

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY SİNİR AĞLARI VE MAKİNE ÖĞRENMESİ İLE
OTOMOBİL SATIŞ TAHMİNİNİN YAPILMASI VE ZAMAN
SERİLERİ ANALİZİ İLE KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

Beyza KURTGERİ

Endüstri Mühendisliği Anabilim Dalı

SUBAT 2024

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YAPAY SİNİR AĞLARI VE MAKİNE ÖĞRENMESİ İLE
OTOMOBİL SATIŞ TAHMİNİNİN YAPILMASI VE ZAMAN
SERİLERİ ANALİZİ İLE KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

Beyza KURTGERİ

Endüstri Mühendisliği Anabilim Dalı

Tez Danışmanı: Doç. Dr. Merve CENGİZ TOKLU

ŞUBAT 2024

Beyza KURTGERİ tarafından hazırlanan “Yapay Sinir Ağları ve Makine Öğrenmesi ile Otomobil Satış Tahmininin Yapılması ve Zaman Serileri Analizi ile Karşılaştırılması” adlı tez çalışması 02.02.2024 tarihinde aşağıdaki jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı :	Dr. Öğr. Üyesi Enes Furkan ERKAN Sakarya Üniversitesi	
Jüri Üyesi :	Doç. Dr. Merve CENGİZ TOKLU Sakarya Üniversitesi	
Jüri Üyesi :	Dr. Öğr. Üyesi Seda Hatice GÖKLER Kahramanmaraş Sütçü İmam Üniversitesi	



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “YAPAY SİNİR AĞLARI VE MAKİNE ÖĞRENMESİ İLE OTOMOBİL SATIŞ TAHMİNİNİN YAPILMASI VE ZAMAN SERİLERİ ANALİZİ İLE KARŞILAŞTIRILMASI” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(05/01/2024)

Beyza KURTGERİ





Anneme ve Babama...



TEŞEKKÜR

Tez çalışmamın hazırlanmasında desteğini esirgemeyen sayın danışman hocam Doç. Dr. Merve Cengiz Toklu'ya teşekkürlerimi sunarım.

Bugüne kadar benden maddi ve manevi desteklerini esirgemeyen babam Hamdi Kurt'a, annem Fikriye Kurt'a, eşim Mehmet Akif Kurtgeri'ye çalışmam sürecinde yanımda oldukları için çok teşekkür ederim.



Beyza KURTGERİ



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	ix
İÇİNDEKİLER	xi
KISALTMALAR	xiii
SİMGELER	xv
TABLO LİSTESİ	xvii
ŞEKİL LİSTESİ	xix
ÖZET	xxi
SUMMARY	xxiii
1.GİRİŞ	1
1.1. Literatür Taraması	3
2. TALEP TAHMİNİ	9
2.1. Talep Tahmininin Önemi	10
2.2. Talep Tahminlerinin Sınıflandırılması	11
2.3. Talep Tahmin Yöntemleri	11
2.3.1. Nitel yöntemler (Kalitatif)	12
2.3.2. Nicel yöntemler (Kantitatif)	13
2.3.2.1. Nedensel modeller	13
2.3.2.2. Zaman serileri modelleri	15
2.3.3. Yapay zeka tabanlı yöntemler	17
2.3.3.1. Yapay sinir ağları (ANN-YSA)	17
2.3.3.2. Gradyan artırma yöntemi (Gradient Boosting)	19
2.3.3.3. Kategorik artırma yöntemi (CatBoost)	20
2.3.3.4. Rastgele ormanlar yöntemi (Random Forest - RF)	21
2.3.3.5. Destek vektör regresyonu (SVR)	21
2.4. Talep Tahmin Aşamaları	22
3. MATERYAL VE YÖNTEM	25
3.1. Veri Setinin Oluşturulması ve Analiz Edilmesi	25
3.2. Makine Öğrenmesi ve Yapay Sinir Ağları için Veri Ön İşleme	31
3.2.1. Aykırı değer analizi	31
3.2.2. Standardizasyon	35
3.2.3. Makine öğrenmesi ile talep tahmini	37
3.2.3.1. Veri setindeki bağımlı ve bağımsız değişkenlerin ayrılması	37
3.2.3.2. Veri setinin eğitim ve test verisi olarak ayrılması	37
3.2.3.3. Makine öğrenmesi modellerinin performanslarının karşılaştırılması ve talep tahmin modelinin seçilmesi	38
3.3. Performansı Yüksek Makine Öğrenmesi Modelleri ile Uygulama	39
3.4. Zaman Serisi Analizi ile Talep Tahmini	40
3.4.1. Mevsimsel otoregresif entegre hareketli ortalama yöntemi	40
4. UYGULAMA	59
4.1. Tahmin Yöntemleri ve Uygulama	59

4.1.1. CatBoost ile talep tahmini	59
4.1.2. Gradient boosting ile talep tahmini	60
4.1.3. Rastgele ormanlar (Random Forest) ile talep tahmini	61
4.1.4. Destek vektör regresyonu (SVR) ile talep tahmini	62
4.1.5. Yapay sinir ağları (YSA) ile talep tahmini.....	63
4.1.5.1. YSA parametrelerinin belirlenmesi.....	63
4.1.5.2. YSA ağ tasarımı	64
4.1.5.3. YSA ağ eğitimi.....	65
4.1.5.4. YSA ağ performans sonuçları	66
4.1.6. SARIMA modeli ile talep tahmini	67
4.1.7. Yöntemlerin değerlendirilmesi.....	69
4.1.7.1. Determinasyon katsayısı (R2)	70
4.1.7.2. Ortalama kare hata karekökü (RMSE)	70
5. SONUÇ VE ÖNERİLER	73
KAYNAKLAR.....	75
EKLER.....	83
ÖZGEÇMİŞ	85

KISALTMALAR

LOF	: Yerel aykırı değer faktörü
RF	: Rassal karar ormanları
SVR	: Destek vektör regresyonu
YSA	: Yapay sinir ağları
ANN	: Yapay sinir ağları
SARIMA	: Mevsimsel otoregresif entegre hareketli ortalama
SVM	: Destek vektör makineleri
GBM	: Gradyan arttırma makineleri
CatBoost	: Kategorik artırma yöntemi
XGB	: Aşırı gradyan arttırma yöntemi
R²	: Determinasyon katsayısı
RMSE	: Ortalama kare hata karekökü
TÜİK	: Türkiye İstatistik Kurumu
TCMB	: Türkiye Cumhuriyeti Merkez Bankası
EVDS	: Elektronik veri dağıtım sistemi
ODD	: Otomotiv Distribütörleri Derneği
RKGE	: Reel kesim güven endeksi
TGE	: Tüketici güven endeksi
ACF	: Otokorelasyon fonksiyonu
PACF	: Kısmi otokorelasyon fonksiyonu
ADF	: Augmented Dickey Fuller birim kök testi
KPSS	: Kwiatkowski-Phillips-Schmidt-Shin birim kök testi



SİMGELER

Z	: Z-Score puanı
μ	: Ortalama
σ	: Standart sapma
p	: Modele dahil edilen gecikmelerin sayısı
d	: Farklılaşma derecesi
q	: Hata gecikme sayısı
P	: Otoregresif sırası
D	: Fark sırası
Q	: Hareketli ortalama sırası
m	: Her sezondaki periyot sayısı
Q1	: 1. çeyreklik
Q3	: 3. çeyreklik
IQR	: Çeyrekler açıklığı
ρ_k	: Otokorelasyon katsayısı
cov	: Kovaryans
var	: Varyans
ρ_{kk}	: Kısmi otokorelasyon katsayısı
ρ_k	: k. gecikmedeki otokorelasyon katsayısı
J	: Kısmi otokorelasyon



TABLO LİSTESİ

Sayfa

Tablo 1.1. Literatür özet tablosu.	5
Tablo 3.1. Bağımsız değişkenler özet tablosu.	25
Tablo 3.2. Örnek uygulama veri seti.	30
Tablo 3.3. Orijinal veriler korelasyon tablosu.	33
Tablo 3.4. Aykırı değer analizi yapılmış veriler korelasyon tablosu.	35
Tablo 3.5. Değişkenlerin ortalama ve standart sapma değerleri.	36
Tablo 3.6. Aykırı değer analizi ve standardizasyon işlemi sonucunda oluşan örnek uygulama veri seti	36
Tablo 3.7. Makine öğrenmesi yöntemleri performans analizi tablosu.	38
Tablo 3.8. Zaman serisi örnek veri seti.	42
Tablo 3.9. Zaman serisinin tanımlayıcı istatistikleri.	43
Tablo 3.10. Box-plot yöntemi ile belirlenen aykırı değerler.	45
Tablo 4.1. YSA parametre kombinasyonları.	64
Tablo 4.2. Model performans sonuçları.	71



ŞEKİL LİSTESİ

Sayfa

Şekil 2.1. Yapay sinir ağı model yapısı (Mijwil, 2017).....	19
Şekil 2.2. Gradient boosting algoritmasının çalışma yapısı (Gradient Boosting, 2019).	20
Şekil 2.3. RF algoritmasının çalışma yapısı (Ekelik ve Altaş, 2019).	21
Şekil 2.4. SVR Hiper Düzlem Grafiği (Support Vector Regression, 2023).	22
Şekil 2.5. Talep tahmin aşamaları.....	24
Şekil 3.1. İthalat miktar endeksi.	27
Şekil 3.2. Reel kesim güven endeksi.	28
Şekil 3.3. Tüketici güven endeksi.	28
Şekil 3.4. Otomobil ortalama faiz oranları.....	29
Şekil 3.5. Türkiye otomobil üretim sayısı.....	29
Şekil 3.6. Otomobil satış sayısı.....	30
Şekil 3.7. Veri setinin zamana bağlı grafiği.....	40
Şekil 3.8. Otomobil satış sayılarına ait histogram grafiği.....	43
Şekil 3.9. Örnek Box-Plot grafiği (yalindunya, 2020).....	43
Şekil 3.10. Ham veriler üzerinde oluşturulan box-plot grafiği.	44
Şekil 3.11. Aykırı değerlerin grafik üzerinde noktasal gösterimi.	45
Şekil 3.12. Veri setine ait ayrıştırma grafikleri.....	46
Şekil 3.13. Aylık ortalama otomobil satış grafiği.	46
Şekil 3.14. Ham veri ACF grafiği.....	48
Şekil 3.15. Ham veri PACF grafiği.....	49
Şekil 3.16. Ham veri üzerinde ADF birim kök testi.	50
Şekil 3.17. Ham veri üzerinde KPSS birim kök testi.....	51
Şekil 3.18. 1.dereceden mevsimsel farkı alınmış serinin ACF grafiği.	52
Şekil 3.19. 1.dereceden mevsimsel farkı alınmış serinin PACF grafiği.	52
Şekil 3.20. 1.dereceden mevsimsel farkı alınmış verinin ADF birim kök testi.	53
Şekil 3.21. 1.dereceden mevsimsel farkı alınmış verinin KPSS birim kök testi.	53
Şekil 3.22. 1. dereceden mevsimsel olmayan farkı alınmış verinin ACF grafiği.	54
Şekil 3.23. 1. dereceden mevsimsel olmayan farkı alınmış verinin PACF grafiği.	54
Şekil 3.24. 1.dereceden mevsimsel olmayan farkı alınmış verinin ADF birim kök testi.	55
Şekil 3.25. 1.dereceden mevsimsel olmayan farkı alınmış verinin KPSS birim kök testi.	55
Şekil 3.26. 2. dereceden mevsimsel olmayan farkı alınmış ve durağanlaşmış verinin ACF grafiği.	56
Şekil 3.27. 2. dereceden mevsimsel olmayan farkı alınmış ve durağanlaşmış verinin PACF grafiği.	56
Şekil 3.28. 2.dereceden mevsimsel olmayan farkı alınmış ve durağanlaşmış verinin ADF birim kök testi.	57
Şekil 3.29. Orijinal ve durağanlaştırılmış veri setinin zaman serisi grafiği.....	57
Şekil 3.30. Durağanlaştırılmış veri özellik grafiği.....	58

Şekil 4.1. CatBoost ile tahmin değerleri ve gerçek otomobil satış sayıları.	60
Şekil 4.2. Gradient boosting ile tahmin değerleri ve gerçek otomobil satış sayıları..	61
Şekil 4.3. RF ile tahmin değerleri ve gerçek otomobil satış sayıları.	62
Şekil 4.4. SVR ile tahmin değerleri ve gerçek otomobil satış sayıları.....	63
Şekil 4.5. YSA ağ yapısı	64
Şekil 4.6. YSA performans grafiği.....	65
Şekil 4.7. YSA tahmin ve gerçek değerler ilişkisi	66
Şekil 4.8. YSA ile tahmin değerleri ve gerçek otomobil satış sayıları.	67
Şekil 4.9. Optimizasyon ile otomobil satış tahmin grafiği.....	69
Şekil 4.10. Çapraz doğrulama ile otomobil satış tahmini grafiği.....	69



YAPAY SİNİR AĞLARI VE MAKİNE ÖĞRENMESİ İLE OTOMOBİL SATIŞ TAHMİNİNİN YAPILMASI VE ZAMAN SERİLERİ ANALİZİ İLE KARŞILAŞTIRILMASI

ÖZET

Talep tahmini, günümüz koşullarını da göz önüne aldığımızda talep üzerinde çeşitli etkenler olması sebebiyle işletmelerin kolay ve etkili kararlar alabilmesini sağlamaktadır. İşletmeler müşteri memnuniyetin artırılması, iyi bir üretim planının gerçekleştirilmesi, kaynakların verimli kullanılabilmesi için güvenilir talep tahminleri gerçekleştirmeyi amaçlamaktadırlar. Belirsizlikler altında, gelecek için öngöründe bulunmak işletmeler için büyük avantajlar sağlayabilmektedir. Otomobil talebini etkileyen birçok faktör bulunmaktadır. Otomobil talebinin ekonomik, çevresel ve sosyal etkenlerin tümünden etkilenebilmesi sebebiyle tahmininin gerçekleştirilmesi kolay olmamaktadır. Gelecek tahmini yapmanın dışında, talep tahmini geçmişte yaşanan ve gelecekte de yaşanabilecek olan olay ve durumların yorumlanmasına fırsat tanımaktadır. Bu sebepler göz önüne alındığında otomobil talep tahminini gerçekleştirmek önemli olmaktadır. Nitel yöntemlerin dışında daha güvenilir sonuçlar elde edilmesi için matematiksel yöntemlere dayanan algoritmalar talep tahmininde kullanılmaktadır. Matematiksel yöntemler literatürde, sıklıkla kullanılan ve yüksek performans gösteren yöntemler olarak yer almaktadır. Özellikle son yıllarda yapılan çalışmalarda, talep tahmini üzerinde yapay zeka yöntemlerinin önemli bir yer tuttuğu görülmektedir.

Bu çalışmada, otomobil talep tahmini uygulaması gerçekleştirilmiş olup makine öğrenmesi algoritmaları, yapay sinir ağları ve zaman serileri analizi yöntemleri tahmin performansları analiz edilmiş ve karşılaştırılmıştır. 2014-2022 yılları arasındaki veriler ile uygulama gerçekleştirilmiştir. Çalışmada, bağımlı veri olarak otomobil satış verileri kullanılmıştır. Bağımsız veriler ise otomobiller için ithalat miktar endeksi, reel kesim güven endeksi, tüketici güven endeksi, ortalama taşıt kredi faiz oranları, Türkiye otomobil üretim adetleri ve zaman olarak belirlenmiştir. Yöntemler ile daha yüksek performans elde edilmesi ve daha güvenilir tahminler yapılması için veri ön işleme uygulamaları gerçekleştirilmiştir. Makine öğrenmesi algoritmaları, yapay sinir ağları ve zaman serileri analizi için veriden güvenilir tahminler elde etmeyi zorlaştıran engellerden olan aykırı veriler için aykırı değer analizi gerçekleştirilmiştir. Çalışmada makine öğrenmesi yöntemlerinden Kategorik Arttırma Yöntemi (CatBoost), Gradyan Arttırma Yöntemi (Gradient Boosting), Rassal Karar Ormanları (Random Forest-RF), Destek Vektör Regresyonu (SVR) algoritmaları, Yapay Sinir Ağları (YSA) ve Zaman Serileri Analizi yöntemleri kullanılmış ve performans analizi gerçekleştirilerek yöntem karşılaştırması yapılmıştır. Sonuç olarak, 0,85 R^2 değeri ile en yüksek performans gösteren yöntemin makine öğrenmesi yöntemlerinden CatBoost algoritması olduğu görülmüştür. Otomotiv sektörü, ülke ekonomisinin etkili sektörlerinden biri olmasının yanında, diğer sektörler için de tedarikçi konumundadır. Bu sebeple otomobil sektöründeki taleplere güvenilir cevap verilmesi bir gereklilik haline

gelmektedir. Bu çalışmada da otomobil talebine doğru ve güvenilir bir yaklaşım sergileyen algoritmanın bulunması amaçlanmıştır.



FORECASTING AUTOMOBILE SALES USING ARTIFICIAL NEURAL NETWORKS AND MACHINE LEARNING AND COMPARISON WITH TIME SERIES ANALYSIS

SUMMARY

Considering today's conditions, demand forecasting enables businesses to make easy and effective decisions as there are various factors on demand. Companies are required to make and implement correct decisions in order to have a competitive edge. Demand forecasting plays a significant role in this aspect. Companies must adapt to rapidly changing demands and desires of consumers. With the advancement of technology, meeting demand, especially in the automotive sector, has become challenging for firms. Due to these reasons, analyzing data effectively and meeting customer expectations has become a necessity for companies. Businesses aim to make reliable demand forecasts to ensure customer satisfaction, realize a good production plan, and use capacity efficiency. Making predictions for the future under uncertainties can provide significant advantages for businesses. Demand forecasting is classified into short, medium, and long-term categories, applicable in daily plans, supply process plans, and large-scale plans requiring capital investment. Initiating demand forecasting involves identifying the dependent variable and the independent variables influencing it. Subsequently, data must be collected and analyzed. For the creation of high-performance models, thorough data analysis and working with complete data are crucial.

In the data analysis stage, prediction periods need to be determined. To create an appropriate prediction model, the correct method must be selected. Finally, the model needs to be analyzed with data that it has not encountered before. There are numerous factors influencing automobile demand, making its prediction challenging due to the impact of economic, environmental, and social factors. Beyond predicting the future, demand forecasting allows for the interpretation of past and potential future events and situations. Given these reasons, predicting automobile demand becomes crucial. Mathematical algorithms based on machine learning methods, artificial neural networks, and time series analysis are used for demand forecasting to obtain more reliable results than traditional methods.

Machine learning methods, artificial neural networks, and time series analysis are frequently used and high-performance techniques in the literature. In recent studies, artificial intelligence methods have played a significant role in demand forecasting. Artificial intelligence methods, including artificial neural networks and machine learning, are employed in various fields. These systems are built to mimic human intelligence, acquiring information, learning, and performing analysis through generalization. Artificial neural networks and machine learning methods are utilized not only for regression problems but also for solving classification problems. ANNs can be utilized in various studies, including regression, classification, and data correlation. ANNs consist of input, output, and hidden layers. The input layer transmits data received from the external environment to the hidden layer. The data

is processed in the hidden layer, and inferences are drawn from this processed data. Subsequently, the data is forwarded to the output layer. The output layer communicates the processed data from the hidden layer to the external environment.

Time series analysis is a demand forecasting method where past data is analyzed to gain insights into future periods. Time series comprises four components: trend, seasonality, cyclical, and irregular components. The trend component illustrates the linear increase or decrease in the time series over time. The seasonality component indicates the repeating pattern of the model at fixed time intervals according to seasons. Stationarity is an important concept in time series. Time series encompass trend, seasonality, and noise components. Therefore, in particular, the condition of stationarity is sought in the SARIMA method commonly used in studies. Time series are detrended from trend and seasonality. There are various methods to make time series stationary. One of these methods frequently used in the literature is differencing. The differencing method is divided into seasonal and non-seasonal differencing operations. If the time series exhibits seasonality, seasonal differencing is performed. If there is a trend within the time series, non-seasonal differencing is done. Unit root tests are applied to check the stationarity of the series. Time series analysis provides more reliable results for short-term predictions.

In this study, an automobile demand forecasting application has been implemented, and the prediction performances of machine learning algorithms, artificial neural networks, and time series analysis methods have been analyzed and compared. The application was conducted using data from the years 2014-2022. The dependent variable used in the study is automobile sales data, and independent variables include the import quantity index for automobiles, real sector confidence index, consumer confidence index, average vehicle loan interest rates, Turkey's automobile production quantities, and time. The import quantity index measures the change in automobile imports. The real sector confidence index is a measure indicating the trend in the manufacturing industry based on the views of top-level managers engaged in production. The consumer confidence index is an indicator showing consumers' financial situations and future spending. Vehicle loan interest rates play a significant role in automobile sales. Economic conditions are among the factors that most influence automobile sales. Data preprocessing applications were performed to achieve higher performance with the methods and make more reliable predictions. Outlier analysis was conducted for outliers, which pose challenges to obtaining reliable predictions, using the local outlier factor (LOF) method, a community learning approach for machine learning algorithms and time series analysis. The LOF method looks at the dataset as a whole rather than variable-wise, as variable-wise outlier analysis may be insufficient for multivariate datasets. During the LOF method, correlation coefficients of the data were compared, and a threshold value needed for the LOF method was determined. Selection of neighborhood coefficient and density parameter is crucial in the LOF method. After outlier analysis, standardization was applied to the data to bring them to the same scale before model creation. The goal is to make the variables comparable. For artificial neural network modeling, standardization was applied to the dataset using the z-score method. The mean of the data is set to zero, and the standard deviations are set to one. For time series analysis, outlier analysis was conducted using the box-plot method, a univariate outlier analysis method. Before creating models, dependent and independent variables of the dataset were separated. Subsequently, variables were split into training and test sets. In machine learning methods, the dataset was used as

80% training and 20% test data. For artificial neural networks, the dataset was divided into 65% training, 15% validation, and 20% test data. Different algorithms were tested in machine learning methods, and the best-performing algorithms were selected for further application. SARIMA method was used for time series analysis. Outlier analysis was performed, and then the dataset with seasonality was differenced to make it stationary. When determining the parameters of the SARIMA model, hyperparameter optimization, cross-validation for time series, and autocorrelation graph analysis were used.

In the study, the Categorical Boosting Method (CatBoost), Gradient Boosting Method, Random Forest (RF), Support Vector Regression (SVR) algorithms from machine learning methods, Artificial Neural Networks (ANN), and the Seasonal Autoregressive Integrated Moving Average (SARIMA) method from time series analysis were used. A performance analysis was conducted, and a method comparison was made. CatBoost is an algorithm which is based on Gradient Boosting. It makes reliable predictions with a small number of data. It can handle numerical, categorical, and text data. It has a high learning rate and can cope with overfitting. Gradient boosting uses the entire dataset, initially forming a single tree and then creating subsequent trees to minimize the errors of the previous ones. Random Forest can be used for both classification and regression problems, forming multiple decision trees and obtaining a performance output from their averages. It provides more reliable results with datasets containing a large number of data points. Support Vector Regression is a machine learning algorithm based on Support Vector Machines, aiming to create a hyperplane encompassing data points and minimizing errors. For the evaluation of methods, determination coefficient and root mean square error performance metrics were used. As a result, the CatBoost algorithm was observed to have the highest performance among the machine learning methods.



1. GİRİŞ

Üretim ve hizmet sektörü içerisinde olan tüm firmaların rekabet edebilme gücü, doğru kararları alma ve uygulama yöntemlerine dayanmaktadır. Buna bağlı olarak, geçmiş verilerden gelecek verileri öngörebilme karar vermede önemli bir rol almaktadır. Teknolojinin büyük bir hızla gelişip, tüketicinin beklenti ve isteklerinin hızlı değiştiği, firmalar arasındaki rekabetin fazla olduğu, ekonomide stratejik yere sahip olan otomotiv sektöründe taleplere karşılık vermek firmalar için zorlaşmaktadır. Bu sebeple, otomobil firmalarının, bu gelişmelerden geri kalmaması için taleplere karşılık vermesi gereklilik haline gelmiştir. Bu bağlamda, otomotiv firmaları için hem verimli bir üretim planlama yapılması hem de müşteri beklenti ve isteklerine karşılık verilmesi adına veri analizi önemli bir konu haline gelmiştir. Otomotiv sektörü, ekonomik, mevsimsel ve çevresel etkenler, kriz dönemleri gibi birçok dalgalanmanın etkisi altında kalabilmektedir. Bu değişimlerin içerisinde, otomobil firmalarının üretim ve pazarlama stratejileri, taleplere karşılık verebilme yetkinlikleri, stok yönetiminde karar verme süreçleri önemli olmaktadır. Talep tahmini, gelecek dönemler hakkında bilgi vermenin yanı sıra geçmiş dönemler için de yaşanmış olay ve durumlar hakkında yorum yapabilmeyi sağlamaktadır.

Otomotiv endüstrisi, ülke kalkınmasında etkili olan önemli bir sektördür. Diğer sektörlerin de tedarikçisi konumunda bulunan otomotiv endüstrisinde taleplere cevap verilmesi gerekmektedir. Sarı (2016) çalışmasında, satış talebinin düşük hatalar ile tahmin edilmesinin, sektör içerisinde bulunan firmalar için bir gereklilik olduğunu belirtmiştir. Kısa veya uzun dönemli tahminde bulunmak, firmaların üretim, lojistik, pazarlama gibi faaliyetlerini etkilemekte ve bu sebeple önem kazanmaktadır. Düşük hatalar ile yapılan talep tahmini hem karar vermede karar verme aşamalarını kolaylaştırmakta hem de zaman ve maliyette verimlilik sağlamaktadır. Talep tahminlerinin, talep tahmin yöntemlerinden destek alınarak uygulanması, işletmeler için önemli kararların verilmesinde güvenilir olması açısından şart olarak düşünülmektedir (Meydan, 2018). Bu durumlara bağlı olarak talep tahmini için birçok yöntem geliştirilmiştir. Kantitatif metotlar sayısal verilerle işlem yaparak matematiksel yöntemlere dayanan sistemli yaklaşımları kapsamaktadır. Veriye dayalı

tahminleme yapması sebebiyle güvenilir olmaktadır. Makine öğrenmesi algoritmaları, yapay sinir ağları ve zaman serisi analizi kantitatif metotlar arasında yer almaktadır.

Bu çalışmada da, makine öğrenmesi algoritmaları, yapay sinir ağları ve zaman serisi analizi ile talep tahmin uygulaması yapılmıştır. Çalışma 2014 Ocak – 2022 Aralık yılları arasındaki aylık otomobil satış miktarları ile gerçekleştirilmiştir.

Tahmin yöntemi olarak, makine öğrenme yöntemlerinden Kategorik Arttırma Yöntemi (CatBoost), Gradyan Arttırma Yöntemi (Gradient Boosting), Rastgele Karar Ormanları (Random Forest), Destek Vektör Regresyonu (SVR) algoritmaları, Yapay Sinir Ağları (YSA) yöntemi ile birlikte Zaman Serileri Analizi yöntemi kullanılmıştır. Türkiye’de aylık satışı gerçekleşen toplam otomobil miktarları bağımlı veri olarak kullanılmış ve yöntemlerin tahminleri için karşılaştırılmalı performans analizi gerçekleştirilmiştir. Otomobiller için İthalat Miktar Endeksi, Reel Kesim Güven Endeksi, Tüketici Güven Endeksi, Ortalama Taşıt Kredi Faiz Oranları, Türkiye Otomobil Üretim Adetleri ve zaman bağımsız verilerdir. Bağımsız veriler seçilirken, literatürdeki yüksek performans gösteren talep tahmin yöntemleri baz alınmış ve bağımsız değişkenlerin bağımlı değişken ile arasındaki ilişkiye göre değerlendirilmiştir. İstatistiksel yöntemlerden korelasyon yöntemi ile analiz edilmiştir, otomobil satış miktarı ile arasında yüksek korelasyon katsayısına sahip değişkenler çalışmada bağımsız değişken olarak kullanılmıştır. Veriler üzerinde veri ön işleme çalışmaları yapılmıştır. Makine öğrenmesi algoritmalarının ve yapay sinir ağları yönteminin performanslarının daha yüksek olabilmesi için aykırı veri analizi ve verilerin aynı ölçeğe indirgenebilmesi için standardizasyon işlemi gerçekleştirilmiştir. Model aşamasında performansı yüksek sonuçlar elde edebilmek adına hiperparametre optimizasyonu uygulanmıştır. Makine öğrenmesi yönteminde optimal sonuçlara ulaşabilmek için hiperparametre optimizasyonu sırasında çapraz doğrulama (cross validation) kullanılmıştır. Zaman serileri analizi yönteminde modelin aykırı değerlerden etkilenmesini engellemek adına aykırı veri analizi gerçekleştirilmiştir. Zaman serileri analizi ile tahmin yapılırken, Box-Jenkins tahmin modellerinden SARIMA yöntemi kullanılmıştır. Kullanılan yöntemlerin performans analizi yapılmıştır ve yöntemlerin performansları karşılaştırılmıştır. Sonuç olarak, en yüksek performans gösteren yöntemin makine öğrenmesi yöntemlerinden CatBoost algoritması olduğuna karar verilmiştir.

1.1. Literatür Taraması

Bu bölümde, çalışmada uygulanmış yöntemlerden Kategorik Arttırma Yöntemi (CatBoost), Gradyan Arttırma Yöntemi (Gradient Boosting), Rassal Karar Ormanları (RF), Destek Vektör Regresyonu (SVR), Yapay Sinir Ağları (YSA), Zaman Serileri Analizi ile talep tahmini ve otomobil satış tahmini ile ilgili literatürde yer alan çalışmalarının özeti verilmiştir.

Tekin ve Sarı (2022) çalışmalarında, makine öğrenmesi tabanlı algoritmaları kullanarak mağaza zincirleri için satış tahmini yapmışlardır. Algoritmalar üzerinde hiperparametre optimizasyonu ve çapraz doğrulama gerçekleştirmişlerdir. CatBoost algoritmasını en iyi sonucu verdiğini fakat LightGBM algoritması ile uygulamada CatBoost ile çok yakın bir hata değeri aldıklarını, bu sebeple sonuç üretme hızı çok daha yüksek olduğundan LightGBM algoritmasının kullanılmasının çalışmalarda uygulanabilirliğinden bahsetmişlerdir. Ayrıca makine öğrenmesi algoritmaları ile zaman serileri analizini karşılaştırmışlar ve MSE performans değerlendirme ölçütü ile analiz yaptıklarında zaman serileri analizinden ARIMAX ve SARIMAX yöntemlerinin performansı çok daha yüksek sonuçlar çıkardığını vurgulamışlardır. Bu sebeple uygulamada kullanmış oldukları veri setinin mevsimsellikten çok fazla etkilenmediği yorumunu yapmışlardır.

