



Araç Fiyat Tahmininde Makine Öğrenmesi Algoritmalarının Karşılaştırılması ve Performans Analizi



Comparison And Performance Analysis of Machine Learning Algorithms in Vehicle Price Prediction



<https://doi.org/10.25204/iktisad.1494020>

Oğuzhan KIVRAK*

Öz

Makale Bilgileri

Makale Türü:

Araştırma
Makalesi

Geliş Tarihi:

06.12.2024

Kabul Tarihi:

06.05.2025

© 2025 İKTİSAD
Tüm hakları
saklıdır.



Bu çalışma, araç fiyat tahmini problemine odaklanarak farklı makine öğrenmesi algoritmalarının performanslarını karşılaştırmakta ve sektöre yönelik stratejik öneriler sunulması amaçlamıştır. Lineer Regresyon, Ridge Regresyon, Lasso Regresyon, Karar Ağaçları, Rastgele Orman, Gradient Boosting, SVR, XGBoost, Multi-Layer Perceptron (MLP) ve Radyal Tabanlı Fonksiyon (RBF) modelleri Ocak 2022 ile Aralık 2023 dönemini kapsayan, araçlara ait marka, model, motor hacmi, fiyat gibi teknik özellikler ile müşteri bilgileri içeren 23.900 adet veri seti üzerinde test edilmiştir. Veri ön işleme aşamasında eksik veri doldurma, kategorik verilerin dönüştürülmesi ve aykırı değer analizi uygulanmıştır. Lasso Regresyon, %99 R^2 değeri ve düşük hata oranları ile en başarılı model olarak öne çıkmıştır. Rastgele Orman, Gradient Boosting ve MLP algoritmaları da %97 R^2 değerleri ile güçlü alternatifler arasında yer almıştır. Buna karşın, Karar Ağaçları, SVR ve XGBoost algoritmaları yüksek hata oranları ve düşük R^2 değerleri ile zayıf bir performans sergilemiştir. Sonuçlar, veri ön işleme süreçlerinin model başarısı üzerindeki kritik rolünü vurgulamaktadır. Makroekonomik göstergelerin modele entegrasyonu ve hiperparametre optimizasyonu tekniklerinin uygulanması, modellerin doğruluk ve genelleme kabiliyetlerini artırma potansiyeline sahiptir.

Anahtar Kelimeler: Araç fiyat tahmini, makine öğrenmesi, regresyon modelleri, veri analizi, tahmin performansı.

Abstract

This study focuses on the problem of vehicle price prediction and aims to compare the performance of various machine learning algorithms while offering strategic insights for the industry. Linear Regression, Ridge Regression, Lasso Regression, Decision Trees, Random Forest, Gradient Boosting, Support Vector Regression (SVR), XGBoost, Multi-Layer Perceptron (MLP), and Radial Basis Function (RBF) models were tested on a dataset comprising 23.900 records collected between January 2022 and December 2023. The dataset includes technical specifications of vehicles—such as brand, model, engine size, and price—as well as customer-related features. During the data preprocessing phase, missing values were imputed, categorical variables were encoded, and outlier analysis was conducted. Lasso Regression emerged as the most successful model, achieving an R^2 value of 0.99 and low error rates. Random Forest, Gradient Boosting, and MLP also demonstrated strong performance with R^2 values around 0.97. In contrast, Decision Trees, SVR, and XGBoost showed weaker results, characterized by higher error rates and lower R^2 values. The findings highlight the critical importance of data preprocessing in model performance. Moreover, the integration of macroeconomic indicators and the application of hyperparameter optimization techniques have the potential to further enhance model accuracy and generalizability.

Keywords: Vehicle price prediction, machine learning, regression models, data analysis, prediction performance.

Article Info

Paper Type:

Research Paper

Received:

06.12.2024

Accepted:

06.05.2025

© 2025 JEBUPOR
All rights
reserved.



