

PAPER DETAILS

TITLE: TOPLU TASIMA YOLCU TALEBININ SARIMA VE PROPHET YÖNTEMLERİ İLE TAHMINI

AUTHORS: Gizem Küçüker,Özlem Uzun Araz

PAGES: 334-356

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/3077965>



TOPLU TAŞIMA YOLCU TALEBİNİN SARIMA VE PROPHET YÖNTEMLERİ İLE TAHMİNİ

Gizem KAYRAN^{1*}, Özlem UZUN ARAZ²

¹Manisa Celal Bayar Üniversitesi, Mühendislik Fakültesi,
Endüstri Mühendisliği Bölümü, Manisa

ORCID No : <https://orcid.org/0000-0002-4016-1541>

²Manisa Celal Bayar Üniversitesi, Mühendislik Fakültesi,
Endüstri Mühendisliği Bölümü, Manisa

ORCID No : <https://orcid.org/0000-0001-5630-6170>

Anahtar Kelimeler	Öz
<i>Talep tahmini Box-Jenkins Prophet</i>	<i>Taşıma sistemlerinde yolcu talebinin bilinmesi; rotaların belirlenmesi, sefer saatlerinin ve sıklığının planlanması, çalışan sayısının optimize edilmesi gibi birçok operasyon faaliyeti için hayati bir görevdir. Bununla birlikte talebin düzensizliği, trend ve mevsimsellik bileşenlerini barındırması, tahmin edilebilmesini zor bir hale sokmaktadır. Son zamanlarda geleneksel talep tahmin yöntemlerinin yanı sıra modern yapay zekâ teknikleri de kullanılmaya başlanmıştır. Bu çalışmada, literatürde sıklıkla tercih edilen Box-Jenkins metodolojisinden SARIMA yöntemi ile 2017’de Facebook tarafından geliştirilen Prophet yönteminin toplu taşıma araçlarındaki yolcu sayısı tahminindeki performanslarının karşılaştırılması ve yöntemlerin değerlendirilmesi amaçlanmıştır. İki yöntem kullanılarak başarılı birer tahmin modeli geliştirilmiştir. Modelleme sonucunda SARIMA ve Prophet yöntemlerinin MAPE değerleri sırasıyla 0,11 ve 0,13 olarak hesaplanmıştır.</i>

*Sorumlu yazar; e-posta : gizem.kucuker@cbu.edu.tr

doi : <https://doi.org/10.46465/endustrimuhendisligi.1281456>

FORECASTING PUBLIC TRANSPORTATION PASSENGER DEMAND WITH SARIMA AND PROPHET METHODS

Keywords	Abstract
Demand forecasting Box-Jenkins Prophet	<i>In transport systems, knowing passenger demand is a vital task for many operational activities, such as determining routes, planning schedules, and frequency, and optimizing the number of employees. However, the demand's irregularity, trend, and seasonality components make it challenging to forecast. Recently, modern artificial intelligence techniques have been used in addition to traditional demand forecasting methods. This study aims to compare the performance of the SARIMA method, which is one of the most preferred Box-Jenkins methodologies in the literature, and the Prophet method developed by Facebook in 2017, in forecasting the number of passengers in public transportation vehicles and to evaluate the methods. Two successful forecasting models were developed using the two methods. As a result of the modeling, the MAPE values of SARIMA and Prophet methods were calculated as 0.11 and 0.13, respectively.</i>
Araştırma Makalesi	Research Article
Başvuru Tarihi : 12.04.2023	Submission Date : 12.04.2023
Kabul Tarihi : 20.09.2023	Accepted Date : 20.09.2023

1. Giriş

Talep, müşterilerin belli bir zaman aralığında, belli bir fiyattan bir ürünü satın alma istekleri olarak tanımlanmaktadır. Planlama sürecinin sağlıklı bir şekilde yürütülebilmesi için müşteri taleplerinin önceden tahmin edilmesi gerekmektedir.

Geleceğe yönelik karar verme aşamasının en önemli adımı tahmindir. Gerçeğe yakın bir tahmin olmadan verilecek kararların tutarlı olması beklenemez. Gelecekteki olayların ya da değerlerin başarılı tahmini sayesinde işletmeler yapılması gerekenler için önceden hazırlıklı olacaklardır (Küçük Yılmaz, Kafalı, Işık, Yaman, Özen ve Küçük, 2019).

Ulaşım endüstrisinde yolcu talebinin belirlenmesi, rota, sefer sıklığı, personel sayısı gibi birçok önemli planın sağlıklı bir şekilde yapılabilmesi için önemli bir problemdir. Bu problemin gerçeğe yakın çözülmesi ile işletme maliyetleri azaltılıp hizmet kalitesi artırılabilir (Srisaeng, Baxter, Richardson ve Wild, 2015).

Ancak yolcu talebinin belirlenmesi kolay bir görev değildir. Çünkü yolcu talebi düzensizdir, durağan değildir ve mevsimsellik özelliği gösterir (Jin, Li, Sun ve Li, 2020). Dinamik bir endüstri olan ulaşım endüstrisinde, yüksek düzensizlik ve

belirsizlik planlama sürecini zorlaştırmakta, hataya açık hale getirerek yüksek bir risk oluşturmaktadır (Küçük Yılmaz ve diğ., 2019).

Yolcu talebinin düzensizliğinin yanı sıra küresel boyutta gerçekleşen ekonomik kriz, politik olaylar ya da salgın gibi beklenmeyen değişkenler de talebin tahmin edilebilmesini zor bir hale getirmektedir. 2020’de ortaya çıkan ve bu çalışmada kullanılan veri setini de etkileyen COVID-19 pandemisine karşı devletler sınırlarını kapatmış, insanlar gönüllü ya da zorunlu bir şekilde evlerinde kalmışlardır. Hem kısıtlamalar hem insanların toplu taşıma korkusu sebebiyle ulaşım talebinde ciddi bir düşüş meydana gelmiştir. Talepteki düşüşle birlikte personel ve sefer sayısının azaltılması, ekonomik ve fonksiyonel olarak zorunluluk haline dönüşmüştür.

Araştırmacılar geçmişten bu yana tüm bu belirsizlikleri azaltmak, tahmin performansını en iyileyen modeli bulmak için çeşitli çalışmalar yapmışlardır. Mevcut veriler ile geçmiş gözlem değerleri arasındaki ilişkiyi araştıran geleneksel zaman serisi yöntemlerinin yanı sıra daha karmaşık modeller için geliştirilen modern yapay zekâ teknikleri de sıkça kullanılmakta ve yeni modeller ortaya konulmaya devam etmektedir (Jin ve diğ., 2020).

Bu çalışmada, geleneksel bir yöntem olan ve literatürde sıklıkla tercih edilen Box-Jenkins yöntemi ile 2017’de Facebook tarafından geliştirilen Prophet yönteminin mevsimsel zaman serisi tahminindeki performanslarının karşılaştırılması ve nispeten yeni bir yöntem sayılan Prophet’in Box-Jenkins’e karşı avantajlarının ve dezavantajlarının bir uygulama üzerinden değerlendirilmesi amaçlanmıştır.