Huang vd. (2019) çalışmalarında, 2001-2005 yılları Güney Çin meteorolojik verilerini kullanmışlardır. CatBoost, RF ve SVM algoritmaları ile tahmin sonuçlarını RMSE ve MAPE performans ölçütleri ile karşılaştırmışlardır. Üç yöntemin de kullanılabilir olduğunu, başarılı sonuçlar verdiğini göstermişlerdir. CatBoost algoritmasının diğer yöntemlere göre tahmin doğruluğu ve hesaplama hızı konularında daha yüksek bir performans gösterdiğini belirtmişlerdir.

Yeşilyurt ve Dalkılıç (2021) tahmin çalışmalarında, günlük nehir akım verilerini kullanmışlardır. Extreme Gradient Boosting(XGBoost) ve Gradient Boosting(GBM) algoritmaları ile uygulama gerçekleştirmişlerdir. Her iki yöntemin de başarılı sonuçlar verdiğini ve çalışmalarda kullanılabileceğini belirtmişlerdir. En yüksek performansı XGBoost algoritması ile elde etmişlerdir.

Sevgen ve Aliefendioğlu (2020) çalışmalarında, Rassal karar Ormanları (RF) algoritmasını kullanarak gayrimenkul değerlendirme üzerinde tahmin uygulaması gerçekleştirmişlerdir. Veri setinin %75' ini eğitim, %25' ini test verisi olarak

kullanmışlardır. Gerçek ve tahmin değerleri arasında ortalama 25.000 TL fark gözlemlenmiştir. RF algoritmasının sonucunun başarılı olduğunu ve çalışmalarda kullanılabileceğini belirtmişlerdir.

Çemrek ve Demir (2021) çalışmalarında, karbondioksit (CO₂) emisyon miktarı tahminini destek vektör regresyonu(SVR), yapay sinir ağları (YSA) ve Box-Jenkins yöntemi kullanarak gerçekleştirmişlerdir. MSE performans ölçütü ile karşılaştırmaları sonucunda, performansı en yüksek yöntemin YSA olduğuna karar vermişlerdir.

Kadir Sheikh ve Unde (2012) çalışmalarında, elektrik yük tahminini yapay sinir ağları(YSA) yöntemi ile gerçekleştirmişlerdir. Uygulamayı gerçekleştirirken Matlab paket programını kullanmışlardır. Saatlik tüketim miktarlarını kullanmışlardır. Sonuç olarak, MAPE performans ölçütü ile değerlendirmişler ve %0,956 performans ile yüksek başarı elde etmişlerdir.

Çekiç (2023) çalışmasında, hava kargo taşımacılık verilerini kullanarak linear regresyon, karar ağaçları regresyonu, rassal karar ormanları ve aşırı gradyan artırma yöntemlerini kullanmış ve MAE, MSE, RMSE, R² performans ölçütlerine göre performansı en yüksek algoritmanın karar ağaçları olduğunu göstermiştir.

Çiftçi ve Batur Sir (2023) çalışmalarında, acil servise başvuran kişi sayısının tahminini yedi, on beş, otuz günlük olarak gerçekleştirmişlerdir. Zaman serileri yöntemlerinden SARIMA ve Holt-Winters yöntemlerini, makine öğrenmesi yöntemlerinden rastgele karar ormanları (RF) algoritmasını kullanmışlardır. Korelasyon, düzeltilmiş R² ve ortalama mutlak yüzde hatası (MAPE) ile performanslarını karşılaştırmışlar ve sonuç olarak SARIMA yönteminin en yüksek performansı gösterdiğini belirtmişlerdir.

Pongdatu ve Putra (2018) çalışmalarında bir mağaza müşteri tahmini için Holt Winters ve SARIMA zaman serisi analizi modellerini kullanarak model performans analizi gerçekleştirmişlerdir. 2013-2017 mağaza satış verilerini kullanmışlardır. Ortalama mutlak sapma (MAD) ile performans karşılaştırması yapmışlardır. En iyi modelin 5,592 MAD değeri ile SARIMA modeli olduğuna karar vermişlerdir.

Xiang vd. (2022) çalışmalarında, orta ve uzun vadeli enerji yükü tahmin modelleri denemişlerdir. Çin' in bir eyaletinden almış oldukları gerçek veriler ile makine öğrenmesi yöntemleri ile tahmin gerçekleştirmişler ve performans

karşılaştırmalarında Catboost algoritmasının %92,5 R^2 değeri ile en iyi performans gösteren algoritma olduğunu belirtmişlerdir.

Karaatli vd. (2012) yapay sinir ağları yöntemini kullanarak otomobil satış tahmini yapmışlardır. Çalışmanın amacını, otomotiv alanında bazı politikaların belirlenmesine yardımcı olmak olduğunu belirtmişlerdir. Gayri safi yurtiçi hasıla, reel kesim güven endeksi, yatırım harcamaları, tüketim harcamaları, tüketici güven endeksi, dolar kuru ve zaman verilerini bağımsız değişkenler olarak belirlemişlerdir. Bağımlı değişkeni toplam otomobil sayısı olarak belirlemiş ve uygulamalarını gerçekleştirmişlerdir. MAPE (ortalama mutlak yüzde hata) oranının %16,82 olduğunu göstermişlerdir. Sonuca bakıldığında, gerçeğe oldukça yakın bir tahmin gerçekleştirdiklerini belirtmişlerdir.

Akyurt (2015) yerli otomobil sayılarını bağımlı veri olarak kullanmış ve yapay sinir ağları algoritmaları ile uygulamıştır. Uygulamada ileri beslemeli geri yayılım algoritmasını kullanmıştır. En iyi sonuç veren nöron sayısını bulabilmek için MSE (ortalama hata kare) ve MAPE (ortalama mutlak yüzde hata) değerlendirme ölçütleri ile performans karşılaştırması yapmıştır. Sonuç olarak en iyi gecikme uzunluğunun 2, nöron sayısının 7 olduğu modelin en iyi performansı gösterdiğini belirtmiştir. Tablo 1.1’de literatür özet tablosu verilmiştir.

Tablo 1.1. Literatür özet tablosu.

Eser Bilgisi	Çalışmanın Konusu	Yöntem	Bulgular
Tekin ve Sarı (2022)	Mağaza zincirleri verileri ile makine öğrenmesi ve zaman serisi analizi performans karşılaştırması.	CatBoost, LightGBM, Extreme Gradient, ARIMAX, SARIMAX	ARIMAX ve SARIMAX ile performansı en yüksek sonuçları elde etmişlerdir. Uygulamada kullanılmış olan veri setinin mevsimsellikten çok etkilenmediğini belirtmişlerdir.
Huang vd. (2019)	2001-2015 yılları Güney Çin meteorolojik verileri kullanılarak tahmin performans karşılaştırması.	CatBoost, RF, SVM	Her üç yöntemin de iyi sonuçlar gösterdiğini ancak CatBoost algoritmasının hem model doğruluğu hem de tahmin hızı ile en yüksek performans gösteren yöntem olduğunu belirtmişlerdir.

Tablo 1.1. (Devamı) Literatür özet tablosu.

Eser Bilgisi	Çalışmanın Konusu	Yöntem	Bulgular
Yeşilyurt ve Dalkılıç (2021)	Günlük nehir akım verilerini kullanarak Extreme Gradient Boosting(XGBoost) ve Gradient Boosting(GBM) makine öğrenme yöntemleri performans karşılaştırması.	Extreme Gradient Boosting (XGBoost), Gradient Boosting (GBM)	XGBoost ve GBM algoritmalarının karşılaştırmaları yapılmış ve XGBoost algoritmasının en iyi performansı gösterdiği görülmüştür.
Sevgen ve Aliefendioğlu (2020)	Gayrimenkul değerlendirme üzerine random forest(RF) algoritması ile tahmin çalışması.	Rassal Karar Ormanları (RF)	Gerçek ve tahmin değerleri arasında 25.000 TL fark gözlemlenmiş ve başarılı bir sonuç elde etmişlerdir.
Çemrek ve Demir (2021)	CO ₂ emisyon miktarı verilerini kullanarak SVR, YSA ve Box-Jenkins yöntem karşılaştırması.	Destek vektör Regresyonu(SVR), Yapay sinir Ağları(YSA), Box-Jenkins	Sonuçları MSE performans değerlendirme ölçütü ile değerlendirmişler ve en yüksek performansı YSA yönteminin verdiğini gözlemlenmişlerdir.
Kadir Sheikh ve Unde (2012)	Elektrik yük tahmininin YSA ile gerçekleştirilmesi.	YSA	Sonuçlar MAPE ile değerlendirilmiş ve %0,956 ile yüksek performans elde etmişlerdir.
Çekiç (2023)	Hava kargo taşımacılık verilerini kullanarak makine öğrenmesi yöntemleri ile tahmin performans analizi yapılması.	Lineer Regresyon, Karar Ağaçları, Rassal Karar Ormanları, Aşırı Gradyan Arttırma Yöntemi (XGBoost)	MAE, MSE, RMSE ve R ² ölçütleri ile performanslar değerlendirilmiş ve doğruluğu en yüksek modelin karar ağaçları olduğu belirtilmiştir.

Tablo 1.1. (Devamı) Literatür özet tablosu.

Eser Bilgisi	Çalışmanın Konusu	Yöntem	Bulgular
Çiftçi ve Batur Sir (2023)	Acil servise başvuran kişi sayılarının zaman serisi ve makine öğrenmesi ile tahmin edilmesi ve yöntemlerin karşılaştırılması.	SARIMA, Holt-Winters, Random Forest(RF)	Korelasyon, düzeltilmiş R kare ve MAPE ile karşılaştırma sonucunda en yüksek performansı SARIMA yönteminin verdiğini gözlemlemişlerdir.
Pongdatu ve Putra (2018)	Bir mağaza müşteri tahmininin zaman serileri analizi ile gerçekleştirilmesi.	SARIMA, Holt Winters	Zaman serileri analizinde, kullanılan mağaza satış verileri ile yapılan çalışmada SARIMA yönteminin daha yüksek bir performans gerçekleştirdiğini belirtmişlerdir.
Xiang vd. (2022)	Çin' in bir eyaletinden alınan enerji yük verileri ile makine öğrenmesi yöntemlerini kullanarak performans analizi.	CatBoost, XGB, AdaBoost, Linear SVR, Karar Ağaçları, Random Forest, XGB	Kullanılan algoritmalar arasında CatBoost algoritmasının %92,5 performans oranı ile en iyi model olduğunu belirtmişlerdir.
Karaatli vd. (2012)	YSA ile otomobil satış tahmini gerçekleştirilmesi.	YSA	Tahmin ve gerçek değerler ile yapılan analiz sonucunda MAPE değerini %16,82 bularak yüksek bir sonuç elde etmişlerdir.
Akyurt (2015)	YSA ileri beslemeli geri yayılım algoritması ile otomobil satış tahmininin yapılması.	YSA	Otomobil satış sayıları ile yapılan çalışmada MSE ve MAPE ile değerlendirme yapılmış ve gecikme uzunluğu 2, nöron sayısı 7 olan algoritmanın en iyi sonucu verdiğini belirtmiştir.

Literatürde yer alan talep tahmin çalışmaları incelendiğinde matematiksel talep tahmin yöntemlerinin başarılı sonuçlar verdiği aynı zamanda yapay zeka tahmin

yöntemlerinin özellikle son yıllarda çalışmalarda uygulanması ile literatürde büyük bir yer tuttuğu ve çalışmalarda bu yöntemlerin sıklıkla kullanıldığı görülmektedir.



2. TALEP TAHMİNİ

Talep, tüketicinin bir mal veya hizmeti alma isteği ve alım gücünün olmasıyla bu isteği gerçekleştirme ve gerçekleşen isteğin ölçülebilir miktarıdır. Talep, tüketicilerin bir ürün veya hizmeti belirli bir fiyat seviyesinden almaya hazır oldukları miktardır. Ürün ve hizmetin fiyatı, tüketicilerin gelir seviyeleri, ihtiyacın şiddeti, tüketici alışkanlıkları, mevsimsel etkenler talebi etkileyen faktörler arasında sayılabilir (Yazıcıoğlu, 2010).

Tahmin, birçok değişkene bağlı olması sebebiyle öngörülmesi zor bir durumdur. Ancak geleceğe yönelik alınacak olan kararların doğruluğu yapılan tahminin gücüne bağlıdır. Bu sebeple yapılacak olan tahmin doğruluk oranının yüksek olması için bilimsel bir çalışmaya dayalı olması önemlidir. Talep tahmini ise tüketicinin gerçekleştirecek olan isteklerinin gelecek dönemler için yaklaşık olarak öngörülme çalışmasıdır. Talep tahmini, geçmiş dönemlerden belli bir zaman aralığı baz alınarak gelecek dönemin belli bir aralığının talep düzeyinin belirlenmesidir. Yani, sürekli kullanılan bir kelime olan tahmin, bilinmeyen herhangi bir olay veya durum hakkında geçmişe dönük bilgiler ışığında geleceğe yönelik öngörü yapılabilmesidir.

Rekabetin arttığı koşullara direnebilmek ve problemler karşısında etkili kararları verebilmek adına geleceğe yönelik tahmin yapabilme yetisi zorunluluk haline gelmiştir. Geleceğe yönelik alınan kararların başarısı yapılan tahminlerin gücüne bağlıdır. Gelecek dönemlere ilişkin doğru kararlar verebilmek için belirsizliği en aza indirmek yönünde çeşitli tahmin yöntemleri geliştirilmiştir. İşletmelerin, ürünlerini ne zaman ve ne kadar üreteceklerini bilmeleri gerekmektedir. Bu nedenle, talep tahmini işletmeler için kısa ve uzun vadeli gerçekleştirilmesi gereken amaçların başında gelmektedir. Çünkü iyi bir tahmin, işletmelerin devamlılığını sağlamasını, müşteri memnuniyeti kazanmasını, iş yükünün azaltılmasını ve gereksiz üretimin önüne geçilmesini sağlamaktadır. Doğru bir talep tahmini yapmak, işletmeler için önemli olup kaynakların daha verimli kullanılabilmesini sağlamaktadır (Türk ve Kiani, 2019).

2.1. Talep Tahmininin Önemi

Müşteri istek beklentilerinin hızla değiştiği bu zamanda birim maliyetleri azaltmak, müşteri memnuniyetini arttırarak pazar içindeki yerini korumak için üretim miktarını belirlemek, talebi en az hata oranı ile tahmin etmek firmalar için doğru kararlar verebilmek yolunda gereklilik haline gelmiştir. Tüm üretim aşamaları tüketicinin talebine dayanır. Talep miktarını etkileyecek toplumun istekleri göz önüne alınmadan yapılan plansız üretimler sonucunda, talepten az üretim yapılması durumunda talep karşılanmayacak ayrıca birim başına maliyet artacağı için müşteri memnuniyetsizlikleri ile karşılaşılması kaçınılmaz olacaktır. Müşteri talebinden fazla üretim yapıldığında ise sermaye dönme hızı yavaşlayacak ve stoklama problemleri baş gösterecektir. Talep tahmininin yanlış yapılması, sonraki operasyonların da etkilenmesi ile kaynakların etkin ve verimli kullanılmasına engel olacaktır. Üretim planları oluşturulması için talep tahminleri ve pazarlama stratejileri göz önünde bulundurulmaktadır. Talep tahminleri, üretim planlarının temelini oluşturur. Neyin, ne zaman, nasıl ve hangi kaynakla üretileceğini planlamak için talep tahminine ihtiyaç vardır. Tahminlemeyi sezgisel yöntemler yerine matematiksel ve istatistiksel yöntemlerle yapmanın daha güvenilir sonuçlar verdiği literatürdeki birçok çalışmada da kanıtlanmıştır. Ancak, hiçbir tahmin yöntemi tamamıyla doğru sonuç verir yorumu yapılmamalıdır.

Lojistik planları için de önemli bir girdi oluşturan talep tahminleri, ne kadar ham madde alınıp nereye ne zaman taşınacağını ve ne kadar stok bulundurulacağını, mamulün nereye ve ne zaman taşınacağını belirlenmesinde etkilidir. Talep tahmini alınan ve uygulanan birçok kararın temelini oluşturur. Kararların seçimleri yüksek oranda talep tahminlerine göre yapılır. Talep tahminlerinden yola çıkarak uygulanan planlara örnek olarak; üretim planlaması, pazarlama planlaması, finansal planlar ve iş gücü planlaması gösterilebilir. İşletme yönetimleri için gelecekteki belirsizlikleri tahmin etmek ve riskleri en aza indirebilmek üretim seviyesini koruyabilmek ve arttırabilmek adına önemli olmaktadır. Bu nedenle işletmeler, geleceğe yönelik öngörü sahibi olmak için tahmin tekniklerinden yararlanmaktadırlar (Korucuk ve Tatlı, 2017).

2.2. Talep Tahminlerinin Sınıflandırılması

Zaman aralığına göre:

- Çok kısa vadeli tahminler
- Kısa vadeli tahminler
- Orta vadeli tahminler
- Uzun vadeli tahminler

olarak sınıflandırılırlar (Haliloğlu ve Tutu, 2018).

Çok Kısa Vadeli Tahminler: Haftalık hatta günlük olarak yapılan işletme içi planların oluşturulmasını sağlayan tahminlerdir.

Kısa Vadeli Tahminler: Genellikle 3-6 aylık süreci kapsayan tahminlerdir. Kısa vadeli tahminlerde üstel düzeltme, hareketli ortalamalar yöntemi gibi daha kolay ve basit matematiksel yöntemler kullanılabilir.

Orta Vadeli Tahminler: 6 ay-5 yıl gibi bir süreyi kapsayan, tedarik sürecinin de planlamaya dahil edildiği tahminlerdir. Mevsimsel dalgalanmalara göre talebin değişkenlik gösterebileceği üretimler için kullanılabilir.

Uzun Vadeli Tahminler: 5 yıldan daha uzun bir süreyi kapsayan bu tahminler, işletmeler için donanım ile ilgili sermaye yatırımı gerektiren planlamaların karar aşamasında kullanılabilir.

Kısa vadeli tahminlerde uzun vadeli tahminlere göre hata payı daha az ve doğruluğu daha yüksek tahminler elde edilmektedir. Tahmin edilmek istenen dönemin uzaması, tahmin sonucunun doğruluğu da azalmaktadır.

2.3. Talep Tahmin Yöntemleri

Talep tahminini, sezgisel yöntemlerle gerçekleştirmek her zaman mümkün olmamaktadır. Ayrıca sezgisel yöntemlerin güvenilirlik düzeyinin düşük olması sebebiyle literatürde farklı talep tahmin yöntemleri karşımıza çıkmaktadır. Tahmin çalışmaları için geliştirilmiş birçok yöntem vardır. Hizmet ya da ürünün talebini tahmin ederken kullanılacak yöntemin doğruluk derecesi, ürün ya da hizmetin özelliklerine, müşteri istek ve beklentilerine, sosyal ve ekonomik nedenler gibi farklı sebeplere bağlı olarak değişkenlik gösterebilmektedir. Talep tahmininde yöntem seçimi yapılırken, üretilen ürün ya da hizmet için kullanılan talep tahmin yönteminin

hata oranının az olması gerekmektedir. Talep tahmininde kullanılacak her yöntem için bir doğruluk derecesi söz konusudur. Farklı yöntemler denenerek talep tahmininde en az hata oranına sahip olan yöntem seçilmelidir. Yapılacak olan tahminin hata payının en az olması için dikkat edilmesi gereken konulardan birisi ise incelenecek zaman periyodunun doğru olarak seçilmesidir. Tahmin yapılırken yeterli olmayan zaman aralığı hata payını yükseltebildiği gibi sezgisel yöntemlerin kullanılmasını gerektirebilir. Tahminleme yapılırken herhangi bir tahmin çeşidinin tamamıyla doğru sonuç vermesini beklemek yanlıştır. Fakat sayısal yöntemlerle tahminin daha etkin olduğu literatürdeki birçok çalışmada kanıtlanmıştır. Tahmin yöntemleri temelinde nitel ve nicel yöntemler olarak ayrılmaktadır. Nitel yöntemler genellikle uzman kişilerin görüşlerine dayanırken, nicel yöntemler matematiksel ve istatistiksel hesaplamalara dayanmaktadır. Bu iki yöntemin yetersiz kaldığı durumlarda, yapay zeka tabanlı yöntemler kullanılmaktadır (Akbaş, 2022).

2.3.1. Nitel yöntemler (Kalitatif)

Tahminde matematiksel ve istatistiksel analiz olmayan, sayısal veriler yerine uzman görüş ve hislerine dayanan tahmin yöntemidir. Nitel yöntemler uzman ve yönetim görüşlerini dikkate alan tahmin yöntemleridir. Sözel veriler ile çalışmalar gerçekleştirilmektedir (Ulucan ve Kızılırmak, 2018). Tahmini yapan kişinin görüş ve tecrübeleri bu yöntem için yüksek derecede önem arz etmektedir. Matematiksel olarak ölçülebilir olmadığından performansı ve güvenilirlik oranı diğer yöntemlere göre daha düşüktür. Delphi yöntemi, görüş toplama yöntemi, nominal grup yöntemi, pazar araştırması yöntemi nitel yöntemler arasındadır.

Delphi yöntemi: Konu ile ilgili uzman kişiler tarafından anket şeklinde yapılan bir tahmin yöntemidir. Amaç, tahminin gerçekleşmesi için görüş birliğine varılmasıdır. Genellikle sayısal tahmin yöntemlerini uygulamak için geçmişe ait yeteri kadar veri bulunmadığında kullanılan bir tahmin yöntemidir. Konu ile ilgili uzman bir grup seçilmektedir. Uzman kişilere anket şeklinde sorular yöneltilmekte olup cevaplar analiz edildikten sonra kişiler kendi cevaplarını değiştirebilmektedir ve ortak bir sonuca bağlanana kadar süreç devam etmektedir. Karamustafa ve Kılıçhan (2019) çalışmalarında, turist rehberlerinin sorunlarının tespiti amacıyla delphi yöntemi kullanmışlardır. Sorun tespiti sonrasında çözüm önerileri geliştirmeye çalışmışlardır.

Nominal grup yöntemi: Bu yöntemde, yöneticiler ve üst düzey uzmanlardan oluşan bir grubun fikir ve görüşleri ile ortak bir tahminleme yapılmaktadır. Nominal grup

yöntemi, uzun vadeli tahminlerde veya geçmişe dönük veri bulunmadığı durumlarda kullanılabilir. Kolay ve maliyetsiz bir yöntem olması sebebiyle avantaj sağlarken, hiçbir sayısal veriye dayanmaması ve grup psikolojisinin etkili olması sebebiyle literatürde dezavantajı yüksek bir yöntem olarak geçmektedir. Atabay ve Dikmeoğlu (2018), çalışmasında kullandığı niteliklerin, önem sıralamalarını belirlemek amacı ile nominal grup yöntemini kullanmışlardır.

Pazar araştırması yöntemi: Bu yöntemle birlikte, tüketici üzerinde sayısal ve niteliksel olarak meydana gelen değişimlerin, talep miktarlarındaki azalış veya artışın, yeni ürün bazında oluşabilecek tahmini talep miktarının gözlemlenebilmesi amaçlanmaktadır. Örneklem ile belirlenmiş olan kitlelere yapılan anketler sonucunda tahmin yapılmaktadır. Tüketicinin soruları yanlış anlaması ya da anlayamaması, dikkatli bir şekilde cevap vermeme durumu pazar araştırması yönteminin dezavantajları arasındadır. Keleş vd. (2018) çalışmalarında, e-ticaret hizmetleri için müşteri beklenti analizinin önemli olduğunu ve pazar araştırması yönteminin bu konuda yararlı olabileceğini savunmuşlardır.

2.3.2. Nicel yöntemler (Kantitatif)

Sayısal tahmin yöntemleri olarak adlandırılan nicel yöntemler, geçmiş dönem değerlerinden hareketle gelecek dönem değerlerinin tahmininde verilerin sistematik bir şekilde kullanıldığı tahmin yöntemleridir. Veriler analiz edilerek matematiksel yöntemler kullanılır. Sonuçların kişiye ve kişilerin görüşlerine bağlı olmayan tahmin yöntemleridir (Avni ES vd., 2014). Kalitatif tahmin yöntemlerine kıyasla objektiftir. Sayısal veriler kullanılarak tahminleme işlemi gerçekleştirilmektedir. Nedensel modeller ve zaman serileri modelleri olarak çeşitlenmektedir. Nedensel modellerde basit doğrusal regresyon yöntemi, korelasyon analizi gibi yöntemler vardır. Zaman serileri analizinde, aritmetik ortalama yöntemi, hareketli ortalama yöntemi, ağırlıklandırılmış hareketli ortalama, mevsimsel değişkenlik yöntemi, üssel düzeltim yöntemi box-jenkins yöntemi gibi yöntemler yer almaktadır.

2.3.2.1. Nedensel modeller

Bağımsız değişkenler arasındaki, bağımlı değişken ile bağımsız değişken arasındaki ilişkiyi ve bu ilişkinin gücünü, birbirlerini nasıl etkilediklerini bulmak için kullanılan nedensel modeller talep tahmininde istatistiksel bir yöntemdir. Geçmiş değerlerden yararlanarak, değişkenler arasındaki ilişkiyi bulmaktadır. İlişkiyi açıklayan bir model

kurulmaktadır (Özdemir ve Özdemir, 2006). Gelecek dönemler için güçlü bir tahmin yapılma olanağı sağlayan bu yöntemde, bağımlı değişkeni etkileyen bağımsız değişkenler ile bağımlı değişken arasındaki ilişkinin gücünün ölçülmesi amaçlanmaktadır. Regresyon analizi ve korelasyon analizi en çok tercih edilen nedensel modellerdir.

Regresyon analizi: Regresyon analizinde bağımlı ve bağımsız değişken ya da bağımsız değişkenler arasındaki ilişki incelenerek bağımsız değişken ya da değişkenlerin bağımlı değişkeni nasıl etkilediğine karar vermek ve bir çıkarım yapmak için sıklıkla kullanılan ve literatürde yerini alan bir yöntemdir. Hem değişkenler arasındaki ilişkiler hakkında fikir vermekte hem de geçmiş verilerden yola çıkarak gelecek verileri tahmin etme avantajı bulunmaktadır.

Basit doğrusal regresyon yöntemi: Basit doğrusal regresyon yönteminde bağımlı ve bağımsız değişken arasında bir ilişki olduğu kabul edilerek tahminde bulunmaktadır. Geçmiş verilere bakılarak bağımsız değişkenin almış olduğu değer karşısında bağımlı değişkenin aldığı değerden yola çıkılarak gelecek dönemler için bağımlı değişkenin alabileceği değerlerin en az hata oranı ile tahmin edilmesi istenmektedir. Nedensel modeller içerisinde en sık kullanılan yöntem olarak literatürde yerini almıştır. Bağımlı değişkeni etkileyen bağımsız değişkenlerin tahminde kullanılması ve bağımlı değişkeni nasıl ve ne kadar etkilediğinin analizi uygulanacak olan stratejilerde hangi değişkenlerin kritik olduğunun karar verilmesinde önemli olmaktadır. Regresyon analizi, değişkenlerde oluşacak artış ya da azalışların bağımlı veri üzerinde ne kadar etkili olduğunu gösterebilmektedir (Zorlu, 2021). Talep tahmininde ise, talep bağımlı değişken olmakla beraber talebi etkileyen faktörler de bağımsız değişkenlerdir. Bağımsız değişkenler, farklı bir değişkenden etkilenmeyen, bağımlı değişkeni etkilediği öngörülen değişkenlerdir. Bağımlı değişken, bağımsız değişken ya da değişkenlerden etkilenen, bağımsız değişkene göre değişebilen değişkenlerdir.

Bağımsız değişken sayısı tek olduğundan tek değişkenli regresyon analizi olarak adlandırılmaktadır. Bağımsız değişken sayısı birden fazla ise, çoklu doğrusal regresyon yöntemi kullanılmaktadır. Tek değişkenli ve çok değişkenli regresyon yöntemlerinin çalışma sistemleri aynıdır. Değişken sayısı farkı sebebiyle birbirlerinden ayrılmaktadır.

Korelasyon analizi: Korelasyon analizi, iki deęiřken arasındaki ya da bir deęiřken ile dięer deęiřkenler arasındaki iliřkiyi grmek ve derecelendirmek iin kullanılan bir yntemdir. Korelasyon analizinde ama, baęımlı deęiřkenin baęımsız deęiřkenden ne kadar ve hangi ynde etkilenebileceęini lmektir. Korelasyon analizi ile deęiřkenler arasındaki iliřki grlebilmekte ve llebilmektedir. Kaya (2013) alıřmasında, deęiřkenler arasındaki iliřkinin kuvvetini analiz edebilmek iin korelasyon analizi yntemini kullanmıřtır. Bu derecelendirme -1 ile +1 arasında deęiřmektedir. Derecelendirme “-“ olduęunda iki deęiřken arasında negatif ynl bir iliřki olduęu, “+” olduęunda iki deęiřken arasında pozitif ynl bir etkileřim olduęu grlebilmektedir. 0 olduęunda, iki deęiřken arasında iliřki olmadıęı, 0’a yaklařtıķ deęiřkenler arasındaki iliřkinin az olduęu anlamına gelmektedir.

2.3.2.2. Zaman serileri modelleri

Zaman serisi verileri, saatlik, gnlk, haftalık, aylık, eyreklik, yıllık gibi belirli zaman aralıklarından oluřan gzlemlerdir. Verilerde kronolojik sıralamanın nemli olduęu zaman serileri analizi, gemiř dnemlerden hareketle gelecek dnemler iin istatistiksel tahmin yapmaktadır. Zaman serileri, farklı alıřma alanlarında verilerin sıklıęını ifade etmektedir. Dzenli zaman aralıklarında llmektedir. Zaman serileri analizinde ama, gemiřte ne olduęunu anlayarak gelecekteki hareketlerin tahminini gerekleřtirmektir. Zaman serileri verilerinin eřitli gruplara ayrılması zaman serilerindeki amacın gerekleřmesine olanak saęlamaktadır (Arslankaya, 2019).

Zaman serileri trend bileřeni, mevsim bileřeni, evrimsel bileřen, dzensiz bileřen olmak zere drt bileřenenden oluřmaktadır. Trend bileřeni, zaman serisinin zaman iinde doęrusal artan veya azalan durumunu gstermektedir. Mevsim bileřeni, zaman serisinin mevsimlere gre, sabit zaman aralıklarında modelin tekrarlaması durumunu belirtmektedir. evrimsel bileřen, dnemlerde mevsimsel deęiřimler dıřındaki deęiřimleri gstermektedir. Dzensiz bileřen, hata terimini ifade eden deęiřimlerdir. Zaman serileri modelleri, kısa dnemde tutarlı sonular vermesi, veri ihtiyaının dięer yntemlere gre daha az olması, kolay modellenenebilir olması sebebi ile sıklıkla kullanılmaktadır. Literatrde birok rneęi bulunmaktadır.

Bu blmde, zaman serileri modellerinden, aritmetik ortalama yntemi, hareketli ortalama yntemi, aęırlıklandırılmıř hareketli ortalama yntemi, ssel dzeltim yntemi, box-jenkins yntemi anlatılacaktır.