Atıf/ to Cite (APA): Kıvrak, O. (2025). Araç fiyat tahmininde makine öğrenmesi algoritmalarının karşılaştırılması ve performans analizi. *İktisadi İdari ve Siyasal Araştırmalar Dergisi*, 10(27), 454-474. <https://doi.org/10.25204/iktisad.1494020>

*ORCID Dr. Öğr. Üyesi, Bandırma Onyediy Eylül Üniversitesi, Bandırma Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü
Bilgisayar Programcılığı Programı, oguzhan.kivrak@gmail.com

Extended Abstract

Background:

The automotive industry plays a vital role in both developed and emerging economies by driving economic transformation and technological innovation. In 2023, over 93 million vehicles were produced globally, highlighting the sector's impact on supply chains and market dynamics. With rapid digitalization and shifting consumer expectations, accurate vehicle price estimation has become crucial—not only for manufacturers and sellers, but also for consumers, insurers, and financial institutions. The used car market adds further complexity due to diverse buyer profiles and vehicle conditions. Therefore, effective pricing strategies supported by reliable prediction models are essential for market stability and economic efficiency.

Research Purpose:

This study aims to fill two major gaps in vehicle price prediction research. First, it systematically compares various machine learning algorithms—including regression models, decision trees, ensemble methods, and neural networks—to evaluate their accuracy. Second, it examines how preprocessing and model optimization strategies affect prediction performance. Using a large dataset from January 2022 to December 2023, the study explores how different algorithms handle complex vehicle and consumer features. Unlike previous studies with limited scope, this research incorporates robust metrics and hyperparameter tuning. The ultimate goal is to support strategic decisions in pricing, production, and marketing.

Methodology:

This study utilized multiple machine learning algorithms to predict monthly vehicle sales, including Linear, Ridge, and Lasso Regression, Decision Trees, Random Forest, Gradient Boosting, XGBoost, Support Vector Regression (SVR), Multi-Layer Perceptron (MLP), and Radial Basis Function (RBF) networks. The dataset, sourced from Kaggle, contains 23,900 entries from the U.S. market between January 2022 and December 2023, incorporating both vehicle specifications and customer attributes. Data preprocessing involved handling missing values, applying One-Hot Encoding for categorical variables, standardizing numerical values, and removing or capping outliers. The dataset was split into 80% training and 20% test sets. Feature importance analysis identified vehicle price, customer income, and brand as the most influential predictors. Model performance was evaluated using MAE, MAPE, and R^2 metrics. Hyperparameter optimization was conducted using Grid Search and Randomized Search CV. All modeling was implemented in Python using the Google Colab environment. The methodology aimed to ensure high accuracy and generalizability in sales prediction.

Findings:

The study revealed significant fluctuations in average monthly vehicle prices between 2022 and 2024, with noticeable peaks during the summer months and declines toward the year-end, influenced by seasonal demand and market dynamics. Gender-based analysis showed a clear imbalance, with male buyers significantly outnumbering females. Certain brands such as Chevrolet and Dodge were more popular among men, while female buyers prioritized safety, fuel efficiency, and technology. SUV and hatchback body types were the most preferred overall. In terms of model performance, Lasso Regression achieved the highest accuracy with an R^2 of 0.99 and the lowest error rates, benefiting from feature selection via L1 regularization. Linear and Ridge Regression also performed well, followed by Random Forest, Gradient Boosting, and MLP (all with $R^2 \approx 0.97$). In contrast, Decision Tree, XGBoost, and SVR models showed poor predictive performance. The results highlight the importance of model choice, feature relevance, and proper preprocessing in forecasting tasks.

