

DESTEK VEKTÖR REGRESYONU, RIDGE REGRESYON VE ÇOKLU DOĞRUSAL REGRESYON YÖNTEMLERİYLE TURİZM TALEP TAHMİNİ

TOURISM DEMAND FORECASTING WITH SUPPORT VECTOR REGRESSION, RIDGE REGRESSION, AND MULTIPLE LINEAR REGRESSION METHODS

Serkan KARDEŞ^{ID}
Burcu ÖNGEN BİLİR^{ID}

Öz

Günümüzde değişen ve küreselleşen dünya koşulları, insan ihtiyaçlarını ve taleplerini şekillendirmektedir. Değişen koşullar, işletmelerin faaliyetlerini ve gelecek planlamalarını gerçekleştirmek için zorluklar oluşturmaktadır. İşletmeler bu zorlukların üstesinden gelebilmek için doğru ve güvenilir tahminler yapması gerekir. Son yıllarda makine öğrenmesi gibi gelişmiş veri analiz yöntemleri, işletmelerin daha doğru tahminler yapabilmelerine yardımcı olmaktadır. Bu bağlamda çalışmada geleneksel yöntemlerden çoklu doğrusal regresyon yöntemi ile makine öğrenmesi tekniklerinden destek vektör regresyonu ve ridge regresyon yöntemleri kullanılarak bu yöntemlerin tahmin güçleri karşılaştırılmıştır. Turizm talebinin tahmin edildiği çalışmada 2004-2019 dönemi için aylık veriler kullanılmaktadır. Literatürde turizm talebi tahmin edilirken genellikle talep yönü incelenirken çalışmada turizmin konaklama kapasiteleri ve acente sayıları gibi arz yönü modele dahil edilmiştir. Analizler sonucu en iyi tahmin gücüne sahip yöntemin destek vektör regresyonu olduğu sonucuna ulaşılmıştır. Turizm sektöründeki yöneticiler için karar verme ve planlama süreçlerinde geleneksel tahmin yöntemleri yerine makine öğrenmesi yöntemlerinin kullanılabilmesi ortaya koyulmaktadır.

Anahtar Kelimeler: Çoklu doğrusal regresyon, destek vektör regresyonu, ridge regresyon, turizm talep tahmini.

JEL Sınıflandırılması: Z3, C13, C45, C53

* **Sorumlu Yazar:** Arş. Gör., Bursa Teknik Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi, İşletme Bölümü, serkan.kardes@btu.edu.tr, ORCID: 0000-0002-4021-1975.

** Doç. Dr., Bursa Teknik Üniversitesi, İnsan ve Toplum Bilimleri Fakültesi, İşletme Bölümü, burcu.bilir@btu.edu.tr, ORCID: 0000-0003-2581-0125.

To cite this article: Kardeş, S. & Öngen Bilir, B. (2024). Destek vektör regresyonu, ridge regresyon ve çoklu doğrusal regresyon yöntemleriyle turizm talep tahmini. *Journal of Research in Business*, 9(1), 194-218. DOI: 10.54452/jrb.1395182

Abstract

Changing and globalizing world conditions have been shaping human needs and demands recently. These changing conditions generate challenges for businesses to carry out their activities and future planning. Businesses need to make accurate and reliable predictions in order to overcome these challenges. In recent years, advanced data analysis methods such as machine learning have been helping businesses to make more accurate predictions. In this study, the forecasting powers of these methods were compared using the multiple linear regression method from traditional methods and the support vector regression and ridge regression methods from machine learning techniques. In the study, which estimates tourism demand, for the 2004-2019 period monthly data are used. While in the literature, the demand aspect of tourism is usually examined when it comes to demand forecasting, in this study, the supply aspect such as tourism accommodation capacities and agency numbers was included in the model. As a result of the analysis, it was concluded that the support vector regression method has the best forecasting power. It is revealed that machine learning methods can be used instead of traditional forecasting methods in decision-making and planning processes for tourism sector managers.

Keywords: Multiple linear regression, ridge regression, support vector regression, tourism demand forecasting

JEL Classification: Z3, C13, C45, C53

Extended Summary

Text Tourism has become one of the fastest-growing sectors in the world economy. In 2022, global tourism revenues reached 2 trillion US dollars. This revenue constitutes approximately 10% of the world trade (World Tourism Organization, 2023). In addition to being a social-cultural activity area for society, the tourism sector offers various contributions to national economies. Some of these contributions include an increase in foreign exchange reserves, an improvement in the balance of payments, and an increase in employment. Tourism directly or indirectly affects countries' gross domestic product (GDP). It also supports economic development by creating employment and income in rural areas. It strengthens international cultural and social interaction. It promotes international understanding and tolerance by increasing communication between cultures (Oskay, 2012).

Forecasting the developments in the tourism sector is important for managing the sector more effectively. This situation will facilitate decision-makers to make the right decisions and be useful in resource planning (Karahan, 2015). The high level of variability and uncertainty in the tourism sector, which involves human behavior, causes traditional time series methods to be inadequate in demand forecasting. Recently, there has been an increase in studies on tourism demand forecasting. In these studies, methods based on both qualitative and quantitative analysis are used. Analyzing tourism demand with econometric analyses is economically important. Determining the factors affecting tourism demand and the strength of these factors in interaction is a situation that needs to be explained.

When the literature is analyzed, it is revealed that machine learning algorithms based on learning are more successful than traditional methods. In this context, this study aims to predict the number of foreign tourists coming to Turkey with Support Vector Regression, Ridge Regression, and Multiple

Linear Regression methods and to determine the most successful prediction model. The reason for choosing these methods is to compare the predictive power of linear and non-linear machine learning techniques. After the analysis with the most successful prediction method, it aims to provide accurate results for managers and decision-makers in the tourism industry.

Tourism demand forecasting studies conducted in this field generally analyze the demand side. The supply side of tourism, such as accommodation capacities and the number of agencies, needs to be addressed as well. The fact that the study considers both the demand and supply side of tourism demand shows its importance in literature. Accordingly, the study analyzes Turkey's tourism data between 2004 and 2019. In the course of analyzing the variables in the model, the number of foreign tourist arrivals to Turkey series is used as the dependent variable to represent tourism demand. The explanatory variables are the bed capacity of accommodation facilities with tourism business certificates in Turkey, services inflation, which is one of the special comprehensive CPI indicators, the number of tourism agencies in Turkey, the dollar/tl exchange rate, the Brent oil price, the London selling price of 1 ounce of gold and tourism expenditures in Turkey.

In the model created by considering the supply side of tourism demand, R2 values were calculated as 94.89% with Support Vector Regression, 88.83% with Ridge Regression, and 83.79% with Multiple Linear Regression. When the error statistics were analyzed, MAE values were calculated as 0.0402 with Support Vector Regression, 0.0628 with Ridge Regression, and 0.0621 with Multiple Linear Regression. MAPE values were calculated as %8.90 with Support Vector Regression, %19.99 with Ridge Regression, and %21.89 with Multiple Linear Regression. RMSE values were calculated as 0.0528 with Support Vector Regression, 0.0776 with Ridge Regression, and 0.0791 with Multiple Linear Regression. When the results are evaluated, it is seen that the best prediction methods are Support Vector Regression, Ridge Regression, and Multiple Linear Regression, respectively. According to the results, nonlinear machine learning techniques show better prediction performance than linear methods. In addition, among the linear methods, it is seen that the Ridge Regression method gives better results than the Multiple Linear Regression method in the case of multicollinearity.

Regarding the study's limitations, as the number of agencies in Turkey and service inflation data started to be measured from the beginning of the 21st century, the starting year of the study was limited to 2004. In addition, the fact that the pandemic period in 2020 contained outliers and outliers determined the end year of the study as 2019. The limitation of the period analyzed led to the exclusion of some variables from the model.

1. Giriş

Turizm, günümüzde dünya ekonomisinde en hızlı gelişen sektörlerden biri haline gelmiştir. 2022 yılında küresel turizm gelirleri, 2 trilyon ABD dolarına ulaşmıştır. Bu gelir, dünya ticaretinin yaklaşık %10'unu oluşturmaktadır (Dünya Turizm Örgütü, 2023). Turizm sektörü, toplum için sosyal-kültürel bir faaliyet alanı olmasının yanı sıra ülke ekonomileri için çeşitli katkılar sunmaktadır. Bu katkılardan bazıları döviz rezervlerinde artış, ödemeler dengesinde iyileşme ve istihdam artışıdır.

Turizm, ülkelerin gayri safi yurt içi hasılasını (GSYİH) doğrudan ve dolaylı olarak etkilemektedir. Ayrıca kırsal alanlarda istihdam ve gelir yaratarak ekonomik kalkınmaya destek olmaktadır. Uluslararası kültürel ve sosyal etkileşimi güçlendirmektedir. Kültürler arasında iletişimi artırarak uluslararası anlayışı ve hoşgörüyü sağlamaktadır (Oskay, 2012).

Turizm sektöründe gerçekleşecek gelişmelerin önceden tahmin edilmesi, sektörünün daha etkin bir şekilde yönetilmesi açısından önemlidir. Bu durum karar vericilerin doğru kararlar almalarını kolaylaştırarak, kaynak planlaması konusunda faydalı olacaktır (Karahana, 2015). İnsan davranışlarını içeren turizm sektöründe değişkenliğin ve belirsizliğin fazla olması talep tahminlerinde geleneksel zaman serisi yöntemlerinin yetersiz kalmasına yol açmaktadır. Makine öğrenmesi algoritmaları, ortaya çıkabilecek olası problemleri çözecek öğrenme algoritmalarını otomatik bir şekilde oluşturduğu için geleneksel yöntemlere göre daha başarılı sonuçlar ortaya koymaktadır (Çelebi vd., 2023).

Bu bağlamda çalışmanın amacı, Türkiye'ye gelen yabancı turist sayılarını daha doğru bir şekilde tahmin etmek ve turizm endüstrisindeki yöneticiler ile karar vericilere bu tahminler üzerinden stratejik kararlar alabilme imkanı sağlamaktır. Bu kapsamda, Destek Vektör Regresyonu, Ridge Regresyon ve Çoklu Doğrusal Regresyon gibi çeşitli istatistiksel ve makine öğrenmesi yöntemleriyle turizm talebini en iyi tahmin eden yöntem belirlenmiştir. Bu yöntemlerin seçilme nedeni, doğrusal ve doğrusal olmayan veri yapılarını etkin bir şekilde modelleyebilme kapasitelerinin karşılaştırılması ve en başarılı tahmin modelinin tespit edilmesidir.

İlgili literatürde yer alan çalışmalarda genellikle turizm talebinin gelir ve fiyat gibi talep yönü incelenmektedir. Konaklama kapasiteleri ve acente sayıları gibi turizmin arz yönü göz ardı edilmiştir. Çalışmanın turizm talebinin hem talep hem de arz yönünü dikkate alması literatür açısından önemini göstermektedir. Bu yaklaşım, turizm talebinin daha derinlemesine anlaşılmasını ve turizm politikalarının daha etkin bir şekilde şekillendirilmesini sağlayacaktır.