Aritmetik ortalama yöntemi: Aritmetik ortalama yöntemi, talep tahminlerinde kullanılan en yaygın ve en basit yöntemlerden biridir. En büyük avantajı, kolay uygulanabilir olmasıdır fakat dezavantajları da bulunmaktadır. En büyük dezavantajı, veri sayısı arttığında, beklenmedik olaylar gerçekleştiğinde, trend değişip mevsimsel dalgalanmalar olduğunda tutarlı ve hata sapmasının az olduğu tahminler yapmaktan uzaklaşmaktadır. Aritmetik ortalama yöntemine göre tahmin edilmek istenen dönem, kendinden önceki dönemlerin ortalaması alınarak hesaplanır. Bu yöntem, kullanılmak istenirse, çok uzun dönemleri kapsamayacak şekilde kullanılmalıdır. Tan (2010) çalışmasında, aritmetik ortalama yöntemini öğrencilerin sınav başarılarının öngörülmesinde kullanmıştır.

Hareketli ortalama yöntemi: Hareketli ortalama yöntemi, anlık veya beklenmedik durumlarda oluşabilecek aykırı verilere karşı daha dayanıklı olduğundan aritmetik ortalama yöntemine göre doğru tahmin yapabilme yeteneği daha yüksektir. Hareketli ortalama yönteminde, bir değeri tahmin edebilmek için belirlenen gözlem sayısı kadar, tahminden önceki verilerin ortalaması alınır ve böylece tahmin edilmek istenen veri oluşmaktadır. Gözlem sayısı, tahmini yapacak kişiye göre değişmektedir (Karakaş, 2019).

Ağırlıklandırılmış hareketli ortalama yöntemi: Ağırlıklandırılmış hareketli ortalama yönteminde, aritmetik ortalama ve hareketli ortalama yönteminden farklı olarak geçmiş verilere ağırlık verilir (Kılınç ve Çakır Aydın, 2016). Hangi verilere ne kadar ağırlık verileceği ile ilgili bir hesaplama yoktur. İsteğe ve problemin çözüm şekline göre değişebilmektedir. Aritmetik ortalama ve hareketli ortalama yöntemi uygulama olarak daha kolay olsa da ağırlıklandırılmış hareketli ortalama yönteminde hata oranı daha düşük sonuçlar görülebilmektedir.

Üstel düzeltim yöntemi: Üstel düzeltim yönteminde tahmin, bir önceki tahminden yola çıkarak bulunmaktadır. Bir önceki dönemin tahmin değeri ile bir önceki dönem yapılan hata değerlerine dayalı bir hesaplama yapılmaktadır. Tesadüf ile meydana gelen etkilerin ortadan kaldırılmasını amaçlayan bir yöntemdir. Serinin içerisindeki verilere uygun ağırlıklar verilerek tahmin gerçekleştirilmektedir (Karakaş, 2019). Verilere ağırlık verilerek tahmin yapılmaktadır ve mevsimsellik, trend barındıran tüm veri setleri ile çalışabilmektedir (Yağımlı ve Ergin, 2017).

Box-Jenkins yöntemi: Tahmin metotlarından biri olan Box-Jenkins yöntemi tek değişkenli zaman serilerinde kullanılmaktadır. Box-Jenkins yöntemi ile tahminde serinin durağan olması önemli bir varsayımdır. Kısa dönem tahminlerinde başarı oranının yüksek olması literatürde birçok çalışmada görülmektedir. Box-Jenkins Yöntemi, mevsimsel ve mevsimsel olmayan serilerin tümüyle çalışabilmektedir. Seride, mevsimsellik görüldüğünde durağan bir seri haline getirildikten sonra uygulama yapılabilmektedir (Bircan & Karagöz, 2003).

2.3.3. Yapay zeka tabanlı yöntemler

1950’li yıllarda bilgisayarların gelişmesi ile bulunan teknikler ve algoritmalar sonrasında Yapay zeka teknikleri kullanılmaya başlanmıştır. Birçok farklı alanda sıklıkla ve yenilenerek kullanılmaktadır (Güven, 2020). Günümüzde, tıp, sağlık, tarım, askeri, bankacılık, otomotiv, ekonomi, finans, üretim gibi birçok alan ve sektörde kullanılmaktadır. Çalışma sistemleri, insan zekasını taklit ederek verilen bilgileri analiz edip değerlendirmek üzerine kurulmaktadır. İnsan zekasındaki öğrenmenin nöronlar sayesinde olmasından yola çıkılarak yapay zeka uygulamalarında da bu yöntem uygulanmaktadır. İnsan zekasındaki nöronların çalışma düzeni ile benzer yapıya sahip olan yapay zeka uygulamalarında, çok sayıda veri ve bilgi girdisinin işlenmesi ve analiz edilip değerlendirilmesi sonucunda problem çözme yeteneği oluşturulmaktadır. Yapay zeka ile çalışan sistemler, kendilerine yüklenen görevler dışında hafızalarındaki veri birikimi, analizi ve değerlendirmeleri sonucunda edindikleri tecrübeler ile karar verme yeteneğine sahip olabilmektedirler.

Bu bölümde, yapay zeka tabanlı yöntemlerden yapay sinir ağları (YSA) ve makine öğrenme yöntemleri algoritmalarından catboost, gradient boosting, random forest (RF), destek vektör regresyonu (SVR) anlatılacaktır.

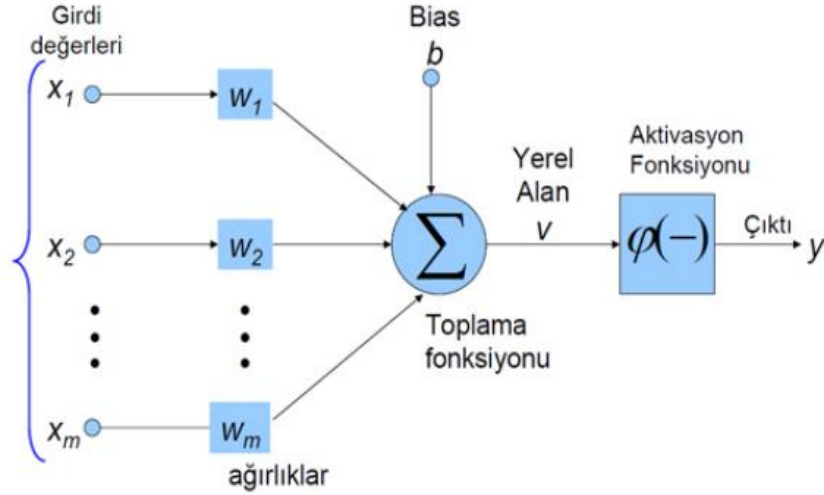
2.3.3.1. Yapay sinir ağları (ANN-YSA)

Geleneksel programlamalardan farklı olan yapay sinir ağları (YSA), insan beyninin özelliklerinden esinlenilerek öğrenme sürecinin matematiksel olarak modellenmesi sonucu ortaya çıkmıştır. Yapay zeka, bilgisayarlara veya makinelere, insana özgü yetiler olarak varsayılan anlam çıkarma, akıl yürütme, yargı, karar verme ve insanın en önemli özelliklerinden olan öğrenme gibi yeteneklerin kazandırılmasıdır. Yapay sinir ağları, insan beyninin öğrenme yeteneği özelliği ile bilgileri keşfedebilme,

hafızaya alabilme, bilgi ve verileri işleyerek yeni bilgiler üretebilme yolu ile problem çözme kabiliyeti oluşturan bilgi işleme teknolojisidir. Yapay sinir ağları, karmaşık problemlere daha hızlı, daha pratik ve daha az hata toleransına sahip çözümler sunmakta ve tahminde bulunmaktadır. Yapay sinir ağları tahminleme ile birlikte kümeleme, sınıflandırma problemlerinde de kullanılmakta olup analiz ve ilişkilendirme de yapabilmektedir. YSA öğrenme yapısı, insan beynine benzer özellikle taşımaktadır. Yöntemle birlikte, modelin öğrenmesi, eğitilmesi ve karar vermesi amaçlanmaktadır. İnsan beyni nöronları da, bilgileri öğrenme, hafızasında tutma ve veriler arasındaki ilişkiyi çözebilme yeteneğine sahiptirler (Keskenler ve Keskenler, 2017).

Yapay sinir ağı sisteminin temelini atmayı Amerikalı Warren McCulloch gerçekleştirmiştir. Basit bir sinir hücresi modelini öğrencisi ile birlikte 1943 tarihinde oluşturmuştur. İlk adımdan esinlenerek Frank Rosenblatt 1957’de bir YSA modeli geliştirmiştir ve algılayıcı olarak adlandırmıştır. 2 katmanlı ağların problemleri çözmede yetersiz kaldığını savunan Marvin Minsky ve John Hopfield, tüm problemleri çözemeyeceğinden bahsetmişlerdir. Paul Werbos tarafından 1974 yılında geri yayılım öğrenme yöntemi ilk kez formülleştirilmiştir. Bugün kullanılan YSA için önemli bir adım olmuştur (Mijwil, 2017).

YSA, öğrenme yetenekleri ile farklı algoritmalar ile çalıştırılabilmektedirler. YSA sistemleri, tahmin etmenin yanı sıra, sınıflandırma, veri ilişkilendirme, veri yorumlama, teşhis gibi farklı problemlere de çözüm getirebilmektedirler. Girdi, çıktı ve gizli katmanlara sahip olan YSA, öğrenme, veriler arasındaki ilişkiyi anlamlandırma ve sonuçlandırma gibi matematiksel işlemleri içerisinde barındırdığı nöronlar ile gerçekleştirmektedir. YSA yapısı, Şekil 2.1.’de gösterilmiştir.



Şekil 2.1. Yapay sinir ağı model yapısı (Mijwil, 2017).

Girdi katmanı, dış dünyadan gelen gerçek verilerin, gizli katmana iletiildiği katmandır. Görevi X_n değişkeni olarak belirtilen değerlerin, gizli katmana taşınmasıdır. Gizli katmanı, toplam fonksiyonu olarak gösterilmiştir. Girdi katmanından gelen verilerin işlendiği katmandır. Sinir ağının eğitilmesi, veriler arasındaki ilişkinin ortaya çıkarılması bu katmanda olmaktadır. Gizli katmanlar, birden fazla olabilmektedir. Çıktı katmanı, gizli katmanın işlediği veri çıktılarını, dış dünyaya iletmektedir. İşlevi, Y değişkeni olarak belirtilen çıktıların, dış dünyaya iletilmesidir. Önemli bir parametre olan gizli katmandaki nöron sayısı, ağın performansını etkilemektedir (Çuhadar ve Kayacan, 2005). En yüksek performans gösteren nöron sayısı modelde parametre olarak kullanılmaktadır.

Yapay sinir ağlarının avantajları olabildiği gibi dezavantajları da vardır. Örneklerden öğrenebilme, hızlı işlem yapabilme ve paket programlarla kolay uygulanabilme özellikleri avantajları arasındadır. Bazı dezavantajları ise ağıdaki performansı etkileyen parametreler belirlenirken bir kuralı olmayıp kullanıcıya bağlı olması, ağın eğitiminin bitirilmesi için sistemin önerdiği bir zamanın olmaması ve belirli hata oranına göre eğitimin sonlanmasına kullanıcının karar vermesi, ağın kullandığı fonksiyonların nasıl ve nereden üretildiğinin açıkça gözlemlenememesi ve ağın öğrenme zamanının uzun olmasıdır.

2.3.3.2. Gradyan arttırma yöntemi (Gradient Boosting)

Gradient Boosting algoritması veri setini tümüyle kullanmaktadır. Veri setini alt veri setlerine bölme yaklaşımında bulunmamaktadır. Veri setini kullanarak bir karar ağacı

oluşturmaktadır. Oluşturulan karar ağacının hataları göz önünde bulundurularak yeni bir karar ağacı daha oluşturulmaktadır. Yöntemde, oluşan ağaçta bir hata hesabı yapılmaktadır. Bu hatayı minimize edecek yeni ağaçlar oluşturulmaktadır. Her yeni ağaç, bir önceki ağacın hata oranını düşürme amacı ile oluşturulmaktadır. Gerçek ve tahmin arasındaki farkın sıfırlanmasını istenmektedir (Kelle & Yüce, 2022). Çalışma yapısı Şekil 2.2’de gösterilmiştir.

Gradient Boosting algoritması aşamaları (Bayrakçı vd., 2021);

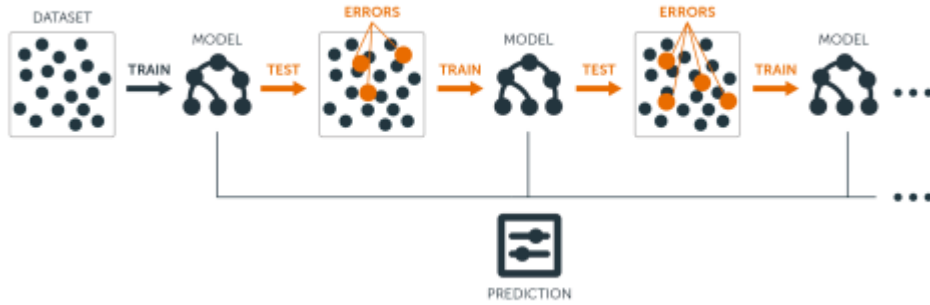
1-Regresyon ağacı oluşturulmaktadır.

2-Oluşturulan ağaçta hedef değerler ve tahmin değerleri arasındaki fark alınarak hata hesabı yapılmaktadır.

3- Bulunan hata oranları yeni gözlem değeri olarak kullanılmaktadır.

4-Her yeni oluşturulan ağaçta hata oranının düşürülmesi istenmektedir.

5-Hata oranının 0’a yaklaşması istenmektedir. Hata oranı minimize edilene kadar süreç devam etmektedir.



Şekil 2.2. Gradient boosting algoritmasının çalışma yapısı (Gradient Boosting, 2019).

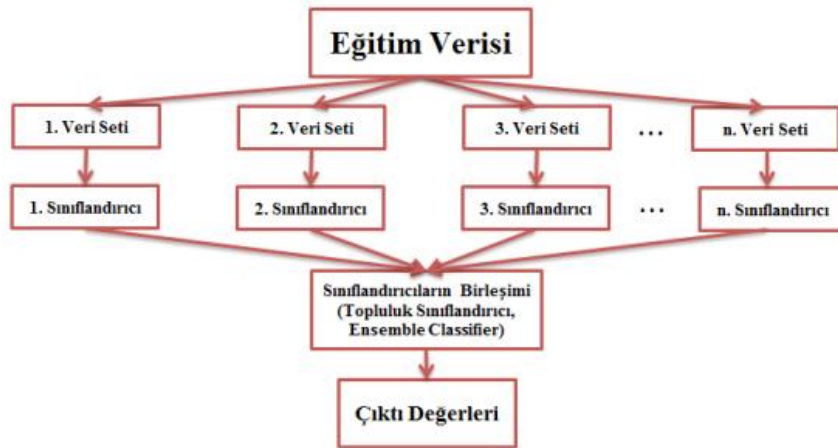
2.3.3.3. Kategorik arttırma yöntemi (CatBoost)

Catboost, Gradient Boosting tabanlı bir makine öğrenmesi algoritmasıdır. Gradient Boosting’in performansının artırılması amacı ile Yandex tarafından geliştirilmiş bir algoritmadır. XGBoost ve LightGBM’e alternatif olarak geliştirilmiştir. Az sayıdaki veri ile de yüksek tahmin performansı oluşturabilmektedir (Kuş vd., 2021). Yüksek öğrenme hızına sahip olması ile birlikte avantajları arasında sayısal, kategorik, metin verilerinin tümü ile çalışabilmesi yer almaktadır (Zhou vd., 2021). Catboost

algoritması derin olmayan, simetrik ağaçlar kurmaktadır. Tahmin performans oranı yüksektir. Aşırı öğrenme sorunu ile başa çıkabilmektedir.

2.3.3.4. Ratgele ormanlar yöntemi (Random Forest - RF)

Random forest, temeli karar ağaçlarına dayanan denetimli makine öğrenmesi algoritmasıdır. Breiman (2001) tarafından geliştirilmiştir. Sınıflandırma ve regresyon problemlerinde kullanılabilen topluluk öğrenme yaklaşımlarındandır. Torbalama yönteminin geliştirilmiş şekli olarak değerlendirilmektedir (Ekelik ve Altaş, 2019). Birden fazla karar ağacının oluşumunu kullanmaktadır. Oluşturulan modelde farklı ağaçların ortalamalarını alarak performans çıktısı vermektedir. Yüksek veriye sahip veri setleri ile uygulamalarda daha yüksek performans göstermektedir. Verinin sahip olduğu özelliklerin belirlenmesi, aykırı değerlere karşı baş gösterebilme yeteneği RF algoritmasının avantajlarındandır. Model oluşturma sırasında ağaç sayısı, ağaçların derinliği, özellik seçimi ile model performansının artırılması sağlanabilmektedir (Akdağ, 2023). Şekil 2.3'te çalışma yapısının genel şeması verilmiştir.



Şekil 2.3. RF algoritmasının çalışma yapısı (Ekelik ve Altaş, 2019).

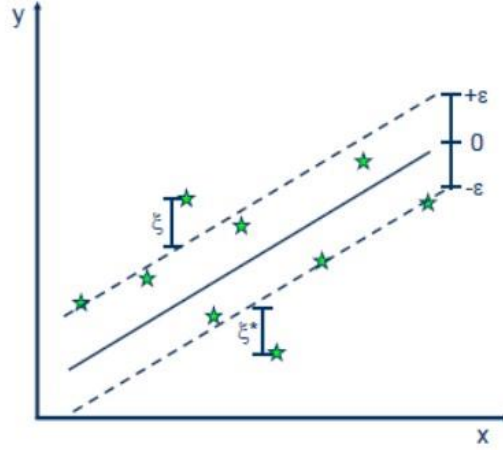
2.3.3.5. Destek vektör regresyonu (SVR)

Vapnik tarafından ortaya çıkan destek vektör makineleri (DVM), istatistiksel öğrenme yöntemlerindendir. DVM sınıflandırma ve regresyon problemlerini çözmek için oluşturulmuştur (Çoban ve Demir, 2021). Sonrasında, regresyon problemlerinde, tahmin etmeye dayalı DVM tabanlı SVR yöntemi tasarlanmıştır. Lineer ve lineer olmayan denklemlerde kullanılabilmektedir (Özcan, 2021).

SVR, veri noktalarını mümkün olduğunca fazla kapsayan bir hiper düzlemi bulmayı hedefler. İki vektör arasındaki uzaklık marjin olarak adlandırılmaktadır. Hataların

minimize edilmesini ve veri noktalarının marjin içerisinde bulunmasını amaçlamaktadır.

Şekil 2.4'te, ε ifadeleri hiper düzlemin oluşmasındaki noktalar olan destek vektörlerini, 0 çizgisi hiper düzlemi ifade etmektedir.



Şekil 2.4. SVR Hiper Düzlem Grafiği (Support Vector Regression, 2023).

2.4. Talep Tahmin Aşamaları

Talep tahmini, talebi etkileyen değişkenlerin belirlenmesi, verilerin toplanması, tahmin edilecek periyodun tespiti, tahmin yönteminin seçimi, tahmin sonuçlarının analiz edilmesi olarak temelde beş aşamadan oluşmaktadır (Yazıcıoğlu, 2010). Talep tahmin aşamaları iş akış şeması Şekil 2.5'te verilmiştir.

Talebi etkileyen değişkenlerin belirlenmesi: Tahmin gerçekleştirilmeden önce, tahmini yapılmak istenen bağımlı değişkenin talebini etkileyen değişkenler belirlenmelidir. Bağımlı ve bağımsız değişkenler arasında yüksek korelasyon olması modelin hata oranının düşük, performansının yüksek olması için gerekli olmaktadır.

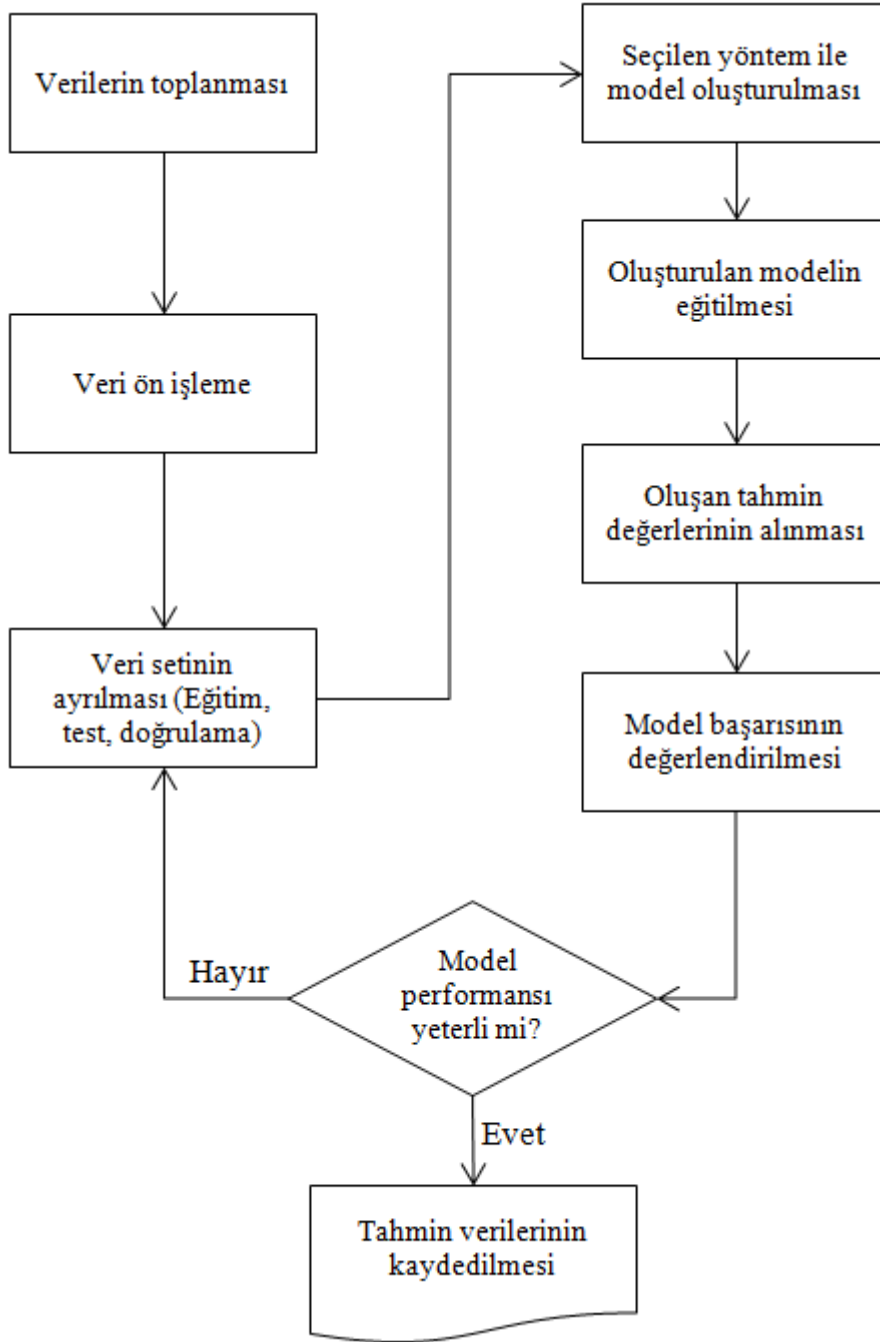
Verilerin toplanması: Bağımlı değişken tahminini etkileyecek olan değişken verilerinin içeriği talep tahmin oranının yüksek olması için önemli olmaktadır. Kullanılacak verilerin gerçek veriler olması, eksiksiz ve doğru olması gerekmektedir. Ayrıca, uygulanacak modele uyum sağlaması ve performansının artırılması amacıyla verilerin iyi analiz edilmesi, ön işlemlerden geçmesi gerekmektedir.

Tahmin edilecek periyodun tespiti: Kısa süreli yapılan tahminlerin performanslarını, uzun süreli yapılan tahminlerin performansına göre daha yüksek olmaktadır. Kullanılan yöntemle göre de değişkenlik gösterebilen periyot tespitinin iyi yapılması

gerekmektedir. Materyal ve yöntem bölümünde, makine öğrenmesi yöntemlerinde veriler test ve eğitim verileri, yapay sinir ağları ile talep tahmininde eğitim, doğrulama ve test olarak ayrılmıştır.

Tahmin yönteminin seçimi: Talep tahmini etkenlerde dalgalanmaların olması sebebi ile uygulanması zor olmaktadır. Bu sebeple verilere uyum sağlayabilen yöntemlerin uygulanması gerekmektedir. Talep tahmini için literatürde nitel (kalitatif), nicel (kantitatif), yapay zeka tabanlı yöntemler bulunmaktadır. Kantitatif ve yapay zeka tabanlı yöntemler, matematiksel sistemlere dayanarak işlem gerçekleştirdiklerinden doğruluk oranları daha yüksek olmaktadır.

Tahmin sonuçlarının analiz edilmesi: Tahmin yöntemleri ile oluşturulan modelin, veri setine uyum sağlayıp sağlamadığının tespitinin yapılması gerekmektedir. Modelin hiç görmediği veriler ile modelin tahmin verileri karşılaştırılarak performans analizinin yapılması gerekmektedir.



Şekil 2.5. Talep tahmin aşamaları.

3. MATERYAL VE YÖNTEM

Bu bölümde, oluşturulan veri setinin bağımlı ve bağımsız verileri tanıtılmış, veriler üzerinde makine öğrenmesi, yapay sinir ağları ve zaman serileri analizi yöntemleri için veri ön işleme çalışmaları yapılmış ve veri setinin oluşturulma aşaması açıklanmıştır. Veriler analiz edilmiş ve farklı algoritmalar ile tahmin çalışmalarının aşamaları anlatılmıştır. Otomobil satış verileri üzerinde makine öğrenmesi, yapay sinir ağları ve zaman serileri analizi yöntemleri için Python programlama dilinde aykırı veri analizi yapılmıştır. Sonrasında veriler üzerinde makine öğrenmesi, yapay sinir ağları yöntemleri için standardizasyon işlemi uygulanmıştır. Python programlama dilinde uygulaması gerçekleştirilen makine öğrenmesi yöntemlerinden CatBoost, Gradient Boosting (GBM), Rastgele Karar Ormanları (Random Forest), Destek Vektör Regresyonu (SVR) algoritmaları, Zaman Serileri Analizi ve Matlab paket programında uygulaması gerçekleştirilen Yapay Sinir Ağları (YSA) hakkında bilgiler verilmiştir.

3.1. Veri Setinin Oluşturulması ve Analiz Edilmesi

Literatürde birçok talep tahmin uygulaması yer almaktadır. Uygulanan satış talep tahmini çalışmalarında farklı bağımsız veriler kullanılmıştır. Literatürde yer alan çalışmalarda kullanılan bağımsız değişken örnekleri ve kullanıldıkları çalışmalar Tablo 3.1’de verilmiştir.

Tablo 3.1. Bağımsız değişkenler özet tablosu.

Bağımsız Değişkenler	Eser Bilgisi
Aylar	Karaatlı vd. (2012), Raizada ve Saini (2021)
Gayri Safi Yurtiçi Hasıla (GSYH)	Brühl vd. (2009), Hülsmann vd. (2011), Karaatlı vd. (2012)
Sanayi Kullanım Oranı	Demir ve Özcan (2023)
İthalat Miktarı	Ülkü ve Yalpır (2021)
İhracat Miktarı	Karahan (2011), Ülkü ve Yalpır (2021)

Tablo 3.1. (Devamı) Bağımsız değişkenler özet tablosu.

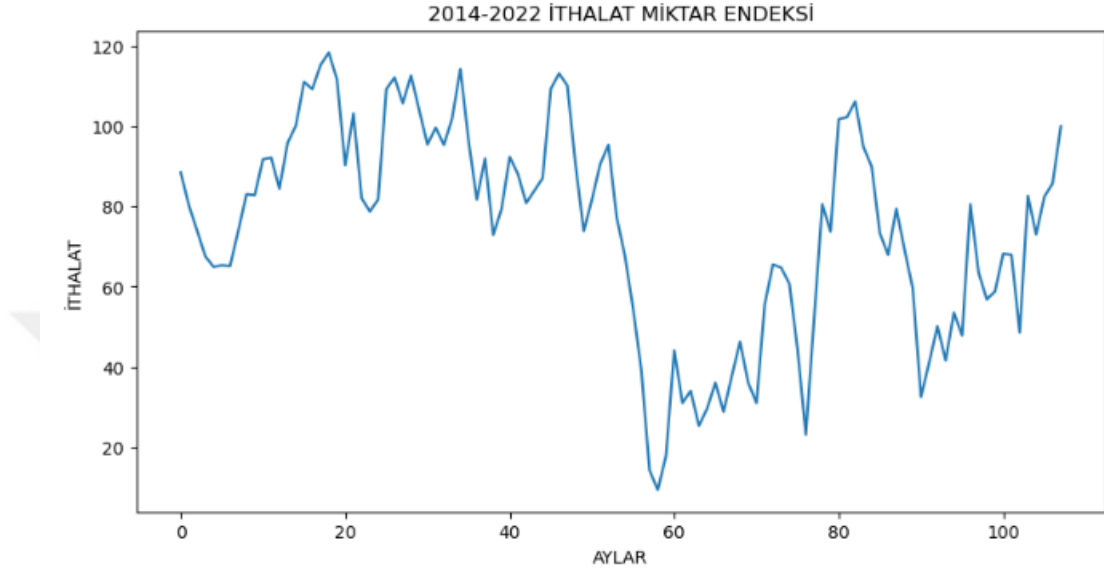
Tüketici Fiyat Endeksi (TÜFE)	Brühl vd. (2009), Raizada ve Saini (2021), Sharma ve Sinha (2012), Gökcel (2009)
Üretici Fiyat Endeksi (ÜFE)	Yılmaz ve Çilengiroğlu (2022)
Reel Kesim Güven Endeksi (RKGE)	Hülsmann vd. (2011), Karaatlı vd. (2012)
Tüketici Güven Endeksi (TGE)	Karaatlı vd. (2012), Yılmaz ve Çilengiroğlu (2022)
Kredi Oranları	Brühl vd. (2009), Hülsmann vd. (2011), Çınar (2022)
Dolar Kuru	Sohrabpour vd. (2021), Karahan (2011), Karaatlı vd. (2012)

Literatürdeki çalışmalardan destekle çalışmada kullanılacak bağımsız değişkenler belirlenmiştir. Bağımsız değişkenler belirlenirken Tablo 3.1’de bulunan bağımsız değişkenler ile bağımlı değişken olan toplam otomobil satış sayısı arasındaki korelasyon incelenmiş ve bağımlı değişken ile yüksek korelasyona sahip bağımsız değişkenler çalışmada kullanılmıştır.