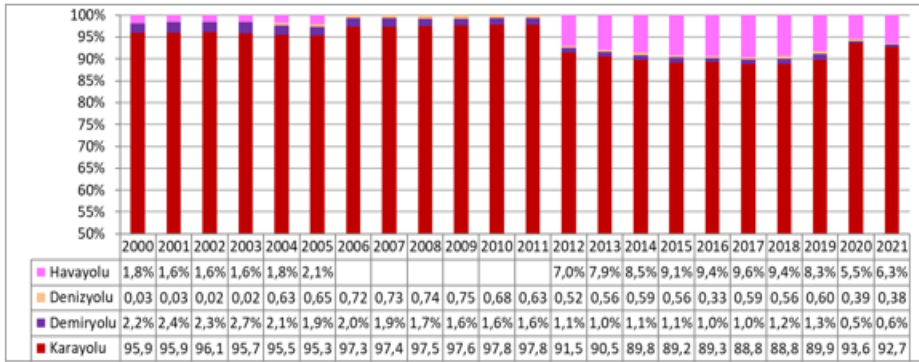
Çalışmanın sonraki bölümlerinde sırasıyla, ulaşım endüstrisinde yolcu talebinin önemi anlatılarak literatür açıklanacak, kullanılan materyal ve yöntemler tanıtılacak, zaman serisine Box-Jenkins ve Prophet yöntemi uygulanacak, öngörü yapılacak ve sonuçlar tartışılacaktır.

2. Ulaşım Endüstrisinde Yolcu Talebi

Ulaşım ve ulaştırma kavramları insanoğlunun ilk yıllarından beri var olan, insanlık devam ettiği sürece de var olacak kavramlardır (Küçük Yılmaz ve diğ., 2019). İnsanların gerek kendilerini gerekse yüklerini ve postalarını bir noktadan başka bir noktaya taşımaları insanlığın ve sosyal yaşamın gerekliliklerindendir.

Teknolojinin ve ulaştırma araçlarının gelişmesiyle birlikte ulaşım kişisel amaçlarla tek kişilik bir yolculuk olmaktan çıkarak farklı sistemleri bünyesinde barındıran bir sektör haline gelmiştir. Karayolu, havayolu, denizyolu ve demiryolu alt sistemlerinin tamamı ulaşım sistemini oluşturmaktadır.

Şekil 1’de Çevre, Şehircilik ve İklim Değişikliği Bakanlığı’ndan alınan 2000 ve 2021 yılları arasındaki ulaştırma türlerine göre yolcu taşıma oranları yer almaktadır.



Şekil 1. Yurt İçi Yolcu Taşıma Oranları (Çevre, Şehircilik ve İklim Değişikliği Bakanlığı)

Şekil 1'e bakıldığında karayolu ulaşım türünün yolcu taşıma oranları içerisinde en büyük paya sahip olduğu görülmektedir. Bununla birlikte 2000 yılında %1,8 olan havayolu yolcu payı 2017 yılında %9,6'ya çıkmıştır. 2020 ve 2021 yıllarında yaşanan pandemi sebebiyle yıllık bazda düşüş göstermiş ancak tüm ulaştırma türleri içerisindeki ikinciliğini korumuştur. Yurtiçinde olduğu gibi dünya çapında da kullanımı en yaygın ulaşım sistemi karayolu ulaşımıdır. Karayolu ulaşımının bu denli yoğun kullanılması iyi bir ulaşım sisteminin gerekliliğini ortaya koymaktadır.

Ulaşım yalnızca sosyal ve kültürel amaçlarla değil aynı zamanda ülke ekonomisine katkıları nedeniyle de büyük önem arz etmektedir. İç ve dış ticaretin geliştirilmesi, turizm faaliyetlerinin gerçekleştirilmesi, mal ve hizmetlerin taşınması ancak ulaşımın etkili ve verimli bir planlanması ve ulaştırma ağının oluşturulması ile mümkündür (Küçük Yılmaz ve diğ., 2019).

Planlama süreci tüm sistemlerin temelinde yer almaktadır. Planlama süreci, işletme amaçları ve bu amaçlara ulaşmak için gerekli araç ve yöntemlerin belirlenme aşamalarını kapsamaktadır (Gerede, Orhan, Orhan, Erdoğan ve Ateş, 2019). Ulaştırma sektörünün her bir alt sisteminin kendine özgü araç ve birimleri olmasına rağmen belli başlı ortak problemleri bulunmaktadır. Bunlar; araçların sefer saatleri ve sıklıklarının, rotaların, her vardiyada çalıştırılacak personel miktarları gibi operasyonel faaliyetlerin belirlenmesine yöneliktir. Tüm bu sorulara cevap aramadan önce talebin tahmin edilmesi gerekmektedir. Talep tahmininin gerçeğe yakın bir şekilde çözülmesi ile işletme amaçlarının gerçekleştirilmesi için önemli bir adım atılmış olacaktır. Bununla birlikte talep tahmini kolay bir süreç değildir. Bilet fiyatı, ekonomik koşullar, hizmet sunumu gibi değişkenlerin yanı sıra meteorolojik koşullar ya da SARS, 11 Eylül terör saldırısı gibi sıra dışı gelişen olaylar da talebi etkilemektedir.

Talep tahmininin gerçek değerlere en yakın şekilde tespit edilebilmesi için literatürde çeşitli yöntemler bulunmaktadır ve günümüzde yeni yöntemler oluşturulmaya devam edilmektedir.

Phyoe, Guo ve Zhong (2016), Gayrisafi Yurtiçi Hasıla (GSYİH) ile havayolu uçuş miktarı arasındaki ilişkiyi araştırmak amacıyla geleneksel yöntemlerden olan Üstel Trend, ARIMA ve ARIMAX yöntemlerini, Srisaeng ve diğ. (2015) ise, Avustralya'nın iç hat yolcu miktarlarını tahmin etmek için Genetik Algoritma yöntemini tercih etmişlerdir.

Makine öğrenimi yöntemlerinden Yapay Sinir Ağları'nın kullanıldığı çalışmalara ise Koç'un (2017) havayolu yolcu talebini tahminlemesi ve Liyanage, Abduljabbar, Dia ve Tsai'nin (2022) otobüs yolcularını tahminlemek için farklı Yapay Sinir Ağları modellerini karşılaştırdığı çalışmalar örnek verilebilir.

Farklı yöntemlerin özelliklerini birleştiren hibrit yöntemler de literatürde sıkça tercih edilmektedir. Dantas, Oliveira ve Repolho (2017), Bagging ve Holt Winters yöntemlerinin birleşiminden meydana gelen Bagging Holt Winters ile, Wei ve Chen (2012) ise, metro yolcularının tahmini için Ampirik Mod Ayrıştırması ve Yapay Sinir Ağları'nın birleşiminden oluşan hibrit modelle çalışmışlardır.

Bunlarla birlikte, geleneksel ve makine öğrenimi yöntemlerini karşılaştıran çalışmalar da mevcuttur. Halyal, Mulangi ve Harsha (2022) otobüs istasyonlarındaki yolcu sayılarını tahminlemek için Mevsimsel ARIMA(SARIMA) ve Yapay Sinir Ağları'nı, Nurmi (2021), COVID-19 pandemisi öncesi ve pandemi dönemindeki tahminleme yöntemlerinin performanslarını incelemek amacıyla SARIMA, TBATS, Prophet ve Yapay Sinir Ağları yöntemlerini karşılaştırmışlardır. Wen, Zhao, He, Ma ve Zhang (2022), demiryolu bilet satış miktarlarını SARIMA ile modellemiş, LSTM, Random Forest, XGBoost, GCN ve Decom yöntemleri ile sonuçlarını kıyaslamışlardır.

Tüm talep tahmin yöntemlerinin kendilerine özgü avantajları ve dezavantajları bulunmaktadır (Jin ve diğ., 2020). Araştırmacılar yöntemleri farklı veri setleriyle deneyerek, çeşitli yöntemlerle kıyaslayarak, yöntemlerin bu avantaj ve dezavantajlarını tespit ederek çözümler bulmaya çalışmışlardır. Bu çalışmada ise geleneksel yöntemlerden biri olan Box-Jenkins metodolojisinden SARIMA ve Facebook tarafından geliştirilen Prophet yöntemleri karşılaştırılacaktır. Bu çalışma ile sınırlı sayıda olan Prophet yönteminin literatürüne katkıda bulunulması ve araştırmacılara bu yeni yöntemin tanıtılması amaçlanmıştır.