Çalışma, 2004-2019 yılları arasında Türkiye'ye yönelik turizm verilerini kullanarak, turizm talebinin zaman içindeki değişimlerini ve bu değişimleri etkileyen faktörleri detaylı bir şekilde analiz etmeyi amaçlamaktadır. Elde edilen bulgular, turizm endüstrisi yöneticileri, politika yapıcılar ve diğer karar vericilere yönelik stratejik planlama ve kaynak dağıtım konusunda önemli bilgiler sunacak ve Türkiye turizm sektörünün sürdürülebilir büyümesine destek olacak niteliktedir. Böylelikle, bu çalışma, turizm talebinin daha iyi anlaşılması ve yönetilmesine, sektörel politikaların daha bilinçli bir şekilde formüle edilmesine ve bilimsel temelli politika önerilerinin geliştirilmesine katkıda bulunmayı hedeflemektedir.

Çalışmanın yapısal akışı, derinlemesine bir literatür incelemesi ve kapsamlı analitik metodolojileri içerecek şekilde dört ana bölümden oluşmaktadır. İlk bölüm, turizm talep tahminleme çalışmaları üzerine kapsamlı bir literatür taraması sunarak, konunun teorik ve kavramsal çerçevesini detaylandırmaktadır. Bu kısımda, geçmişten günümüze turizm talebinin nasıl ölçüldüğü, hangi faktörlerin etkili olduğu ve farklı coğrafyalarda yapılan benzer çalışmaların sonuçları derinlemesine incelenmektedir. Ayrıca, turizm tahminleme metodolojileri ve bu alanda kullanılan çeşitli yaklaşımlar ele alınmaktadır.

İkinci bölüm, çalışmanın temelini oluşturan Destek Vektör Regresyonu, Ridge Regresyon ve Çoklu Doğrusal Regresyon yöntemlerinin teorik temellerini ve bu yöntemlerin nasıl uygulandığını, içerdikleri matematiksel formülleri ile birlikte ayrıntılı bir şekilde anlatılmaktadır. Bu yöntemlerin tercih edilme nedenleri, avantajları ve sınırlılıkları üzerinde durulmaktadır. Ayrıca, bu tekniklerin tahmin performanslarını değerlendirmek için kullanılan çeşitli istatistiksel performans kriterleri hakkında bilgiler verilmekte ve bu kriterlerin her bir tahmin modeli üzerindeki etkileri tartışılmaktadır.

Üçüncü bölümde veri seti ve model detaylandırılmaktadır. Bu kısımda, Türkiye'ye yönelik turizm verileri, toplanma yöntemleri ve veri kaynakları açıklanmaktadır. Modelin kurulum aşamasında kullanılan veri ön işleme teknikleri, değişken seçimi ve model doğrulama süreçleri anlatılmakta, her bir tahmin yönteminin analiz sonuçları betimsel istatistiklerle desteklenerek sunulmaktadır. Her bir modelin sonuçları ayrı ayrı değerlendirilerek, hangi yöntemin daha başarılı olduğu ve nedenleri üzerinde durulmaktadır.

Son bölümde ise, literatür taraması ve analiz bulguları ışığında elde edilen sonuçlar tartışılmakta ve bu bulguların turizm endüstrisi, politika yapıcılar ve literatür için olan önemi ve etkileri üzerinde durulmaktadır. Çalışmanın sınırlılıkları, gelecek çalışmalar için öneriler ve analiz sonuçlarının turizm sektörüne sağlayabileceği potansiyel katkılar detaylı bir şekilde ele alınmaktadır. Bu bölüm, çalışmanın genel bir değerlendirmesi ile sonlandırılmakta ve turizm tahmini alanında ileriye yönelik perspektifler sunulmaktadır.

2. Literatür Taraması

Son yıllarda turizm talep tahmini üzerine yapılan çalışmalarda artış olduğu gözlemlenmektedir. Bu çalışmalarda hem nitel hem de nicel analizlere dayanan yöntemler kullanılmaktadır. İlgili literatürün oldukça geniş çaplı olması sebebiyle literatür taramasında, sadece nicel yöntemlere dayanan çalışmalar ele alınmaktadır. Bu çalışmalar kullanılan yöntemlere göre zaman serisi analizleri, ekonometrik analizler ve makine öğrenmesi teknikleri olmak üzere üç kategoride incelenmektedir.

Bu kategorilerden ilki olan zaman serileri analizinde genellikle Box ve Jenkins (1970) tarafından geliştirilen entegre otoregresif hareketli ortalama (ARIMA) yöntemi kullanılmaktadır. Ayrıca turizm verilerinin mevsimsel etkiler içermesi literatürde mevsimsel otoregresif hareketli ortalama (SARIMA) modellerinin de tercih edilmesine yol açmıştır. Chen (2000), Goh ve Law (2002), Loganathan ve Yahaya (2010) çalışmalarında turizm talebini ARIMA modelleri ile tahmin etmişlerdir. Bu çalışmalar turizm talebi tahmini için ARIMA yönteminin en iyi performansı verdiğini ve tahminlerin güvenilir olduğunu vurgulamışlardır. Yüksel (2007), Ankara'da bulunan bir oteli örneklem seçerek turizm talebini mikro ölçekte hem Holt Winters hem de yargısal tahmin yöntemleri ile öngörmeye çalışmıştır. Soysal ve Ömürgönülşen (2010), çalışmalarında zaman serileri tekniklerinden Holt Winters yönteminden yararlanarak Türkiye'ye gelen turist sayılarını tahmin etmeye çalışmışlardır. Lim, Chang ve McAleer (2009), Yeni Zelanda'nın turizm talep tahminini Holt Winters ve ARIMA modelleri ile tahmin etmiştir. Her iki yöntemin başarılı ve güvenilir sonuçlar verdiğini vurgulamışlardır.

Turizm talebinin ekonometrik analizler ile incelenmesi iktisadi açıdan önem arz etmektedir. Çünkü turizm talebini etkileyen faktörlerin belirlenmesi ve bu faktörlerin etkileşimdeki gücü açıklanması gereken bir konudur. Bu yöntemlerle ilk çalışmalar 1990'lı yıllara dayanmaktadır. Law ve Au (1999), 1967-1996 döneminde Hong Kong'a gelen turist sayılarını çoklu doğrusal regresyon yöntemleri ile tahmin etmişlerdir. Çalışmada hizmet fiyatları, döviz kuru, pazarlama giderleri, ekonomik büyüme ve nüfus gibi değişkenler modelde açıklayıcı değişken olarak yer almıştır. Bu değişkenler ile gelen turist sayısı arasındaki ilişkiler incelenmiştir. Turizm talep tahminlerinde Song, Wong ve Chon (2003) çalışmalarında gecikmesi dağıtılmış otoregresif sınır testleri (ARDL) modelini kullanırken, Shan ve Wilson (2001) çalışmalarında vektör otoregresyon (VAR) modelini kullanmışlardır.

Bu yöntemler ile yapılan tahminlerin makine öğrenmesi teknikleri ile karşılaştırmalı analizleri literatürde sıklıkla çalışılan konular haline gelmiştir. Uysal ve El Roubi (1999), ABD'ye gelen Kanadalı turistlerin turizm harcamalarını yapay sinir ağları ve çoklu doğrusal regresyon yöntemleri ile tahmin etmiştir. Bu iki modelin birbirine yakın tahminler elde ettiğini ifade etmişlerdir. Burger, Kathrada ve Law (2001), Durban şehrine gelen ABD'li turist sayılarını tahmin etmek için üstel düzeltme, ARIMA, regresyon ve yapay sinir ağları yöntemlerini kullanmışlardır. Analizler sonucu elde edilen sonuçlara göre yapay sinir ağları yönteminin en doğru tahminleri yaptığı görülmüştür.

Çuhadar (2013), 1987-2012 yılları arası Türkiye'ye gelen yabancı turist sayılarını kullanarak Türkiye'nin dış turizm talebini çok katmanlı ileri beslemeli (MLP), zaman gecikmeli (TDNN) ve radyal tabanlı fonksiyon (RBF) yapay sinir ağları aracılığıyla modelleyerek tahmin etmiştir. Elde edilen tahminlerin doğruluğunu hata istatistiklerini karşılaştırarak değerlendirmiştir. Yapılan değerlendirmeler sonucunda en doğru tahminin çok katmanlı ileri beslemeli yapay sinir ağı modeli ile yapıldığı sonucuna varmıştır. Law, Li, Fong ve Han (2019), çalışmalarında 2011-2018 dönemi verilerini kullanarak Çin Halk Cumhuriyeti Makao Özel İdari Bölgesi'ne gelen yabancı turist sayılarını tahmin etmek için derin öğrenme yaklaşımı, destek vektör regresyonu ve yapay sinir ağları yöntemlerini kullanmıştır. Çalışmada yiyecek, konaklama, giyim, ulaşım ve rekreasyon gibi kelimelerin Google arama motorundaki istatistikleri kullanılmıştır. Analizler sonucunda derin öğrenme yaklaşımının diğer yöntemlere göre daha başarılı sonuçlar ürettiği tespit edilmiştir.

Çuhadar (2020) yılında yaptığı bir çalışmada Türkiye'nin dış aktif turizm gelirleri için, Box-Jenkins, Yapay Sinir Ağları ve Üstel Düzeltme gibi yöntemler kullanarak modeller oluşturmuş ve 2020 için aylık turizm geliri tahminleri yapmıştır. Modellerin tahmin başarısı MAPE hata istatistiği ile değerlendirmiş, en başarılı model olarak [5:7:1] mimarisine sahip YSA modelini belirlemiş ve bu model ile 2020 için aylık turizm geliri tahminleri üretmiştir. Çuhadar (2020) yaptığı diğer bir çalışmada Türkiye'nin turizm gelirlerini tahmin etmek için üç farklı yöntemin (Exponential Smoothing, Box-Jenkins ve Yapay Sinir Ağları) performanslarını karşılaştırmayı amaçlamıştır. Modellerin tahmin başarısını değerlendirmek için MAPE hata istatistiğini kullanmıştır. Yapılan analizler sonucunda, [4:5:1] mimarisine sahip Yapay Sinir Ağı modelinin en iyi performansı sergilediğini tespit etmiştir.

Erdoğan vd. (2021) tarafından yapılan çalışmada, Akdeniz Bölgesi'nde önemli destinasyon merkezlerine sahip olan Türkiye'yi konaklama amacıyla en çok tercih eden ülkelerden biri olan

Almanya'dan gelen turist sayısını tahmin etmişlerdir. Geleneksel çoklu doğrusal regresyon yöntemi ile yapay sinir ağı ve destek vektör regresyonu gibi makine öğrenmesi teknikleri kullanılarak üç farklı model oluşturulmuş ve bu modellerin tahmin gücü karşılaştırılmıştır. Araştırmada, Almanya'nın turizm talebini etkileyen makro ekonomik ve nüfus gibi faktörler bağımsız değişken olarak modele dahil edilmiştir. 1998-2019 dönemini kapsayan bu çalışmada yapay sinir ağı tekniği ile oluşturulan tahmin modelinin, diğer tekniklere göre en yüksek tahmin başarısına sahip olduğu belirlenmiştir. Bu yöntemi çoklu doğrusal regresyon ve destek vektör regresyonu teknikleri ile yapılan tahmin modelleri izlemektedir.