Bu çalışmada kullanılan, makine öğrenmesi ve yapay sinir ağları yöntemlerinin uygulamasında, veri setini oluşturan bağımsız veriler, 2014-2022 yıllarına ait, otomobiller için ithalat miktar endeksi, reel kesim güven endeksi, tüketici güven endeksi, ortalama taşıt kredi faiz oranları, Türkiye otomobil üretim sayısı ve zamandır. Makine öğrenmesi, yapay sinir ağları ve zaman serileri analizi yöntemlerinin uygulamasında kullanılan, bağımlı veri seti ise 2014-2022 yılları içerisinde Türkiye’de aylık satılan otomobil adetleri ile oluşturulmuştur. Otomobiller için ithalat miktar endeksi, tüketici güven endeksi verileri Türkiye İstatistik Kurumu (TÜİK) İstatistik Veri Portalı’ndan (TÜİK - Veri Portalı, 2023) alınmıştır. Reel kesim güven endeksi, ortalama taşıt kredi faiz oranları, Türkiye otomobil üretim sayısı verileri Türkiye Cumhuriyeti Merkez Bankası (TCMB), Elektronik Veri Dağıtım Sistemi (EVDS)’nden (EVDS | Tüm Seriler, 2023) alınmıştır. Bağımlı veri olan toplam otomobil satış sayısı verileri ise Otomotiv Distribütörleri Derneği (ODD) web sitesinden (Pazar - Perakende Satışlar, 2023) alınmıştır.

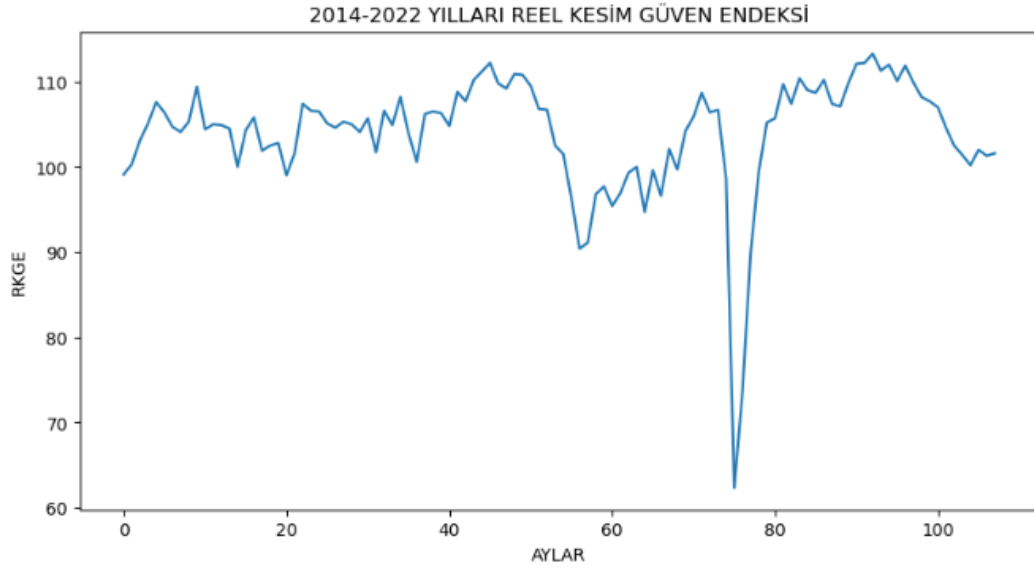
Otomobiller için İthalat Miktar Endeksi: İthalat Miktar Endeksi, ithalat miktarında meydana gelen değişimi ölçmektedir. Çetin (2020) çalışmasında, sepet kurun

düzeyinde değişiklik olduğunda ithal otomobil ve hafif ticari araç perakende satışlarında ters değişikliğin olduğunu görmüştür. Binek otomobil ithalatı, toplam satış sayısını etkileyen bir faktör olarak literatüre geçmiştir. Tablolarda İthalat kısaltması ile kullanılmıştır. İthalat miktar endeksi aylık grafiği Şekil 3.1’de verilmiştir.



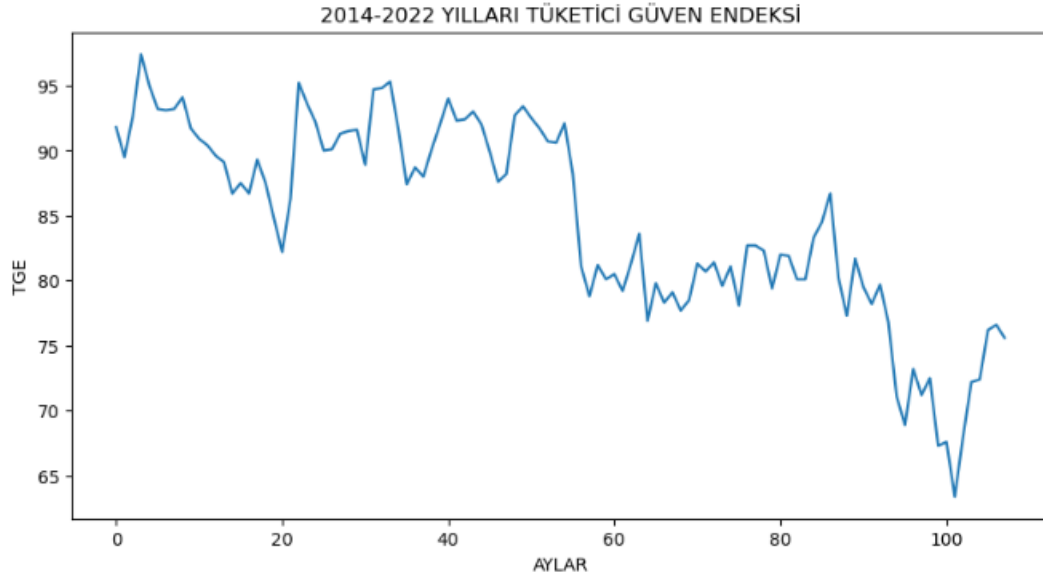
Şekil 3.1. İthalat miktar endeksi.

Reel Kesim Güven Endeksi (RKGE): TCMB verilerine göre, reel kesim güven endeksi, imalat sanayinde faaliyet göstermekte olan üst düzey yöneticilerinin de geçmiş, mevcut durum ve gelecek hakkındaki görüşleri alınarak, kısa dönemli imalat sanayindeki eğilimleri yansıtan bir göstergedir. İmalat sanayindeki artış ile otomobil satış sayıları arasında etki olduğu literatür çalışmalarında görülmektedir. Demir ve Özcan (2023) çalışmalarında, kapasite kullanım oranı ve sanayi üretimi arasındaki ilişkiyi açıklamışlardır. Reel kesim güven endeksi aylık grafiği Şekil 3.2’de verilmiştir.



Şekil 3.2. Reel kesim güven endeksi.

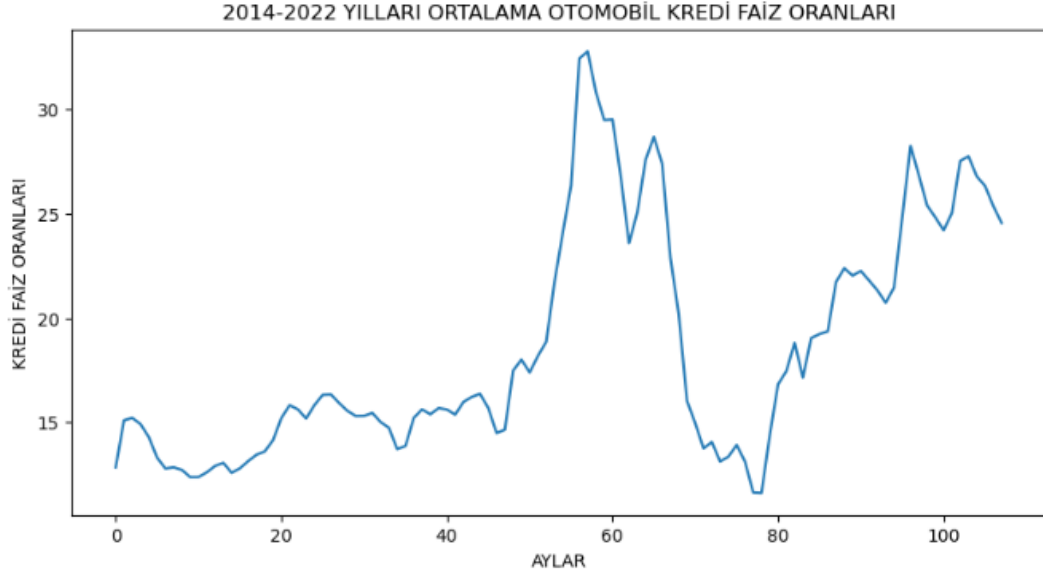
Tüketici Güven Endeksi (TGE): Tüketici Güven Endeksi, tüketicilerin ekonomi ile ilgili düşüncelerini, tüketicilerin mali durumlarını, yakın gelecekteki harcama eğilimlerini ölçen bir göstergedir (TÜİK Kurumsal, 2023). Tüketicilerin, dayanıklı tüketim mallarına harcama yapma etkilerinin önemli bir göstergesi olmaktadır. Otomobil satış sayılarına etki eden bir faktör olabilmesi sebebi ile bağımsız veri seti içerisinde kullanılmıştır. Tüketici güven endeksi aylık grafiği Şekil 3.3'te verilmiştir.



Şekil 3.3. Tüketici güven endeksi.

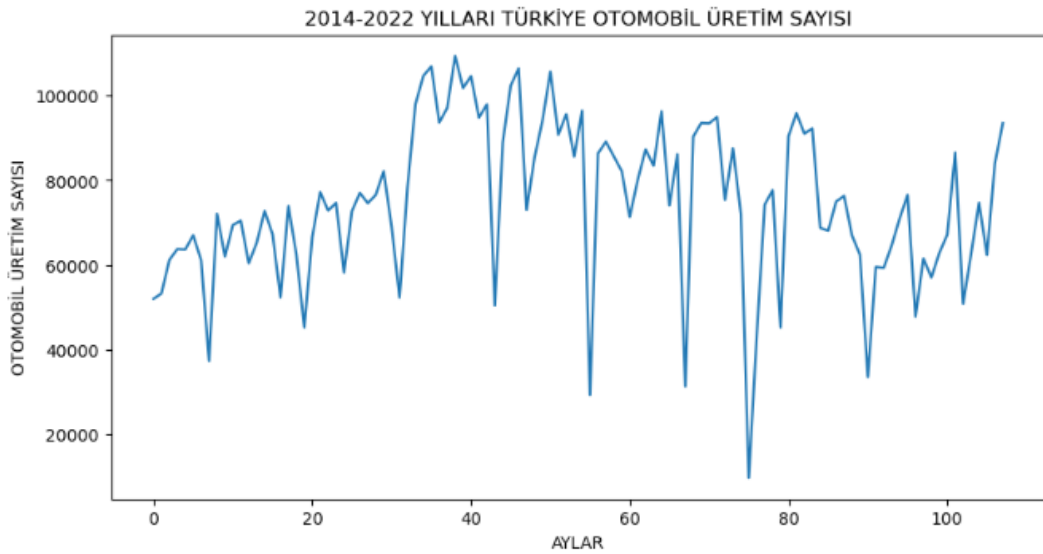
Ortalama Taşıt Kredi Faiz Oranları: Taşıt kredi oranları arttıkça, tüketicilerin araç alımındaki eğiminin azalması beklenmektedir. Batırlık vd. (2023) çalışmasında da

otomobil ve faiz oranı arasındaki ilişki görülmektedir. Otomobil satışları ile faiz oranı arasında tek yönlü nedensellik ilişkisi saptanmıştır. Tablolarda kredi kısaltması ile kullanılmıştır. Otomobil kredi faiz oranları aylık grafiği Şekil 3.4'te verilmiştir.



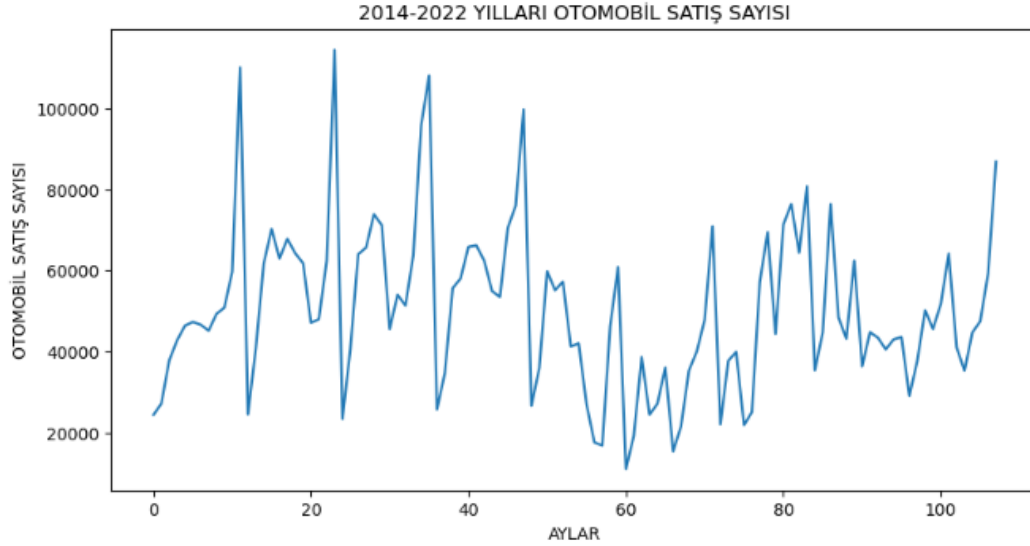
Şekil 3.4. Otomobil ortalama faiz oranları.

Türkiye Otomobil Üretim Sayısı: Otomobil Üretim Sayısı ile tüketici talebi arasında doğrusal bir ilişki olması beklenmektedir. Toplam otomobil satışını etkileyebilmesi sebebi ile otomobil üretim verileri bağımsız veri seti içerisinde kullanılmıştır. Tablolarda üretim kısaltması ile kullanılmıştır. Türkiye otomobil üretim sayısı aylık grafiği Şekil 3.5'te verilmiştir.



Şekil 3.5. Türkiye otomobil üretim sayısı.

Toplam Otomobil Satış Sayısı: Tahmin yapılmak istenen verilerdir. 2014-2022 yılları arasındaki aylık satış sayıları uygulama içerisinde bağımlı veri olarak kullanılmıştır. Tablolarda toplam kısaltması ile kullanılmıştır. Türkiye otomobil üretim sayısı aylık grafiği Şekil 3.6’da verilmiştir.



Şekil 3.6. Otomobil satış sayısı.

Örnek veri seti Tablo 3.2’de verilmiştir. Veri setinin tümü, ekler bölümünde gösterilmiştir.

Tablo 3.2. Örnek uygulama veri seti.

Yıl	Aylar	İthalat	RKGE	TGE	Kredi	Üretim	Toplam
2014	1	88,4	99,1	91,8	12,84	52.006,00	24.368,00
2014	2	80,2	100,3	89,5	15,09	53.250,00	27.167,00
2014	3	73,9	103,1	92,6	15,21	61.203,00	37.812,00
2014	4	67,5	105,1	97,4	14,91	63.761,00	42.769,00
2014	5	64,9	107,6	95	14,29	63.658,00	46.379,00
2014	6	65,3	106,4	93,2	13,31	67.013,00	47.278,00
2014	7	65,1	104,7	93,1	12,79	61.151,00	46.602,00
2014	8	73,9	104,1	93,2	12,84	37.330,00	45.131,00
2014	9	83	105,3	94,1	12,72	72.120,00	49.262,00
2014	10	82,8	109,4	91,7	12,37	62.001,00	50.814,00
2014	11	91,7	104,4	90,9	12,38	69.466,00	59.695,00
2014	12	92,1	105	90,4	12,6	70.480,00	110.054,00
2022	1	80,5	111,9	73,2	28,25	47.778,00	29.020,00
2022	2	63,5	109,9	71,2	26,85	61.544,00	37.641,00
2022	3	56,8	108,2	72,5	25,4	57.041,00	50.173,00
2022	4	58,8	107,7	67,3	24,82	62.837,00	45.564,00
2022	5	68,2	107	67,6	24,2	67.162,00	51.750,00
2022	6	67,9	104,6	63,4	25,04	86.585,00	64.134,00

Tablo 3.2. (Devamı) Örnek uygulama veri seti.

Yıl	Aylar	İthalat	RKGE	TGE	Kredi	Üretim	Toplam
2022	7	48,6	102,5	68	27,53	50.795,00	41.031,00
2022	8	82,6	101,4	72,2	27,76	62.560,00	35.230,00
2022	9	73	100,2	72,4	26,79	74.704,00	44.681,00
2022	10	82,4	102	76,2	26,33	62.363,00	47.440,00
2022	11	85,7	101,3	76,6	25,37	83.999,00	59.222,00
2022	12	99,9	101,6	75,6	24,55	93.521,00	86.774,00

3.2. Makine Öğrenmesi ve Yapay Sinir Ağları için Veri Ön İşleme

Veri ön işleme, oluşturulmuş olan veri setinin analiz edilmesi ve özellikle makine öğrenme algoritmaları ve yapay sinir ağı yönteminde daha iyi bir performans sonucu alınması için veri setinin, yapılan çalışmalara uygun ve kullanışlı hale getirilme sürecidir. Çetin ve Yıldız (2022) çalışmalarında, verilerin veri analizinde önemli olduğunu fakat veri analiz başarısının verinin özelliklerine bağlı olduğunu vurgulamışlardır. Bu sebeple, tahmin işleminden önce verileri ön işlemden geçirmenin önemli olduğundan ve performansı arttırdığından bahsetmişlerdir. Makine öğrenmesi ve yapay sinir ağı yöntemleri ile model oluşturmadan önce aykırı veri analizi Local Outlier Factor (LOF) ve sonrasında standardizasyon işlemi Z-Score yöntemi ile gerçekleştirilmiştir. Zaman serileri analizi ile model oluşturmadan önce ham veriler üzerinde aykırı veri analizi yapılmış ve tek değişkenli veri setlerinde sıklıkla kullanılan kutu grafiği (box-plot) yöntemi kullanılarak veri ön işleme uygulamaları gerçekleştirilmiştir.

3.2.1. Aykırı değer analizi

Diğer verilerden önemli ölçüde farklı olan veri, istatistikte aykırı değer olarak adlandırılmaktadır. İstatistiksel analizlerde, sorun oluşturabilmektedirler (Ovula ve Taşdelen, 2012). İstatistiksel analiz ve makine öğrenme algoritmalarının çoğu aykırı değerlere hassasiyet gösterirler. Bu sebeple, veri seti içerisindeki aykırı gözlemlerin tespiti ve analizi önemlidir. Veri setindeki aykırı değerlerin analiz edilmeden makine öğrenmesi algoritmaları ile işlenmesi sonucunda eksik öğrenme (underfitting) ya da aşırı öğrenme (overfitting) gibi problemlerle karşılaşılabilir. Aykırı veriler istatistiksel testlerin gücünü azaltabilmektedir.

Aykırı değer türleri 3'e ayrılmaktadır. Bunlar;

Nokta Aykırı Değerler: Veri setindeki bir noktanın veri setindeki diğer değerlerden farklı olduğu durumlarda gözlenmektedir.

Bağılamsal Aykırı Değerler: Gözlemin, meydana gelen bir olay sonucunda etkilenmesi ile diğer noktalardan ayrıldığı durumlarda gözlenmektedir.

Toplu Aykırı Değerler: Veri setindeki belirli bir grup verinin kendi içerisinde normal fakat veri setinin bütününe bakıldığında diğer noktalardan ayrıldığı durumlarda gözlenmektedir.

Aykırı gözlem tespitinde, tek değişkenli ve çok değişkenli veri setleri için farklı yöntemler bulunmaktadır. Tek değişkenli aykırı gözlem yöntemlerinde, her değişken kendi içinde incelenmektedir ve aykırı veriler her değişken için ayrı gözlemlenmektedir. Tek değişkenli aykırı gözlem yöntemlerine, grafiksel yaklaşımlardan, Box-Plot Yöntemi, Histogram Grafiği Yöntemi; istatistiksel yaklaşımlardan, Çeyrekler Açıklığı Yöntemi, Z-Score Yöntemi gibi uygulamalar örnek verilebilmektedir. Çok değişkenli aykırı gözlem yöntemlerinde, verideki değişkenler tek tek değil bir bütün olarak analiz edilmektedir. Çok değişkenli aykırı gözlem yöntemlerine, grafiksel yaklaşımlardan Saçılım Grafiği (Scatter Plot) Yöntemi; istatistiksel yaklaşımlardan LOF (Local Outliers Factor), İzolasyon Ormanı (Isolation Forest) Yöntemi gibi uygulamalar örnek verilebilmektedir.

Atan (2016) çalışmasında, çok değişkenli yöntemler ile aykırı veri analizinin gerekliliğini vurgulamıştır. Çalışmasında, verilerin ayrı ayrı ele alındığında aykırılık göstermeyip iki ya da daha fazla parametre ele alındığında aykırı değer olarak gözlemlenen verilerin olduğuna değinmiştir.

Bu çalışmada, çok değişkenli veri seti üzerinde, aykırı veri analizi yöntemlerinden veri setine bütüncül bakarak analiz edilmesine olanak sağlayan yerel aykırı değer faktörü (LOF) yöntemi kullanılmıştır. Öncelikle ham veriler Python programına tanıtılmıştır. Sonrasında istatistiksel olarak analiz edilmiş ve değişkenler arasındaki korelasyon incelenmiştir. Programa tanıtılan ham verilerin korelasyon tablosu Tablo 3.3'te gösterilmiştir. Korelasyon, iki değişken arasındaki ilişkinin gücünü gösteren bir katsayıdır. -1 ile 1 arasında değişmektedir. 1'e yaklaşırken pozitif yönde, artan kuvvette bir ilişki olduğunu; -1'e yaklaşırken negatif yönde, artan kuvvette bir ilişki olduğunu göstermektedir. 0'a yaklaşması durumunda, aradaki ilişkinin zayıfladığı ve

zayıf olduğu yorumu yapılabilmektedir. 0 olduğunda ise iki değişken arasında ilişki olmadığı anlamına gelmektedir.

Tablo 3.3. Orijinal veriler korelasyon tablosu.

	Aylar	İthalat	RKGE	TGE	Kredi	Üretim	Toplam
Aylar	1	0,00	0,07	-0,10	0,06	0,16	0,50
İthalat	0,00	1	0,36	0,47	-0,58	0,16	0,56
RKGE	0,07	0,36	1	0,13	-0,12	0,26	0,35
TGE	-0,10	0,47	0,13	1	-0,64	0,15	0,24
Kredi	0,06	-0,58	-0,12	-0,64	1	0,00	-0,38
Üretim	0,16	0,16	0,26	0,15	0,00	1	0,34
Toplam	0,50	0,56	0,35	0,24	-0,38	0,34	1

Aykırı değer analizinde LOF yöntemi 2000 yılında Markus M. Breunig, Han-Peter Kriegel, Raymond T. Ng ve Jörg Sander tarafından önerilmiştir. Belirlenen bir veri noktasının en yakın komşularını dikkate alarak aykırı değer tespiti yapan bir algoritmadır (Ceyhan, 2019). Komşuluk derecelerine göre çalışmaktadır. Literatürde, topluluk öğrenme yaklaşımı olarak adlandırılabilir.

Çok değişkenliler ile aykırı veri analizinde LOF, gözlem değerlerinin bulundukları noktada yoğunluk tabanlı skor oluşturarak aykırı gözlem olabilecek değerleri görmemize imkan sağlar. Skor oluştururken komşuluk yaklaşımı ile çalışmaktadır. Bir noktanın local yoğunluğu, komşularının yoğunluklarından anlamlı bir şekilde düşük olduğunda, seyrek bir bölgede olduğu anlaşılmakta ve aykırı gözlem değeri olarak kabul edilmektedir. Skorun 1 değerine yakın olması, gözlem değerinin komşularına benzer yoğunlukta olduğunu göstermektedir. 1 değerinden büyük ise komşularından daha düşük yoğunlukta olduğunu göstermekte ve aykırı değer olarak kabul edilmektedir.

$LOF_{(k)} \sim 1$ ise gözlem değeri komşularına benzer yoğunluktadır.

$LOF_{(k)} > 1$ ise gözlem değeri komşularına düşük yoğunluktadır ve aykırı gözlem değeri olarak kabul edilebilmektedir.

Aykırı veri analizinde, LOF yönteminin avantaj ve dezavantajları vardır. Avantajı, yöntem ile aykırı veri tespiti için verilerin bir bütün olarak analiz edilmesidir. Dezavantajı ise; gözlem değerinin aykırı değer olarak tanımlanacağı belirli bir eşik değeri yoktur. Eşik değerinin belirlenmesi soruna ve kullanıcıya bağlı olmaktadır.

Literatürde, çok değişkenli aykırı veri analizi yöntemi olarak LOF, birçok çalışmada kullanılmıştır ve başarılı sonuçlar verdiği gözlemlenmiştir. Yarat ve Orman da çalışmalarında, anomali tespiti için kullanmışlar ve başarılı sonuçlar almışlardır (Yarat ve Orman, 2023).

Bu çalışmada da, LOF ile aykırı veri analizi uygulamasında, PYTHON programlama dili ve JUPITER NOTEBOOK ara yüzü kullanılmıştır. Python programında, “sklearn” kütüphanesinden “LocalOutlierFactor” fonksiyonu kullanılarak skorlama işlemi gerçekleştirilmiştir.

Uygulamada komşuluk katsayısı ve yoğunluk parametreleri kullanılmıştır. Komşuluk katsayısı, faktör değerini direkt olarak etkilemektedir. Literatürdeki çalışmalarda, komşuluk katsayısının 10-20 arasında seçilmesinin denemeler sonucunda başarılı sonuçlar verdiği görülmüştür (Breunig vd., 2000).

Bu çalışmada da, komşuluk sayısı 10, yoğunluk 0,1 alınarak fonksiyon çalıştırılmıştır. Yoğunluk skorları çıktı olarak alınmış ve yoğunluk puanları sıralanmıştır. LOF ile aykırı değer tespitinde yoğunluk puanları sıralandıktan sonra eşik değer belirlenmesi gerekmektedir. Yöntemde, gözlem değerinin aykırı değer olarak tanımlanacağı belirli bir eşik değeri yoktur. Aykırı değer belirlenmesi soruna ve kullanıcıya bağlıdır. Eşik değeri 5 olarak belirlenmiştir. Eşik değeri belirlenirken 3 (veri setinin %3) ile 10 (veri setinin %9) arasındaki skor değerlerin tümü denenmiş ve aykırı veri analizi yöntemiyle oluşturulan veri setinin korelasyon katsayıları karşılaştırılmıştır. Eşik değeri 5 olarak belirlendiğinde oluşturulan yeni veri setinin korelasyon katsayılarının en iyi değerleri verdiği görülmüştür. Programda uygulanan kodlar verilmiştir. Aykırı değer analizi sonucunda oluşturulan yeni veri setinin korelasyonu da Tablo 3.4’teki gibidir. LOF yöntemi ile aykırı değer analizi çalışma kodları açıklamaları ile birlikte Ekler bölümünde verilmiştir.

Tablo 3.4. Aykırı değer analizi yapılmış veriler korelasyon tablosu.

	Aylar	İthalat	RKGE	TGE	Kredi	Üretim	Toplam
Aylar	1	0,03	0,06	-0,10	0,05	0,22	0,49
İthalat	0,03	1	0,39	0,49	-0,60	0,21	0,64
RKGE	0,06	0,39	1	0,12	-0,22	0,14	0,38
TGE	-0,10	0,49	0,12	1	-0,66	0,19	0,26
Kredi	0,05	-0,60	-0,22	-0,66	1	-0,08	-0,42
Üretim	0,22	0,21	0,14	0,19	-0,08	1	0,45
Toplam	0,49	0,64	0,38	0,26	-0,42	0,45	1

3.2.2. Standardizasyon

Çok değişkenli analiz yöntemlerinde, özellikle makine öğrenmesi ve yapay sinir ağları yöntemlerinde, algoritmaların optimal şekilde çalışabilmesi için verilerin aynı ölçeğe getirilmesi önemlidir. Değişkenlerin standartlaştırılması doğru analiz için önemli rol oynamaktadır. Farklı ölçeklerde ölçülen değişkenler analize eşit katkıda bulunamaz. Standardizasyon işleminde varyasyon korunarak ölçek değiştirilmektedir. Veri setinin yapısı korunarak, verilerin karşılaştırılabilir duruma getirilmesi hedeflenmektedir. Farklı değişkenlerin, farklı yapıları ve ölçekleri olması sebebiyle değişkenleri karşılaştırılabilir ve ölçeklenebilir duruma getirmek, algoritmaların verimli sonuçlar vermesi için gereklidir. Standardizasyon işleminde değişkenlerin ortalaması 0'a, standart sapmaları ise 1 değerine getirilmektedir.

Akyiğit ve Taşçı (2022) da çalışmalarında, değişkenlerin standartlaştırılmasının öneminden bahsetmişlerdir. Değişkenlerdeki farklı birimlerin karşılaştırma durumlarını ortadan kaldırmak için yöntemi kullanmışlardır. Ayrıca, makine öğrenme algoritmaları için standartlaştırmanın bir gereklilik olduğunu vurgulamışlardır.

Bu çalışmada, çok sık kullanılan ve literatürde yerini almış olan standardizasyon yöntemlerinden Z-Score standardizasyon yöntemi ile standartlaştırma işlemi gerçekleştirilmiştir.

Standardizasyonda, veri setinin yapısı korunarak, verilerin karşılaştırılabilir duruma getirilmesi hedeflenmektedir. Farklı değişkenlerin, farklı yapıları ve ölçekleri olması sebebiyle değişkenleri karşılaştırılabilir ve ölçeklenebilir duruma getirmek,

algoritmaların verimli sonuçlar vermesi için gereklidir. Standardizasyon işleminde değişkenlerin ortalaması 0'a, standart sapmaları ise 1 değerine getirilmektedir.

Öncelikle, Tablo 3.5' te gösterildiği gibi 2014-2022 yılları arasındaki veriler için her bir değişkenin ortalama ve standart sapma değeri bulunmuştur.

Tablo 3.5. Değişkenlerin ortalama ve standart sapma değerleri.

	Aylar	İthalat	RKGE	TGE	Kredi	Üretim	Toplam
Ortalama	6,5	74,41	104	84,67	18,58	74.567,21	50.285,40
Standart Sapma	3,47	26,42	6,99	7,63	5,53	18.902,68	20.426,25

Değişken ortalama ve standart sapmaları ile Denklem (3.1) üzerinden Z-Score standardizasyon işlemi gerçekleştirilmiştir. Çalışmada, kullanılan veri setinin karşılaştırılabilirlik kaygısından kaynaklı veriler belirli bir değer aralığına Z-Score yöntemi ile indirgenmiştir. Her bir değişkenin, ortalaması 0, standart sapması 1 değerini almıştır. Standardizasyon işlemi sonucu oluşan örnek veri değerleri Tablo 3.6'daki gibidir.

$$Z = \frac{X_i - \mu}{\sigma} \quad (3.1)$$

Denklem (3.1)'de;

Z, değişken içerisindeki verinin z-score puanını;

X_i , z-score' u alınan değişken içerisindeki i'inci veriyi;

μ , değişkenin ortalamasını,

σ , değişkenin standart sapmasını ifade etmektedir.