3. Materyal

Ulaşım, binlerce yıldır var olan ve gelişen teknolojiyle birlikte yeni bir boyut kazanan bir kavramdır. Ulaşımın genel amacı, insanların ve yükün bir noktadan başka bir noktaya en kısa ve en az maliyetle taşınmasıdır. Ulaşım kapasitesini ekonomik ve toplumsal gelişmelerin yarattığı ihtiyaçları karşılayabilecek şekilde oluşturmak, ülke ve toplum çıkarları çerçevesinde hareket eden sistemleri

kurmak ve koordine etmek ise devletin görevlerindendir (Küçük Yılmaz ve diğ., 2019).

Çalışmada Balıkesir Büyükşehir Belediyesi'nin açık veri platformundan yararlanılmıştır. 2020 yılında vatandaşların kullanımı için yayına alınan "Açık Veri Portalı" ile belediye ve iştirakleri tarafından üretilen çeşitli kategorilerden verilere ulaşmak mümkündür. Ulaşım Planlama ve Raylı Sistemler Dairesi Başkanlığı tarafından platformda paylaşılan "BTT-Günlük Taşınan Yolcu Sayıları" veri setine ait 127 numaralı Erdek-Bandırma hattının verileri bu çalışmada kullanılmıştır.

Veri seti, Excel dosyası olarak indirilmiş, ardından tarih ve taşınan yolcu sayılarını içeren yeni bir tablo oluşturulmuştur. İlgili veri seti, 1 Nisan 2020 ile 31 Temmuz 2022 tarihleri arasındaki 757 günün yolcu değerlerini içermektedir.

Eksik veri analizi kapsamında veri seti içerisinde 77 günün verisinin girilmediği görülmüştür. Tarihler incelendiğinde, sokağa çıkma yasağının bulunduğu günlerde yolcu sayısının girilmediği fark edilmiş olup, o günlerde hattın durdurulduğu düşünülmektedir. Tahmin modellerini kötü yönde etkilememesi adına bu günler ve veri setine aykırılık gösteren değerler veri setinden çıkartılmıştır. Ardından yöntemlerde kullanılmak üzere Excel belgesi .csv formatına çevrilmiştir.

Uygulama aşamasında Anaconda platformunda yer alan Jupyter Notebook yazılımı kullanılmıştır. Kodlama dili olarak sade ve kolay anlaşılır yapısıyla sıklıkla tercih edilen Python dili seçilmiştir. Gerekli kütüphaneler uygulama yapılırken eklenmiştir.

4. Yöntem

Bu bölümde öncelikle Box-Jenkins ve Prophet yöntemlerinin matematiksel arka planı anlatılacak, ardından yöntemlerin değerlendirilmesine yardımcı olacak metrikler hakkında bilgi verilecektir.

4.1 Box-Jenkins Yöntemi

Zaman serileri, değişkenlerin geçmiş gözlemlerinin incelenerek aralarındaki ilişkinin analiz edildiği önemli bir tahmin alanıdır (Zhang, 2003). Zamana bağlı serilerin tahminlemede sıkça kullanılan yöntemlerden biri 1970 yılında Box ve Jenkins tarafından ortaya konulan Box-Jenkins yöntemidir. Yöntem alternatif modeller arasından en iyi modeli seçerek tahmin işlevini uygulamaya yöneliktir (Bozkurt, 2013).

Box-Jenkins yönteminde üç modelleme mevcuttur. Bunlar; Otoregresif (AR) süreci, Hareketli Ortalamalar (MA) ve hem AR(p) hem MA(q) bileşeninden oluşan Otoregresif Hareketli Ortalamalar (ARMA) sürecidir (Bozkurt, 2013).

AR(p), MA(q) ve ARMA(p,q) modelleri yalnızca durağan serilere uygulanmaktadır. Varyans ve kovaryansı sabit olmayan yani durağan olmayan serilerin durağan hale geçene kadar farkının alınması gerekmektedir. Belirli sayıda (d) farkı alınarak durağan hale getirilen zaman serilerine ise ARIMA(p,d,q) yöntemi uygulanmaktadır.

ARIMA(p,d,q) modellerinde, incelenen değişkenin bugünkü değeri, geçmiş değerlerinin ağırlıklı toplamı ile rassal şokların birleşiminden oluşmaktadır (Akgül, 2003).

Mevsimsel olmayan ARIMA(p,d,q) modeli

$$f(B)(1-B)^d y_t = c + q(B)\varepsilon_t \quad (1)$$

eşitliği ile ifade edilmektedir (Wen ve diğ., 2022).

Eşitlikteki $f(B)$; p dereceli otoregresyon parametresini, $q(B)$ ise q dereceli hareketli ortalama parametresini belirtmektedir (Akgül, 2003). Sırasıyla y_t ve ε_t gerçek değer ve rastgele hatalar olup, rastgele hataların ortalaması 0 ve varyansı σ^2 olan dağılımdan geldiği varsayılmaktadır (Zhang, 2003).

Belirli sayıda (d) farkı alınarak trend durağan hale gelen ancak korelasyonlar arasında anlamlı mevsimselliğin görüldüğü zaman serilerinde, ARIMA(p,d,q) modelleri yerine Mevsimsel ARIMA(p,d,q)(P,D,Q)_s yönteminin kullanılması önerilmektedir. Mevsimsel ARIMA(SARIMA) modeli

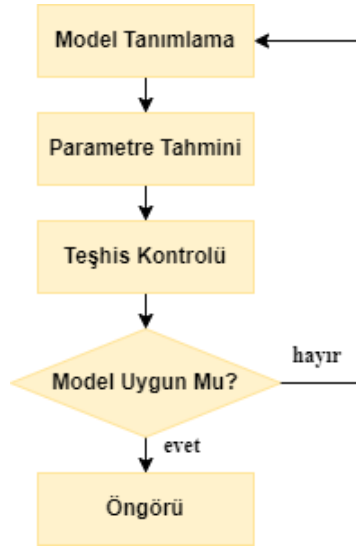
$$f(B)\Phi(B^s)(1-B)^d(1-B^s)^D y_t = c + q(B)\theta(B^s)\varepsilon_t \quad (2)$$

eşitliği ile ifade edilmektedir (Wen ve diğ., 2022).

Eşitlikteki $\Phi(B^s)$ mevsimsel otoregresyon parametresini, $\theta(B^s)$ ise mevsimsel hareketli ortalama parametresini belirtmektedir. ARIMA(p,d,q)(P,D,Q)_s modelinin uygulanabilmesi için; mevsimsel olmayan p,d,q ve mevsimsel P,D,Q parametrelerinin, ayrıca s mevsim periyodunun belirlenmesi gerekmektedir.