Ercan ve Irmak (2022) yaptıkları çalışmalarında 119.390 otel rezervasyon kaydını içeren bir veri setini kullanarak rezervasyon iptallerini tahmin eden makine öğrenmesi modelleri geliştirmişlerdir. Yapay Sinir Ağları, Destek Vektör Makineleri, Sınıflama ve Regresyon Ağaçları ve Rastgele Orman gibi dört farklı makine öğrenmesi yöntemini kullanmışlardır. Geliştirilen modeller karşılaştırmalı olarak incelenmiştir. En yüksek doğruluk oranı %84,97 ile Destek Vektör Makineleri modeliyle elde edilmiştir. Modeller ayrıca Kesinlik, Duyarlılık, Özgüllük ve Negatif Tahmin Edicilik Değeri kriterleri bakımından da değerlendirilmiştir. Modellerden elde edilen yüksek doğruluk ve hassasiyet değerleri, makine öğrenmesi yöntemlerinin rezervasyon süreçlerini tahmin etme potansiyelini göstermektedir.

3. Yöntem

Çalışmada destek vektör regresyonu, çoklu doğrusal regresyon ve ridge regresyon yöntemleri kullanılmaktadır. Bu bölümde ilgili yöntemlerin teorik çerçevesi incelenmektedir.

3.1. Destek Vektör Regresyonu

Destek vektör makineleri, Vapnik (1995) tarafından geliştirilmiş olup Vapnik-Chervonenkis teorisine dayanan bir denetimli makine öğrenmesi tekniğidir. Bu yöntemin temeli istatistiksel öğrenme teorisine dayanmaktadır. Öncelikle sınıflandırma ve regresyon problemlerini çözmek adına tasarlanan bu yöntem, sonrasında model tahminleri için destek vektör regresyonu (DVR) olarak geliştirilmiştir (Drucker vd., 1997).

Destek vektör regresyonu yöntemi tahmin hatalarını en aza indirmeye çalışır ve eğitim verisine yakın bir fonksiyon elde etmeyi hedefler. Bu süreçte, yerel değerlere gelme riskini azaltmak adına kurulan fonksiyonun düzgünlüğü maksimize edilir (Durgun vd., 2023).

$x_i \in \mathbb{R}^n$ gelecek vektörü ve $y_i \in \mathbb{R}$ hedef çıktısı olan bir $\{(x_1, y_1), \dots, (x_l, y_l)\}$ eğitim noktaları veri seti olsun. Teorik olarak girdi ve çıktı verileri arasındaki doğrusal olmayan ilişkiler doğrusal bir kalıp ile formüle edilmelidir. Bu durumu işaret eden fonksiyon aşağıdaki gibidir (Awad vd., 2015):

$$f(x) = w^T \phi(x) + b \quad (1)$$

Burada $f(x)$ öngörülen tahmin değerlerini, \square ise doğrusal olmayan haritalama fonksiyonunu göstermektedir. Ayrıca w , ($w \in \mathbb{R}^n$) ve b ($b \in \mathbb{R}$) ise ayarlanabilir katsayıları temsil etmektedir. Destek vektör regresyonu yönteminin $C > 0$ ve $e > 0$ olmak üzere standart formu ve kısıtları şu şekilde gösterilebilir (Awad vd., 2015; Zhang ve O'Donnell, 2020):

$$\text{Min}_{w,b,z} z^* \frac{1}{2} w^T w + C \sum_{i=1}^l z_i + z_i^* \quad (2)$$

Kısıtlar;

$$\begin{aligned} & \square w^T \square (x_i) + b \square y_i \square \square + z_i \\ & \square y_i \square w^T \square (x_i) \square b \square \square + z_i^* \\ & \square z_i, z_i^* \square 0; i = 1, 2, \dots, l \end{aligned} \quad (3)$$

Burada z_i e altındaki eğitim hatalarını gösterirken z_i^* ise e üzerindeki eğitim hatalarını göstermektedir (Basak ve Patranabis, 2007; Awad vd., 2015; Zhang ve O'Donnell, 2020).

Kısıtlamalar dikkate alınarak ikinci dereceden optimizasyon problemi çözüldükten sonra denklem (2)'deki parametre vektörü olan w denklem (5) ile hesaplanır.

$$w = \sum_{i=1}^l (l_i^* \square l_i) \square (x_i) \quad (4)$$

Burada l_i ve l_i^* Lagrange çarpanlarını göstermektedir (Basak ve Patranabis, 2007).

Böylece destek vektör regresyon formülü denklem (6)'daki gibi gösterilebilir:

$$f(x) = \sum_{i=1}^l (l_i^* \square l_i) K(x_i, x_j) + b \quad (5)$$

Burada $K(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right)$ fonksiyonu radyal temelli kernel fonksiyonunu göstermektedir (Basak ve Patranabis, 2007; Awad vd., 2015; Zhang ve O'Donnell, 2020; Çoban ve Demir, 2021).

3.2. Çoklu Doğrusal Regresyon

Çoklu doğrusal regresyon analizi iki veya daha fazla değişken arasındaki ilişkinin incelenmesinde kullanılmaktadır. Bu ilişkiler fonksiyonel olarak belirlenmekte ve bir model ile açıklanmaktadır. Bir başka ifadeyle mevcut veriler kullanılarak açıklayıcı değişkenler ile bağımlı değişkeni açıklayacak bir model kurma sürecidir. Bu yöntemde bağımlı değişkeni etkileyen katsayılar hesaplanmakta,

açıklayıcı değişkenlerin bağımlı değişken üzerindeki etkileri irdelenebilmekte ve bağımlı değişkenin ileriye dönük öngörülere yapılabilir. Çoklu doğrusal regresyon modelinde (ÇDRM) tahminler gerçekleştirilirken en küçük kareler yönteminden yararlanır. Bu yöntemin amacı hataların kareler toplamını minimum yapacak katsayıları hesaplamaktır (Arslan ve Ertuğrul, 2022).

Çoklu doğrusal regresyon modelinde yer alan iki tip değişken vardır. Birincisi açıklayıcı değişkenler tarafından açıklanmaya çalışılan bağımlı değişkendir. İkincisi ise bağımlı değişkeni açıklayan açıklayıcı değişkenlerdir. Bu modelde genellikle bağımlı değişken Y ile gösterilirken açıklayıcı değişkenler X ile gösterilir. Buradan hareketle ÇDRM aşağıdaki gibi gösterilebilir (Tranmer ve Elliot, 2008; Pandis, 2016):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (6)$$

Burada Y_i bağımlı değişkeni, β_0 kesme terimini, β_k model parametrelerini, X_{ik} açıklayıcı değişkenleri ve ε_i hata terimini göstermektedir (Tranmer ve Elliot, 2008; Pandis, 2016).

3.3. Ridge Regresyon

Çoklu doğrusal regresyon modelinin iyi bir tahmin yapabilmesi için bazı varsayımların sağlanması gerekmektedir. Bu varsayımlardan bir tanesi modelde yer alan açıklayıcı değişkenler arasında çoklu bağlantının olmamasıdır. Çoklu doğrusal bağıntı problemi olduğu durumlarda bu problemin çözülebilmesi için geliştirilmiş alternatif yöntemler mevcuttur. Bunlardan en önemlisi modelde yer alan değişkenleri dışlamadan parametre katsayılarını yanlı olarak tahmin eden ridge regresyon yöntemidir. Bu yöntem modele dahil edilmesi gereken tüm değişkenleri modele alarak en küçük kareler tahminlerinden daha küçük varyanslı parametre tahminleri hesaplaması ve modelde bulunmaması gereken değişkenlerin çıkarılmasını hedeflemektedir (Marquardt ve Snee, 1975).

Bu yöntem çoklu doğrusal bağıntı probleminin yaşandığı durumlarda en küçük kareler yöntemine alternatif olarak 1970 yılında Hoerl ve Kennard tarafından geliştirilmiştir. Açıklayıcı değişkenlerin birbirlerine olan etkilerini en aza indirmek ve kararlı parametre tahminleri gerçekleştirmek amacıyla kullanılmaktadır (Hoerl ve Kennard, 1970).

En küçük kareler yönteminin matris notasyonu ile regresyon modeli gösterimi şu şekildedir:

$$Y = bX + e \quad (7)$$

Burada Y: bağımlı değişken, X: bağımsız değişken b: regresyon katsayılarını, ve e: hata terimini göstermektedir (Hoerl ve Kennard, 1970; McDonald, 2009).

Denklem (8)'te dönüşümler yapılarak regresyon tahmin denklemi aşağıdaki gibidir:

$$\hat{b} = (X^T X)^{-1} X^T Y \quad (8)$$

Modelde yer alan değişkenler arasında çoklu doğrusal bağıntı problemi olduğu durumlarda $X'X$ matrisinin varyansları artacaktır. Bu problemi çözmek için denklem (9)'de bulunan $X'X$ matrisinin köşegen elemanlarına pozitif k sabiti eklenmeli ve bu sayede matrisin varyansı azaltılmalıdır (Hoerl ve Kennard, 1970; McDonald, 2009).

Ridge regresyon yönteminde k sabiti eklenerek oluşturulan regresyon tahmin denklemi şu şekildedir:

$$\hat{b} = (X'X + kI)^{-1} X'Y \quad 0 \leq k < \infty \quad (9)$$

Burada \hat{b} ; $(p-1) \times 1$ boyutlu Ridge Regresyon katsayıları vektörü, I ; $(p-1) \times (p-1)$ boyutlu birim matris ve k sabit değeri göstermektedir. Ridge parametresini ifade eden k değeri (yanlılık sabiti) en küçük kareler yöntemine göre daha küçük hata kareler ortalaması ortaya çıkarmaktadır. Optimum k değerinin belirlenmesi için genellikle ridge regresyonun grafiksel gösteriminden yararlanılarak ridge izi ve varyans şişirme değeri kullanılır. Ridge iz grafiğinde dikey ekseninde standartlaştırılmış regresyon katsayıları, yatay ekseninde yanlılık sabiti(k) yer almaktadır. K değerinin sıfıra eşit olması en küçük kareler yöntemi ile aynı sonuçları vermesi anlamına gelmektedir. Ancak k değerinin artırılması varyansı önemli ölçüde azaltmaktadır (Hoerl ve Kennard, 1970; McDonald, 2009).

3.4. Tahmin Performans Kriterleri (R^2 , MAE, MAPE, RMSE)

Tahmin edilen modeller arasında modelin öngörü başarısının karşılaştırılması için geliştirilmiş çeşitli istatistikler mevcuttur. Öngörünün doğruluğunun tespit edilebilmesi için tahmin edilen modeldeki öngörü değerleri ile gerçek değerler arasındaki farklara dayanan bazı formüller ile öngörü başarısı standartlaştırılır. Buradaki farklar u ile gösterilir ve öngörü hatası olarak tanımlanır. Hatayı en aza indiren modelin tahmin performansının yüksek olduğu söylenebilir (Chicco vd., 2021). Bu kapsamda geliştirilmiş bazı hata istatistikleri mevcuttur. Çalışmada modelin öngörü başarısının ölçümünde kullanılan hata istatistikleri hatanın mutlak ortalaması (MAE), hatanın ortalama mutlak yüzdesi (MAPE) ve hata kareler ortalamasının karekökü (RMSE)'dir.