Tablo 3.6. Aykırı değer analizi ve standardizasyon işlemi sonucunda oluşan örnek uygulama veri seti

Aylar	İthalat	RKGE	TGE	Kredi	Üretim	Toplam
-1,61027	0,46266	-0,94074	0,90534	-1,03446	-1,30430	-1,34031
-1,31897	0,16168	-0,73154	0,60533	-0,62758	-1,23756	-1,19684
-1,02766	-0,06957	-0,24342	1,00969	-0,60635	-0,81086	-0,65120
-0,73636	-0,30449	0,10524	1,63580	-0,66054	-0,67362	-0,39712
-0,44505	-0,39992	0,54107	1,32274	-0,77290	-0,67914	-0,21208
-0,15374	-0,38524	0,33187	1,08795	-0,95047	-0,49914	-0,16600

Tablo 3.6. (Devamı) Aykırı değer analizi ve standardizasyon işlemi sonucunda oluşan örnek uygulama veri seti

Aylar	İthalat	RKGE	TGE	Kredi	Üretim	Toplam
0,13756	-0,39258	0,03551	1,07491	-1,04440	-0,81365	-0,20065
0,42887	-0,06957	-0,06909	1,08795	-1,03446	-2,09170	-0,27605
0,72017	0,26445	0,14011	1,20535	-1,05704	-0,22514	-0,06430
1,01148	0,25711	0,85487	0,89230	-1,11900	-0,76805	0,01525

Makine öğrenmesi ve yapay sinir ağları ile talep tahmin uygulamasında LOF yöntemi ile aykırı değer analizi ve z-score standardizasyonu yapılmış veriler, zaman serileri analizi yöntemi ile talep tahmin uygulamasında ise box-plot yöntemi ile aykırı değer analizi gerçekleştirilmiş veriler kullanılmıştır. Box-plot yöntemi tek değişkenli veri setleri için aykırı değer analiz yöntemlerinden biridir. Veri seti üzerinde uygulaması, zaman serileri analizi ile talep tahmini bölümünde verilmiştir.

3.2.3. Makine öğrenmesi ile talep tahmini

Veri ön işleme yapıldıktan sonra doğrusal ve doğrusal olmayan gözetimli makine öğrenme yöntemleri ile algoritmalar oluşturulmuş ve karşılaştırmaları yapılmıştır. Makine öğrenmesi ile talep tahmini uygulamalarında Python programı kullanılmıştır. Makine öğrenme algoritmalarını uygulayabilmek için Sklearn, CatBoost, XGBoost kütüphaneleri içe aktarılarak uygulamaların zemini hazırlanmıştır. 2014-2022 yıllarından oluşan veri seti programa tanıtılmıştır. Toplu makine öğrenmesi yapabilmek için kod yazılmıştır. 12 algoritma için çalıştırıldıktan sonra hata sonuçları karşılaştırılmıştır.

3.2.3.1. Veri setindeki bağımlı ve bağımsız değişkenlerin ayrılması

Veri setinin bağımlı ve bağımsız değişkenleri ayrılmıştır. Bağımlı değişken olan Toplam Otomobil Sayısı Y adındaki değişkene, diğer değişkenler ise bağımsız değişkenler olarak X adındaki değişkene atanmıştır.

3.2.3.2. Veri setinin eğitim ve test verisi olarak ayrılması

Eğitim ve test veri setinin ayrılması makine öğrenmesi algoritmalarının başarı hedefleri için önemli bir aşamadır. Eğitim veri seti ile model eğitilirken, test veri seti ile modelin eğitiminden sonra sistemin daha öncesinde görmemiş olduğu veriler ile modelin doğruluğu test edilmektedir. Aslan ve Yıldız (2022) çalışmalarında, modelin performans ölçümünün doğru şekilde yorumlanması için veri setlerinin

eğitim ve test veri seti olarak ayrılmasının öneminden bahsetmişler ve çalışmalarında kullanmışlardır.

Bu çalışmada, Python programında model_selection kütüphanesinin modülü olan train_test_split() fonksiyonuna X adını verdiğimiz bağımsız değişkenler ile Y adını verdiğimiz bağımlı değişken verilmiştir. Verilerin %20 i test, geri kalan %80 i eğitim verisi olarak kullanılmıştır.

3.2.3.3. Makine öğrenmesi modellerinin performanslarının karşılaştırılması ve talep tahmin modelinin seçilmesi

Veri ön işleme uygulamaları yapılmış veri seti ile 12 farklı model uygulanmış ve sonuçları performans değerlendirme ölçütlerinden RMSE (Ortalama Kare Hata Karekökü) ve R^2 (Determinasyon Katsayısı) ile karşılaştırılmıştır. Sonuçların karşılaştırılması Tablo 3.7'deki gibidir. R^2 değerinin 0-1 arasında sonuç vermesi beklenmektedir. 1' yakın olması istenmektedir. R^2 değeri, bağımsız değişkenlerin bağımlı değişkeni ne kadar iyi açıklayabildiğinin göstergesidir. RMSE değerinin ise düşük olması istenmektedir. RMSE, verilerin tahmin değerlerinin gerçek değerlerinden ne kadar uzak olduğunu gösteren bir ölçüttür.

Tablo 3.7. Makine öğrenmesi yöntemleri performans analizi tablosu.

Model Adı	R^2 Eğitim	RMSE Eğitim	R^2 Test	RMSE Test
CatBoost Regressor	0,99	0,02	0,86	0,410
Gradient Boosting Regressor	0,99	0,09	0,86	0,412
Random Forest Regressor	0,96	0,2	0,80	0,484
Bagging Regressor	0,93	0,26	0,78	0,515
SVR	0,81	0,43	0,74	0,557
XGB Regressor	0,99	0,001	0,73	0,566
Ridge	0,69	0,54	0,71	0,582
Linear Regression	0,69	0,54	0,71	0,587
KNeighbors Regressor	0,72	0,51	0,67	0,626
Decision Tree Regressor	1	0	0,56	0,723
Elastic Net	0,08	0,93	0,08	1,043
Lasso	0	0,97	0,00	1,089

Modellerin performansları karşılaştırıldığında CatBoost, Gradient Boosting, Random Forest, SVR algoritmalarının daha iyi sonuçlar verdiği görülmektedir. Performansları daha yüksek olan algoritmaların ayrıntılı çözümleri gerçekleştirilmiştir.

3.3. Performansı Yüksek Makine Öğrenmesi Modelleri ile Uygulama

Bu bölümde, seçilen modellerin uygulanabilmesi için öncelikle Python programında sklearn ve catboost kütüphaneleri içe aktarılmış ve uygulamanın temeli atılmıştır. Çalışmada kullanılan gözetimli makine öğrenme algoritmaları açıklanmış ve uygulamaları Bölüm 4'te gerçekleştirilmiştir.

Makine öğrenme algoritmaları için, 2014-2022 yıl verilerini içeren ve veri ön işleme uygulamaları yapılmış olan veri seti programa tanıtılmıştır. Verilerin %20 i test, geri kalan %80 i eğitim verisi olarak kullanılmıştır.

Makine Öğrenme algoritmalarında, model düzenleme bölümlerinde en yüksek performansı gösteren değişkenleri bulabilmek için hiperparametre optimizasyonu yapılmıştır. Ayrıca, GridSearchCv, parametreleri daha güvenilir değerlendirmek adına Cross Validation (Çapraz Doğrulama) kullanmaktadır. Hiperparametre optimizasyonu, Cross Validation'ın ikinci bir kullanım alanı olarak da adlandırılmaktadır. Amaç optimal hiperparametreleri bulabilmektir. Utku ve Can (2022) da çalışmalarında, GridSearchCV fonksiyonunu kullanarak parametre optimizasyonu gerçekleştirmişlerdir. Hiperparametre optimizasyonu, makine öğrenme algoritmalarının doğruluk derecesini arttırmak için en iyi performansı veren değişken değerin tespit edilme işlemidir.

Hiperparametre belirleme, makine öğrenmesi yönteminde optimal sonucu veren modelin parametrelerinin oluşturulmasıdır. Farklı parametre kombinasyonlarının denenmesi ve test edilmesi gerekmektedir. GridSearch, Random Search, Bayesian Optimization hiperparametre belirleme yöntemlerindendir (Murat ve Özcanhan, 2023).

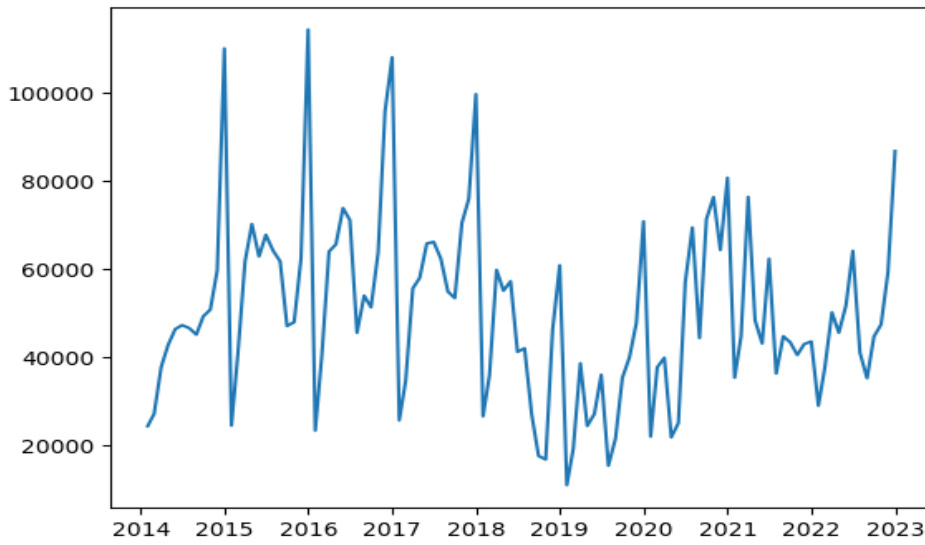
Duran ve Bakır (2023) çalışmalarında, Grid Search, Random Search ve Bayesian Optimization yöntemlerinin karşılaştırmasını yapmışlar ve çalışmalarında, kullanılan makine öğrenme algoritmaları çok fazla hiperparametrelere sahip olmadığından dolayı ve en iyi hiperparametreleri Grid Search algoritması ile tespit ettikleri için Grid Search hiperparametre yönteminin başarı oranının yüksek olduğunu belirtmişlerdir.

GridSearchCv yöntemi, algoritmaya verilen tüm parametre kombinasyonlarını tek tek deneyerek oluşturulan modellerin performanslarını incelemektedir. Sonucunda, en iyi model performansı gösteren parametreleri vermektedir. Model düzenleme ve

hiperparametre optimizasyon işlemleri makine öğrenme algoritmaları için farklılık oluşturabilmektedir. Model düzenleme ve hiperparametre optimizasyon işlem farklılıkları algoritmalar ile talep tahmini bölümlerinde anlatılacaktır.

3.4. Zaman Serisi Analizi ile Talep Tahmini

Bu çalışmada, Box-Jenkins yöntemlerinden olan ve mevsimsellik içeren verilerde mevsimsel eğilimleri içine alarak modelleme imkanı sağlayan SARIMA (Mevsimsel Otoresif Entegre Hareketli Ortalama) modeli kullanılmıştır. Uygulamada kullanılan serinin, zamana bağlı grafiği Şekil 3.7’de gösterilmektedir.



Şekil 3.7. Veri setinin zamana bağlı grafiği.

3.4.1. Mevsimsel otoresif entegre hareketli ortalama yöntemi

SARIMA, ARIMA modeline mevsimsel bileşen eklenerek uyarlanan bir modeldir. Literatürde mevsimsel ARIMA modeli olarak da geçmektedir. Sarıma modeli ile modelleme yapılırken, mevsim etkisi taşıyan ve mevsim etkisi taşımayan bölümler birlikte var olmaktadır. Mevsimler arasındaki sabit periyotlardaki değişimleri mevsimsel modeller, mevsim içerisindeki değişimleri ise mevsimsel olmayan modeller formüle etmektedir (Güneş, 2010). Mevsimsel ve mevsimsel olmayan parametrelerin bir arada kullanımı söz konusudur. SARIMA modeli, mevsimsellik bulunduran modellerdir. Modelin temel gösterimi, $SARIMA(p,d,q)(P, D, Q)_m$ şeklindedir.

(p,d,q) mevsimsel olmayan bölümü ifade ederken; (P, D, Q)_m mevsimsel bölümü ifade etmektedir.

p: modele dahil edilen gecikmelerin sayısını,

d: farklılaşma derecesini,

q: hata gecikme sayısını,

P: otoregresif sırasını,

D: fark sırasını ve

Q: hareketli ortalama sırasını

m: her sezondaki periyot sayısını temsil etmektedir.

Zaman serilerinde durağanlık durumu:

Durağanlık zaman serileri analizinde önem taşıyan bir kavramdır. Bir çok istatistiki zaman serileri tahmin analizi serinin durağanlığını gerektirir. Serinin durağan olmadığı durumlarda, durağan hale getirmek için logaritmik dönüşüm, trend ekleme, box-cox dönüşümü, fark alma gibi yöntemler mevcuttur.

Altın ve Çelik Eroğlu (2020) çalışmalarında seriyi durağanlaştırmak için fark alma yöntemini kullanmışlardır. Seride durağanlık olmadığında serinin durağanlığı sağlanan kadar seri üzerinde fark alma işleminin gerçekleşmesi ve model tespitinde durağan seri üzerinden uygun olan modelin belirlenmesi gerektiğini belirtmişlerdir.

Zaman serilerinde durağanlığı bozan durumlar trend, mevsimsellik ve birim kökler olmaktadır. Durağanlığı bozan yapılar için geliştirilmiş testler mevcuttur. Seri de trend ve mevsimsellik söz konusu olduğunda serinin trend ve mevsimsellikten arındırılması gerekmektedir. Yılmaz (2021) çalışmasında, seride hem trend hem de mevsimsellik olması durumunda, önce mevsimsellik için daha sonra trend için fark alınması gerektiğini belirtmiştir.

Bu çalışmada, zaman serileri analizi ile talep tahmininde, Box-Jenkins istatistiksel zaman serisi analiz yöntemlerinden SARIMA kullanılmıştır. SARIMA, ARIMA modeline mevsimsel bileşen eklenerek uyarlanan bir modeldir. Literatürde mevsimsel ARIMA modeli olarak da geçmektedir.

Bu bölümde, seçilen modellerin uygulanabilmesi için öncelikle Python programında SARIMA modeli kullanabilmek için statsmodel, model parametrelerini belirleyebilmek için pmdarima kütüphaneleri içe aktarılmış ve uygulamanın temeli atılmıştır.

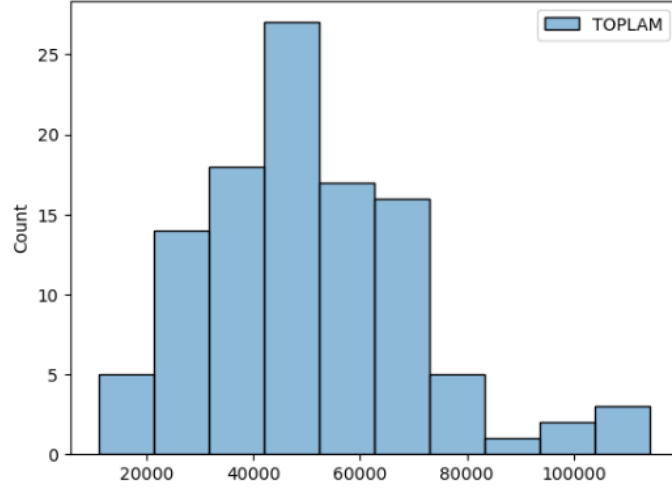
Zaman serisi verisi olarak, binek otomobil satış sayısı verileri içe aktarılmıştır. Veri seti içerisinde bulunan aylar değişkeni, zaman olarak tanımlanmış ve tarih indeksi olarak kullanılmıştır. Uygulamada kullanılan veri örneği ve indeks sütunu olarak tarih verisi sütunu Tablo 3.8’de gösterilmiştir.

Tablo 3.8. Zaman serisi örnek veri seti.

	Toplam Satış Sayısı
Tarih	
2014-01-31	24368
2014-02-28	27167
2014-03-31	37812
2014-04-30	42769
2014-05-31	46379
...	...
2022-08-31	35230
2022-09-30	44681
2022-10-31	47440

Sarıma modeli için veri ön işleme: Zaman serisi analizinde, veri ön işleme adımları, modele uygun ve doğru sonuçlar elde etmek için önem taşımaktadır. Veri setinin özelliklerine ve analiz yapılmak istenilen modele göre, veri üzerinde veri ön işleme gerçekleştirmek önemlidir.

Aykırı değer tespiti: Veri seti içerisinde, zaman serisi analizinden önce aykırı değerlerin incelenmesi gerekmektedir. Zaman serisi verilerinde aykırı değerler, analizi ve modelleme sürecini etkileyebilir. Aykırı değerleri belirlemek ve aykırı değer analizi ile değer kontrolü yapmak önemlidir. Toplam Otomobil Satış serisine ait histogram grafiği Şekil 3.8’de, serinin tanımlayıcı istatistikleri Tablo 3.9’da verilmiştir.

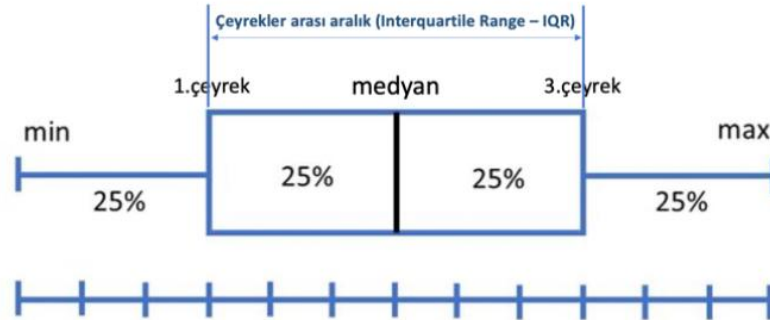


Şekil 3.8. Otomobil satış sayılarına ait histogram grafiği.

Tablo 3.9. Zaman serisinin tanımlayıcı istatistikleri.

Toplam Satış Sayısı	
Veri Sayısı	108
Ortalama	50.285,40
Standart Sapma	20.426,25
Minimum Değer	10.779
25%	37.705,5
50%	47.183
75%	62.517,25
Maximum Değer	114.340

Bu bölümde, aykırı değer analiz yöntemlerinden Box-Plot (Kutu Grafiği) yöntemi kullanılarak aykırı veri kontrolü sağlanmak istenmiştir. Örnek Box-Plot grafiği Şekil 3.9’da gösterilmiştir.



Şekil 3.9. Örnek Box-Plot grafiği (yalindunya, 2020).

Box-Plot grafiđi John Tukey tarafından kutu grafiđi adıyla veri analiz aracı olarak geliştirilmiřtir (Rousseeuw vd., 1999). Kutu grafiđi verilerin istatistiksel olarak analiz edilmesi ve görsel olarak özetlenmesine olanak sağlamaktadır. Verilerin sıklıđını göstermek için kullanılmaktadır. Deđiřkenlerdeki verilerin yayılım durumunu analiz ederek aykırı deđerlerin tespitinde kullanılmaktadır.

Q1, birinci kuartil, ve Q3,üçüncü kuartil üzerinden işlemler gerçekleştirilmektedir. Belirlenen katsayıya göre, alt sınır ve üst sınır dışında kalan deđerler aykırı veri olarak tanımlanmaktadır. Çeyrek aralıđı katsayısı genelde 1,5 olarak alınmaktadır.

$Q1 = \text{quantile}(0.25)$: Skorların %25' lik bölümünü,

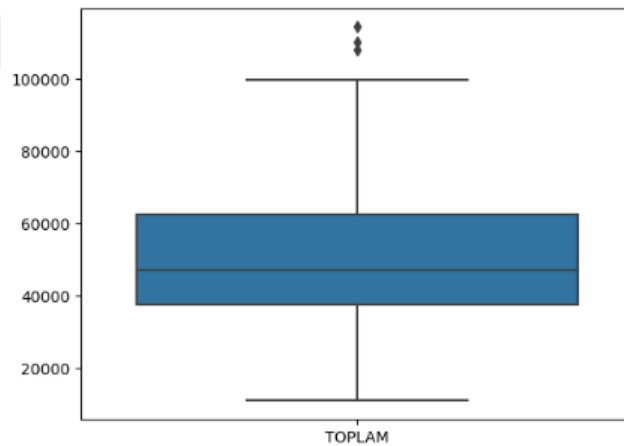
$Q3 = \text{quantile}(0.75)$: Skorların %75' lik bölümünü,

$IQR = Q3 - Q1$: Çeyrekler arası açıklıđı,

$\min = Q1 - 1.5 * IQR$: Aykırı deđerler hariç en küçük veriyi,

$\max = Q3 + 1.5 * IQR$: Aykırı deđerler hariç en büyük veriyi ifade etmektedir.

Oluřturulan box-plot ve aykırı deđerler řekil 3.10'da gösterilmiřtir.

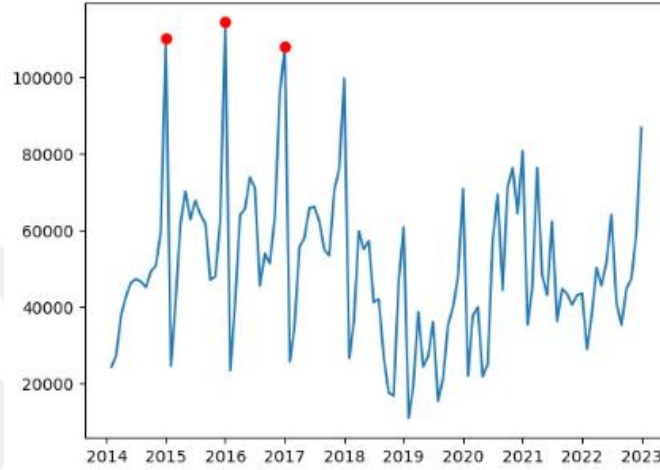


řekil 3.10. Ham veriler üzerinde oluřturulan box-plot grafiđi.

Aykırı deđer analizinden çıkan sonuçlara göre, Tablo 3.10'da gösterilen veriler, aykırı deđer olarak kabul edilmiřtir. Grafik üzerinde noktasal gösterimi řekil 3.11'deki gibidir.

Tablo 3.10. Box-plot yöntemi ile belirlenen aykırı değerler.

Tarih	Toplam Satış Sayısı
2014-12-31	110054
2015-12-31	114340
2016-12-31	108044



Şekil 3.11. Aykırı değerlerin grafik üzerinde noktasal gösterimi.

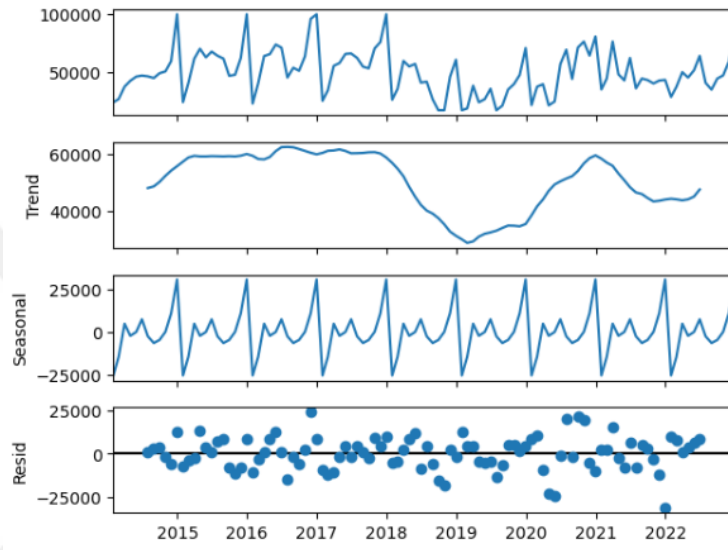
Aykırı değer kontrolünü sağlayabilmek için, baskılama yöntemi olarak kullanılan zaman serisi analizi veri ön işleme için Winsorize yöntemi kullanılmıştır. Aykırı değerler, üst sınır veri değeri ile eşitlenmiştir. Üst sınır veri değeri, 99734,875 olarak bulunmuştur. Aykırı değerler üst sınır değerine eşitlenmiştir.

Winsorize yöntemi Barnett ve Lewis tarafından 1978 yılında önerilen aykırı gözlemlerin düzeltilmesine yönelik bir yöntemdir. Veri seti içerisinde veri setinin yapısını bozan çarpık verilerin ortalama ve varyans üzerindeki etkisini azaltarak daha iyi modellerin kurulmasını sağlayan bir yöntemdir (İçen, 2022). Ovla ve Taşdelen (2012) çalışmalarında, winsorize yöntemini, aykırı değerlerin örnekleme olan etkisini azaltan güçlü bir yöntem olarak belirtmişlerdir. Çobanoğlu ve Gümrah (2023), uç değerlerin silinmesi yerine winsorize işlemi ile düzenlenmesinin veri kaybını önlemekte olduğunu vurgulayarak çalışmalarında winsorize yöntemini kullanmışlardır.

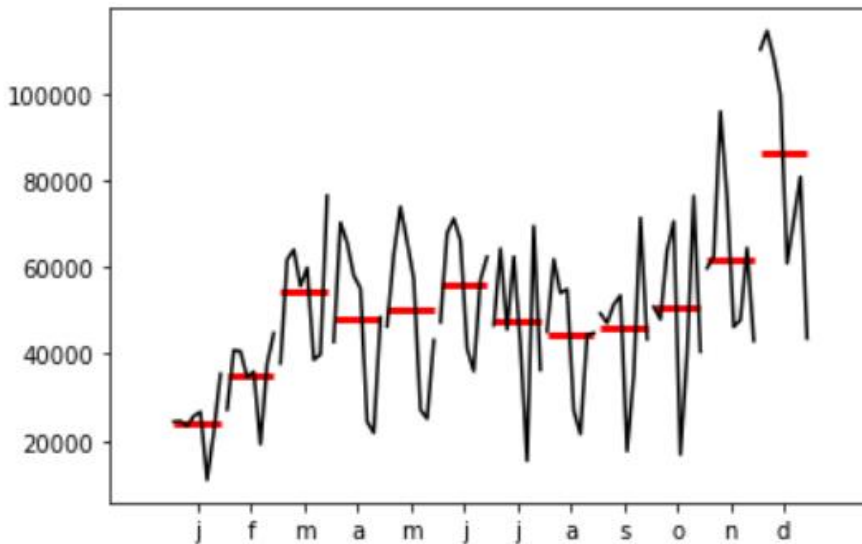
Seriye Durağanlaştırma: Zaman serileri genellikle trend, mevsimsellik ve gürültü bileşenlerine sahiptir. Bu bileşenleri analiz etme ve uygun olmayan bileşenleri

özmlleme, durađan olmayan veri setini durađanlařtırma modelleme srecinde nem tařımaktadır. Durađan veri, sabit ortalama ve sabit varyansa sahiptir. Box-Jenkins modellerinden olan SARIMA modeli kurulabilmesi iin, serinin trend etkisinden ve mevsimsel dalgalanmadan arındırılmasına yani serinin durađan olmasına dikkat edilmelidir.

Veri setinde, ayrılmıř mevsimsellik ve trend faktrlerinin grseli, řekil 3.12’de gsterilmiřtir. řekil 3.13’te ay bazında ortalama otomobil satıř grafiđi verilmiřtir.



řekil 3.12. Veri setine ait ayrıřtırma grafikleri.



řekil 3.13. Aylık ortalama otomobil satıř grafiđi.

Görsel olarak, mevsimselliğin varlığı gözlemlenmektedir. Durağan bir seri olmadığı varsayımı vardır. İstatistiksel olarak birim kök testleri ile kontrol etmek gerekmektedir.

Zamanla artan ya da azalan bir ortalama ve varyansa ya da herhangi bir tanesine sahip olan zaman serisi durağan olmamakta ve birim köke sahip olmaktadır. Seride durağanlığı bozan bir unsur birim köklerdir. Varlığının tespit edilmesi gerekmektedir. Seride birim kök olması durumunda seride fark alma işlemi gerçekleştirilerek serinin birim kökten arındırılması gerekmektedir (K. Yılmaz, 2021).

Seriye durağanlaştırmak için farklı yöntemler bulunmaktadır. Bu çalışmada seriyi durağanlaştırmak için fark alma yöntemi kullanılmıştır. İstatistiksel olarak durağanlık tespitinde ACF ve PACF fonksiyonları kullanılmaktadır.

ACF (otokorelasyon fonksiyonu) grafiği, serinin kendisi ile gecikmeleri arasındaki korelasyon katsayılarının bir çubuk grafiğidir ve ilişkinin boyutunu belirlemektedir. Otokorelasyon denklemi Denklem (3.2)'de verilmiştir.

$$\rho_k = \frac{cov(y_t, y_{t+k})}{\sqrt{var(y_t)var(y_{t+k})}} \quad (3.2)$$

Denklem (3.2)'de;

ρ_k : otokorelasyon katsayısını,

cov: değerlerin kovaryansını,

var: değerlerin varyansını ifade etmektedir.