ARIMA modellerinde tüm parametrelerin doğru şekilde belirlenebilmesi için kullanılan zaman serisinin incelenmesi ve özelliklerinin bilinmesi gerekmektedir. Bu nedenle diğer zaman serisi yöntemlerine kıyasla ARIMA yöntemlerinde daha büyük bir çaba gösterilmesi gerekmektedir. Alternatif modeller arasından en iyi modeli seçmeye yönelik uygulanan yöntem için “model kurma stratejileri” geliştirilmiştir (Akgül, 2003). Box-Jenkins yönteminin akış şeması Şekil 2’de verilmiştir. Yöntem üç ana adımdan meydana gelmektedir; model tanımlama, parametre tahmini ve teşhis kontrolü (Zhang, 2003). İlk adım olan model tanımlama adımı, otokorelasyon fonksiyonu yardımıyla zaman serisinin özellikleri tespit edilerek uygun olabilecek geçici bir Box-Jenkins modeli seçilir. İkinci adımda, seçilen geçici modelin parametreleri belirlenir ve son olarak, belirlenen modelin seriye uygunluğu kontrol edilir. Geçici modelin istatistiksel test değerleri istenilen sonuçları veriyor ise tahmin aşamasına geçilir,

eğer vermiyorsa ilk adıma geri dönülür ve uygun model seçilene kadar döngü tekrarlanır.



Şekil 2. Box-Jenkins Metodunun Akış Şeması (Kovačić, 1995)

4.2 Prophet Yöntemi

Zaman serileri tahmininde kullanılan bir makine öğrenmesi modeli olan Prophet yöntemi, Facebook'un "Veri Bilimi" ekibi tarafından oluşturulan açık kaynak kodlu bir yazılımdır (Facebook, Prophet, Forecasting at Scale, 2023). Bu yöntem, özellikle iş dünyasında ve finans sektöründe talep tahmini, satış tahmini, stok yönetimi, bütçe tahmini, pazarlama planlaması gibi birçok alanda kullanılabilmektedir.

Klasik zaman serisi yöntemlerine kıyasla Prophet yönteminde eksik ve aykırı veriler başarılı bir şekilde kullanılabilmektedir. Prophet'in en büyük avantajlarından biri ise, kullanıcının sezgisel parametreler kullanarak tahminleme modelini hızlı bir şekilde oluşturabilmesidir. Bu sayede, istatistik bilgisi olmayan kullanıcılar bile Prophet'i kullanarak başarılı tahminler yapabilirler (Taylor ve Letham, 2017).

Yöntem, Python ve R yazılımlarına Prophet kütüphanesi ile aktararak hızlı bir şekilde kullanılabilmektedir. Prophet, zaman serisi verilerini yumuşatmak ve

tahmin etmek için Bayes tabanlı bir eğri uydurma yöntemi uygulamaktadır (Zhao, Liu, Vanos ve Cao, 2018).

Prophet'in üç ana bileşenden oluşan eşitliği Eşitlik 3'te verilmiştir.

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \quad (3)$$

Eşitlikte yer alan $g(t)$; trend fonksiyonunu, $s(t)$; mevsimsellik etkisini, $h(t)$; tatillerin etkilerini temsil etmektedir. Hata terimi $\varepsilon(t)$ ise model tarafından ana bileşenlerle bağdaştırılamayan değişkenleri belirtmektedir (Taylor ve Letham, 2017).

Ayrıca Prophet, ön tanımlı parametrelerin değiştirilmesine ve özelleştirilmesine de imkân vermektedir. Örneğin, haftalık (weekly_seasonality) ve yıllık (yearly_seasonality) örüntüler mevsimsellik parametreleri ile ayarlanabilir (Navratil ve Kolkova, 2019).

Mevsimsellikler, kısmi bir Fourier toplamı kullanılarak tahmin edilmektedir. Yıllık mevsimsellik için varsayılan Fourier serisi 10 olarak ayarlıdır ancak mevsimselliğin daha yüksek frekanslı değişikliklere uyması gerektiğinde artırılabilir (Facebook, Prophet, Forecasting at Scale, 2023).

Prophet yönteminin işleyişindeki sezgisel yaklaşım, zaman serisi tahminlerini daha anlaşılır ve anlaması kolay hale getirmektedir. Bu yaklaşım, verinin detaylarına boğulmadan, daha yüksek doğrulukta tahminler oluşturmayı sağlamaktadır.

4.3 Performans Metrikleri

Tahmin yöntemlerinin performanslarının karşılaştırılması amacıyla çeşitli istatistikler kullanılmaktadır. Bu çalışmada, üç farklı performans metriğinden faydalanılmıştır. Sırasıyla Ortalama Mutlak Yüzde Hata (MAPE), Hata Kareler Ortalamasının Karekökü (RMSE) ve Belirleme Katsayısı (R²) metriklerinin eşitlikleri (Eşitlik 4-5-6) verilmiştir (Chicco, Warrens ve Jurman, 2021).

$$MAPE = \frac{1}{m} \sum_{i=1}^m \left| \frac{Y_i - X_i}{Y_i} \right| \quad (4)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (X_i - Y_i)^2} \quad (5)$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2} \quad (6)$$

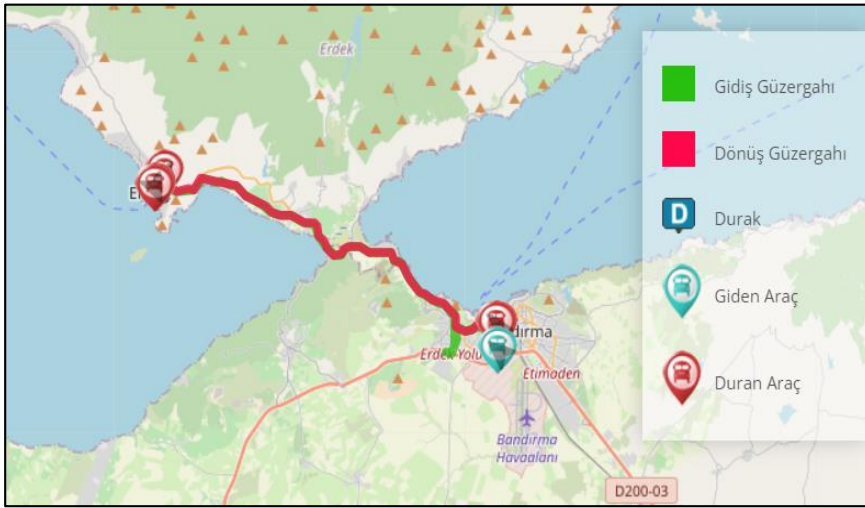
Eşitlikteki m ; örneklem büyüklüğünü, Y_i ; gerçek değerleri, X_i ; tahmin edilen değerleri, $\bar{Y}$ ise gerçek değerlerin ortalamasını belirtmektedir.

Çalışmada, veri setine ilk olarak Box-Jenkins ve Prophet yöntemleri Jupyter Notebook üzerinden Python programlama dili kullanılarak uygulanacak, ardından MAPE, RMSE ve R2 metrikleri ile yöntemlerin tahmin performansları değerlendirilecek ve yorumlanacaktır.

Bu çalışmada araştırma ve yayın etiğine uyulmuştur.

