], modeli en düşük hata payı ile en başarılı tahmini sunmuştur. 2024 yılına kadar gerçek veriler mevsimsel dalgalanmalar gösterirken, 2020 yılındaki keskin düşüş pandemi gibi olağan dışı bir etkinin yansımasıdır. ELM modeli mevsimsel desenleri takip etmekle birlikte bazı dönemlerde daha stabil bir yapı sergilemiştir. MLP modeli ise tahminlerinde dalgalı yapıyı koruyarak veriye esnek bir uyum sağlamıştır. SARIMA(3,0,2)(0,1,1)[12], modeli, dönemsel yapıyı en belirgin şekilde tahmin ederek mevsimsel bileşenlerin devam edeceğini öngörmüştür. Şekil incelendiğinde, tahmin modellerinin 2025 sonrası yıllarında genel olarak birbirine yakın sonuçlar verdiği, ancak SARIMA modelinin dönemsel desenleri daha net öngördüğü görülmüştür. MLP modeli ise daha geniş sapmalar ile tahmin yapmıştır. Sonuç olarak, SARIMA(3,0,2)(0,1,1)[12], modeli daha tutarlı bir performans sergilemekle birlikte her model farklı avantajlar sunmaktadır.

Bu bulgular doğrultusunda, çalışmanın kuramsal katkısı, turizm talebi tahmini konusunda klasik zaman serisi yöntemleri ile makine öğrenmesi tabanlı modellerin birlikte incelenmesiyle, iki yaklaşımın da güçlü ve zayıf yönlerinin uygulamalı bir bağlamda karşılaştırılması sağlanmış, bu da literatürdeki sınırlı sayıda çalışmaya özgün bir katkı sunmuştur. Metodolojik katkı olarak ise, farklı eğitim-test oranlarında modellerin performanslarının değerlendirilmesiyle sadece mutlak başarı düzeyleri değil, veri bölme stratejilerinin tahmin gücüne etkisi de tartışmaya açılmıştır. Ayrıca ELM gibi görece daha az kullanılan bir algoritmanın zaman serisi bağlamında uygulanması, metodolojik çeşitlilik açısından çalışmanın özgünlüğünü artırmaktadır. Turizm sektöründe faaliyet gösteren işletmeler için stratejik karar alma sürecine rehberlik edebilir. Konaklama talebinin doğru tahmin edilmesi, işletmelerin kapasite planlamasını daha etkin bir şekilde yapmasına, oda ve hizmet fiyatlandırma stratejilerini optimize etmesine ve operasyonel verimliliği artırmasına katkı sağlayacaktır. Ayrıca, dönemsel dalgalanmaların ve beklenmedik krizlerin etkilerini öngörerek proaktif önlemler alınmasına imkân tanıyacaktır. Bu sayede, turizm işletmeleri rekabet avantajı elde ederken, kaynaklarını en verimli şekilde kullanarak sürdürülebilir büyümeyi destekleyebilecektir.

5 Kaynakça

- [1] UNWTO. "Glossary of tourism terms". <https://www.unwto.org/glossary-tourism-terms> (18.06.2025).
- [2] Ulucan E, Kızıllırmak İ. "Konaklama işletmelerinde talep tahmin yöntemleri: yapay sinir ağları ile ilgili bir araştırma". *Seyahat ve Otel İşletmeciliği Dergisi*, 15(1), 9–101, 2018.
- [3] OECD. *OECD Tourism Trends and Policies 2024*. Paris, Fransa, OECD Publishing, 2024. <https://doi.org/10.1787/80885d8b-en>
- [4] Yalçın A, Gelen Mert MB. "Estimating the occupancy rate of an accommodation business using artificial neural networks". *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 47, 209–218, 2021.
- [5] Çuhadar M, Kayacan C. "Yapay sinir ağları kullanılarak konaklama işletmelerinde doluluk oranı tahmini: Türkiye'deki konaklama işletmeleri üzerine bir deneme". *Anatolia: Turizm Araştırmaları Dergisi*, 16(1), 24–30, 2005.
- [6] Caicedo-Torres W, Payares F. "A machine learning model for occupancy rates and demand forecasting