PACF (kısmi otokorelasyon fonksiyonu) grafiği ise, 2 gecikmeli değer arasındaki gecikmeleri veren ve diğer gecikmelerin önemli olmadığı bir çubuk grafikdir. Diğer değişkenler sabit iken iki değişken arasındaki ilişkinin miktarını gösteren bir grafikdir. X_t ve X_{t+k} arasındaki korelasyon, X_t ve X_{t+k} dışındaki değerlerin etkisi çıkarılarak gözlemlenmektedir. “k”, gecikme sayısını ifade etmektedir. Kısmi otokorelasyon denklemi Denklem (3.3)'te verilmiştir.

$$\rho_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} (\rho_{k-j}) (\rho_j)}{1 - \sum_{j=1}^{k-1} (\rho_{k-j}) (\rho_j)} \quad (3.3)$$

Denklem (3.3)'te;

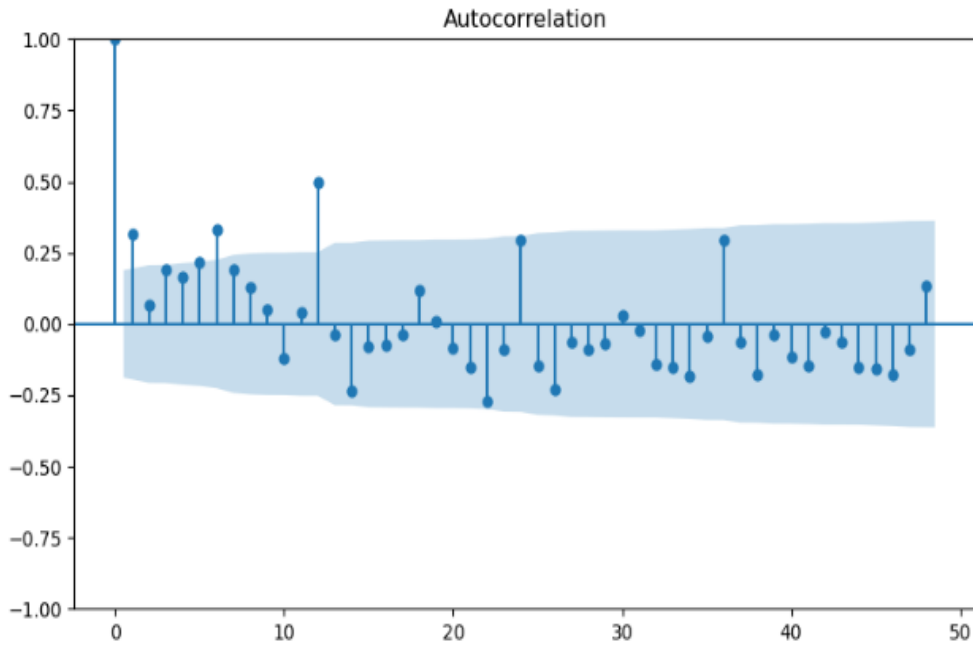
ρ_{kk} : kısmi otokorelasyon katsayısını,

ρ_k : k. gecikmedeki otokorelasyon katsayısını,

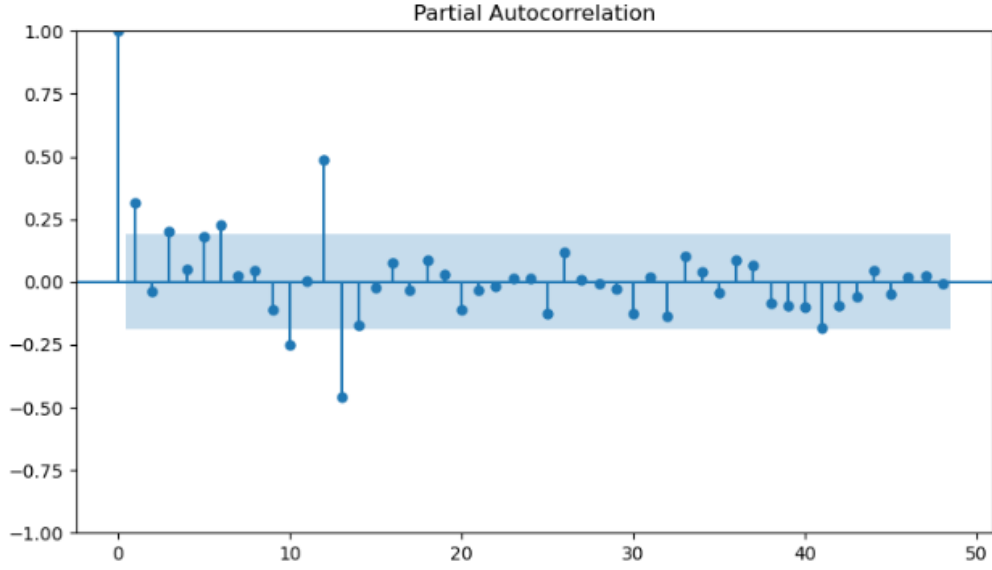
j: kısmi otokorelasyonu hesaplanan değerler dışında kalan gecikmedeki etkileri temsil etmektedir.

ACF ve PACF, model seçiminde yardımcı olabilecek fonksiyonlardır. Genellikle AR terimi için PACF, MA terimi için ACF grafiğine bakılmaktadır. AR ve MA düzeyini belirlerken, güven sınırının dışında kalan otokorelasyon sayısına bakılmaktadır.

Şekil 3.14 ve Şekil 3.15'te, serinin ACF ve PACF grafikleri gösterilmiştir.



Şekil 3.14. Ham veri ACF grafiği.



Şekil 3.15. Ham veri PACF grafiği.

Birim Kök Testleri:

Bu çalışmada, birim kök analizi için, Augmented Dickey Fuller (ADF) ve Kwiatkowski-Phillips-Schmidt-Shin (KPSS) testlerinden yararlanılmıştır.

ADF Testi Denklem (3.4)'teki test regresyonunun tahmin edilmesine dayanmaktadır:

$$y_t = Dt + \phi y_{t-1} + \sum_{j=1}^p \phi_j \Delta y_{t-j} + \varepsilon_t \quad (3.4)$$

Denklem (3.4)'te, Dt , t anındaki sabit ve trend gibi terimlerin deterministik fonksiyonu olarak tanımlanmaktadır. ε_t ise hata terimleridir. p değeri ise, ε_t serisinin korelasyonsuz olacağı şekilde ayarlanmış gecikme değeridir. ADF testinin hipotezleri aşağıdaki gibi tanımlanmaktadır:

H_0 : Seri, birim kök içerir (Seri, durağan değildir.)

H_1 : Seri, birim kök içermez (Seri, durağandır.)

$\hat{\phi}$, Denklem (3.5)'te yer alan ϕ katsayısının en küçük kareler yöntemiyle elde edilmiş tahmini, SE de standart hatayı ifade etmektedir. ADF testinin test istatistiği Denklem (3.5)'te verilmiştir.

$$ADF = \frac{\hat{\phi}}{SE(\hat{\phi})} \quad (3.5)$$

KPSS testi, ADF testi aksine;

H_0 : Seri, birim kök içermez (Seri, durağandır.)

$H1$: Seri, birim kök içerir (Seri, durağan değildir.)

Hipotezlerini ifade etmektedir.

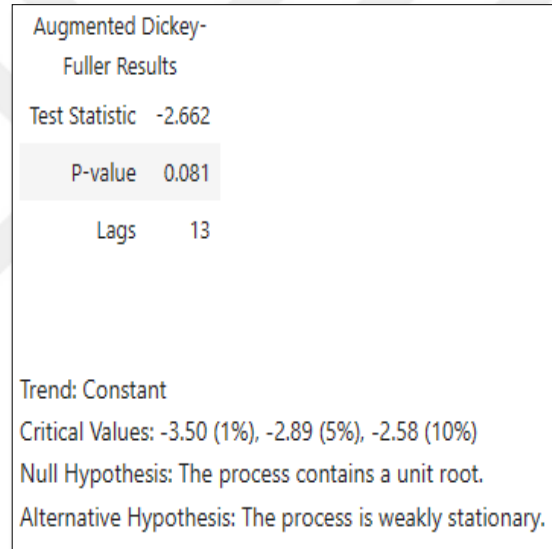
KPSS test modeli Denklem (3.6) ve Denklem (3.7)'daki gibidir.

$$y_t = \beta' D_t + \mu_t + u_t \quad (3.6)$$

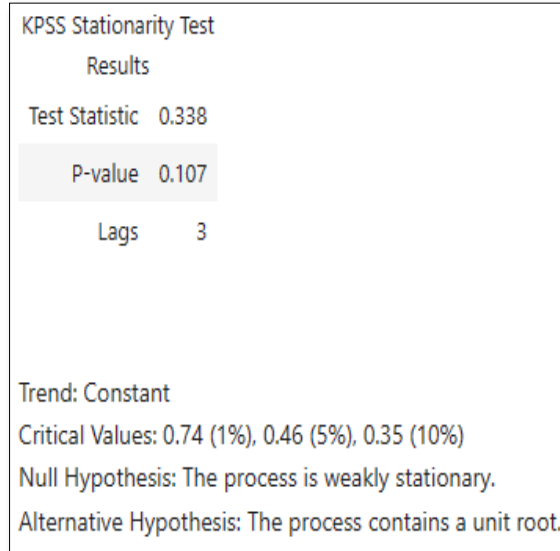
$$\mu_t = \mu_t + \varepsilon_t \quad (3.7)$$

D_t , t anındaki sabit veya sabitle trend gibi deterministik bileşen içeren fonksiyondur, β' , D_t fonksiyonunun katsayısıdır. u_t durağandır. μ_t , temiz dizi sürecidir (K. Yılmaz, 2021).

Şekil 3.16 ve Şekil 3.17'de seri üzerinde gerçekleştirilen ADF ve KPSS birim kök testlerinin sonucu gösterilmiştir.



Şekil 3.16. Ham veri üzerinde ADF birim kök testi.

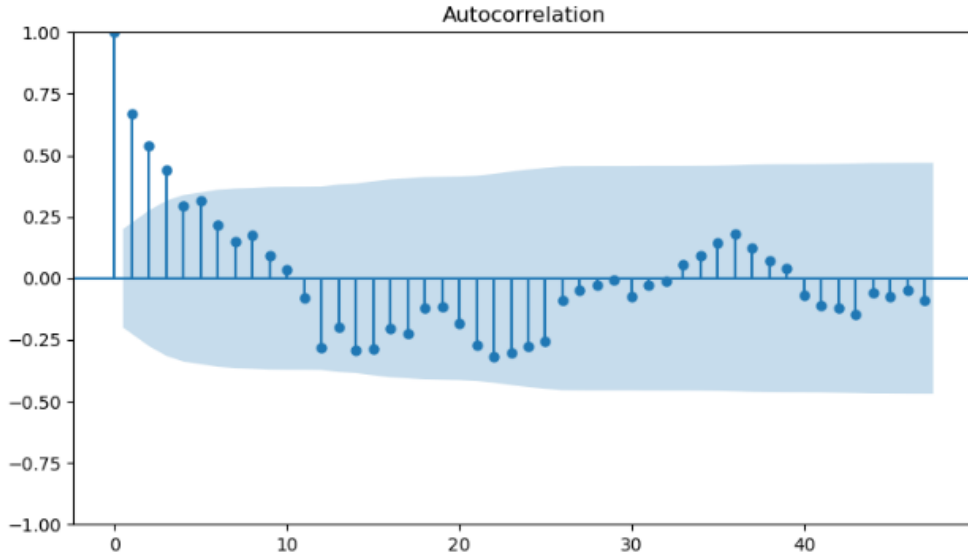


Şekil 3.17. Ham veri üzerinde KPSS birim kök testi.

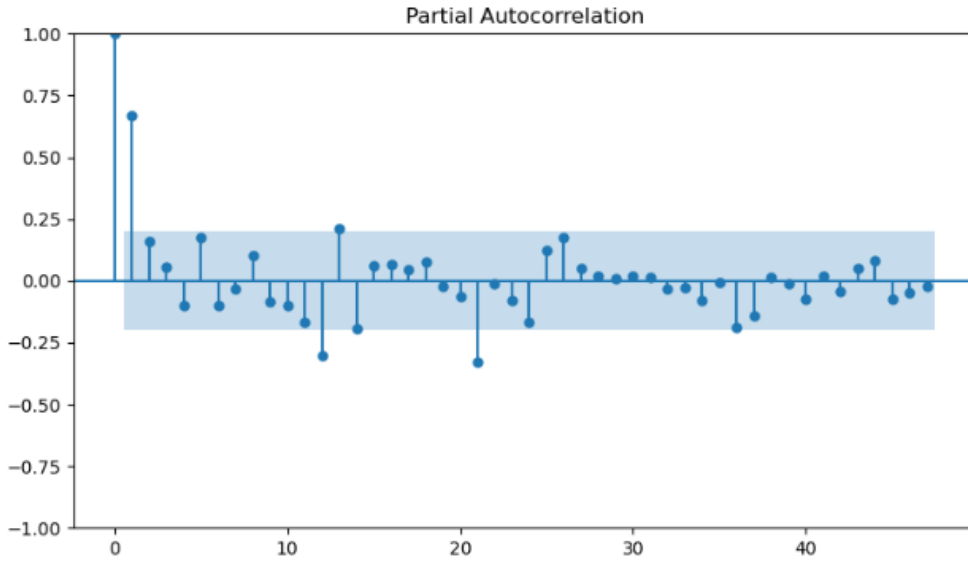
ADF testi, serinin durağan olmadığını, birim kök barındırdığını gösterirken, KPSS testi serinin durağan olduğunu ifade etmektedir. Birim kök testlerinin birbiri ile çeliştiği noktada seriyi durağan kabul etmek doğru sonuçlar vermemektedir. Seri üzerinde fark alma işlemi gerçekleştirilerek, serinin birim kökten arındırılması gerekmektedir.

Yılmaz çalışmasında, hem trend hem mevsimsellik birlikte varsa önce mevsimsellik için, sonra trend için fark alınmasının gerekliliğini vurgulamıştır (K. Yılmaz, 2021). Serinin ayrıştırma grafiği, ACF ve PACF grafiklerinden de gözlemlendiği üzere seride mevsimsellik söz konusudur. Mevsimsellik için fark alma işlemi gerçekleştirilmiştir.

Ham veriler ile oluşturulan ACF ve PACF grafiklerine bakılarak mevsimsellikte periyot 12 olarak belirlenmiştir. 12 periyodunda, 1. dereceden mevsimsel fark alınmıştır. Mevsimsel fark alınması sonucunda oluşan ACF ve PACF grafikleri Şekil 3.18 ve Şekil 3.19’da gösterilmiştir.



Şekil 3.18. 1.dereceden mevsimsel farkı alınmış serinin ACF grafiği.



Şekil 3.19. 1.dereceden mevsimsel farkı alınmış serinin PACF grafiği.

ACF grafiğindeki ilk 4 gecikmenin sınırlar dışında kalması serinin trende sahip olduğunun göstergesidir. Birim kök testlerini de seri üzerinde uygulayarak birim kök testleri ile seri durağanlaşana kadar test etmek gerekmektedir. Şekil 3.20 ve Şekil 3.21’de ADF ve KPSS birim kök testleri sonuçları verilmiştir. KPSS testi, serinin durağan olduğunu gösterirken, ADF testi, serinin birim kök barındırdığını göstermektedir.

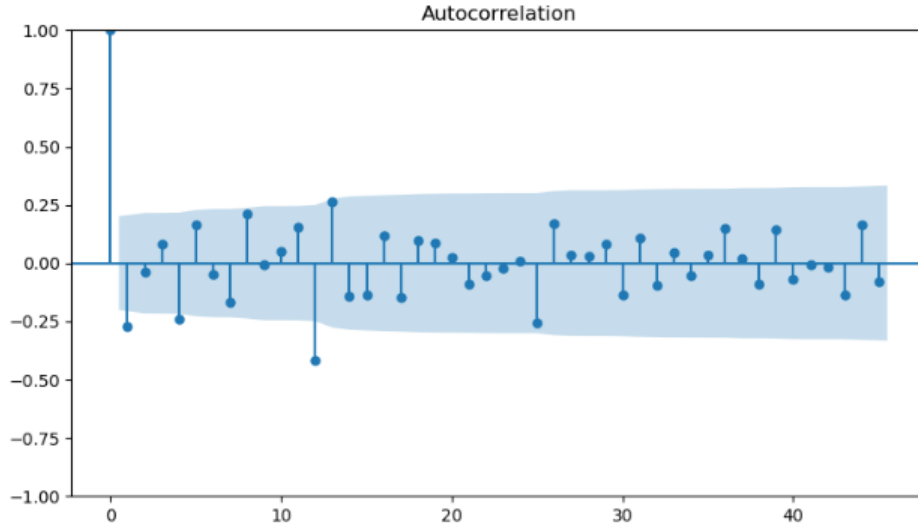
Augmented Dickey-Fuller Results	
Test Statistic	-2.517
P-value	0.111
Lags	12
Trend: Constant	
Critical Values: -3.51 (1%), -2.90 (5%), -2.59 (10%)	
Null Hypothesis: The process contains a unit root.	
Alternative Hypothesis: The process is weakly stationary.	

Şekil 3.20. 1.dereceden mevsimsel farkı alınmış verinin ADF birim kök testi.

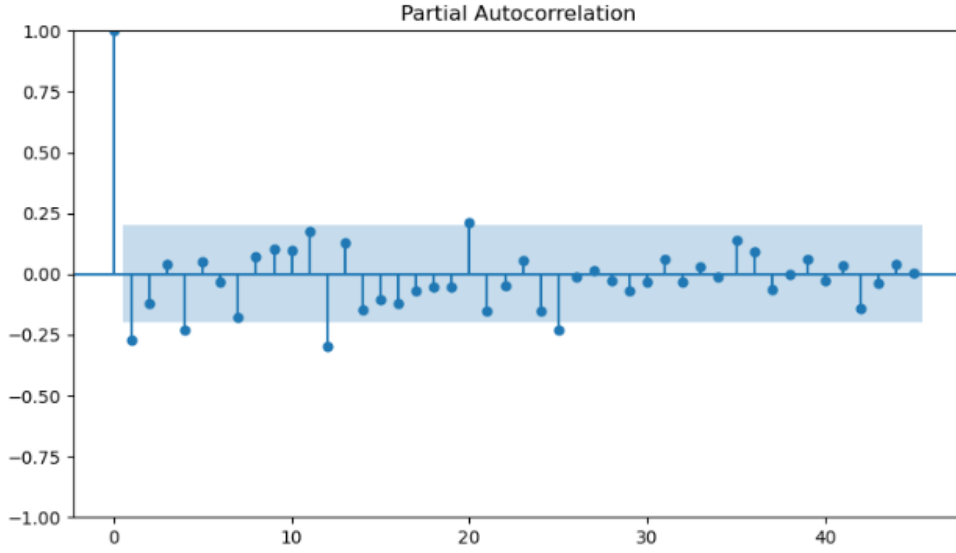
KPSS Stationarity Test Results	
Test Statistic	0.126
P-value	0.474
Lags	5
Trend: Constant	
Critical Values: 0.74 (1%), 0.46 (5%), 0.35 (10%)	
Null Hypothesis: The process is weakly stationary.	
Alternative Hypothesis: The process contains a unit root.	

Şekil 3.21. 1.dereceden mevsimsel farkı alınmış verinin KPSS birim kök testi.

Mevsimsel fark alındıktan sonra, birim kök testlerinin yanı sıra, ACF grafiğine bakıldığında seride trend olduğunun gözlemlenmesi sebebi ile, serinin 1.dereceden mevsimsel olmayan farkı alınması gerekmektedir. Serinin 1.dereceden mevsimsel olmayan farkı, alınmış verinin ACF ve PACF grafikleri Şekil 3.22 ve Şekil 3.23'te verilmiştir.



Şekil 3.22. 1. dereceden mevsimsel olmayan farkı alınmış verinin ACF grafiği.



Şekil 3.23. 1. dereceden mevsimsel olmayan farkı alınmış verinin PACF grafiği.

Serinin 1.dereceden mevsimsel olmayan farkı alındıktan sonra ADF ve KPSS birim kök testleri yapılmıştır. Sonuçları, Şekil 3.24 ve Şekil 3.25'te gösterilmiştir.

Augmented Dickey-Fuller Results	
Test Statistic	-2.606
P-value	0.092
Lags	11
Trend: Constant	
Critical Values: -3.51 (1%), -2.90 (5%), -2.59 (10%)	
Null Hypothesis: The process contains a unit root.	
Alternative Hypothesis: The process is weakly stationary.	

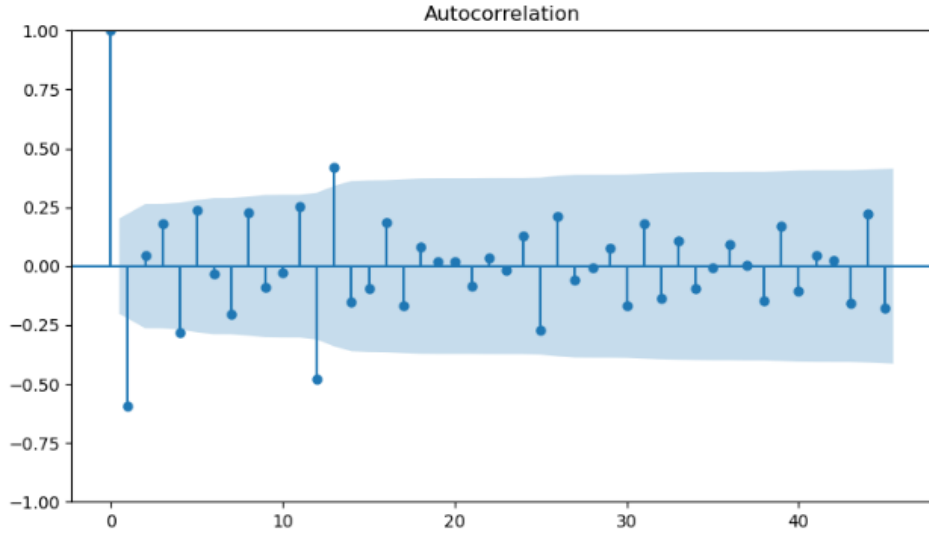
Şekil 3.24. 1.dereceden mevsimsel olmayan farkı alınmış verinin ADF birim kök testi.

KPSS Stationarity Test Results	
Test Statistic	0.119
P-value	0.499
Lags	7
Trend: Constant	
Critical Values: 0.74 (1%), 0.46 (5%), 0.35 (10%)	
Null Hypothesis: The process is weakly stationary.	
Alternative Hypothesis: The process contains a unit root.	

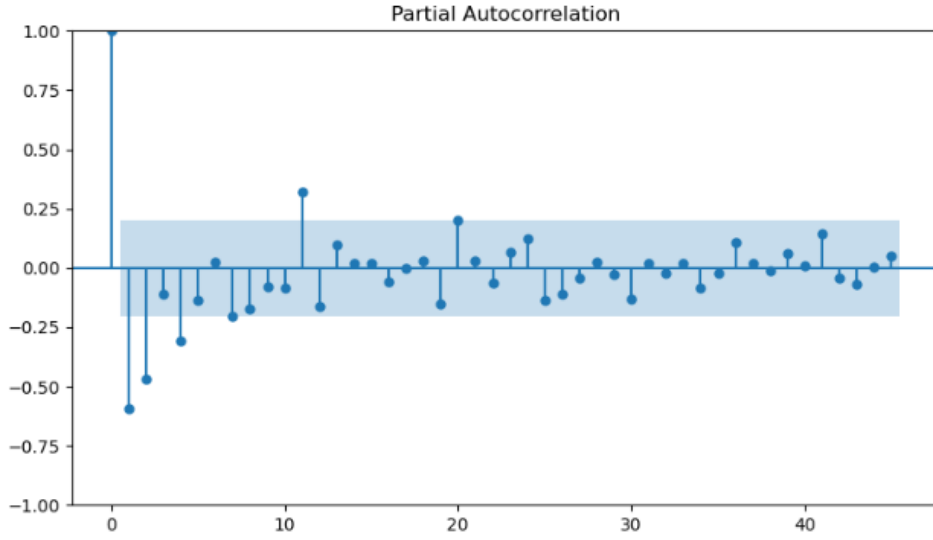
Şekil 3.25. 1.dereceden mevsimsel olmayan farkı alınmış verinin KPSS birim kök testi.

KPSS testi, serinin durağan olduğunu gösterirken, ADF testi, serinin birim kök barındırdığını göstermektedir. 1.dereceden farkı alınan serinin durağan durumuna ulaşamaması sonucunda serinin 2.dereceden farkını almak gerekmektedir.

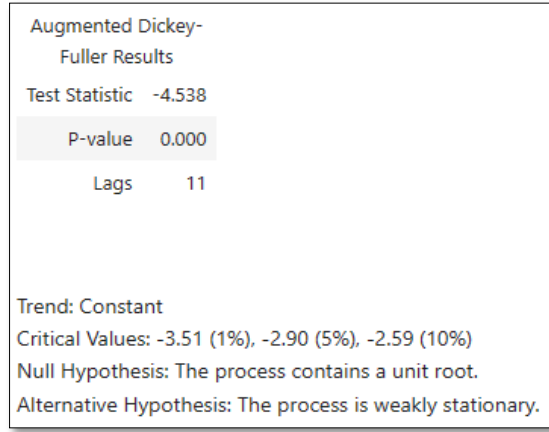
2.dereceden farkı alınan serinin, ACF ve PACF grafikleri Şekil 3.26 ve Şekil 3.27’de, ADF birim kök testi Şekil 3.28’de gösterilmiştir.



Şekil 3.26. 2. dereceden mevsimsel olmayan farkı alınmış ve durağanlaşmış verinin ACF grafiği.

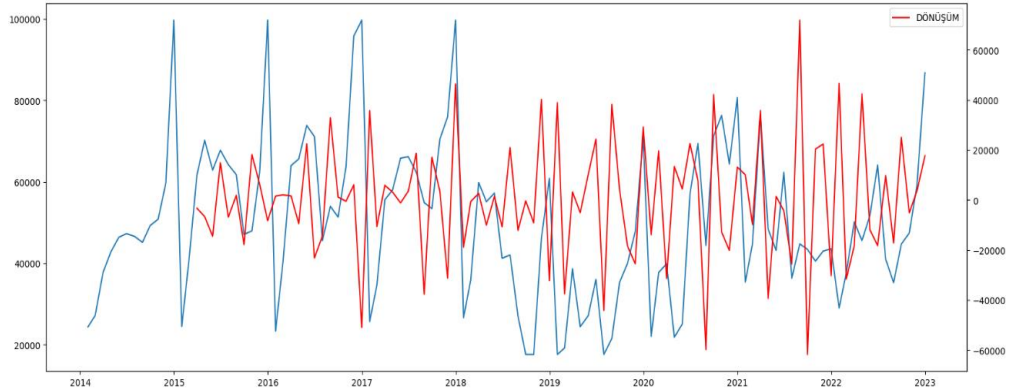


Şekil 3.27. 2. dereceden mevsimsel olmayan farkı alınmış ve durağanlaşmış verinin PACF grafiği.



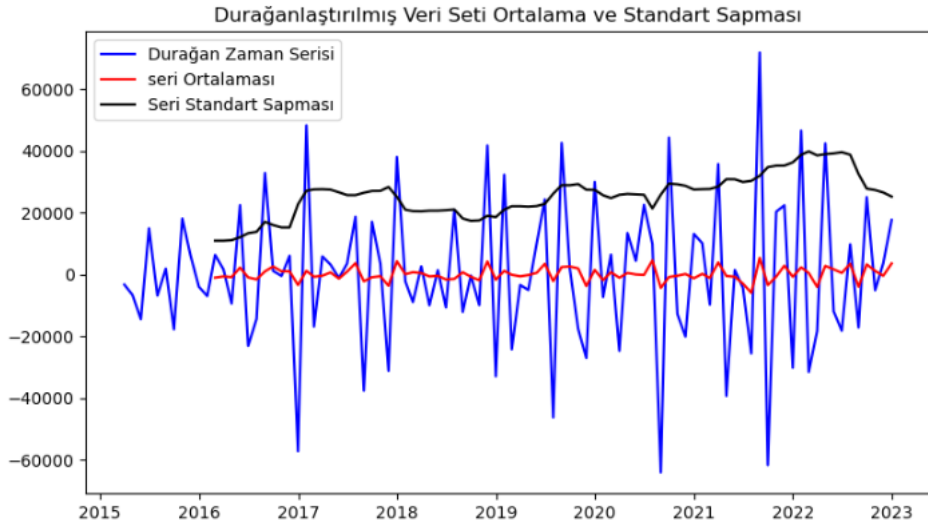
Şekil 3.28. 2.dereceden mevsimsel olmayan farkı alınmış ve durağanlaşmış verinin ADF birim kök testi.

2. dereceden farkı alınan seri, durağanlaşmıştır ve analiz yapılmaya, model kurulmaya hazır hale getirilmiştir. Durağanlaştırma işlemi yapılan serinin orijinal ve dönüşüm yapılması ile oluşan grafiği Şekil 3.29’da birlikte gösterilmiştir.



Şekil 3.29. Orijinal ve durağanlaştırılmış veri setinin zaman serisi grafiği.

Durağanlaştırılmış zaman serisinin ortalaması sabit bir ortalamaya gelmiştir. Durağanlaştırılmış veri özellik grafiği Şekil 3.30’da gösterilmiştir.



řekil 3.30. Durađanlařtırılmıř veri zellik grafiđi.

4. UYGULAMA

Bu bölümde yöntemler ile uygulamalar gerçekleştirilmiştir. Sonrasında bu algoritma sonuçları, performans değerlendirme ölçütlerinden Determinasyon Katsayısı (R^2) ve Ortalama Kare Hata Karekökü (RMSE) ile karşılaştırılmıştır.

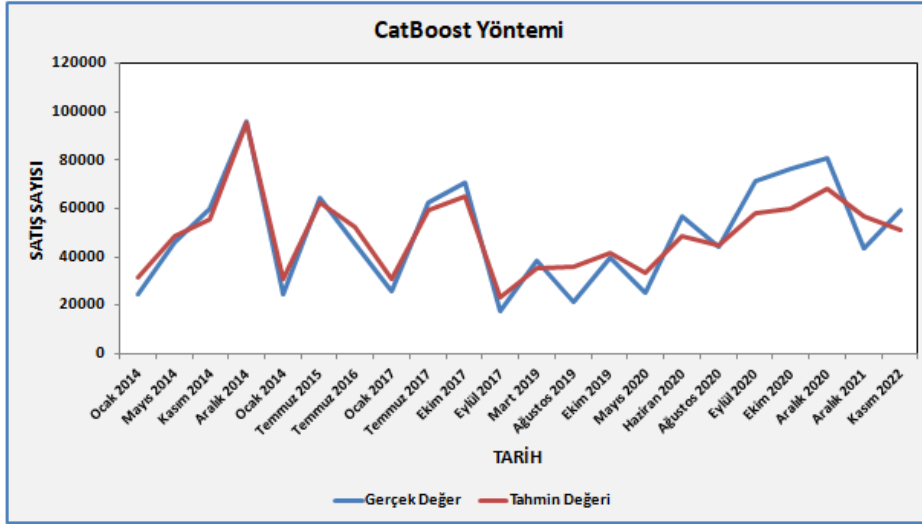
4.1. Tahmin Yöntemleri ve Uygulama

Makine öğrenme yöntemlerinden CatBoost, Gradient Boosting, Random Forest, Destek Vektör Regresyonu algoritmaları, Yapay Sinir Ağları ve Zaman Serileri Analizi ile talep tahmin uygulamaları gerçekleştirilmiş ve sonuçları gösterilmiştir.

4.1.1. CatBoost ile talep tahmini

Bu çalışmada, Catboost kütüphanesi üzerinden CatBoostRegressor algoritması ile işlem yapılmıştır. CatBoost ile temel model kurulmuştur. Sonrasında modelin düzenleme işlemi gerçekleştirilmiştir. CatBoost algoritmasında, iterations, learning rate, max_depth parametre kombinasyonlarının en iyi performansları istenmiştir. Cross validation katsayısı da 5 olarak seçilmiştir. Iterations parametresi, ağaç sayısını ifade etmektedir. Learning rate parametresi 0-1 arası değer alır ve kurulan ağaçları ölçeklendirir. Küçük olması daha iyi sonuç verirken, optimum seviyede olmaması aşırı öğrenmeye neden olabilmektedir. Max_depth ise ağacın dallarının uzamasının ölçeğidir. Ağacın derinliği olarak da ifade edilmektedir.

Düzenlenmiş model üzerinden, modelin daha önce görmediği test verileri ile tahminleme yapılması istenmiştir. Model tahmin verileri ile gerçek veriler karşılaştırılarak performans analizi gerçekleştirilmiştir. Veriler standardizasyon işlemi gerçekleştirilmiş değerlerden dönüştürülmüş ve performans analizi için performans değerlendirme ölçütlerinden RMSE (Ortalama Kare Hata Karekökü) ve R^2 (Determinasyon Katsayısı) kullanılmıştır. Sonuçta RMSE değeri 8156,15, R^2 değeri ise 0,85 olarak bulunmuştur. Gerçek ve tahmin değerleri karşılaştırmalı olarak Şekil 4.1’de verilmiştir.