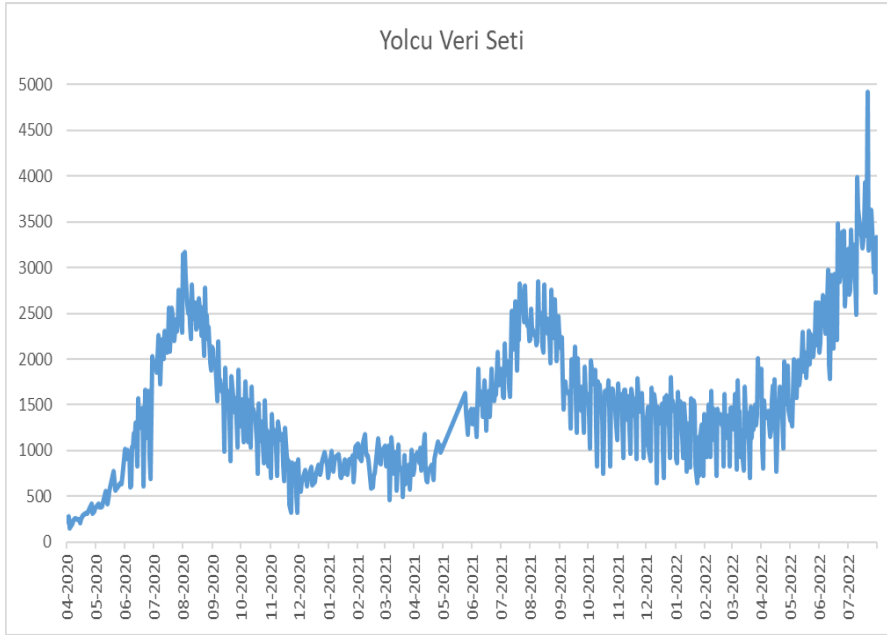
5. Uygulama

Çalışmada, Balıkesir Büyükşehir Belediyesi'ne bağlı 127 numaralı Erdek-Bandırma hattına ait veriler kullanılmıştır. Erdek-Bandırma hattı, Erdek ve Bandırma ilçelerine ulaşımı sağlayan ilçeler arası otobüs hattıdır. Hat güzergahı Şekil 3'te yer almaktadır.



Şekil 3. Erdek-Bandırma Hat Güzergahı
(<https://www.balikesirulasim.com.tr/hat/ERDEK-BANDIRMA>)

Erdek-Bandırma hattı şehirdeki yoğun kullanılan hatlardan biri olması sebebiyle bu çalışmada kullanılmak üzere tercih edilmiştir. Açık veri platformundan (<https://www.balikesirulasim.com.tr/hat/ERDEK-BANDIRMA>) alınan 2020 Nisan ayı ile 2022 Temmuz ayları arasındaki 757 günün yolcu değerlerine ilişkin zaman grafiği Şekil 4'te verilmiştir.



Şekil 4. Erdek-Bandırma Hattının Yolcu Sayısı Grafiği

Veri setine ait grafik incelendiğinde yolcu sayılarının kış ve ilkbahar aylarında yıl içindeki en düşük değerlerini aldığı, en yüksek değerlerine ise yaz ve sonbahar aylarında ulaştığı görülmektedir. Grafikten zaman serisinde belirgin bir mevsimsellik etkisi bulunduğu söylenebilir.

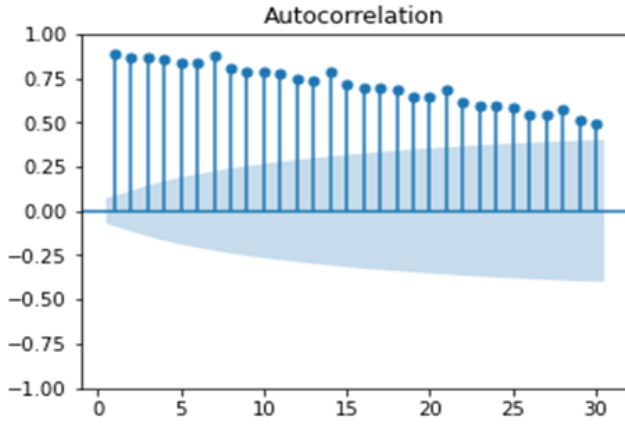
Grafik analizinin ardından tahminleme modeli için sırasıyla Box-Jenkins ve Prophet yöntemleri uygulanacaktır.

5.1 Box-Jenkins Yöntemi

Model Tanımlama

Box-Jenkins yöntemi için zaman serisinin ayrıntılı incelemesi ve seri özelliklerinin tespiti gerekmektedir. İlk adım olan model tanımlama adımında öncelikle serinin durağanlık koşulları hakkında bilgi alınacaktır. Durağanlık hakkında bilgi almak için zaman serisinin otokorelasyon fonksiyonuna bakılmaktadır. Otokorelasyon fonksiyonu (ACF), sürecin geçmiş dönemlerle ilişkisini göstermektedir. Bu nedenle gözlemler arasındaki ilişkinin modellenmesinde önemli bir rol oynamaktadır. Seride birim kök bulunuyor ise durağanlaştırmak için öncelikle birinci dereceden farkının alınması gerekmektedir.

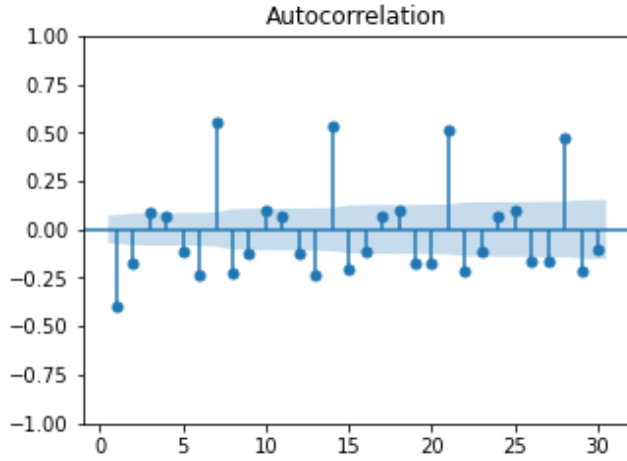
Şekil 5'te Python'un statsmodels kütüphanesindeki plot_acf modülü ile oluşturulan ACF grafiği verilmiştir.



Şekil 5. Yolcu Serisinin ACF Grafiği

Zaman serisine ait ACF grafiğine bakıldığında otokorelasyon değerlerinin yüksek olduğu görülmektedir. Değerler yavaş bir şekilde azalmakta ve sıfır eksenine yakın zamanda kesilmemektedir. Bu bilgiler zaman serisinin durağan olmadığını göstergesidir; dolayısıyla incelenen zaman serisi durağan değildir denilebilir.

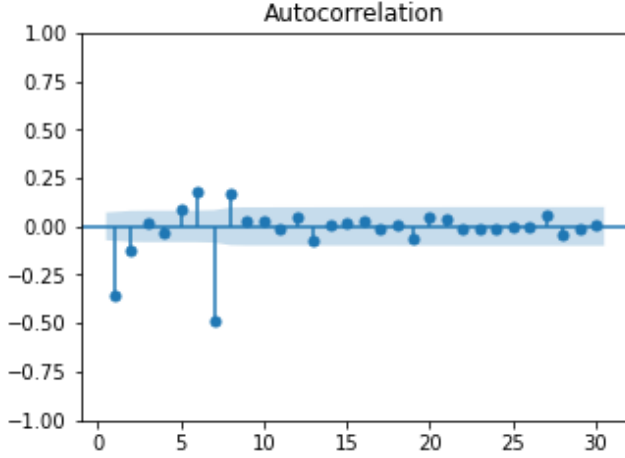
Model tanımlama adımında durağan olmadığı bilinen zaman serisinin kaçınıcı mertebeden birim köke sahip olduğunun da tespit edilmesi gerekmektedir. Bu sayede Box-Jenkins modeli de belirlenecektir. Bu amaçla, durağan olmayan yolcu zaman serisinin birinci mertebeden farkı alınmıştır. Farkı alınan serinin yeni ACF grafiği Şekil 6'da verilmiştir.