Hatanın mutlak ortalaması (MAE), tahmin edilen değerler ile gerçek değerler arasındaki farkı ifade eder. Gerçek değerler ile tahmin edilen değerler arasındaki farkların mutlak değerce ortalaması alınarak hesaplanmaktadır. Hata kareler ortalamasının karekökü (RMSE), çokça kullanılan regresyon kaybı fonksiyonudur. Veri kümesindeki tüm örnekler için kare kayıplarının toplanıp örnek sayısına bölünmesi ile elde edilen değer karekökü alınarak hesaplanmaktadır (Chai ve Draxler, 2014). Hatanın ortalama mutlak yüzdesi (MAPE), gerçek değerler ile tahmin edilen değerler arasındaki yüzdesel hataların ortalamasını alır. Genellikle yüzde olarak ifade edilir. Bu nedenle tek başına bir anlam ifade etmemektedir. MAPE değerinin %0 olması mükemmel bir tahmin, %10 veya daha az olması çok iyi bir tahmin, %10 – %20 arasında olması iyi bir tahmin, %20 – %50 arasında olması orta seviye bir tahmin ve %50'den fazla olması düşük tahmin performansı sergilediğini göstermektedir (Makridakis, 1993; Vivas vd., 2020).

Modelin öngörü başarısının ölçümünde kullanılan bir başka ölçüm determinasyon katsayısıdır (R^2) Determinasyon katsayısı modelde yer alan açıklayıcı değişkenlere göre bağımlı değişkendeki değişkenliğin model tarafından açıklanma gücünü göstermektedir. Bir başka ifade ile verilerin regresyon doğrusuna yakınlığının istatistiksel bir ölçüsüdür. Determinasyon katsayısındaki artış gerçekleşen değerler ile tahmin edilen değerler arasındaki farkın azaldığı anlamına gelmektedir. Bu katsayısı yüzdesel olarak ifade edilir ve $0 \leq R^2 \leq 1$ aralığında değer alır. R^2 ve hata istatistiklerine ait formüller Tablo 1'de gösterilmektedir (Chicco vd., 2021).

Tablo 1: Determinasyon Katsayısı ve Hata İstatistiklerine Ait Formüller

Determinasyon Katsayısı (R^2)	Hatanın Mutlak Ortalaması (MAE)	Hatanın Ortalama Mutlak Yüzdesi (MAPE)	Hata Kareler Ortalamasının Karekökü (RMSE)
$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$	$MAE = \frac{1}{n} \sum_{i=1}^n u_i $	$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right * 100$	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n u_i^2}$

Burada $u_i = Y_i - \hat{Y}_i$, Y_i gerçekleşen değerleri, \hat{Y}_i tahmin edilen değerleri, \bar{Y} ortalamaları göstermektedir (Willmott ve Matsuura, 2005; Vivas vd., 2020).

3.5. Veri Seti

Çalışmada Türkiye'ye gelen yabancı turist sayısı ile belirlenen açıklayıcı değişkenler arasındaki ilişkiler incelenerek turizm talebi tahmin edilmektedir. Bu tahmin 3 farklı makine öğrenmesi tekniği ile gerçekleştirilmektedir. Bu teknikler Destek Vektör Regresyonu, Ridge Regresyon ve Çoklu Doğrusal Regresyon yöntemleridir. Buradan hareketle çalışmanın amacı bahsedilen 3 yöntem ile turizm talep tahmini yapılarak en iyi tahmin yapan yöntemin belirlenmesidir.

Çalışma Ocak 2004 ve Aralık 2019 tarihleri arası 191 aylık frekansta veriden oluşmaktadır. Başlangıç olarak 2004 yılının seçilmesinin sebebi bu dönemde turizm politikalarında yapılan yapısal değişiklikler ve Türkiye'de bulunan acente sayılarının ve hizmet enflasyonu verilerinin bu tarihten itibaren yayınlanmaya başlamasıdır. Bitiş olarak 2019 yılının seçilmesinin sebebi pandemi dönemi verilerinin aykırı ve uç değerler içermesidir. Tahminlerin daha doğru yapılabilmesi için pandemi öncesi dönem tercih edilmiştir. Çalışmada kullanılan veriler T.C. Kültür ve Turizm Bakanlığı, Türkiye İstatistik Kurumu ve Türkiye Cumhuriyet Merkez Bankası resmi internet sitelerinden elde edilmiştir.

Çalışmada yayınlanmış istatistiklerin kullanılması yöntemi ile ikincil veriler tercih edilmiştir. İkincil verilerin tercih edilmesinin sebebi zaman kısıtlılığına uygunluk, objektiflik, karşılaştırma yapma fırsatı sağlaması, daha az kaynak kullanımı ve maliyetlerdir. Ancak en önemlisi periyodik çalışmalara imkân kılması bir başka ifade ile dönemlik analizlere uygun olmasıdır. Bu zaman serisi analizinin yapılabileceği anlamına gelmektedir. Zaman serisi, değişkenlerin değerlerinin bir dönemden diğerine

ardışık bir biçimde gözlemlendiği veri türüdür. Zaman serileri değişkenler arasındaki öncelleştirme, geciktirme ve geri besleme ilişkilerini ortaya koymaktadır. Dinamik yapının sağladığı bilgi, geleceğe ilişkin doğru tahminlerin üretimini sağlamakta ve optimal kontrol şemalarını tasarlamaktadır. Zaman serileri analizi, bir serinin yapısını özetler ve serinin öne çıkan yapısını ortaya koymaya çalışır (Sevüktekin ve Çınar, 2017).

3.6. Model

Turizm talep tahmini yapılırken önceki çalışmalarda genellikle turizm talebinin gelir ve fiyat gibi talep yönü incelenirken konaklama kapasiteleri ve acente sayıları gibi arz yönü göz ardı edilmiştir. Bu bağlamda çalışma turizm talebinin arz yönünü de dikkate almaktadır. Turizm talebini etkileyen faktörlerin incelendiği en kapsamlı çalışmalardan birisi Johnson ve Ashworth (1990) tarafından gerçekleştirilmiştir. Modelde yer alan açıklayıcı değişkenler bu çalışma ve ilgili literatür ışığında belirlenmiştir.

Modelde yer alan değişkenler incelendiğinde bağımlı değişken olarak turizm talebini temsilen Türkiye'ye gelen yabancı turist sayısı serisi kullanılmıştır. Açıklayıcı değişkenler ise Türkiye'de turizm işletmesi belgeli konaklama tesislerinin yatak kapasitesi, özel kapsamlı tüfe göstergelerinden hizmet enflasyonu, Türkiye'de bulunan turizm acente sayısı, dolar/tl döviz kuru, brent petrol fiyatı, 1 ons altın Londra satış fiyatı ve Türkiye turizm gideridir.

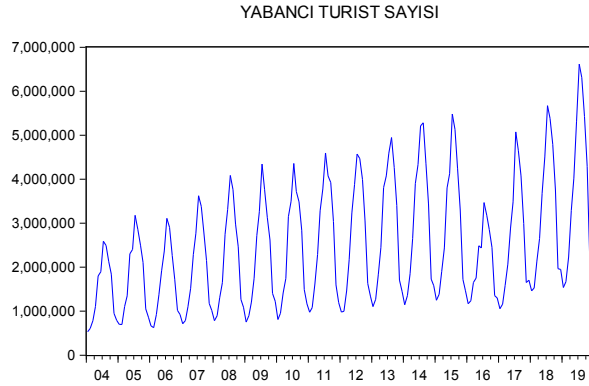
Çalışmada ortaya koyulan model aşağıdaki gibi gösterilebilir:

$$yts_t = \beta_1 + \beta_2 ytk_t + \beta_3 enf_t + \beta_4 acn_t + \beta_5 dvz_t + \beta_6 ptr_t + \beta_7 alt_t + \beta_8 tgd_t + \varepsilon_t \quad (10)$$

Burada b_1 kesme terimini, yts_t yabancı turist sayısını, $ytkt$ yatak kapasitesini, enf_t hizmet enflasyonunu, acn_t acente sayısını, dvz_t döviz kurunu, ptr_t petrol fiyatını, alt_t altın fiyatını, tgd_t turizm giderini ve e_t rassal şoku göstermektedir.

4. Araştırma Bulguları

Turizm talebi denkleminde bağımlı değişken Türkiye'ye gelen yabancı turist sayılarıdır. Yabancı turist sayıları serisine ait zaman grafiği Şekil 1'de gösterilmiştir. Zaman grafiği incelendiğinde serinin artan bir trende sahip olduğu ve mevsimsel dalgalanmaların etkisi gözlemlenmektedir. Dalgalanmaların birbirini takip eden yıllarda yaz aylarında yüksek, kış aylarında düşük değerlere ulaştığı görülmektedir. Bu sistematik yapı mevsimselliğe işaret etmektedir. Yapılan mevsimsellik testi sonucu mevsimsellik tespit edilmiş ve seri mevsimsel etkilerden arındırılarak modele dahil edilmiştir.



Şekil 1: Türkiye'ye Gelen Yabancı Turist Sayıları Serisi Zaman Grafiği

Veri ön işleme sürecinde eğitim ve test için gruplandırılacak veri setinin belirlenmesi için çeşitli denemeler gerçekleştirilmiştir. Bu denemeler sonucu başarılı sonuçlar veren %80-20 ayırımı yapılması öngörülmüştür. Ocak 2004 ve Aralık 2019 tarihleri arası 191 aylık veriden oluşan seri %80'i eğitim, %20'ü test kümesi şeklinde tesadüfi olarak gruplandırılmıştır. Bu doğrultuda ele alınan verilerden rastgele bir biçimde 152 adeti eğitim, 39 adeti test veri seti olarak belirlenmiştir. Verilerin eğitim sürecinde hızlı yakınsayabilmesi ve tahmin performansının artırılması için değişkenlere $\frac{x_0 - x_{\min}}{x_{\max} - x_{\min}}$

formülü yardımıyla Min-Max normalizasyon işlemi uygulanarak veriler [0,1] aralığında ölçeklendirilmiştir. Bu formülde x_0 orijinal veriyi, x_{\min} veri setinde bulunan en küçük değeri ve x_{\max} veri setinde bulunan en büyük değeri ifade etmektedir. Analiz sonuçları R programı ile elde edilmiştir. Veri setini eğitim ve test verisi olarak ayırabilmek için “caret”, model performanslarını karşılaştırmak üzere gerekli ölçümleri hesaplamak için “metrics”, Destek Vektör Regresyonu için “e1071”, Ridge Regresyon için “glmnet” kütüphaneleri kullanılmıştır. Çoklu Doğrusal Regresyon için ayrı bir kütüphaneye ihtiyaç olmadığından programın içerisinde yer alan “lm()” ve “predict()” fonksiyonları kullanılmıştır.