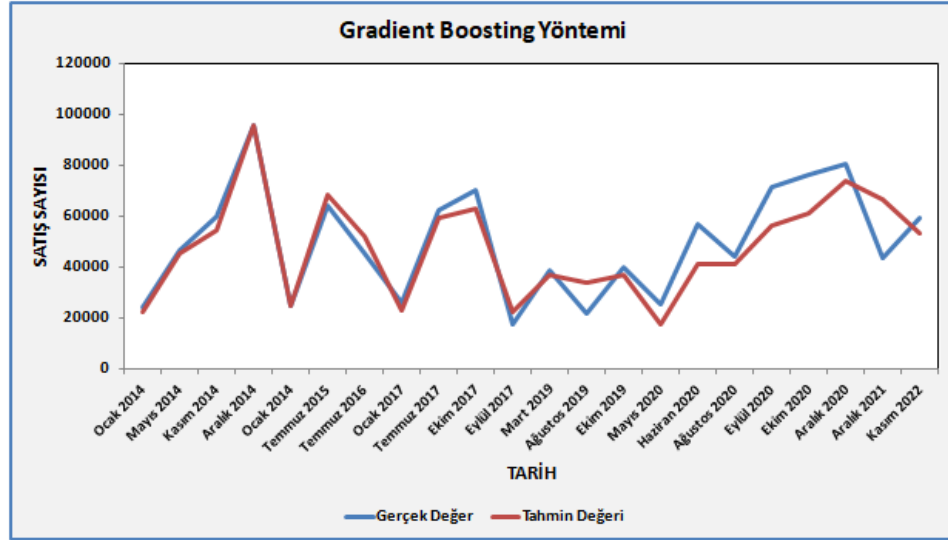


Şekil 4.1. CatBoost ile tahmin değerleri ve gerçek otomobil satış sayıları.

4.1.2. Gradient boosting ile talep tahmini

Bu çalışmada, Sklearn kütüphanesi üzerinden Gradient Boosting Regressor algoritması ile işlem yapılmıştır. Gradient Boosting Regressor ile temel model kurulmuştur. Sonrasında modelin düzenleme işlemi gerçekleştirilmiştir. Gradient Boosting Regressor algoritmasında, learning rate, max_depth, n_estimators ve subsample parametre kombinasyonlarının en iyi performansları istenmiştir. n_estimators parametresi, ağaç sayısını ifade etmektedir. Learning rate parametresi 0-1 arası değer alır ve kurulan ağaçları ölçeklendirir. Küçük olması daha iyi sonuç verirken, optimum seviyede olmaması aşırı öğrenmeye neden olabilmektedir. Max_depth ise ağacın dallarının uzamasının ölçeğidir. Ağacın derinliği olarak da ifade edilmektedir. Subsample, eğitim örneklerinin satır oranı olarak adlandırılmaktadır. 0-1 arasında değer almaktadır. Değerinin 0,5 olması aşırı öğrenmeyi engellemektedir. Cross validation katsayısı da 10 olarak seçilmiştir.

Düzenlenmiş model üzerinden, modelin daha önce görmediği test verileri ile tahminleme yapılması istenmiştir. Model tahmin verileri ile gerçek veriler karşılaştırılarak performans analizi gerçekleştirilmiştir. Veriler standardizasyon işlemi gerçekleştirilmiş değerlerden dönüştürülmüş ve performans analizi için performans değerlendirme ölçütlerinden Sonuçta RMSE değeri 8840,93, R^2 değeri ise 0,83 olarak bulunmuştur. Gerçek ve tahmin değerleri karşılaştırmalı olarak Şekil 4.2’de verilmiştir.

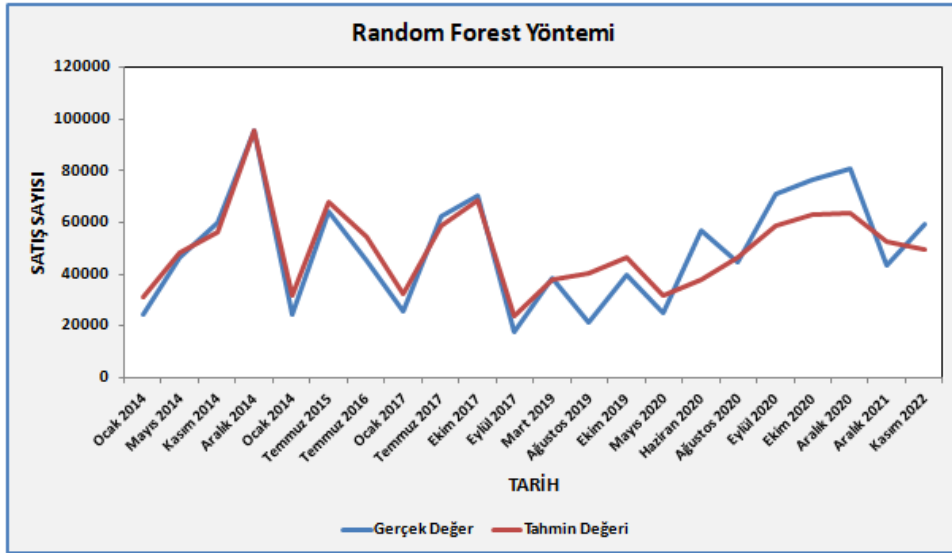


Şekil 4.2. Gradient boosting ile tahmin değerleri ve gerçek otomobil satış sayıları.

4.1.3. Rastgele ormanlar (Random Forest) ile talep tahmini

Bu çalışmada, Sklearn kütüphanesi üzerinden RandomForestRegressor algoritması ile işlem yapılmıştır. RandomForestRegressor ile temel model kurulmuştur. Sonrasında modelin düzenleme işlemi gerçekleştirilmiştir. Random Forest algoritmasında, max_depth, max_features, n_neighbors parametre kombinasyonlarının en iyi performansları istenmiştir. Cross validation katsayısı da 10 olarak seçilmiştir. Max_depth parametresi, ağacın dallarının uzamasının ölçeğidir. Max_features parametresi kullanılacak en fazla özellik sayısını belirtmektedir. n_neighbors parametresi ağaç sayısını belirtmektedir.

Düzenlenmiş model üzerinden, modelin daha önce görmediği test verileri ile tahminleme yapılması istenmiştir. Model tahmin verileri ile gerçek veriler karşılaştırılarak performans analizi gerçekleştirilmiştir. Veriler standardizasyon işlemi gerçekleştirilmiş değerlerden dönüştürülmüş ve performans analizi için performans değerlendirme ölçütlerinden RMSE(Ortalama Kare Hata Karekökü) ve R^2 (Determinasyon Katsayısı) kullanılmıştır. Sonuçta RMSE değeri 9329,12, R^2 değeri ise 0,81 olarak bulunmuştur. Gerçek ve tahmin değerleri karşılaştırmalı olarak Şekil 4.3'te verilmiştir.

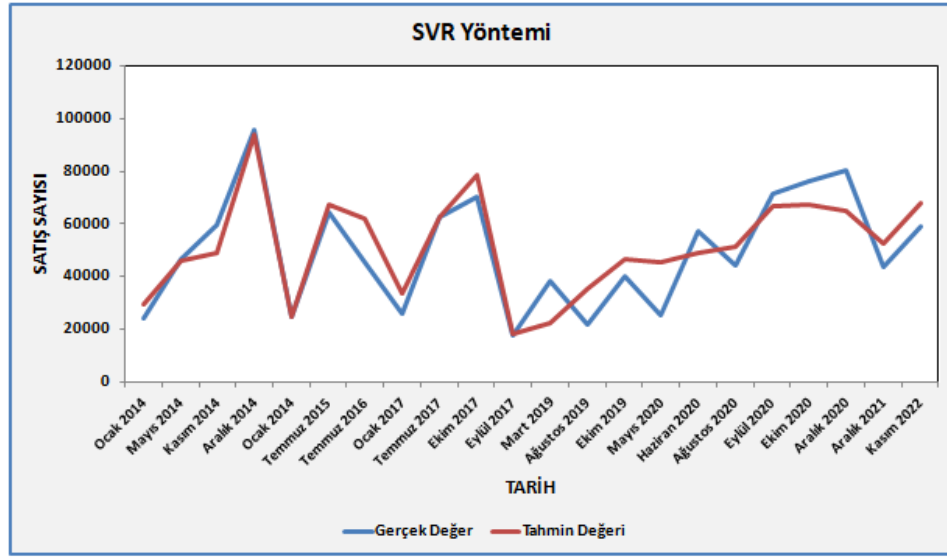


Şekil 4.3. RF ile tahmin değerleri ve gerçek otomobil satış sayıları.

4.1.4. Destek vektör regresyonu (SVR) ile talep tahmini

Bu çalışmada, Sklearn kütüphanesi üzerinden SVR algoritması ile işlem yapılmıştır. Veriler içe alındıktan sonra, SVR ile temel model kurulmuştur. Temel modelde test hataları hesaplanmış ve modelin düzenleme aşamasına geçilmiştir. Sklearn kütüphanesi üzerinden GridSearchCV nesnesi içe alınmıştır. Amaç optimal hiperparametreleri bulabilmektir. SVR algoritmasında, C parametre kombinasyonlarının en iyi performansları istenmiştir. C parametresi, destek vektörleri arasında kalan alanı ifade etmek için kullanılan bir parametredir. C, ceza katsayısının ifade eder ve hataları sınırlandırmak için kullanılmaktadır. Çok yüksek olması overfite (aşırı öğrenme) neden olabilmektedir. Cross validation katsayısı da 10 olarak seçilmiştir.

Düzenlenmiş model üzerinden, modelin daha önce görmediği test verileri ile tahminleme yapılması istenmiştir. Düzenlenmiş model üzerinden model tahmin verileri ile gerçek veriler karşılaştırılarak performans analizi gerçekleştirilmiştir. Veriler standardizasyon işlemi gerçekleştirilmiş değerlerden dönüştürülmüş ve performans değerlendirilmesi yapılmıştır. Performans değerlendirme ölçütlerinden RMSE (Ortalama Kare Hata Karekökü) ve R^2 (Determinasyon Katsayısı) ile model performans analizinde sonuç olarak RMSE değeri 9763,45, R^2 değeri ise 0,79 olarak bulunmuştur. Gerçek ve tahmin değerleri karşılaştırmalı olarak Şekil 4.4'te verilmiştir.



Şekil 4.4. SVR ile tahmin değerleri ve gerçek otomobil satış sayıları.

4.1.5. Yapay sinir ağları (YSA) ile talep tahmini

Bu çalışmada, Yapay Sinir Ağları ile talep tahmini için MATLAB paket programı kullanılmıştır. LOF yöntemi ile aykırı değer analizi gerçekleştirilen ve Z-Score yöntemi ile standartlaştırılan, Ocak 2014 – Aralık 2022 tarihleri arasındaki aylık veriler programda içe aktarılmıştır. YSA uygulamasında, Matlab programında NFTOOL Toolbox kullanılarak tahmin sonuçları üretilmiştir. NFTOOL kullanılmasının sebebi, verilen yüzdeliklere göre eğitim ve test verilerini algoritmanın belirlemesidir.

4.1.5.1. YSA parametrelerinin belirlenmesi

NFTOOL algoritmasında, yöntemin performansı yüksek bir model oluşturabilmesi için belirlenmesi gereken farklı parametreler mevcuttur. Deneme yapmak için parametreler belirlenirken literatürdeki mevcut çalışmalar dikkate alınmış ve parametre kombinasyonları oluşturulmuştur. Veri setinden eğitim, doğrulama, test verilerinin oranları ile ağıın nöron sayısı için farklı parametreler ile uygulama gerçekleştirilmiş ve sonuçları kaydedilmiştir. En iyi performans gösteren parametreler ağıın eğitimi için kullanılmıştır. Ağı üzerinde, veriler ile deneme yapılan parametreler Tablo 4.1’de verilmiştir.

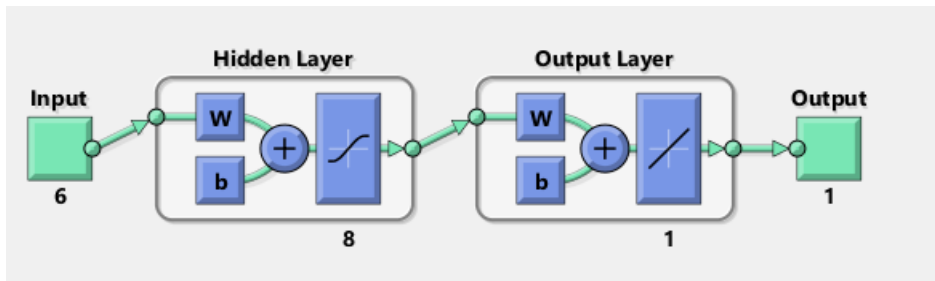
Tablo 4.1. YSA parametre kombinasyonları.

Parametreler	1	2	3	4
Nöron Sayısı	6-8-10	6-8-10	6-8-10	6-8-10
Eğitim Seti(%)	80	70	70	65
Doğrulama Seti(%)	10	15	10	15
Test Seti(%)	10	15	20	20

Denemeler sonucunda, en iyi performansı gösteren ağ parametreleri, 8 nöron sayısı ile uygulanmış 4. Denemedeki gibidir. Verilerin, %65'i eğitim, %15'i doğrulama, %20'si test olarak; 108 verinin 70 tanesi ağın eğitimi (training), 16 tanesi ağ genellemesini ölçmek ve genellenmenin gelişmesi durduğunda ağı durdurmak (validation), 22 tanesi de eğitim sonrasında ağ performansının bağımsız ölçüsünü sağlamak için (testing) kullanılmıştır.

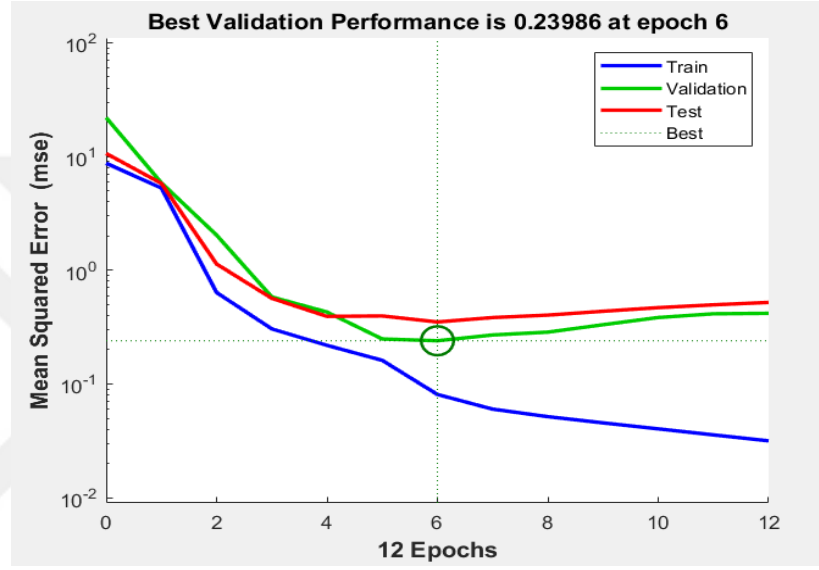
4.1.5.2. YSA ağ tasarımı

Bu çalışmada tek katmandan oluşan çok katmanlı algılayıcılar kullanılmıştır. Tasarlanan ağda bulunan yapay sinir ağı hücreleri için birleştirme fonksiyonu olarak toplama fonksiyonu kullanılmıştır. Aktarma fonksiyonu olarak ara katmanda bulunan yapay sinir hücreleri için Hiperbolik Tanjant fonksiyonu, çıktı katmanında bulunan yapay sinir hücreleri içinse doğrusal fonksiyon kullanılmıştır. Aktarma fonksiyonu ile girdi değerlerine tanjant fonksiyonu uygulanması sonucunda çıktı elde edilmektedir (Güven, 2020). Ağın eğitiminde eğitmenli öğrenme dizisi ile hatayı geriye doğru yayan Levenberg-Marquardt Geri Yayılım Algoritması kullanılmıştır. Tasarlanan ağın yapısı Şekil 4.5'te gösterilmiştir.

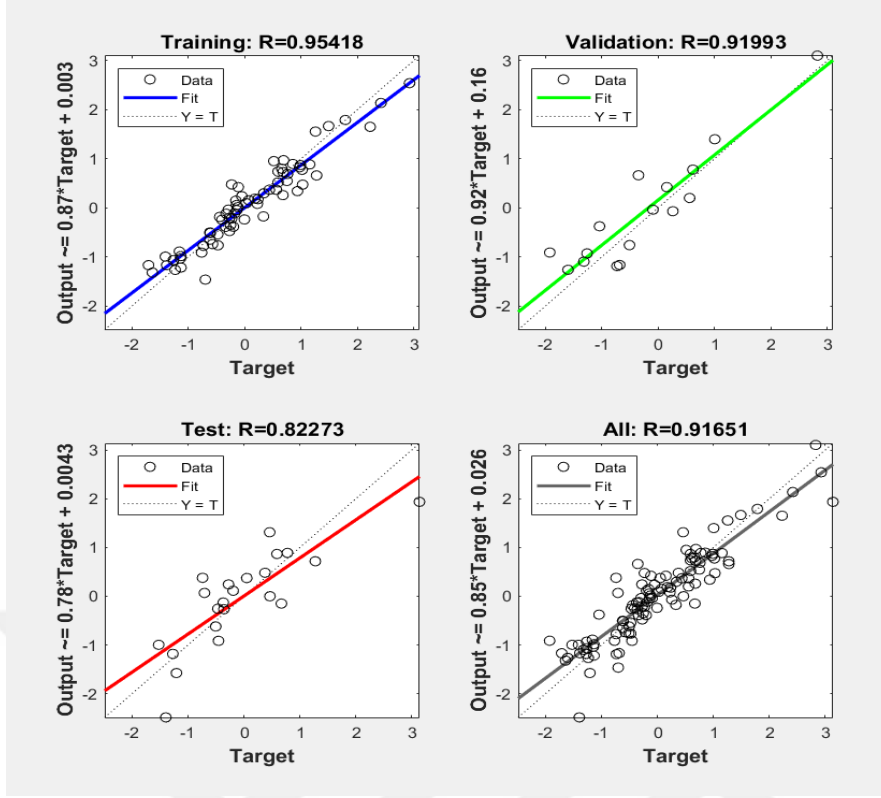
**Şekil 4.5.** YSA ağ yapısı

4.1.5.3. YSA ađ eđitimi

Yapay sinir ađı modelinin eđitim, test ve dođrulama sũreçlerine ait Őekil 4.6'da gũsterilmiő olan performans eđrisi oluőmuőtur. Nihai ortalama kare hatasının kũçük olması, test seti hatası ve dođrulama seti hatasının benzer ȳzelliklere sahip olması, en iyi dođrulama performansının gerçekteőtiđi yerde ȳnemli bir aőırı uyum meydana gelmemiő olması iyi bir performans grafiđinin gũstergeleridir (Verileri Sıđ Bir Sinir Ađıyla Uydurmak - MATLAB ve Simulink, 2023). Eđitim sonrası, tahmin ve gerçekte deđerler iliőkisi Őekil 4.7'de gũsterilmiőtir.



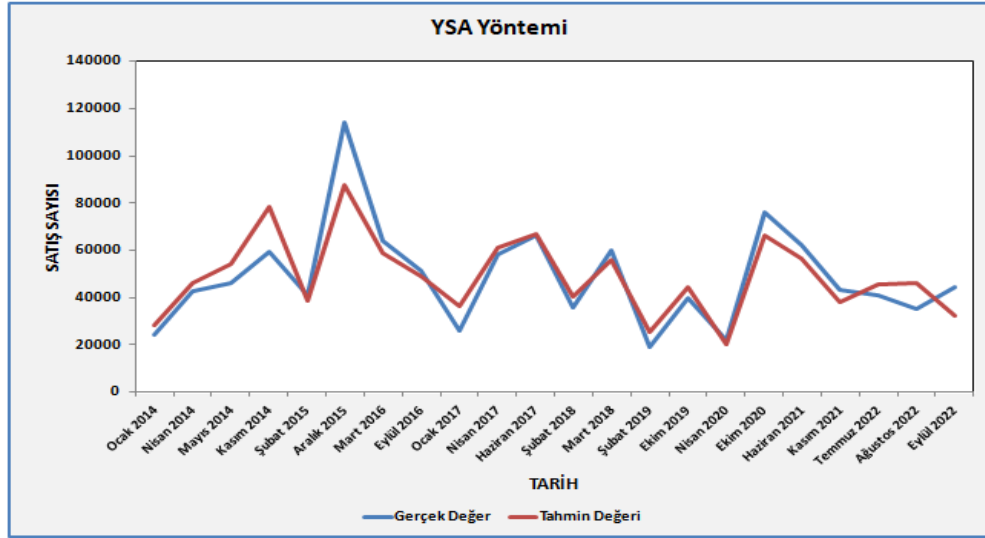
Őekil 4.6. YSA performans grafiđi



Şekil 4.7.YSA tahmin ve gerçek değerler ilişkisi

4.1.5.4. YSA ağ performans sonuçları

Ağ eğitiminden sonra oluşan tahmin ve gerçek karşılaştırmalı gözlem değerleri Şekil 4.8’de gösterilmiştir. Model tahmin verileri ile gerçek veriler karşılaştırılarak performans analizi gerçekleştirilmiştir. Veriler standardizasyon işlemi gerçekleştirilmiş değerlerden dönüştürülmüş ve performans analizi için performans değerlendirme ölçütlerinden RMSE(Ortalama Kare Hata Karekökü) ve R^2 (Determinasyon Katsayısı) kullanılmıştır. Sonuçta RMSE değeri, 9173,02, R^2 değeri ise 0,81 olarak hesaplanmıştır.



Şekil 4.8. YSA ile tahmin değerleri ve gerçek otomobil satış sayıları.

4.1.6. SARIMA modeli ile talep tahmini

Sarima modelinin belirlenmesi: Zaman serileri analizinde iyi bir model belirlenmesi ve modelin veri setine uyması önemli olmaktadır. Modelin iyi sonuçlar vermesi için modelin de doğru belirlenmesi gerekmektedir. Oluşturulan modelin doğruluğu ve veri setine uygun olup olmama durumunun test edilmesi gerekmektedir (Çelik, 2013).

Box-Jenkins yöntemlerinde durağanlık şartı aranmaktadır. Çelik (2013)'in, çalışmasında bahsettiği gibi zaman serileri ile analizin gerçekleşebilmesi için serinin durağan olması ve model ile oluşan hata terimlerinin beyaz gürültüye (white noise) sahip olması gerekmektedir. Beyaz gürültü, çoğunluk olarak otokorelasyon göstermeyen zaman serileri olarak adlandırılmaktadır. Geçmiş veriler ile arasında doğrusal bir ilişki göstermeyen, rastgele dalgalanmaları olan zaman serisi olarak değerlendirilmektedir (Demir, 2018). Oluşturulan modelin yeterli olup olmadığına karar verebilmek için model analiz kısmında beyaz gürültüye sahip olup olmadığına bakılması gerekmektedir. Beyaz gürültü sürecine sahip olması modelin doğruluğunun bir göstergesi olmaktadır (Güneş, 2010).

Ljung-Box testinde;

H_0 : Veriler bağımsız olarak dağılmıştır, aralarında korelasyon bulunmamaktadır.

H_1 : Verilerin bağımsız olarak dağılmamaktadır, zaman serisi korelasyon sergilemektedir.

ARIMA model analizlerinde sıklıkla görülen alan Ljung-Box testi, modelin kalıntılarının otokorolesyona sahip olmadığının göstergesi olarak kullanılmaktadır (Özer ve Yavuz, 2014).

Çalışmada, birim kök testleri bölümünde gösterildiği gibi durağanlığın sağlanması için 1 kez mevsimsel fark 2 kez ise mevsimsel olmayan fark alınmış ve durağanlık sağlanmıştır. SARIMA parametrelerinden d parametresi 2 değerini alırken, D ise 1 değerini almıştır.

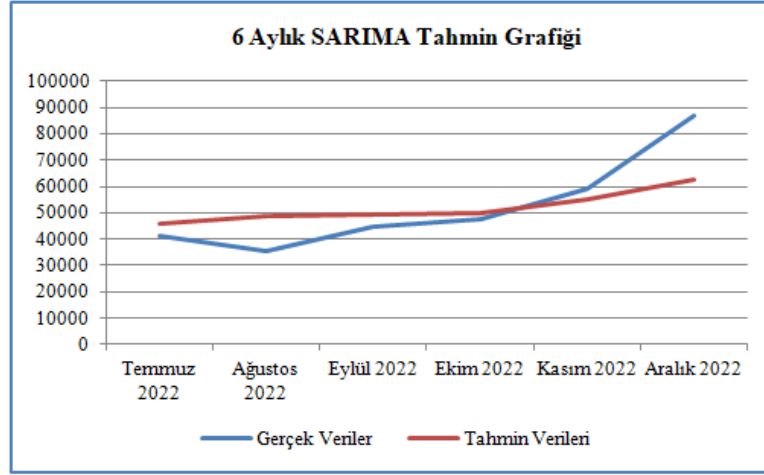
2014-2022 yılları arasındaki aylık otomobil satış verileri kullanılmıştır. Veri seti 108 adet satış verisinden oluşmaktadır. Zaman serileri analizi SARIMA modeli ile uygulama gerçekleştirilirken farklı modeller denenmiştir.

Auto_arima fonksiyonu ile model tahmini: İlk olarak veri seti 102 adet eğitim, 6 adet test verisi olarak ayrılmış ve 6 aylık bir tahmin modeli oluşturulmaya çalışılmıştır. Test verileri veri setinin son dönemini kapsamaktadır. Python programında auto_arima fonksiyonu ile sarıma modeli için en iyi parametrelerin bulunması amaçlanmıştır. Auto_arima ile model kombinasyon performanslarını analiz edebilmek için akaike bilgi kriteri (AIC) ölçüt olarak kullanılmıştır. AIC ölçütü, model karşılaştırmalarında güvenilir bir istatistiksel ölçüdür. Akaike Bilgi Kriteri, modele dahil edilen değişkenlerin yükselmesine sınırlama getirmektedir (Ucal, 2006). AIC bilgi ölçütünün daha küçük olduğu modellerin daha yüksek performans göstermesi beklenmektedir. AIC denklemi, Denklem (4.1)'de verilmiştir.

$$AIC = 2k - 2\ln(L) \quad (4.1)$$

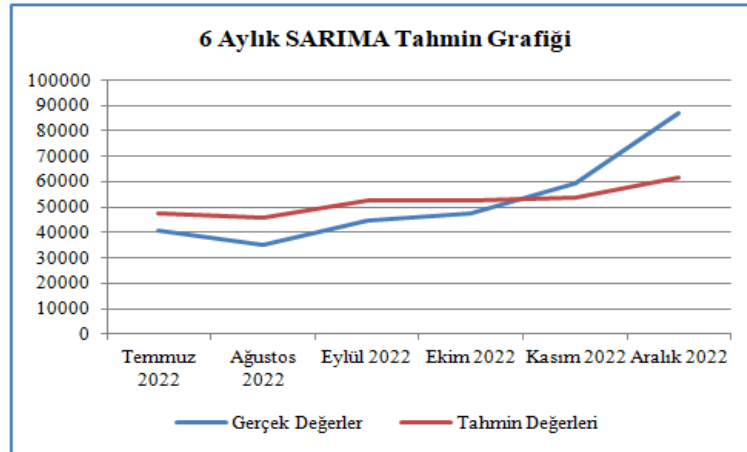
Denklem (4.1)'de; k, modeldeki parametre sayısını; L ise model olabilirlik fonksiyonu maksimum değerini temsil etmektedir.

6 aylık satış tahmini için, farklı d ve D parametreleri üzerinde deneme yapılmıştır. Periyot sayısı (m) 12, fark dereceleri (d ve D) 1 parametreleri ile beraber en düşük AIC değeri 1950,03 olarak model oluşturulmuştur. Ljung-Box testi ile test edilmiş ve hataların beyaz gürültülü olduğu tespit edilmiştir. ARIMA (0,1,1)(1,1,1)₁₂ parametreleri ile 11733,96 RMSE, 0,52 R² değerine ulaşılmıştır. Tahmin ve gerçek değerler grafiği Şekil 4.9'da verilmiştir.



Şekil 4.9. Optimizasyon ile otomobil satış tahmin grafiği.

Auto_arima fonksiyonu ve çapraz doğrulama (cross validation) ile model tahmini: Python sklearn kütüphanesi içerisinde Time Series Split komutu ile cross validation zaman serisi üzerinde gerçekleştirilmiş ve n (bölünme sayısı) 15 olarak alınmıştır. 6 aylık bir tahmin işlemi gerçekleştirilmiştir. Durağanlaşma fark dereceleri bilindiğinden fark derecesi (d) 2, mevsimsel fark derecesi (D) 1 olarak alınmıştır. Ljung-Box testi ile test edilmiş ve hataların beyaz gürültülü olduğu tespit edilmiştir. Auto_arima ile en iyi performans gösteren parametre belirlenmiş ve RMSE değeri 12195,02, R^2 değeri 0,49 olarak bulunmuştur. Tahmin ve gerçek değerler grafiği Şekil 4.10'da verilmiştir.



Şekil 4.10. Çapraz doğrulama ile otomobil satış tahmini grafiği.

4.1.7. Yöntemlerin değerlendirilmesi

Tahmin yönteminin değerlendirmesini yapabilmek için çok sayıda performans değerlendirme ölçütü kullanılabilmektedir. Tahmin hatalarının analiz edilmesi,

gerçek gözlem değeri ile tahmin edilen değer arasındaki farkın ölçülebilmesi, tahmin yöntemlerinin doğruluğunu değerlendirebilmek için önemli olmaktadır. Çalışmada, tahmin modelleri ile uygulama sonucunda modellerin göstermiş olduğu performans, RMSE ve R^2 performans değerlendirme ölçütleri ile karşılaştırılmıştır.

4.1.7.1. Determinasyon katsayısı (R^2)

Modelde kullanılan gerçek girdi verileri ve modelin ürettiği çıktı verileri arasındaki ilişkinin ölçülmesinde kullanılır. R^2 , değeri 1 olduğunda tahminin kusursuz olduğu anlamına gelirken, 1'e yakın değerlerin bulunması model performansının iyi olduğunu göstermektedir. Değer 0 olduğunda ise algoritmanın tüm çıktı değerlerine ortalama değeri verdiği anlamına gelmektedir. R^2 , negatif (-) bir değer ise, algoritmanın çıktı değerlerine çok kötü değerler verdiği ve yapılan analizde mutlak bir hatanın olduğu anlamına gelmektedir. Ayrıca, R^2 tek başına yeterli olmamakla beraber, aynı zamanda diğer performans ölçütleri (MSE, RMSE vb.) ile birlikte değerlendirmeye alınması gerekmektedir. R^2 , formülü Denklem (4.2)'de verilmiştir.

$$R^2 = 1 - \sum_i^n \frac{(y_{i\text{gerçek}} - y_{i\text{tahmin}})^2}{(y_{i\text{gerçek}} - \text{ortalama}(y_{i\text{gerçek}}))^2} \quad (4.2)$$

Denklem (4.2)'de;

R^2 , determinasyon katsayısını,

n, değişken içerisindeki veri sayısını,

$y_{\text{gerçek}}$, gerçek çıktı verilerini,

y_{tahmin} , tahmin çıktı verilerini,

ortalama $y_{\text{gerçek}}$, gerçek çıktı değişkenlerinin ortalamasını temsil etmektedir.