Şekil 6. Birinci Derece Farkı Alınmış Yolcu Serisinin ACF Grafiği

Birinci dereceden farkı alınmış serinin trend durağan hale geldiği görülmektedir. Ancak gecikmeler arasındaki ilişki devam etmektedir. Serinin grafiğinin yorumlanma aşamasında görüldüğü üzere kullanılan zaman serisi, mevsimsel bir zaman serisidir. Mevsimsel bir zaman serisi olduğunun belirlenmesinin ardından Box-Jenkins metodolojisinden $SARIMA(p,d,q)(P,D,Q)_s$ yönteminin kullanılmasına karar verilir. Ayrıca gecikmeler 7 günde bir zirve yaptığı için mevsim periyodu s , 7 olarak belirlenir.

Zaman serisinin mevsimsel derecesinin bulunması için mevsimsel farkı alınır. Birinci dereceden trend farkı ve birinci dereceden mevsimsel farkı alınan serinin ACF grafiği Şekil 7’de verilmiştir.

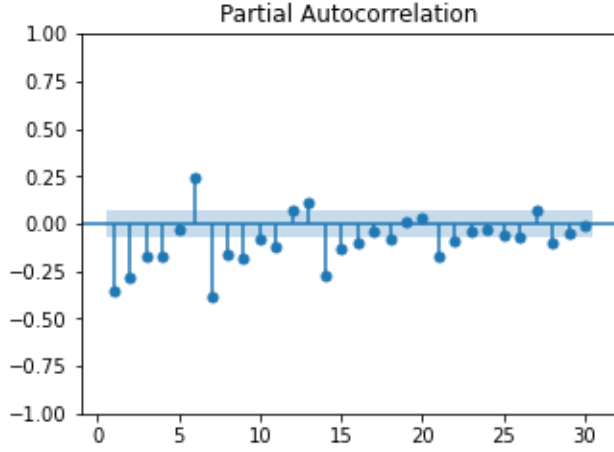


Şekil 7. Mevsimsel Farkı Alınan Zaman Serisinin ACF Grafiği

Şekil 7'ye bakıldığında trend farkı ve birinci dereceden mevsimsel farkı alınan zaman serisinin gecikmeleri arasındaki ilişkinin sona erdiği ve serinin durağan hale geldiği görülmektedir. $ARIMA(p,d,q)(P,D,Q)_s$ modelinin trend ve mevsimsel; d ve D gecikme değerleri 1 olarak belirlenir.

Parametre Tahmini

Box-Jenkins yönteminin parametre tahmini adımı Otoregresif (p) ve Hareketli Ortalamalar (q) parametreleri belirlenecektir. Tahmin modelinin p ve q parametrelerinin belirlenmesi aşamasında da otokorelasyon ve kısmi otokorelasyon (PACF) grafiklerine bakılmaktadır. Şekil 8'de durağan hale getirilmiş serinin PACF grafiği verilmiştir.



Şekil 8. Birinci Derece Farkı Alınmış Yolcu Serisinin PACF Grafiği

Şekil 7 ve Şekil 8 incelendiğinde birden fazla otokorelasyon değerlerinin anlamlı olduğu görülmektedir. Bu bağlamda p ve q değerlerinin sıfırdan farklı olduğu söylenebilir. En uygun parametreleri belirlemek için deneme yapmak gerekmektedir ancak bunun yerine `pmdarima` paketindeki `auto_arima()` fonksiyonu kullanılarak da model hızlı bir şekilde belirlenebilir.

`Auto_arima` fonksiyonu; çeşitli ARIMA modellerini otomatik olarak deneyerek modellerin Akaike Bilgi Kriteri(AIC) değerlerine göre veri setine en uygun modeli belirlemektedir. AIC, istatistiksel modellerin uyum iyiliğini göstermektedir. Daha küçük bir AIC değeri, daha iyi bir model anlamına gelmektedir. Tablo 1’de `auto_arima()` fonksiyonunun denediği ve AIC değerlerine göre sıralanmış SARIMA(p,d,q)(P,D,Q)_s modelleri yer almaktadır.

Tablo 1

SARIMA Modellerinin AIC Değerleri

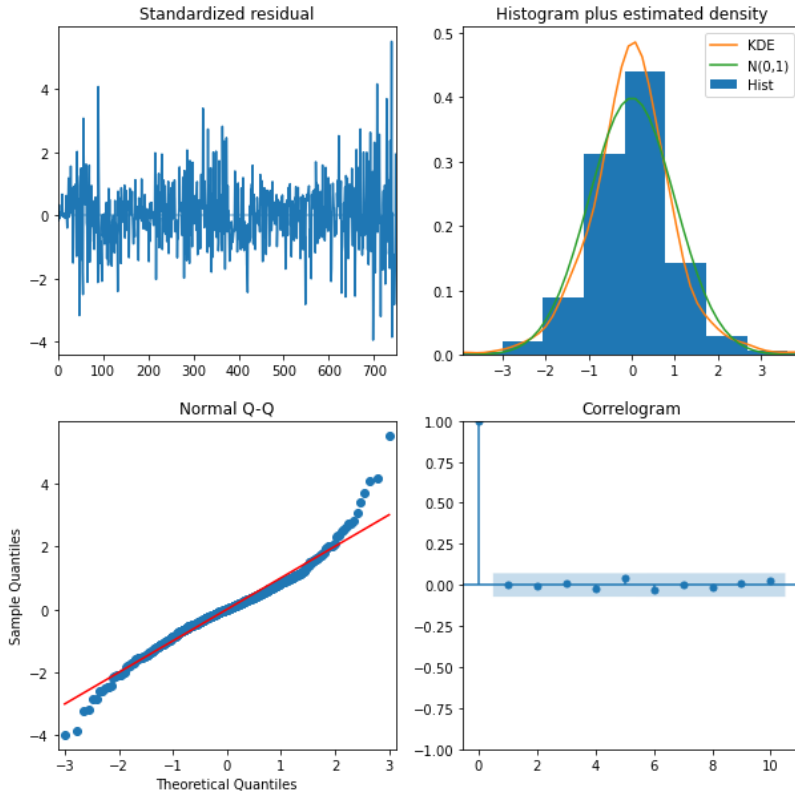
SARIMA(p,d,q)(P,D,Q)	AIC
SARIMA(1,1,1)(0,1,1)	10184.49
SARIMA(0,1,2)(0,1,1)	10184.54
SARIMA(1,1,1)(1,1,1)	10186.36
SARIMA(1,1,1)(0,1,2)	10186.37
SARIMA(2,1,1)(0,1,1)	10186.43
SARIMA(2,1,2)(0,1,2)	10187.91
SARIMA(2,1,2)(1,1,1)	10187.96
SARIMA(2,1,1)(0,1,2)	10188.32
SARIMA(0,1,1)(0,1,1)	10193.53
SARIMA(2,1,0)(0,1,1)	10240.22
SARIMA(1,1,0)(0,1,1)	10291.63
SARIMA(1,1,1)(1,1,0)	10318.95
SARIMA(0,1,0)(0,1,1)	10406.83
SARIMA(1,1,0)(1,1,0)	10443.81
SARIMA(2,1,2)(0,1,0)	10498.29
SARIMA(1,1,1)(0,1,0)	10499.85
SARIMA(0,1,0)(0,1,0)	10756.99

Auto_arima() fonksiyonu ile 17 farklı SARIMA modeli denenmiştir. AIC değerlerine göre en iyi sonucu veren model 10184.49 ile SARIMA(1,1,1)(0,1,1)₇ olarak bulunmuş ve SARIMA(1,1,1)(0,1,1)₇ modeli geçici olarak seçilmiştir.

Teshis Kontrolü

Model belirlendikten sonra seriye uygunluğunu görmek amacıyla kontrol testlerinden geçmesi gerekmektedir. Kontrol testlerinden geçilmesinin ardından geçici modelin kullanılmasına ya da değiştirilmesine karar verilecektir.