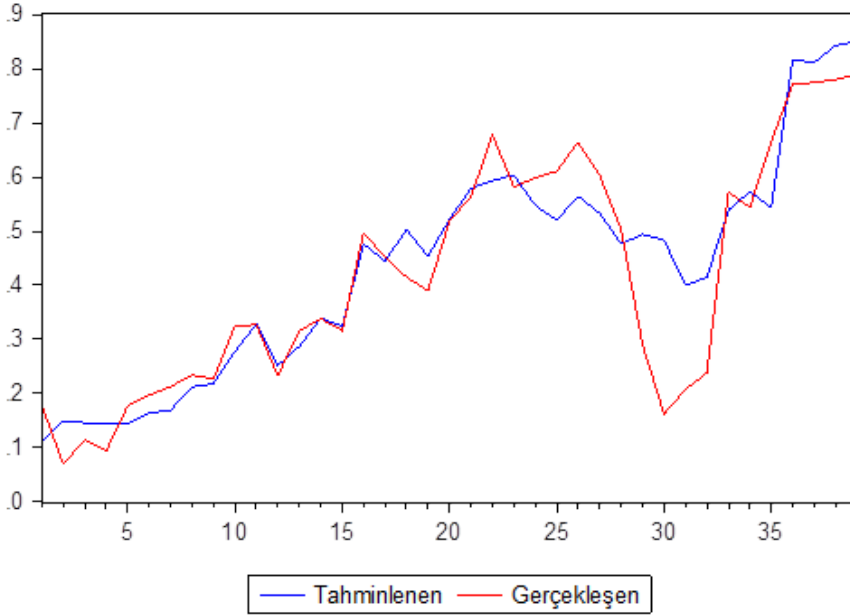
Modelde yer alan değişkenlere ait tanımlayıcı istatistikler Tablo 2'de özetlenmektedir.

Tablo 2: Modelde Yer Alan Değişkenlere Ait Tanımlayıcı İstatistikler

Değişkenler	N	Ortalama	Standart sapma	En küçük değer	En büyük değer
Turist sayısı	192	2486401	1371518	533694	6617380
Acente sayısı	192	7019,216	1978,600	4613,240	11410,00
Yatak kapasitesi	192	694195,5	175487,4	423517,2	992341,0
Turizm gideri	192	3,57E+08	98555668	1,69E+08	6,26E+08
Hizmet enflasyonu	192	9,469427	3,002218	3,560000	16,38000
Döviz kuru	192	2,335938	1,318172	1,180000	6,380000
Altın fiyatı	192	1091,683	380,9505	384,0000	1766,000
Petrol fiyatı	192	74,27010	25,94300	29,53000	138,4000

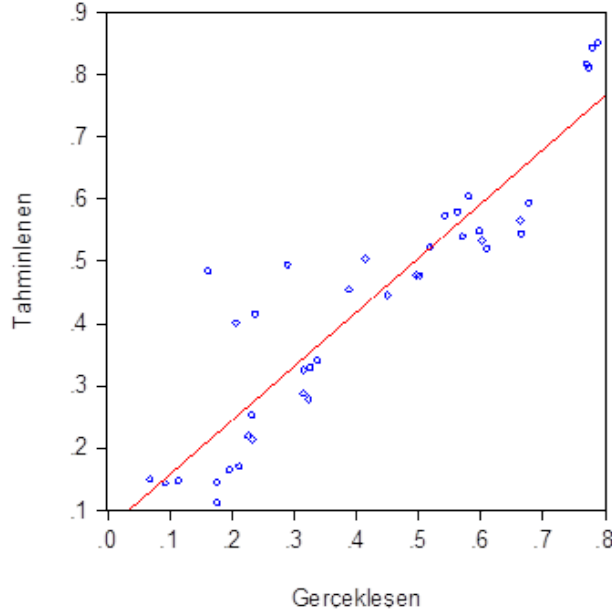
4.1. Destek Vektör Regresyon Bulguları

Destek Vektör Regresyonu (DVR) yöntemi doğrusal ve doğrusal olmayan iki çeşit modele sahiptir. Çalışmada doğrusal olmayan model kullanılmaktadır. Destek Vektör Regresyonu kernel tipi çekirdek fonksiyonları olan doğrusal, radyal, polinom ve sigmoid modellerine göre cost, epsilon ve gamma parametreleri dikkate alınarak birçok deneme gerçekleştirilmiştir. Bu denemeler sonucu en başarılı sonuçları veren kernel tipi radyal tabanlı fonksiyon tercih edilmiştir. Radyal tabanlı fonksiyon parametrelerinin optimize edilmesi ve aşırı öğrenme probleminin engellenmesi için k-katlı çapraz geçerlilik (k-fold Cross Validation) yöntemi kullanılmıştır. Çalışmada sapma ve varyans arasındaki dengenin sağlanması ve modelin aşırı öğrenme probleminin önlenmesi için k değeri 10 olarak alınmıştır. Yapılan işlemler sonucu kernel tipi radyal fonksiyonun cost parametresi 8, epsilon parametresi 0,1 ve gamma parametresi 0,1428 olarak tespit edilmiştir. Bu parametreler ile model eğitim işlemleri gerçekleştirilmiş ve eğitim-test kümeleri tahmin edilmiştir. Destek Vektör Regresyonu yöntemi ile gerçekleştirilen test kümesine ait tahmin ve serpilme grafikleri sırasıyla Şekil 2 ve Şekil 3'te gösterilmektedir.



Şekil 2: DVR ile Gerçekleştirilen Test Kümesine Ait Tahmin Grafiği

Bu grafik, model yardımıyla tahmin edilen değerlerin gerçekleşen değerler ile karşılaştırıldığı ve model tahmin performansını görsel olarak ifade etmektedir. Grafik incelendiğinde yatak kapasitesi, hizmet enflasyonu, acente sayısı, petrol fiyatı, döviz kuru, altın fiyatı ve turizm gideri değerlerinin açıklayıcı değişken olarak yer aldığı modele ait gerçekleşen ve tahminlenen değerler arasında belirli dönemlerde sapmalar olduğu ancak birlikte hareketliliğin söz konusu olduğu gözlemlenmektedir.



Şekil 3: DVR ile Gerçekleştirilen Test Kümesine Ait Serpilme Grafiği

Destek Vektör Regresyonu ile elde edilen modelin tahmin edilen ve gerçekleşen değerlerinin karşılaştırmalı grafikleri görsel açıdan tahmin performansının incelenmesine olanak sağlamaktadır. Ancak tahmin performansının ortaya konulabilmesi için çeşitli hata istatistikleri hesaplanması gerekmektedir. Bu kapsamda yapılan testler sonucu R^2 0,9489; MAE 0,0402; MAPE 0,0890 ve RMSE 0,0528 olarak elde edilmiştir. Modelin açıklanma oranı olarak da bilinen R^2 'nin 1'e yakın olması modelin verilerdeki varyansı çok iyi bir seviyede açıkladığını göstermektedir. Bir başka ifade ile modelin uyumu yüksektir. Buradan hareketle bağımlı değişken olan yabancı turist sayısındaki değişimlerin %94,89'u modelde yer alan açıklayıcı değişkenlerdeki değişimler tarafından açıklanmaktadır.

MAE değerine bakıldığında, gerçekleştirilen tahminlerin ortalama olarak gerçek değerden 0.0402 birim sapma göstereceği görülmektedir. Bu değer düşük olması modelin doğru tahmin yaptığını işaret etmektedir. MAPE, tahminlerin gerçek değerlerden ne kadar sapma gösterdiğini yüzdesel olarak ifade etmektedir. %8,90 olarak hesaplanan bu değer model tahminlerinin ortalama olarak gerçek değerden %8,90 oranında sapma gösterdiğini ifade etmektedir. Bu oranın %10'un altında olması çok iyi bir tahmin gerçekleştirildiğini işaret etmektedir. RMSE, hataların karelerinin ortalamasının karekökünü alarak büyük hataları ağırlıklı olarak cezalandırır. Bu değer 0,0528 olması doğru tahmin yapıldığını göstermektedir.

Sonuç olarak hesaplanan hata istatistiklerinin tümü değerlendirildiğinde değerlerin sıfıra yakın olması modelin oldukça yüksek bir doğrulukla tahminler gerçekleştirdiğini ve veri seti üzerinde iyi bir performans sergilediğini göstermektedir.

4.2. Çoklu Doğrusal Regresyon Bulguları

En küçük kareler yöntemi ile elde edilen regresyon sonuçları Tablo 3 ve Tablo 4'te verilmiştir. Tahmin incelendiğinde modelin genel olarak anlamlı olduğu ve uyum iyiliğinin yüksek olduğu görülmektedir. Ancak açıklayıcı değişkenlerin bireysel anlamlılıklarının sınındığı t istatistik değerlerinin düşük olduğu tespit edilmiştir. Bununla birlikte modelin tanımında ve tahmininde kullanılan gözlem sayısında küçük değişiklikler yapıldığında tahmin değerlerinin ve standart hataların büyük ölçüde değiştiği gözlemlenmiştir. Bu durumlar çoklu doğrusal regresyon modelinin varsayımlarından tam çoklu doğrusal bağıntı probleminin varlığına işaret etmektedir. Çoklu doğrusal bağıntı (ÇDB), modelde yer alan açıklayıcı değişkenler arasında tam ya da yüksek dereceden doğrusal bir ilişkinin olduğu anlamına gelmektedir. Bu problemin varlığında regresyon katsayıları, varyanslar ve kovaryanslar artacak, hipotez testlerinin gücü zayıflayacaktır. Tahminler denklem tanımlanmasına çok duyarlı hale gelecektir. Bu yüzden bu problemin varlığı test edilmeli ve problem çözümlenmelidir (Hongehang, 2005).

Tablo 3: Tahmin Edilen Regresyon Katsayıları

Değişkenler	Regresyon katsayısı	Standart hata	t değerleri	p değerleri
Sabit terim	-339473,7	247834,5	-1,3697	0,1724
Acente sayısı	-244,7304	135,1156	-1,8112	0,0717
Yatak kapasitesi	3,590516	1,358680	2,6426	0,0089
Turizm gideri	0,001088	0,000239	4,5486	0,0000
Hizmet enflasyonu	31308,11	13445,04	2,3285	0,0210
Döviz kuru	223101,4	54332,08	4,1062	0,0001
Altın fiyatı	290,8605	128,3526	2,2661	0,0246
Petrol fiyatı	7020,351	1101,876	6,3712	0,0000

Tahmin performansının ortaya konulabilmesi için R^2 ve çeşitli hata istatistikleri hesaplanması gerekmektedir. Bu kapsamda yapılan testler sonucu değerler Tablo 4'teki gibi hesaplanmıştır.

Tablo 4: Determinasyon Katsayısı ve Hata İstatistikleri

R^2	R_d^2	MAE	MAPE	RMSE	F istatistiği
0,8494	0,8437	0,0621	0,2189	0,0791	148,333 (0,000)

Modelin açıklanma oranı olarak da bilinen R^2 'nin 1'e yakın olması modelin verilerdeki varyansı çok iyi bir seviyede açıkladığını göstermektedir. Bir başka ifade ile modelin uyumu yüksektir. Buradan hareketle bağımlı değişken olan yabancı turist sayısındaki değişimlerin %84,37'u modelde yer alan açıklayıcı değişkenlerdeki değişimler tarafından açıklanmaktadır.