4.1.7.2. Ortalama kare hata karekökü (RMSE)

Model performansının ölçülmesinde kullanılan istatistiksel yöntemlerden biri de RMSE performans değerlendirme ölçütüdür. RMSE, modelin tahmin ettiği değerler ile veri setinin gerçek değerlerinin farkını değerlendirerek hata miktarını ölçmektedir. RMSE değerinin küçük olması model performansının daha iyi olduğu anlamına gelmektedir (Asil ve Alptekin, 2022). RMSE, tek bir yöntem değerlendirildiğinde yorum yapılabilir bir değerlendirme ölçütü değildir. Modellerin birlikte karşılaştırılmasında kullanılmaktadır. RMSE, formülü Denklem (4.3)'te verilmiştir.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_{i\text{tahmin}} - y_{i\text{gerçek}})^2}{n}} \quad (4.3)$$

Denklem (4.3)'te,

RMSE, ortalama kare hata karekökünü,

ygerçek, gerçek çıktı verilerini,

ytahmin, tahmin çıktı verilerini,

n, değişkendeki veri sayısını ifade etmektedir.

Gerçek değerler ve tahmin modelleri ile elde edilen tahmin değerleri analiz edilmiştir. Karşılaştırmalı olarak, performans değerleri Tablo 4.2'de gösterilmiştir.

Tablo 4.2. Model performans sonuçları.

Yöntemler	RMSE	R²
CatBoost	8156,15	0,85
Gradient Boosting	8840,93	0,83
Random Forest	9329,12	0,81
SVR	9763,45	0,79
YSA	9173,02	0,81
SARIMA (Parametre belirleme fonksiyonu ile 6 aylık)	11733,96	0,52
SARIMA (Parametre belirleme fonksiyonu ve çapraz doğrulama ile 6 aylık)	12195,02	0,49



5. SONUÇ VE ÖNERİLER

Otomobil satışlarında talep, farklı sebeplerle oluşan dalgalanmalardan etkilenmektedir. Bu sebeple değişken bir grafiğe sahiptir. 2019 yılında adını duyuran ve 2020 Mart ayında Türkiye’de etkisini göstermeye başlayan Covid-19 pandemisi, dünya genelinde krizlere sebep olmuştur. Pandemi etkisi ile artış gösteren çip krizi ve dünya genelindeki ekonomik dalgalanmalar otomotiv sektöründeki satışları etkilemiştir. Talepte farklı sebeplerde oluşan dalgalanmalardan kaynaklı talepleri tahmin etmek zorlaşmaktadır. Talep tahmininin bahsedilen koşullar altında daha güvenilir yapılabilmesi için geleneksel yöntemlerin dışında yapay zeka yöntemleri literatürde yerini almıştır.

Otomotiv sektörünün ülke ekonomisinde etkili bir sektör olmasının yanında, diğer sektörlerin de tedarikçisi konumunda bulunması sebebi ile otomotiv sektöründe taleplere doğru ve güvenilir cevap verilmesi bir gereklilik haline dönüşmektedir. Bu çalışmada da otomobil taleplerine en doğru ve güvenilir yaklaşım sergileyen algoritmanın bulunması amaçlanmıştır.

Bu çalışmada, makine öğrenmesi, yapay sinir ağları ve zaman serileri analizi ile 2014-2022 otomobil satış verileri kullanılarak talep tahmin analizi gerçekleştirilmiştir. Tahmin modelleri ile uygulama gerçekleştirildikten sonra modellerin göstermiş olduğu performans, performans değerlendirme ölçütlerinden RMSE ve R^2 ile analiz edilmiştir. Gerçek değerler ve tahmin modelleri ile elde edilen tahmin değerleri karşılaştırılmıştır. Uygulama sonucunda CatBoost algoritması 8156,15 RMSE, 0,85 R^2 ; Gradient Boosting algoritması 8840,93 RMSE, 0,83 R^2 ; Random Forest algoritması 9329,12 RMSE, 0,81 R^2 ; SVR algoritması 9763,45 RMSE, 0,79 R^2 ; YSA yöntemi 9173,02 RMSE, 0,81 R^2 ; zaman serileri analizi SARIMA yöntemlerinde parametre belirleme fonksiyonu ile oluşturulan 6 aylık tahmin periyoduna sahip model 11793,96 RMSE, 0,52 R^2 ; parametre belirleme fonksiyonu ve çapraz doğrulama ile oluşturulan 6 aylık tahmin periyoduna sahip model 12195,02, 0,49 R^2 performans değerlerine sahiptir.

Makine öğrenmesi algoritmaları ve yapay sinir ağı yöntemi ile oluşturulan modellerde yapılan tahminlerde uzun dönemli tahminlerin de başarılı sonuçlar verdiği gözlemlenmiştir. Buna karşılık zaman serileri analizi ile oluşturulan modellerde uzun dönemli tahminlerde başarılı sonuçlar vermediği görülmüştür. Zaman serileri analizi ile kısa dönemli tahmin çalışması gerçekleştirilmiştir.

Çalışmada kullanılan yöntemler R^2 değerlerine göre sıralandığında, CatBoost, Gradient Boosting, Random Forest, YSA, SVR, SARIMA; RMSE değerlerine göre sıralandıklarında CatBoost, Gradient Boosting, YSA, Random Forest, SVR ve SARIMA olarak sıralanmaktadır. Sonuç olarak, modelin oluşturduğu tahmin değerleri ve gerçek değerler karşılaştırıldığında en yüksek performansı gösteren model, makine öğrenmesi yöntemlerinden CatBoost algoritmasıdır.

Çalışmada, talep tahmini için oluşturulan modellerin daha geniş bir bakış açısıyla yorumlanabilmesi ve performanslarının arttırılarak karşılaştırılabilmesi için, daha fazla veriye sahip olan bir veri seti ve farklı yöntemler kullanılarak çalışma genişletilebilir.

KAYNAKLAR

- Akbaş, B. C. (2022). *Tedarik zinciri Yönetiminde Yapay Zeka Tabanlı Talep Tahmini: Bir Tekstil Firmasında Uygulama* [Yüksek Lisans Tezi]. Akdeniz Üniversitesi.
- Akdağ, S. (2023). *Spor Bilimlerinde Kullanılan Y-Denge Verilerinin Makine Öğrenimi Yöntemleri İle Analizi* [Yüksek Lisans Tezi]. Karabük Üniversitesi.
- Akyiğit, H. E., ve Taşçı, T. (2022). Sigortacılık Sektöründe Makine Öğrenmesi ile Müşteri Kaybı Analizi. *FBU-DAE* 2022, 2(1), 66-79.
- Akyurt, I. (2015). Talep Tahmininin Yapay Sinir Ağlarıyla Modellenmesi: Yerli Otomobil Örneği. *Ekonometri ve İstatistik Dergisi*, 23, 147-157.
- Altın, F. G., ve Çelik Eroğlu, Ş. (2020). Gri Tahmin ve Box-Jenkins Yöntemleri ile Antalya Limanı için Aylık Konteyner Talep Tahmini. *Mehmet Akif Ersoy Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 7(3), 540-562. <https://doi.org/10.30798/makuiibf.689532>
- Arslankaya, S. (2019). Bir Lojistik Firmasında Zaman Serileri Analizi ve Yapay Sinir Ağları ile Talep Tahmininin Karşılaştırılması. *4th International Symposium on Innovative Approaches in Engineering and Natural Sciences Proceedings*, 239-245. <https://doi.org/10.36287/setsoci.4.6.068>
- Asil, M., ve Alptekin, G. I. (2022). Pırlanta Fiyat Tahmini İçin Regresyon Modellerinin Karşılaştırmalı Analizi. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 11(4), 838-845.
- Aslan, S., ve Yıldız, T. (2022). Makine Öğrenmesinde Rastgele Oran ve Sıralı Küme Örneklemesi Yöntemlerinin Doğrusal Regresyon Modellerine Etkisi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 24(70), 29-36.
- Atabay, Ş., ve Dikmeoğlu, N. E. (2018). İnşaat Sektöründe Değer Mühendisliği Yöntemi ile Malzeme Seçimi. *Teknik Bilimler Dergisi*, 8(3), Article 3.
- Atan, S. (2016). Veri, Büyük Veri ve İşletmecilik. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 19(35), 137-154. <https://doi.org/10.31795/baunsobed.645312>
- Avni ES, H., Kalender, F., ve Hamzaçebi, C. (2014). Yapay Sinir Ağları İle Türkiye Net Enerji Talep Tahmini. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 29(3), 495-504. <https://doi.org/10.17341/gummfd.41725>
- Batırlık, S. N., Zeren, F., ve Genç, Y. G. (2023). Faiz Oranı, Döviz Kuru ve Ekonomik Büyümenin Otomobil Satışları Üzerindeki Etkisinin İncelenmesi: Türkiye Örneği. *Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 16(3), 862-869. <https://doi.org/10.25287/ohuiibf.1270906>

- Bayrakçı, H. C., Çiçekdemir, R. S., ve Özkahraman, M. (2021). Tarım Arazilerinde Harcanan Su Miktarını Yapay Zekâ Teknikleri Kullanarak Belirlenmesi. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 9(6), Article 6. <https://doi.org/10.29130/dubited.1015690>
- Bircan, H., ve Karagöz, Y. (2003). Box-Jenkins Modelleri ile Aylık Döviz Kuru Tahmini Üzerine Bir Uygulama. *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 6(2), 49-62.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Breunig, M., Kröger, P., Ng, R., ve Sander, J. (2000). LOF: Identifying Density-Based Local Outliers. İçinde *ACM Sigmod Record* (C. 29, s. 104). Proc. ACM SIGMOD 2000 Int. Conf. On Management of Data. <https://doi.org/10.1145/342009.335388>
- Brühl, B., Hülsmann, M., Borscheid, D., Friedrich, C. M., & Reith, D. (2009). A Sales Forecast Model for the German Automobile Market Based on Time Series Analysis and Data Mining Methods. İçinde P. Perner (Ed.), *Advances in Data Mining. Applications and Theoretical Aspects* (ss. 146-160). Springer. https://doi.org/10.1007/978-3-642-03067-3_13
- Ceyhan, N. (2019). *Akan Verilerde Aykırı Değer Tespiti Yaklaşımları* [Yüksek Lisans Tezi]. Ege Üniversitesi.
- Çekiç, E. (2023). *Increasing Firm's Efficiency with Machine Learning Algorithms: An Application in the Logistics Industry* [Yüksek Lisans Tezi]. Marmara Üniversitesi.
- Çelik, Ş. (2013). Zaman Serileri Analizi ve Trafik Kazası Verilerine Uygulanması. *Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 3(4), 43-51.
- Çemrek, F., ve Demir, Ö. (2021). Estimating CO2 Emission Time Series with Support Vector Machines Regression, Artificial Neural Networks, and Classic Time Series Analysis. *Turkish Journal of Forecasting*, 05(2), Article 2. <https://doi.org/10.34110/forecasting.1035912>
- Çetin, A. C. (2020). Türkiye' deki İthal Otomobil ve Hafif Ticari Araç Perakende Satışlarının Döviz Kuru ile İlişkisi. *Sinop Üniversitesi Sosyal Bilimler Dergisi*, 4(2), 401-432. <https://doi.org/10.30561/sinopusd.806649>
- Çetin, V., ve Yıldız, O. (2022). Veri Analizinde Veri Ön İşleme Teknikleri Üzerine Kapsamlı Bir İnceleme. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 28(2), 299-312. <https://doi.org/10.5505/pajes.2021.62687>
- Çınar, M. (2022). Fiyat, Gelir ve Faiz Oranlarının Konut Talebi Üzerindeki Etkisi: Panel Veri Yaklaşımı. *International Journal of Social Inquiry*, 15(2), Article 2. <https://doi.org/10.37093/ijsi.1095419>
- Çiftçi, S., ve Batur Sir, G. D. (2023). Acil Servise Başvuru Sayısının Zaman Serisi Analiz ve Makine Öğrenmesi Yöntemleri ile Tahmin Edilmesine Yönelik Bir Uygulama. *Pamukkale University Journal of Engineering Sciences*, 29(7), 667-679. <https://doi.org/10.5505/pajes.2022.18488>

- Çoban, F., ve Demir, L. (2021). Yapay Sinir Ağları ve Destek Vektör Regresyonu ile Talep Tahmini: Gıda İşletmesinde Bir Uygulama. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 23(67),327-338. <https://doi.org/10.21205/deufmd.2021236729>.
- Çobanoğlu, C., ve Gümrah, Ü. (2023). Düşük Nakit Akış Seviyelerinde İşletme Sermayesinin Kaynak Rolü ve Sabit Yatırımın Nakit Akış Duyarsızlığı. *Anadolu Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 24(1), 50-74. <https://doi.org/10.53443/anadoluibfd.1168094>
- Çuhadar, M., ve Kayacan, C. (2005). Yapay Sinir Ağları Kullanılarak Konaklama İşletmelerinde Doluluk Oranı Tahmini: Türkiye’deki Konaklama İşletmeleri Üzerine Bir Deneme. *Anatolia: Turizm Araştırmaları Dergisi*, 16(1), 24-30.
- Demir, C., ve Özcan, S. E. (2023). Türkiye’de Sanayi Üretimi, Kapasite Kullanım Oranı ve Üretici Fiyatları Arasındaki Asimetrik İlişki Doğrusal Olmayan ARDL Modeli Yaklaşımı. *Journal of Economic Policy Researches*, 10(2), Article 2. <https://doi.org/10.26650/JEPR1247326>
- Demir, H. (2018, Aralık 24). *R Üzerinde Zaman Serisi Analizi Bölüm 1: Zaman Serisi Regresyonu*. Ravenfo. <https://ravenfo.com/2018/12/24/r-uzerinde-zaman-serisi-analizi-bolum-1-zaman-serisi-regresyonu/> adresinden 30 Aralık 2023 tarihinde alınmıştır.
- Duran, A., ve Bakır, H. (2023). Hiperparametreleri Ayarlanmış Makine Öğrenimi Algoritmalarını Kullanarak Android Sistemlerde Kötü Amaçlı Yazılım Tespiti. *Uluslararası Sivas Bilim ve Teknoloji Üniversitesi Dergisi*, 2(1), Article 1.
- Ekelik, H., ve Altaş, D. (2019). Dijital Reklam Verilerinden Yararlanarak Potansiyel Konut Alıcılarının Rastgele Orman Yöntemiyle Sınıflandırılması. *Journal of Research in Economics*, 3(1), Article 1.
- EVDS|Tüm Seriler (2023). https://evds2.tcmb.gov.tr/index.php?/evds/serieMarket/collapse_15/5959/DataGroup/turkish/bie_rkgema/ adresinden 03 Aralık 2023 tarihinde alınmıştır.
- Gökcel, D. (2009). *Perakende Sektöründe Kategori Bazlı Talep Tahmin ve Sipariş Sistemi Uygulaması*. [Yüksek Lisans Tezi]. Yıldız Teknik Üniversitesi.
- Güneş, M. (2010). *Zaman Serilerinde Müdehale Analiz Teknikleri Üzerine Bir Çalışma* [Yüksek Lisans Tezi]. Namık Kemal Üniversitesi.
- Güven, İ. (2020). *Perakende Hazır Giyim Endüstrisinde Yapay Zeka Yöntemleri ile Talep Tahmini*. [Yüksek Lisans Tezi]. Karabük Üniversitesi.
- Haliloğlu, E. Y., ve Tutu, B. E. (2018). Türkiye İçin Kısa Vadeli Elektrik Enerjisi Talep Tahmini. *Yaşar Üniversitesi E-Dergisi*, 13(51), Article 51.
- Huang, G., Wu, L., Ma, X., Zhang, W., Fan, J., Yu, X., Zeng, W., ve Zhou, H. (2019). Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *Journal of Hydrology*, 574, 1029-1041. <https://doi.org/10.1016/j.jhydrol.2019.04.085>

- Hülsmann, M., Borscheid, D., Friedrich, C., ve Reith, D. (2011). General Sales Forecast Model for Automobile Markets and their Analysis. *International Journal Transaction on Machine Learning and Data Mining*, 5, 65-86.
- İçen, N. M. E. (2022). *Bayesyen Multinomial Lojistik Regresyon ile İşletmelerin Finansal Başarısızlığının Değerlendirilmesi: İmalat Sanayi Uygulaması*. [Doktora Tezi]. İstanbul Üniversitesi.
- Kadir Sheikh, S., ve Unde, M. (2012). Short-term Load Forecasting Using ANN Technique. *International Journal of Engineering Sciences & Emerging Technologies*, 1(2), 97-107. https://doi.org/10.7323/ijeset/v1_i2_12
- Karaatli, M., Helvacioğlu, Ö. C., Ömürbek, N., ve Tokgöz, G. (2012). Yapay Sinir Ağları Yöntemi İle Otomobil Satış Tahmini. *International Journal of Management Economics and Business*, 8(17), 87-100. <https://doi.org/10.11122/ijmeb.2012.8.17.290>
- Karahan, M. (2011). *İstatistiksel tahmin yöntemleri: Yapay sinir ağları ile ürün talep tahmini uygulaması* [Yüksek Lisans Tezi]. Selçuk Üniversitesi.
- Karakaş, E. (2019). Çocuk Yoğun Bakım Ünitesine Olan Talebin Zaman Serisi Yöntemleri ile Tahmin Edilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, 17, Article 17. <https://doi.org/10.31590/ejosat.624407>
- Karamustafa, K., ve Kılıçhan, B. (2019). Delphi Yöntemi ile Profesyonel Turist Rehberliği Mesleğine İlişkin Sorunların Tespit Edilmesi. *Afyon Kocatepe Üniversitesi Sosyal Bilimler Dergisi*, 21(4), Article 4. <https://doi.org/10.32709/akusosbil.597628>
- Kaya, S. S. (2013). Türkiye’de Savunma Harcamalarının İktisadi Etkileri Üzerine Nedensellik Analizi (1970 – 2010). *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 15(2), Article 2.
- Keleş, M. K., Kılıç, Ü. ve Keleş, A. (2018). E-Ticarette Pazar Araştırması ve SWOT Analizi. *El-Cezeri*, 5(2), Article 2. <https://doi.org/10.31202/ecjse.386427>
- Kelle, A. C., ve Yüce, H. (2022). MQTT Trafiğinde DoS Saldırılarının Makine Öğrenmesi İle Sınıflandırılması ve Modelin SHAP İle Yorumlanması. *Journal of Materials and Mechatronics: A*, 3(1), 50-62. <https://doi.org/10.55546/jmm.995091>
- Keskenler, M. F., ve Keskenler, E. F. (2017). Geçmişten Günümüze Yapay Sinir Ağları ve Tarihçesi. *Takvim-i Vekayi*, 5(2), 8-18.
- Kılınç, N., ve Çakır Aydın, M. (2016). Hazır Giyim İşletmelerinde Beden Numaralarına Göre Oluşacak Talebin Belirlenmesi: Örnek Bir Uygulama ve Karşılaştırma. *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, 18(31), 30-35.
- Korucuk, S., ve Tatlı, Y. (2017). Talep Tahminin İşletmelere Sağladığı Yararlar: İmalat İşletmelerinde Bir Araştırma. *Iğdır Üniversitesi Sosyal Bilimler Dergisi*, 12, Article 12.
- Kuş, İ., Keser, S. B., ve Yolaçan, E. (2021). Saldırı Tespit Sistemlerinde Topluluk Öğrenme Yöntemlerinin Kıyaslanması. *Avrupa Bilim ve Teknoloji Dergisi*, 31, Article 31. <https://doi.org/10.31590/ejosat.971875>

- Meydan, Y. A. (2018). *Talep Tahmin Yöntemleri ve Orta Ölçekli Bir İşletmede Uygulanması* [Yüksek Lisans Tezi]. İstanbul Ticaret Üniversitesi.
- Mijwil, M. (2017). *Yapay Sinir Ağlar Yapısı ve Fonksiyonu*. <https://doi.org/10.13140/RG.2.2.32068.14723>
- Murat, E., ve Özcanhan, M. H. (2023). Makine Öğrenmesi Algoritmalarında Hiper Parametre Belirleme. *Mühendislikte Öncü ve Çağdaş Çalışmalar*, 71-98.
- Ovla, H. D., ve Taşdelen, B. (2012). Aykırı Değer Yönetimi. *Mersin Üniversitesi Sağlık Bilimleri Dergisi*, 5(3), Article 3.
- Özcan, E. (2021). *Kükürt Giderme İşlemi İçin Kullanılan Malzeme Miktarının Makine Öğrenme Yöntemleri ile Tahmini* [Yüksek Lisans Tezi]. Karabük Üniversitesi.
- Özdemir, A., ve Özdemir, A. (2006). Talep Tahminlemesinde Kullanılan Yöntemlerin Karşılaştırılması: Seramik Ürün Grubu Firma Uygulaması. *Ege Academic Review*, 6(2), Article 2.
- Özer, O. O., ve Yavuz, G. (2014). *Box-Jenkins Modeli Yardımıyla Fındık Fiyatının Tahmini*. XI. Ulusal Tarım Ekonomisi Kongresi.
- Pazar—Perakende Satışlar. (2023). https://www.odmd.org.tr/web_2837_1/neuralnetwork.aspx?type=36 adresinden 03 Aralık 2023 tarihinde alınmıştır.
- Pongdatu, G. A. N., ve Putra, Y. H. (2018). Seasonal Time Series Forecasting using SARIMA and Holt Winter's Exponential Smoothing. *IOP Conference Series: Materials Science and Engineering*, 407(1), 012153. <https://doi.org/10.1088/1757-899X/407/1/012153>
- Raizada, S., ve Saini, J. R. (2021). Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 12(11), Article 11. <https://doi.org/10.14569/IJACSA.2021.0121112>
- Rousseeuw, P. J., Tukey, J. W., & Ruts, I. (1999). The Bagplot: A Bivariate Boxplot. *The American Statistician*, 53(4).
- Sarı, M. (2016). *Yapay Sinir Ağları ve Bir Otomotiv Firmasında Satış Talep Tahmini Uygulaması* [Yüksek Lisans Tezi]. Sakarya Üniversitesi.
- Sevgen, S. C., ve Aliefendioğlu, Y. (2020). Mass Appraisal With A Machine Learning Algorithm: Random Forest Regression. *Bilişim Teknolojileri Dergisi*, 13(3), Article 3. <https://doi.org/10.17671/gazibtd.555784>
- Sharma, Dr. R., ve Sinha, A. (2012). Sales Forecast of an Automobile Industry. *International Journal of Computer Applications*, 53, 25-28. <https://doi.org/10.5120/8474-2403>
- Sohrabpour, V., Oghazi, P., Toorajipour, R., ve Nazarpour, A. (2021). Export Sales Forecasting Using Artificial Intelligence. *Technological Forecasting and Social Change*, 163, 120480. <https://doi.org/10.1016/j.techfore.2020.120480>
- Tan, Ş. (2010). Öğretim İlke ve Yöntemleri Dersinde Öğrencilerin Akademik Başarılarının Yordanması. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, 23(1), Article 1.

- Tekin, A., ve Sarı, C. (2022). Mağaza Tabanlı Talep Tahmini: Topluluk Öğrenmesi Yaklaşımları ile Zaman Serisi Yaklaşımı Karşılaştırılması (ss. 235-240). International Engineering and Technology Management Summit 2022–ETMS2022.
- TÜİK - Veri Portalı. (2023).
<https://data.tuik.gov.tr/Kategori/GetKategori?p=Ekonomik-Guven-117>
adresinden 23 Aralık 2023 tarihinde alınmıştır.
- TÜİK Kurumsal. (2023, Aralık 20). <https://data.tuik.gov.tr/Bulten/Index?p=Tuketici-Guven-Endeksi-Aralik-2023-49442> adresinden 23 Aralık 2023 tarihinde alınmıştır.
- Türk, E., ve Kiani, F. (2019). Yapay Sinir Ağları ile Talep Tahmini Yapma: Beyaz Eşya Üretim Planlama Örneği. *İstanbul Sabahattin Zaim Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 1(1), Article 1.
- Ucal, M. Ş. (2006). Ekonometrik Model Seçim Kriterleri Üzerine Kısa Bir İnceleme. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 7(2), 41-56.
- Ulucan, E., ve Kızılırmak, İ. (2018). Konaklama İşletmelerinde Talep Tahmin Yöntemleri: Yapay Sinir Ağları ile İlgili Bir Araştırma. *Seyahat ve Otel İşletmeciliği Dergisi*, 15(1), Article 1. <https://doi.org/10.24010/soid.415343>
- Utku, A., ve Can, Ü. (2022). Covid-19'un Yayılım Tahminine Yönelik Makine Öğrenmesi ve Derin Öğrenme Tabanlı Karşılaştırmalı Bir Analiz: Türkiye İçin Örnek Bir Çalışma. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 34(2), 709-717. <https://doi.org/10.35234/fumbd.1125609>
- Ülkü, H., ve Yalpır, Ş. (2021). Enerji Talep Tahmini için Metodoloji Geliştirme: 2030 Yılı Türkiye Örneği. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 10(1), Article 1. <https://doi.org/10.28948/ngumuh.814134>
- Verileri Sığ Bir Sinir Ağıyla Uydurmak—MATLAB ve Simulink. (2023). <https://www.mathworks.com/help/deeplearning/gs/fit-data-with-a-neural-network.html> adresinden 11 Aralık 2023 tarihinde alınmıştır.
- Xiang, W., Xu, P., Fang, J., Zhao, Q., Gu, Z., ve Zhang, Q. (2022). Multi-dimensional data-based medium- and long-term power-load forecasting using double-layer CatBoost. *Energy Reports*, 8, 8511-8522. <https://doi.org/10.1016/j.egyr.2022.06.063>
- Yağımlı, M., ve Ergin, H. (2017). Türkiye’de İş Kazalarının Üssel Düzeltme Metodu ile Tahmin Edilmesi. *Marmara Fen Bilimleri Dergisi*, 29(4), Article 4. <https://doi.org/10.7240/marufbd.305236>
- Yarat, S., ve Orman, Z. (2023). Elektrik Güç Dağıtımında Akıllı Sayaç Verileri için Anomali Tespiti ve Tahminleme. *Bilgisayar Bilimleri ve Teknolojileri Dergisi*, 3(2), 72-85. <https://doi.org/10.54047/bibted.1224628>
- Yazıcıoğlu, N. (2010). *Yapay Zeka ile Talep Tahmini* [Yüksek Lisans Tezi]. Uludağ Üniversitesi.

- Yeşilyurt, S. N., ve Dalkılıç, H. (2021). *Xgboost ve Gradient Boost Machine ile Günlük Nehir Akımı Tahmini*. The 3rd International Symposium of Engineering Applications on Civil Engineering and Earth Sciences 2021 (IEACES2021).
- Yılmaz, K. (2021). *Finansal Zaman Serileri Analizinde Farklı İstatistiksel Modellerin Karşılaştırılması* [Yüksek Lisans Tezi]. Afyon Kocatepe Üniversitesi.
- Yılmaz, M. B., ve Çilengiroğlu, Ö. V. (2022). Talep Tahminleme Değişkenlerinin Üssel Düzeltme Yöntemi ile Belirlenmesi. *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences*, 9(22), Article 22. <https://doi.org/10.5281/zenodo.6948405>
- Zhou, F., Pan, H., Gao, Z., Huang, X., Qian, G., Zhu, Y., ve Xiao, F. (2021). Fire Prediction Based on CatBoost Algorithm. *Mathematical Problems in Engineering*, 2021, 1-9. <https://doi.org/10.1155/2021/1929137>
- Zorlu, M. (2021). *Atölye Tipi Makine İmalatı Yapan Bir Firmada Talep Tahmini Uygulaması* [Yüksek Lisans Tezi]. Kocaeli Üniversitesi.

EKLER

EK A. LOF ile AYKIRI DEĞER ANALİZİ PYTHON KODLARI

```
# Sklearn kütüphanesi içerisinde LOF fonksiyonu import edilir.

from sklearn.neighbors import LocalOutlierFactor

# Komşuluk katsayısı ve yoğunluk belirlenir.

lof=LocalOutlierFactor(n_neighbors=10, contamination=0.1)

# Programa tanıtılan veri setinin adı ile algoritma çalıştırılır.

lof.fit_predict(veri)

#Veri seti aykırı değer analizi için skorlar "negative_outlier_factor_"

#fonksiyonu ile kaydedilir.

veri_scores = lof.negative_outlier_factor_

#Skor değerleri küçükten büyüğe sıralanır.

np.sort(veri_scores)

#Eşik değeri belirlenir.

esik_deger = np.sort(veri_scores)[5]

#Aykırı olmayan gözlem değerleri ayrı bir vektör olarak tutulur.

norm_deger = veri_scores > esik_deger

#Eşik değere karşılık gelen gözlem birimi kaydedilir.

baski_deger = veri[veri_scores == esik_deger]

#Aykırı gözlem değerleri ayrı bir vektör olarak tutulur.

aykiri_deger = veri[~norm_deger]

#Vektör atama işlemi için indexlerden arındırılır.

index = aykiri_deger.to_records(index = False)

#Aykırı değerler, baskı değeri(eşik değeri) ile değiştirilir.
```

```
index[:] = baski_deger.to_records(index = False)

#Düzenlenen aykırı değerler veri setinin içerisine
#index değerleri ile birlikte yerleştirilir.

veri[~norm_deger] = pd.DataFrame(index, index = veri[~norm_deger].index)
```



ÖZGEÇMİŞ

Ad-Soyad : Beyza KURTGERİ

ÖĞRENİM DURUMU:

- **Lisans** : 2019, Sakarya Üniveristesi, Mühendislik Fakültesi, Endüstri Mühendisliği
- **Yükseklisans** : Devam ediyor, Sakarya Üniveristesi, Endütri Mühendisliği Anabilim Dalı, Endüstri Mühendisliği Programı

MESLEKİ DENEYİM VE ÖDÜLLER:

- 2021-2023 yılları arasında Bilkont Dış Tic. ve Tekstil San. A,Ş'de Kalite Yönetim Sorumlusu olarak çalıştı.

TEZDEN TÜRETİLEN ESERLER:

- Kurtgeri B. ve Cengiz Toklu M. 2023. Yapay Sinir Ağları ve Makine Öğrenmesi Yöntemleri ile Karşılaştırılmalı Tahmin Performans Analizi: Otomobil Satış Uygulaması, Ege 10th International Conference on Applied Sciences, December 22 -24, 2023, İzmir, 1119-1132.