- in the hospitality industry". In: Zorrilla M, Mazo R, Cano Á (Eds.). *Advances in Artificial Intelligence – IBERAMIA 2016*, 201–211, Cham, İsviçre, Springer, 2016.
- [7] Burger CJS, Dohnal M, Kathrada M, Law R. "A practitioner's guide to time-series methods for tourism demand forecasting: a case study of Durban, South Africa". *Tourism Management*, 22(4), 403–409, 2001.
- [8] Fernandes P, Teixeira J, Ferreira JM, Azevedo SG. "Modelling tourism demand: a comparative study between artificial neural networks and the Box Jenkins methodology". *Romanian Journal of Economic Forecasting*, 3, 30–50, 2008.
- [9] Çuhadar M, Güngör İ, Göksu A. "Turizm talebinin yapay sinir ağları ile tahmini ve zaman serisi yöntemleri ile karşılaştırmalı analizi: Antalya iline yönelik bir uygulama". *Süleyman Demirel Üniversitesi İİBF Dergisi*, 14(1), 99–114, 2009.
- [10] Önder E, Hasgöl Ö. "Yabancı ziyaretçi sayısının tahmininde Box-Jenkins modeli, Winters yöntemi ve yapay sinir ağlarıyla zaman serisi analizi". *İstanbul Üniversitesi İşletme İktisadi Enstitüsü Dergisi*, 4, 62–83, 2009.
- [11] Çuhadar M. "Türkiye'ye yönelik dış turizm talebinin MLP, RBF ve TDNN yapay sinir ağı mimarileri ile modellenmesi ve tahmini: karşılaştırmalı bir analiz". *Yaşar Üniversitesi E-Dergisi*, 8(31), 5274–5295, 2013. <https://doi.org/10.19168/jyu.54484>
- [12] Zorlutuna Ş, Bircan H. "Türkiye'ye gelen turist sayısı tahmininde zaman serileri analizi ve yapay sinir ağları yöntemlerinin karşılaştırılması". *Cumhuriyet Üniversitesi İİBF Dergisi*, 20(2), 164–185, 2019.
- [13] Güler Ö. *Makine Öğrenmesi Yöntemleriyle Türkiye'de Turizm Gelirine Etki Eden Parametrelerin Belirlenmesi ve Değerlendirilmesi*. Yüksek Lisans Tezi, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Denizli, Türkiye, 2021.
- [14] Duru Ö. *Zaman Serileri Analizinde ARIMA Modelleri ve Bir Uygulama*. Yüksek Lisans Tezi, Sosyal Bilimler Enstitüsü, Türkiye, 2007.
- [15] Koç T, Dünder E, Koç H. "Forecasting of natural gas supplied for EU in Gazprom with ARIMA and machine learning methods". *International Journal of Data Science*, 7(1), 39–47, 2024.
- [16] Çuhadar M. *Turizm Sektöründe Talep Tahmini İçin Yapay Sinir Ağları Kullanımı ve Diğer Yöntemlerle Karşılaştırmalı Analizi (Antalya İlinde Uygulama)*. Doktora Tezi, Akdeniz Üniversitesi, Antalya, Türkiye, 2015. <https://doi.org/10.13140/RG.2.1.4503.7288>
- [17] Akın M. "A novel approach to model selection in tourism demand modeling". *Tourism Management*, 48, 64–72, 2015.
- [18] Kızıldağ M, Abut F, Akay MF. "Development of new electricity system marginal price forecasting models using statistical and artificial intelligence methods". *Applied Sciences*, 14(21), 10011, 2024. <https://doi.org/10.3390/app142110011>
- [19] Kaya MS, İnce K. "Nesnelerin internetinde çok katmanlı algılayıcı kullanarak zamanlama analizi saldırısı ile özel anahtar tahminlemesi". *Computer Science*, (Special), 385–390, 2021.
- [20] Kaynar O, Taştan S. "Zaman serisi analizinde MLP yapay sinir ağları ve ARIMA modelinin karşılaştırılması". *Erciyes Üniversitesi İİBF Dergisi*, (33), 161–172, 2009.
- [21] Öztemel E. *Yapay Sinir Ağları*. Birinci Baskı, İstanbul, Türkiye, Papatya Yayıncılık, 2003.
- [22] Gültepe Y. "Makine öğrenmesi algoritmaları ile hava kirliliği tahmini üzerine karşılaştırmalı bir değerlendirme". *European Journal of Science and Technology*, (17), 8–15, 2019. <https://doi.org/10.31590/ejosat.530347>
- [23] Koç T, Akın P. "Comparison of machine learning methods in prediction Gini coefficient for OECD countries". *International Journal of Data Science and Applications*, 4(1), 16–21, 2021.