Geçici olarak seçilen modelin Statsmodels kütüphanesinde bulunan `plot_diagnostics()` fonksiyonu ile oluşturulan teşhis grafikleri Şekil 9'da yer almaktadır.



Şekil 9. ARIMA(1,1,1)(0,1,1)₇ Modelinin Teşhis Grafikleri

Teşhis grafikleri incelenmiş ve modelin artıklarının 0 etrafında dağıldığı görülmüştür. Ayrıca tüm geçmiş dönemlerin otokorelasyonu sıfırdır. Elde edilen bu sonuçlar, geliştirilen tahmin modelinin zaman serisine uygunluğunu göstermektedir. Elde edilen sonuçlara göre, zaman serisine ARIMA(1,1,1)(0,1,1)₇ modelinin uygulanmasına karar verilmiştir. Öngörü için statsmodels kütüphanesindeki ARIMA modülü kullanılmıştır.

5.2 Prophet Yöntemi

Zaman serisine uygulanacak ikinci yöntem, Prophet yöntemidir.

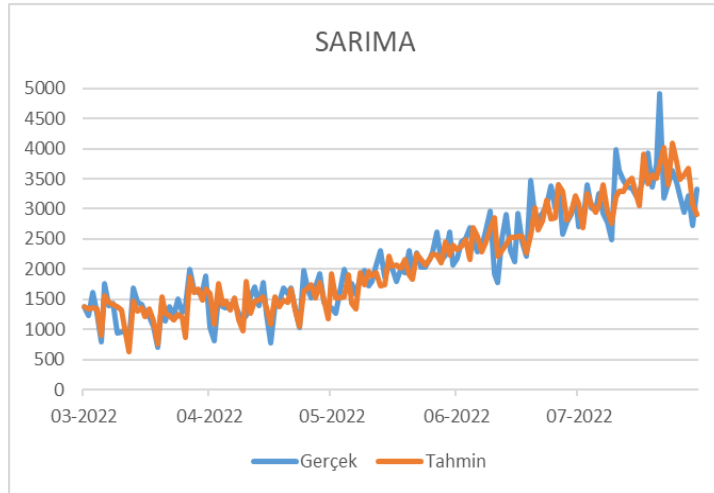
Yöntem için öncelikle gerekli kütüphaneler ve veri seti içeri aktarılmıştır. Ardından yöntemin uygulanabilmesi için gerekli olan veri dönüşümü yapılmış, veri setinin zaman sütunu “ds” ve yolcu değerlerinin bulunduğu sütun “y” olarak yeniden adlandırılmıştır.

Tahmin modelindeki ön tanımlı mevsimsellik parametreleri yerine “yearly_seasonality=30, weekly_seasonality=True” yazılmış, mevsimselliğin zaman içindeki değişimlerinin ve zaman serisine etkisinin modellenmesi sağlanmıştır.

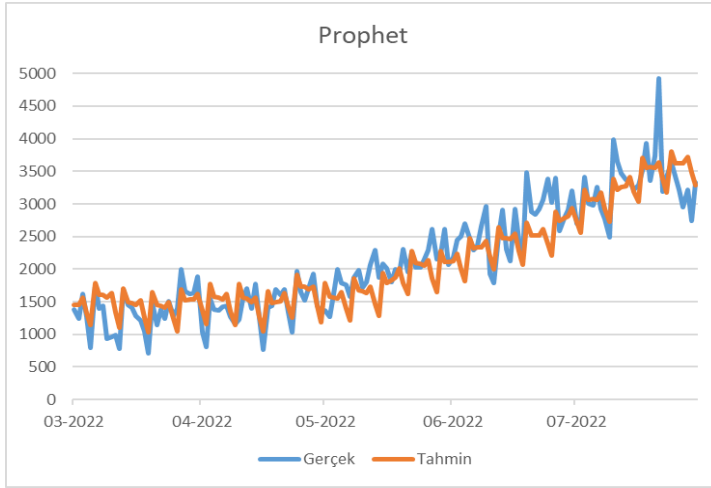
6. Sonuçlar

Öngörü için veri seti %80-%20 oranlarında eğitim ve test kümelerine ayrılmıştır. Bununla birlikte, 1 Nisan 2020 ve 1 Mart 2022 tarihleri arasındaki 605 veri eğitim, 2 Mart 2022 ve 31 Temmuz 2022 tarihleri arasındaki 152 veri ise test amacıyla kullanılmıştır.

SARIMA ve Prophet modelleri eğitim veri seti ile eğitilmiş, test veri seti ile tahmin değerleri karşılaştırılmıştır. Şekil 10’da SARIMA yönteminin, Şekil 11’de ise Prophet yönteminin öngörü grafikleri verilmiştir.



Şekil 10. SARIMA ile Test Verisinin Tahmini



Şekil 11. Prophet ile Test Verisinin Tahmini

SARIMA öngörü grafiğinde tahmin değerleri gerçek değerlere çok yakın bir sonuç göstermiştir; ancak gerçek değerlerle yüzde yüz uyuşmamaktadır. Prophet ise gerçek değerlere yakın ama SARIMA'dan daha kötü bir sonuç çizmiştir. Prophet, hızla değişen dinamiklere ayak uyduramamıştır ancak SARIMA modeli gibi Prophet modeli de trendi başarılı bir şekilde takip edebilmiştir. Modeller sonucunda oluşturulan MAPE, RMSE ve R^2 değerleri Tablo 2'de verilmiştir.

Tablo 2
Yöntemlerin Performanslarının Karşılaştırılması

	SARIMA	Prophet
MAPE	0,11	0,13
RMSE	303,21	330,66
R^2	0,87	0,84

Yöntemler MAPE, RMSE ve R^2 performans metriklerine göre karşılaştırılmıştır. SARIMA yönteminin Prophet yöntemine kıyasla MAPE ve RMSE değerleri daha küçük, R^2 değeri ise daha büyük bulunmuştur. Bu sonuçlar SARIMA yönteminin kullanılan veri setinde daha başarılı olduğunu göstermektedir. SARIMA daha iyi bir sonuç göstermesine rağmen, Prophet yöntemi de gerçek değerlere yakın ve başarılı sayılan bir sonuç göstermiştir. Prophet yönteminin hata değerlerinin daha yüksek olmasının sebebi algoritma tarafından otomatik olarak tanımlanan

parametreler olabilir. Buna karşın SARIMA yönteminde, veri setinin özellikleri incelenerek parametre tanımı elle yapılmıştır.

7. Tartışma

Bu çalışmada Balıkesir Büyükşehir Belediyesi'nin 127 numaralı Erdek-Bandırma hattına ait yolcu talebinin modellenmesi amacıyla Box-Jenkins ve Prophet yöntemleri kullanılmıştır. Box-Jenkins yöntemi uygulanırken, korelogramlara bakılarak zaman serisinin durağanlığı araştırılmış, farkı ve mevsimsel farkı alınarak d ve D parametreleri belirlenmiştir. Mevsimsel bir zaman serisi olduğu belirlenerek SARIMA modelinin uygulanmasına karar verilmiştir. Pmdarima paketindeki `auto_arima()` fonksiyonu ile en iyi otoregresif ve hareketli ortalamalar parametreleri seçilmiştir. Teşhis grafikleri ile modelin uygunluğu kontrol edilmiştir.

Prophet yönteminde veri dönüşümü uygulanmış; sütunlar ds ve y olarak yeniden adlandırılmıştır. Mevsimselliğin daha yüksek frekanslı değişikliklere uyması için ön tanımlı mevsimsellik parametreleri değiştirilmiştir.