MAE değerine bakıldığında, gerçekleştirilen tahminlerin ortalama olarak gerçek değerden 0.0621 birim sapma göstereceği görülmektedir. Bu değer düşük olması modelin doğru tahmin yaptığını işaret etmektedir. MAPE, tahminlerin gerçek değerlerden ne kadar sapma gösterdiğini yüzdesel olarak ifade etmektedir. %21,89 olarak hesaplanan bu değer model tahminlerinin ortalama olarak

gerçek değerden %21,89 oranında sapma gösterdiğini ifade etmektedir. Bu oran verinin uygulama alanı bağlamında değerlendirildiğinde iyi bir tahmin olarak değerlendirilebilir. RMSE, hataların karelerinin ortalamasının karekökünü alarak büyük hataları ağırlıklı olarak cezalandırır. Bu değer 0,0791 olması doğru tahmin yapıldığını göstermektedir.

Sonuç olarak hesaplanan hata istatistiklerinin tümü değerlendirildiğinde değerler sıfıra yakın olsa da değerlerin nispeten yüksek olması modelin doğru tahminler gerçekleştirdiğini ancak iyileştirme yapılması gerektiğini göstermektedir.

Çoklu doğrusal bağıntı probleminin belirlenmesi için çeşitli yöntemler bulunmaktadır. Bu yöntemlerden bazıları şunlardır:

- Açıklayıcı değişkenler arasındaki korelasyonun yüksek olması
- Varyans büyütme faktörü (VIF) değerinin 10'dan fazla olması
- Tolerans katsayısının 0'a yakın olması
- Bir açıklayıcı değişkenin diğer açıklayıcı değişkenlerle regresyonunun determinasyon katsayısının modelin determinasyon katsayısından büyük olması
- En büyük özdeğerin en küçük özdeğere bölümü sonucu elde edilen değer karekökünün 30'dan fazla olması (Neter vd., 1996).

Çalışmada çoklu doğrusal bağıntı probleminin tespit edilebilmesi için çeşitli yöntemler uygulanmıştır. İlk olarak açıklayıcı değişkenler arasındaki doğrusal ilişkiler incelenecektir. Çünkü açıklayıcı değişkenler arasındaki korelasyon katsayılarının yüksek olması çoklu doğrusal bağıntı problemini göstermektedir. Açıklayıcı değişkenler arasındaki ilişkileri gösteren korelasyon matrisi Tablo 5'te gösterilmektedir.

Tablo 5: Açıklayıcı Değişkenler Arasındaki Korelasyon Katsayıları

Değişkenler	Korelasyon matrisi						
	Acente sayısı	Yatak kapasitesi	Turizm gideri	Hizmet enflasyonu	Döviz kuru	Altın fiyatı	Petrol fiyatı
Acente sayısı	1						
Yatak kapasitesi	0,989	1					
Turizm gideri	0,384	0,446	1				
Hizmet enflasyonu	0,041	-0,066	-0,502	1			
Döviz kuru	0,921	0,875	0,179	0,323	1		
Altın fiyatı	0,619	0,684	0,523	-0,587	0,408	1	
Petrol fiyatı	-0,103	-0,033	0,199	-0,526	-0,220	0,535	1

Korelasyon matrisi incelendiğinde acente sayısı ile yatak kapasitesi arasında 0,989, döviz kuru ile acente sayısı arasında 0,921, yatak kapasitesi ile döviz kuru arasında 0,875 gibi yüksek korelasyon değerleri tespit edilmiştir. Bu durum çoklu doğrusal bağıntı probleminin işaret etmektedir. Ancak bu katsayılar tek başına yeterli olmamakla birlikte diğer göstergeler ile desteklenmelidir.

Çoklu doğrusal bağıntı probleminin tespit edilmesinde bir diğer yöntem varyans büyütme faktörüdür. Bu değer 10'a eşit ya da 10'dan büyük olması ÇDB problemini göstermektedir. Aynı zamanda her bir açıklayıcı değişkenin diğer açıklayıcı değişkenlerle regresyonundan elde edilen determinasyon katsayılarının (R_k^2) modelin determinasyon katsayısından yüksek olması ÇDB problemine işaret etmektedir. Hesaplanan bu determinasyon katsayılarından $1 - R_k^2$ formülü ile elde edilen tolerans değerlerinin 0'a yaklaşması ÇDB problemini gösteren bir başka yöntemdir. Hesaplanan varyans büyütme faktörleri, diğer değişkenlerle olan R_k^2 'ler ve tolerans değerleri Tablo 6'da gösterilmektedir.

Tablo 6: VIF, Diğer Değişkenlerle Olan R_k^2 ve Tolerans Değerleri

Değişkenler	VIF	Diğer değişkenlerle olan R_k^2	Tolerans
Acente sayısı	229,7617	0,9956	0,0044
Yatak kapasitesi	182,7583	0,9945	0,0055
Turizm gideri	1,786592	0,4403	0,5597
Hizmet enflasyonu	5,237925	0,8091	0,1909
Döviz kuru	16,48951	0,9394	0,0606
Altın fiyatı	7,685936	0,8699	0,1301
Petrol fiyatı	2,626979	0,6193	0,3807

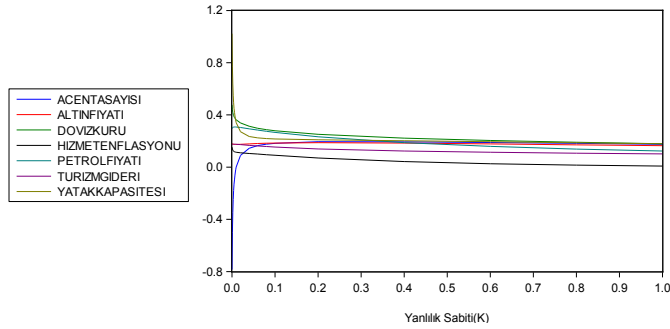
Varyans büyütme faktörleri incelendiğinde acente sayısı, yatak kapasitesi ve döviz kuru değişkenlerinin değerlerinin 10'dan büyük olması değişkenler arasında çoklu doğrusal bağıntı problemi olduğunu göstermektedir. Acente sayısı, yatak kapasitesi, döviz kuru ve altın fiyatı değişkenlerinin diğer değişkenlerle regresyonundan elde edilen determinasyon katsayılarının modelin determinasyon katsayısından yüksek olduğu görülmektedir. Bu doğrultuda ilgili değişkenlerin tolerans değerleri de 0'a yakın hesaplanmıştır.

Bu dört hesaplama ışığında modelde çoklu doğrusal bağıntı problemi olduğu sonucuna varılmaktadır. Hesaplamaların tümü, üç açıklayıcı değişken üzerinde odaklandığı için yöntemler arası iç tutarlılık sağlanmıştır. ÇDB probleminin çözümünde farklı yöntemler bulunmaktadır. Açıklayıcı değişkenlerin birbirleri üzerindeki etkilerini en aza indirmeyi ve kararlı regresyon katsayıları tahmin etmeyi amaçlayan yanlı regresyon yöntemlerinden Ridge regresyon kullanılarak bu problem ortadan kaldırılmaya çalışılmıştır.

4.3. Ridge Regresyon Bulguları

Ridge regresyon, bir durumun ortaya çıkışında etkili olan tüm faktörlerin modele dahil edilerek bütüncül etkinin incelenmesine olanak sağlayan yanlı bir regresyon yöntemidir. ÇDB problemi olduğu durumlarda EKK yöntemine göre daha iyi sonuçlar vermektedir. Ridge regresyon analizinin ilk kısmında optimal yanlılık sabitinin (k) belirlenebilmesi için Şekil 4 Ridge İz grafiği ve Tablo 7 Ridge VIF değerleri verilmiştir. Dikey ekseninde standartlaştırılmış regresyon katsayılarının, yatay ekseninde yanlılık sabitinin yer aldığı Ridge İz grafiğinde katsayıların durağanlaştığı k değeri seçilmektedir.

Ridge VIF değerlerinde ise tüm açıklayıcı değişkenlerin VIF değerlerinin 10'dan küçük olduğu yanlılık sabiti tercih edilmektedir (Marquardt ve Snee, 1975).



Şekil 4: Ridge İz Grafiği

Grafik incelendiğinde 0,0 ile 0,1 arası bir yanlılık sabitinden sonra standartlaştırılmış regresyon katsayılarının durağanlaştığı görülmektedir. Bu aralıkta yanlılık sabiti optimal değer almaktadır. Yanlılık sabitinin tam olarak belirlenebilmesi için Ridge VIF değerlerine bakılmalıdır.

Tablo 7: Ridge VIF Değerleri

Yanlılık sabiti(k)	VIF Değerleri						
	Acenta sayısı	Altın fiyatı	Döviz kuru	Hizmet enflasyonu	Petrol fiyatı	Turizm gideri	Yatak kapasitesi
0,000000	229,7617	7,685936	16,48951	5,237925	2,626979	1,786592	182,7583
0,001000	117,0457	7,497351	14,29656	4,825689	2,532411	1,771044	94,49464
0,002000	71,28584	7,322685	13,13214	4,599750	2,475130	1,758583	58,59890
0,003000	48,27711	7,156538	12,33473	4,439962	2,431178	1,747306	40,50123
0,004000	35,08407	6,997143	11,70933	4,310970	2,393508	1,736650	30,08540
0,005000	26,81223	6,843675	11,18104	4,199246	2,359430	1,726396	23,52340
0,006000	21,27794	6,695641	10,71557	4,098629	2,327718	1,716441	19,10719
0,007000	17,38753	6,552685	10,29504	4,005933	2,297741	1,706730	15,98114
0,008000	14,54402	6,414521	9,909130	3,919335	2,269137	1,697229	13,67809
0,009000	12,39909	6,280906	9,551322	3,837704	2,241682	1,687915	11,92534
0,010000	10,73834	6,151622	9,217223	3,760283	2,215227	1,678774	10,55498
0,020000	4,186801	5,059257	6,730038	3,142322	1,990489	1,594887	4,875235
0,040000	1,713960	3,616506	4,101381	2,374979	1,675214	1,455397	2,329829
0,060000	1,033286	2,731335	2,787021	1,912197	1,459843	1,340379	1,484733
0,080000	0,722863	2,148011	2,033462	1,602193	1,300842	1,242194	1,060008
0,100000	0,548613	1,742502	1,560567	1,379644	1,177084	1,156722	0,808271
0,200000	0,241645	0,818283	0,639639	0,814143	0,808762	0,850664	0,339531
0,400000	0,122547	0,352706	0,264035	0,439093	0,493643	0,532211	0,152032
0,600000	0,089286	0,215996	0,165782	0,294112	0,346378	0,371528	0,102247
0,800000	0,072919	0,154018	0,122328	0,217727	0,261089	0,277080	0,079622
1,000000	0,062551	0,119152	0,097659	0,170843	0,205941	0,216120	0,066328

Tablo incelendiğinde tüm açıklayıcı değişkenlerin VIF değerlerinin 10'unun altına indiği ilk yanlılık sabiti $k=0.02$ 'dir. Bu yanlılık sabiti için Ridge regresyon modelinde yer alan açıklayıcı değişkenlere ait VIF değerleri 4,186, 5,059, 6,730, 3,142, 1,990, 1,594, 4,875 olarak hesaplanmıştır. Bu değerler 10'nun altında olduğu için çoklu doğrusal bağıntı probleminin ortadan kalktığı sonucuna varılmaktadır.

Ridge regresyon analizinin bu kısmında optimal yanlılık sabiti dikkate alınarak Ridge regresyon modeli tahmin edilmiştir. Bu tahmine ilişkin regresyon katsayıları, standartlaştırılmış regresyon katsayıları ve VIF değerleri Tablo 8'de verilmiştir.

Tablo 8: Ridge Regresyon Katsayıları

Açıklayıcı değişkenler	Regresyon katsayıları	Standartlaştırılmış regresyon katsayıları	VIF
Sabit terim	-0,041715		
Acente sayısı	28,66794	0,091989	4,073534
Yatak kapasitesi	283,9167	0,175404	5,020415
Turizm gideri	158188,9	0,338165	6,650475
Hizmet enflasyonu	22427,87	0,109197	3,121141
Döviz kuru	7209,926	0,303341	1,982358
Altın fiyatı	0,001078	0,172256	1,591636
Petrol fiyatı	0,947542	0,269665	4,768424

Ridge regresyon modeliyle elde edilen sonuçlar EKK yöntemine göre farklılık göstermektedir. Parametreler açısından acente sayısının işareti değişmiştir. Bu durum önsel beklentilere uygundur. Ayrıca parametrelerin standart hatalarında ve VIF değerlerinde düşüş gerçekleşmiştir. Bu güvenilir regresyon tahminlerinin elde edildiğinin göstergesidir.

Optimal yanlılık sabitiyle ($k=0,02$) tahmin edilen Ridge regresyon analizi sonucunda modelin açıklama gücünü gösteren determinasyon katsayısı (R^2) %88,83, hata istatistikleri MAE, MAPE ve RMSE sırasıyla 0,0628, 0,1999, 0,0776 olarak hesaplanmıştır. Ridge regresyon modeli %5 anlamlılık düzeyinde genel olarak anlamlıdır.

Modelin açıklanma oranı olarak da bilinen R^2 'nin 1'e yakın olması modelin verilerdeki varyansı çok iyi bir seviyede açıkladığını göstermektedir. Bir başka ifade ile modelin uyumu yüksektir. Buradan hareketle bağımlı değişken olan yabancı turist sayısındaki değişimlerin %88,83'ü modelde yer alan açıklayıcı değişkenlerdeki değişimler tarafından açıklanmaktadır.

MAE değerine bakıldığında, gerçekleştirilen tahminlerin ortalama olarak gerçek değerden 0.0628 birim sapma göstereceği görülmektedir. Bu değer düşük olması modelin doğru tahmin yaptığını işaret etmektedir. MAPE, tahminlerin gerçek değerlerden ne kadar sapma gösterdiğini yüzdesel olarak ifade etmektedir. %19,99 olarak hesaplanan bu değer model tahminlerinin ortalama olarak gerçek değerden %19,99 oranında sapma gösterdiğini ifade etmektedir. Bu oranın %10 – %20 arasında olması iyi bir tahmin gerçekleştirildiğini işaret etmektedir. RMSE, hataların karelerinin ortalamasının karekökünü alarak büyük hataları ağırlıklı olarak cezalandırır. Bu değer 0,0776 olması doğru tahmin yapıldığını göstermektedir.

Sonuç olarak hesaplanan hata istatistiklerinin tümü değerlendirildiğinde değerler sıfıra yakın olsa da değerlerin nispeten yüksek olması modelin doğru tahminler gerçekleştirdiğini ancak iyileştirme yapılması gerektiğini göstermektedir.

5. Sonuç, Tartışma ve Öneriler

Dünya ekonomisinde hızlı gelişme gösteren turizm sektörü, ülkelerin ekonomik büyüme ve kalkınmasında kritik bir role sahiptir. Uluslararası turizmde artan rekabet koşullarında sahip olunan turizm kaynaklarının en etkin biçimde kullanılması ülke ekonomileri açısından önem arz etmektedir. Ülke ekonomisinde birçok sektör için lokomotif görevi gören turizm sektöründe turizm planlamasının etkin yönetimi için önceden turist sayılarının belirlenmesi ve turizm talebi karar vericiler tarafından tahmin edilmek istenmektedir. Bu tahminlerin doğruluğu ve güvenilirliği turizm sektöründe yer alan paydaşların turizm yatırımlarının etkin bir şekilde planlanmasına ve turist beklentilerinin karşılanmasına olanak sağlayacaktır.

Bu çalışmada Türkiye'ye gelen yabancı turist sayısının destek vektör regresyonu, ridge regresyon ve çoklu doğrusal regresyon yöntemleriyle tahmin edilerek en doğru tahmin yapan modelin tespit edilmesi amaçlanmıştır. İlgili yöntemlerin tercih edilmesinin sebebi doğrusal ve doğrusal olmayan makine öğrenmesi tahmin tekniklerinin karşılaştırılmak istenmesidir. Bu doğrultuda 2004-2019 yılları arası döneme ait aylık frekansta veriler incelenmiştir.

Turizm talebinin arz yönü dikkate alınarak oluşturulan modelde, R2 değerleri destek vektör regresyonu ile yapılan analiz sonucunda %94,89, ridge regresyon ile %88,83 ve çoklu doğrusal regresyon ile %83,79 olarak hesaplanmıştır. Hata istatistikleri incelendiğinde MAE değerleri destek vektör regresyonu ile yapılan analiz sonucunda 0,0402, ridge regresyon ile 0,0628 ve çoklu doğrusal regresyon ile 0,0621 olarak hesaplanmıştır. RMSE değerleri ise destek vektör regresyonu ile yapılan analiz sonucunda 0,0528, ridge regresyon ile 0,0776 ve çoklu doğrusal regresyon ile 0,0791 olarak hesaplanmıştır. Son olarak MAPE değerleri destek vektör regresyonu ile yapılan analiz sonucunda %8,90, ridge regresyon ile %19,99 ve çoklu doğrusal regresyon ile %21,89 olarak hesaplanmıştır. Sonuçlar değerlendirildiğinde en iyi tahmini gerçekleştiren yöntemlerin sırasıyla destek vektör regresyonu, ridge regresyon ve çoklu doğrusal regresyon olduğu görülmektedir. Ortaya çıkan sonuçlara göre doğrusal olmayan makine öğrenmesi teknikleri doğrusal yöntemlere göre daha iyi tahmin performans göstermektedir. Bu sonuç Kathrada ve Law (2001), Çuhadar (2013), Kılıç vd. (2018), Law vd. (2019), Sönmez ve Zengin (2019), Yerkök (2020), Çuhadar (2020), Erdoğan (2021), Ercan ve Irmak (2022) çalışmalarıyla benzerlik göstermektedir. Ayrıca doğrusal yöntemler arasında ridge regresyon yönteminin çoklu doğrusal bağıntı durumunda Çoklu doğrusal regresyon yöntemine göre daha iyi sonuçlar verdiği görülmektedir. Bu sonuç Topal vd. (2013), Haworth vd. (2014), Çekerol ve Nalçakan (2015), Caicedo-Torres ve Payares (2016), Alpu vd. (2016), Şamkar (2018), Sun vd. (2021), Pereira ve Cerqueira (2022) çalışmalarıyla benzerlik göstermektedir.

Çalışmanın sınırlılıkları açısından değerlendirildiğinde, Türkiye'de bulunan acente sayılarının ve hizmet enflasyonu verilerinin 21. yüzyılın başından itibaren ölçülmeye başlaması sonucu çalışmada

başlangıç yılı 2004 ile kısıtlanmıştır. Ayrıca 2020 yılında meydana gelen pandemi döneminin aykırı ve uç değerler içermesi çalışmanın bitiş yılını 2019 olarak belirlemiştir. İncelenen periyodun sınırlanması bazı değişkenlerin model dışı kalmasına yol açmıştır.

Turizm sektöründe doğru tahminlerin kritik önemi, bu çalışmanın ortaya koyduğu bulgularla bir kez daha vurgulanmıştır. Doğrusal olmayan modellerin tahmin başarısı açısından doğrusal modellere göre üstünlüğü, özellikle destek vektör regresyonunun en etkin model olarak öne çıkması, turizm işletmeleri ve planlamacılar için stratejik bir bakış açısı sunmaktadır. Ayrıca, Ridge regresyonunun alternatif olarak kullanılması, geniş veri kapsamı ve güncellemelerin tahmin doğruluğunu artırabileceği, yatırım ve kaynak yönetimi ile stratejik planlama süreçlerinde bu tahminlerin kritik bir rol oynayabileceği belirtilmiştir. Bu çıkarımlar, turizm sektöründeki uygulayıcılara ve yöneticilere, daha etkili kararlar alabilmeleri ve rekabet avantajı sağlayabilmeleri için değerli bir yol göstericidir.

Çalışma gelecek araştırmalara turizmi etkileyen diğer arz yönlü faktörlerin modele dahil edilmesi, zaman aralığının genişletilerek veri sayısının artırılması, farklı makine öğrenmesi tekniklerinin kullanılması ve turizm talebini temsil edecek yeni bağımlı değişkenlerin tespit edilmesi hususlarında öneriler sunmaktadır. Ayrıca farklı ülkelerin turizm verilerini kullanarak karşılaştırmalı analizler yapmak ve global trendleri daha iyi anlamak değerli olacaktır. Böylelikle turizm talep tahminlerinin doğruluğunu artırabilir ve sektörel karar verme süreçlerine önemli katkılar sağlayabilir. Çalışmanın ana konusu turizm sektörü olsa da diğer sektörlerin talep tahminlerinde kurulan model ve yöntemlerin doğru sonuçlar üreteceği varsayılmaktadır. Son olarak, araştırmacılar, turizm talebini etkileyen yeni bağımlı ve bağımsız değişkenlerin belirlenmesi üzerine yoğunlaşabilirler. Örneğin, sosyal medya etkileşimleri ve çevresel faktörler gibi yeni nesil veri kaynakları, turizm talep modellerine entegre edilerek sektörün dinamiklerinin daha iyi anlaşılmasına yardımcı olabilir. Bu tür yenilikçi yaklaşımlar, turizm araştırmalarının kapsamını genişletir ve daha etkili politika ve strateji geliştirme imkânı sunar. İlgili literatürde makine öğrenmesi teknikleri ile turizm talep tahmini çalışmalarının kısıtlı sayıda olduğu dikkate alındığında, önerilen fikirlerin literatüre ve sektördeki karar vericilere planlama süreçlerinde destek olacağı söylenebilir.