Veri seti 80:20 oranında eğitim ve test amacıyla ayrılmıştır. Eğitim veri seti ile eğitilen modeller ile 152 günlük öngörü yapılmıştır. Gerçek değerler ve tahmin modelleri ile yapılan öngörüler karşılaştırılmıştır. SARIMA ve Prophet yöntemlerinin MAPE değerleri 0,11 ve 0,13, RMSE değerleri 303,21 ve 330,66 ve R^2 değerleri 0,87 ve 0,84 olarak bulunmuştur. MAPE değeri 0,1 ve 0,2 arasında olan modeller başarılı olarak kabul edilmektedir. Buna göre iki yöntem de kullanılan zaman serisi ile başarılı bir performans göstermişlerdir.

Box-Jenkins yöntemi Prophet yöntemine nispeten daha uzun uğraşlar, istatistiksel analizler ve analiz tecrübesi gerektirmektedir. Oysa Prophet, otomatik olarak sunduğu sezgisel parametreler ile analistlerin yanı sıra zaman serisi tahmini konusunda az deneyimi olan ya da hiç deneyimi olmayan kişilerin de bu yöntemi kullanmasını mümkün hale getirmektedir.

Bununla birlikte, Prophet yöntemi ile çok daha kısa sürede başarılı sonuca ulaşılmasına rağmen tahmin performansı Box-Jenkins yönteminin gerisinde kalmaktadır.

Sonuç olarak, kullanılan iki yöntemin de kendilerine özgü avantajları ve dezavantajları bulunmaktadır. Prophet nispeten yeni bir yöntem olmasına rağmen klasik SARIMA yöntemine çok yakın, başarılı bir sonuç göstermiştir. Zamanla açık kaynak yöntemin daha da iyileştirilebileceği düşünülmektedir.

Bu çalışmada kullanılan veri seti COVID-19'un getirdiği ekstrem koşullardan etkilenmiştir. Farklı bir veri seti ile yöntemlerin sonuçları da farklılık gösterecektir. Sonraki öngörü araştırmalarında Prophet yönteminin de kullanılan yöntemlere eklenmesi tavsiye edilmektedir.

Araştırmacıların Katkısı

Bu araştırmada; Gizem KAYRAN, bilimsel yazın taraması, istatistiksel analizler, modellerin oluşturulması, sonuçların yorumlanması ve makale yazımı; Özlem UZUN ARAZ, çalışmanın yönetilmesi, içerik akışının oluşturulması ve makale yazımı konularında katkı sağlamıştır.

Çıkar Çatışması

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir.

Kaynaklar

- Akgül, I. (2003). *Zaman serilerinin analizi ve ARIMA modelleri*. İstanbul: DER Yayınevi.
- Küçük Yılmaz, A., Kafalı, H., Işık, Y., Yaman, K., Özen, M., Küçük, F. (2019). *Hava taşımacılığı*. Eskişehir: Açıköğretim Fakültesi Yayını, No: 2204.
- Gerede, E., Orhan, G., Orhan, İ., Erdoğan, D., Ateş, S.S. (2019) *Havayolu işletmelerinde operasyonel planlama*. Eskişehir: Açıköğretim Fakültesi Yayını, No: 2649.
- Bozkurt, H. (2013). *Zaman serileri analizi*. Bursa: Ekin Yayınevi.
- Chicco, D., Warrens, M.J. & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput. Doi: <https://doi.org/10.7717/peerj-cs.623>*
- Dantas, T., Oliveira, F. & Repolho, H. (2017). Air transportation demand forecast through Bagging Holt Winters methods. *Journal of Air Transport Management*, 59, 116-123. Doi: <https://doi.org/10.1016/j.jairtraman.2016.12.006>
- Halyal, S., Mulangi, R. H. & Harsha, M. M. (2022). Forecasting public transit passenger demand: with neural networks using APC data. *Case Studies on Transport Policy*, 10, 965-975. Doi: <https://doi.org/10.1016/j.cstp.2022.03.011>
- Jin, F., Li, Y., Sun, S. & Li, H. (2020). Forecasting air passenger demand with a new hybrid ensemble approach. *Journal of Air Transport Management*, 83. Doi: <https://doi.org/10.1016/j.jairtraman.2019.101744>
- Koç, İ. (2017). *Türkiye’de iç hatlarda havayolu yolcu taleplerinin yapay sinir ağları kullanarak tahmini* (Yüksek Lisans Tezi). İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Kovačić, Z. (1995). *The time series analysis*. University of Belgrade. Faculty of Economics, Belgrade.

- Liyanage, S., Abduljabbar, R., Dia, H. & Tsai, P. (2022). AI-based neural network models for bus passenger demand forecasting using smart card data. *Journal of Urban Management*, 11, 365–380. Doi: <https://doi.org/10.1016/j.jum.2022.05.002>
- Navratil, M. & Kolkova, A. (2019). Decomposition and forecasting time series in business economy using Prophet forecasting model. *Central European Business Review*, 8(4). Doi: <https://doi.org/10.18267/j.cebr.221>
- Nurmi, N. (2021). *Forecasting airport passenger traffic in the era of COVID-19 pandemic* (Yüksek Lisans Tezi). Lappeenranta-Lahti University of Technology, School of Business and Management, Finlandiya.
- Phyoe, S.M., Guo, R. & Zhong, Z.W. (2016). An air traffic forecasting study and simulation. *Matter: International Journal of Science and Technology*, 2(3), 55–69. Doi: <https://doi.org/10.20319/Mijst.2016.23.5569>
- Srisaeng, P., Baxter, G., Richardson, S. & Wild, G. (2015). A forecasting tool for predicting Australia's domestic airline passenger demand using a genetic algorithm. *Journal of Aerospace Technology and Management*, 7(4), 476–489. Doi: <https://doi.org/10.5028/jatm.v7i4/475>
- Taylor, S.J. & Letham, B. (2017). Forecasting at scale. *PeerJ Preprints*. Doi: <https://doi.org/10.7287/peerj.preprints.3190v2>
- Wei, Y. & Chen, M. (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. *Transportation Research, Part C* 21, 148–162. Doi: <https://doi.org/10.1016/j.trc.2011.06.009>
- Wen, K., Zhao, G., He, B., Ma, J. & Zhang, H. (2022). A decomposition-based forecasting method with transfer learning for railway short-term passenger flow in holidays. *Expert Systems with Applications*, 189. Doi: <https://doi.org/10.1016/j.eswa.2021.116102>
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159 – 175. Doi: [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- Zhao, N., Liu, Y., Vanos, J. & Cao, G. (2018). Day-of-week and seasonal patterns of pm2.5 concentrations over the United States: time-series analyses using the Prophet procedure. *Atmospheric Environment*, 192, 116–127. Doi: <https://doi.org/10.1016/j.atmosenv.2018.08.050>
- Facebook, Prophet, Forecasting at scale. Erişim adresi: <https://facebook.github.io/prophet/> (Erişim tarihi: 02/02/2023)
- Balıkesir Büyükşehir Belediyesi, Açık Veri Platformu, BTT-günlük taşınan yolcu sayıları. Erişim adresi: <https://acikveri.balikesir.bel.tr/Veriler/VeriSetiTabloGorunum?vsadi=asis-gunluk-yolcu> (Erişim tarihi: 20/09/2022)

T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, Çevresel göstergeler. Erişim adresi: <https://cevresehgostergeler.csb.gov.tr/> (Erişim tarihi: 04/04/2023)

Balıkesir Büyükşehir Belediyesi, Otobüs ve güzergâh bilgisi. Erişim adresi: <https://www.balikesirulasim.com.tr/hat/ERDEK-BANDIRMA> (Erişim tarihi: 10/04/2023)