Yazar Katkısı

KATKI ORANI	AÇIKLAMA	KATKIDA BULUNANLAR
Fikir veya Kavram	Araştırma fikrini veya hipotezini oluşturmak	Serkan Kardeş Burcu Öngen Bilir
Literatür Taraması	Çalışma için gerekli literatürü taramak	Serkan Kardeş Burcu Öngen Bilir
Araştırma Tasarımı	Çalışmanın yöntemini ve desenini tasarlamak	Serkan Kardeş Burcu Öngen Bilir
Veri Toplama ve İşleme	Verileri toplamak, düzenlemek ve raporlamak	Serkan Kardeş Burcu Öngen Bilir
Tartışma ve Yorum	Bulguların değerlendirilmesinde ve sonuçlandırılmasında sorumluluk almak	Serkan Kardeş Burcu Öngen Bilir

Çıkar Çatışması

Çalışmada yazarlar arasında çıkar çatışması yoktur.

Finansal Destek

Bu çalışma için herhangi bir kurumdan destek alınmamıştır.

Kaynakça

- Alpu, Ö., Şamkar, H., & Altan, E. (2016). Sağlam ridge regresyon analizi ve bir uygulama. *Dokuz Eylül Üniversitesi İktisadi İdari Bilimler Fakültesi Dergisi*, 25(2), 137-148.
- Arslan, B., & Ertuğrul, İ. (2022). Çoklu regresyon, ARIMA ve yapay sinir ağı yöntemleri ile Türkiye elektrik piyasasında fiyat tahmin ve analizi. *Yönetim ve Ekonomi Araştırmaları Dergisi*, 20(1), 331-353.
- Awad, M., Khanna, R. (2015). Support Vector Regression. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, 67-80.
- Basak, D., Pal, S., & Patranabis, D. C. (2007). Support vector regression. *Neural Information Processing-Letters and Reviews*, 11(10), 203-224.
- Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis, forecasting and control*. San Francisco: Holden Day.
- Burger, M. D., Kathrada, M., & Law, R. (2001). A practitioners guide to time series methods for tourism demand forecasting a case study of Durban, South Africa. *Tourism Management*, 22(4), 403-409.
- Caicedo-Torres, W., & Payares, F. (2016). A machine learning model for occupancy rates and demand forecasting in the hospitality industry. *In Advances in Artificial Intelligence 2016: 15th Ibero-American Conference on AI* (pp. 201-211). Springer International Publishing.
- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?. *Geoscientific Model Development Discussions*, 7(1), 1525-1534.
- Chen, J. C. (2000). *Forecasting Method Applications to Recreation and Tourism Demand*. (Doctoral dissertation). North Carolina State University, USA.
- 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 computer science*, 7, e623.
- Çekerol, G., & Nalçakan, M. (2015). Lojistik sektörü içerisinde Türkiye demiryolu yurtiçi yük taşıma talebinin ridge regresyonla analizi. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 31(2), 321-344.
- Çelebi, Ö. F., Cavdar Aksoy, N., Kocak Alan, A., & Tümer Kabadayı, E. (2023). İleri teknolojiler, yapay zekâ temelli çözümler: duyu odaklı bir yaklaşım. *Öneri Dergisi*, 18(60), 367-395.
- Çoban, F., & Demir, L. (2021). Yapay sinir ağları ve destek vektör regresyonu ile talep tahmini: gıda işletmesinde bir uygulama. *DEUFMD*, 23(67), 327-338.
- Çuhadar, M. (2013). Türkiye'ye yönelik dış turizm talebinin MLP, RBF VE TDNN yapay sinir ağı mimarileri ile modellenmesi ve tahmini: karşılaştırmalı bir analiz. *Journal of Yasar University*, 8(31), 5274-5295.
- Çuhadar, M. (2020). Türkiye'nin dış aktif turizm gelirlerinin alternatif yaklaşımlarla modellenmesi ve tahmini. *Ankara Hacı Bayram Veli Üniversitesi Turizm Fakültesi Dergisi*, 23(1), 115-141.
- Çuhadar, M. (2020). A comparative study on modelling and forecasting tourism revenues: The case of Turkey. *Advances in Hospitality and Tourism Research*, 8(2), 235-255.
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola A., & Vapnik, V. (1997). Support vector regression machines. *Advances in Neural Information Processing Systems*, 9, 155-161.

- Durgun, H., İnce, E. Y., İnce, M., Çoban, H. O., & Eker, M. (2023). Evaluation of tree diameter and height measurements in UAV data through the integration of remote sensing and machine learning methods. *Gazi Journal of Engineering Sciences*, 9(4), 113-125.
- Ercan, U., & İrmak, S. (2022). Turizm endüstrisinde otel rezervasyon iptallerinin makine öğrenmesi yöntemleri ile tahmini. *Journal of Gastronomy, Hospitality and Travel*, 5(1), 45-54.
- Erdoğan, H., Terzioğlu, M., & Kayakuş, M. (2021). Almanya'dan konaklama amacıyla Türkiye'ye gelen turist sayısının yapay zekâ teknikleri kullanılarak tahmin edilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (27), 961-971.
- Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, 23(5), 499-510.
- Haworth, J., Shawe-Taylor, J., Cheng, T., & Wang, J. (2014). Local online kernel ridge regression for forecasting of urban travel times. *Transportation Research Part C: Emerging Technologies*, 46, 151-178.
- Hongehang, H. (2005). Ridge estimation of a semiparametric regression model. *Journal of Computational and Applied Mathematics*, 176(1), 215-222.
- Hoerl, A. E., & Kennard R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- Johnson P., & Ashworth J. (1990). Modelling tourism demand: a summary review, *Leisure Studies*. 9(2), 145-161.
- Karahan, M. (2015). Turizm talebinin yapay sinir ağları yöntemiyle tahmin edilmesi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 20(2), 195-209.
- Kılıç, F., Akkaya, M. R., & Memili, N. (2018). Yemekhane için yapay zekâ teknikleri kullanımı ile günlük talep tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, (13), 65-71.
- Law, R., & Au, N. (1999). A neural network model to forecast japanese demand for travel to hong kong. *Tourism Management*, 20(1), 89-97.
- Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: a deep learning approach. *Annals of Tourism Research*, 75, 410-423.
- Lim, C., Chang, C., & McAleer, M. (2009). Forecasting H(M) otel guest nights in New Zealand. *International Journal of Hospitality Management*, 28(2), 228-235.
- Loganathan, N., & Yahaya, İ. (2010). Forecasting international tourism demand in Malaysia using box jenkins SARIMA application. *South Asian Journal of Tourism and Heritage*, 3, 50-60.
- Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International journal of Forecasting*, 9(4), 527-529.
- Marquardt, D.W., & Sneec, R.D. (1975). Ridge regression in pratice. *The American Statistician*, 29(1), 3-20.
- McDonald, G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 93-100.
- Neter, J., Kutner, M. H., Nachtsheim, C. J., & William, L. (1996). *Applied linear statistical models* (4th ed.). Chicago, McGraw-Hill Irwin.
- Oskay, C. (2012). Mersin turizminin Türkiye ekonomisindeki yeri ve önemi. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 21(2), 185-202.
- Pandis, N. (2016). Multiple linear regression analysis. *American Journal of Orthodontics and Dentofacial Orthopedics*, 149(4), 581.
- Pereira, L. N., & Cerqueira, V. (2022). Forecasting hotel demand for revenue management using machine learning regression methods. *Current Issues in Tourism*, 25(17), 2733-2750.
- Sevüktekin M., & Çınar M. (2017). *Ekonometrik Zaman Serileri Analizi*. Bursa: Dora Yayıncılık.
- Shan, J., & Wilson, K. (2001). Causality between trade and tourism: empirical evidence from china. *Applied Economics Letters*, 8, 279-283.

- Song, H., Wong, K.K.F., & Chon, K.K.S. (2003). Modelling and forecasting the demand for Hong Kong tourism. *International Journal of Hospitality Management*, 22, 435-451.
- Soysal, M., & Ömürganülşen, M. (2010). Türk turizm sektöründe talep tahmini üzerine bir uygulama. *Anatolia: Turizm Araştırmaları Dergisi*, 21(1), 128-136.
- Sönmez, O., & Zengin, K. (2019). Yiyecek ve içecek işletmelerinde talep tahmini: yapay sinir ağları ve regresyon yöntemleriyle bir karşılaştırma. *Avrupa Bilim ve Teknoloji Dergisi*, 302-308.
- Sun, S., Li, Y., Guo, J., & Wang, S. (2022). Tourism demand forecasting: An ensemble deep learning approach. *Tourism Economics*, 28(8), 2021-2049.
- Şamkar, H. (2018). Modeling of the numbers of tourists staying at hotels in Antalya with the robust ridge regression. *Journal of Tourism & Gastronomy Studies*, 6(3), 295-315.
- Topal, M., Eyduran, E., Yağanoğlu, A. M., Sönmez, A., & Keskin, S. (2013). Çoklu doğrusal bağlantı durumunda ridge ve temel bileşenler regresyon analiz yöntemlerinin kullanımı. *Atatürk Üniversitesi Ziraat Fakültesi Dergisi*, 41(1), 53-57.
- Tranmer, M., & Elliot, M. (2008). Multiple linear regression. *The Cathie Marsh Centre for Census and Survey Research*, 5(5), 1-5.
- Uysal, M., & El Roubi, M. S. (1999). Artificial neural network versus multiple regression in tourism demand analysis. *Journal of Travel Research*, 38(2), 111-118.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer, New York, 314s.
- Vivas, E., Allende-Cid, H., & Salas, R. (2020). A systematic review of statistical and machine learning methods for electrical power forecasting with reported mape score. *Entropy*, 22(12), 1412.
- Willmott C. J., & Matsuura K. (2005). Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in assessing average model performance. *Climate Research*, 30, 79-82.
- World Tourism Organization, (2023). *World Tourism Barometer 2023*. Erişim Adresi: <https://www.e-unwto.org/toc/wtobarometereng/21/2>.
- Yerkök, D. (2020). *Yapay zekâ yöntemleri ile toplu yemek üretiminde günlük talep tahmini*. (Yüksek Lisans Tezi). Mersin Üniversitesi, Mersin.
- Yüksel, S. (2007). An integrated forecasting approach to hotel demand. *Mathematical and Computer Modelling*, 46(7-8), 1063-1070.
- Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. *Machine Learning Methods and Applications to Brain Disorders*, 123-140.

Özgeçmiş

Serkan KARDEŞ (Arş. Gör.), Bursa Teknik Üniversitesi İnsan ve Toplum Bilimleri Fakültesi İşletme Bölümü Sayısal Yöntemler Anabilim Dalında Araştırma Görevlisi olarak görev yapmaktadır. Uludağ Üniversitesi Ekonometri bölümünde doktora öğrencisidir. Ekonometri, makine öğrenmesi ve çok değişkenli istatistiksel analizler gibi konularda araştırmalar yapmaktadır.

Burcu ÖNGEN BİLİR (Doç. Dr.), Bursa Teknik Üniversitesi İnsan ve Toplum Bilimleri Fakültesinde Sayısal Yöntemler Anabilim Dalında Doçenttir. Uludağ Üniversitesinde İstatistik doktorasını tamamlamıştır. Araştırma alanları arasında çok değişkenli veri analizi, bayesyen analiz ve yapısal eşitlik modelleri yer almaktadır. Araştırmaları, İstanbul Üniversitesi Sosyoloji Dergisi, Journal of Economy Culture and Society ve İş Güç, Endüstri İlişkileri ve İnsan Kaynakları Dergisi gibi dergilerde yayınlanmaktadır.