Conclusions:

This study offers a comprehensive evaluation of various machine learning algorithms for vehicle price prediction, revealing that model performance significantly varies depending on data structure and problem characteristics. Lasso Regression demonstrated superior accuracy ($R^2 = 0.99$), while Random Forest, Gradient Boosting, and MLP also showed strong generalization capabilities. In contrast, Decision Tree, XGBoost, and SVR performed poorly due to underfitting or sensitivity to hyperparameters. The findings highlight the importance of appropriate model selection, feature relevance, and preprocessing. Gender-based preference analysis suggests that customized marketing strategies based on safety, efficiency, and design could enhance female representation in the automotive market. The study acknowledges limitations due to dataset scope and geography. Incorporating macroeconomic indicators (e.g., inflation, fuel prices, interest rates) may improve model robustness in future work. Prior literature supports this direction, showing enhanced accuracy through multivariate modeling. Thus, this research contributes both methodologically and practically to predictive modeling in the automotive domain.

1. Giriş

Otomotiv sektörü, yalnızca sanayileşmiş ülkelerde değil, aynı zamanda gelişmekte olan ekonomilerde de büyümenin itici gücü konumundadır. Türkiye’de de otomotiv sanayisi, tarım ağırlıklı ekonomik yapıdan sanayi odaklı üretim modeline geçişte kritik bir rol üstlenmiştir. Sektör, istihdam yaratma, teknolojik gelişmeleri teşvik etme ve diğer sanayi kollarını destekleyerek ekonomik büyümeye önemli katkılarda bulunması nedeniyle küresel ölçekte stratejik bir öneme sahiptir (Eşidir vd., 2022:261). Pazarlama, bakım-onarım, yedek parça satışı, finansman ve sigortacılık gibi birçok yan sektörü kapsayan otomotiv endüstrisi, geniş bir iş hacmi oluşturarak ekonomik ekosistemin temel taşlarından biri haline gelmiştir (Gür vd., 2024:18). 2023 yılı itibarıyla dünya genelinde toplam motorlu taşıt üretimi 93 milyon 547 bin adede ulaşırken, bu üretimin 68 milyon 20 bin adedi otomobillerden oluşmaktadır (Otomotiv Sanayi Derneği, 2024). Otomobil pazarındaki bu yoğunluk göz önüne alındığında, tüketiciler için doğru ve öngörülebilir fiyat tahminleri giderek daha önemli hale gelmektedir. Artan tüketici talepleri ve genişleyen model çeşitliliği hem yeni hem de ikinci el otomobil piyasalarında fiyatlandırma süreçlerinin daha detaylı analiz edilmesini gerektirmektedir.

Dijitalleşmenin hızla yayılmasıyla birlikte otomotiv sektörü, tüketici beklentilerindeki değişimler, teknolojik yenilikler ve ekonomik dalgalanmaların etkisiyle köklü bir dönüşüm sürecinden geçmektedir (Llopis-Albert vd., 2021: 1). Bu dönüşüm, rekabet avantajı elde etmek isteyen üreticiler ve satıcılar için doğru fiyat tahmini yapmayı zorunlu hale getirmiştir. Ancak araç fiyatlarının belirlenmesi, piyasa koşulları, ekonomik krizler, petrol fiyatlarındaki dalgalanmalar, vergi politikaları ve faiz oranları gibi dışsal faktörlerden etkilenerek oldukça karmaşık bir problem olarak karşımıza çıkmaktadır (Yurtsever, 2022: 1482). Tüketici odaklı pazarlarda, doğru fiyatlandırma stratejileri hem yeni hem de ikinci el araç piyasasında satışları optimize etmek ve finansal riskleri en aza indirmek açısından büyük önem taşımaktadır. Kesin ve güvenilir bir değerlendirme süreci, yalnızca alıcılar ve satıcılar arasındaki adil ticareti sağlamakla kalmaz, aynı zamanda sigorta şirketlerinin risk bazlı prim oranlarını belirlemesine ve finans kuruluşlarının kredi değerlerini daha doğru bir şekilde değerlendirmesine olanak tanır. Ayrıca, etkili fiyatlandırma modelleri, üreticiler ve bayiler için stratejik karar alma süreçlerinde önemli bir rehber niteliği taşıyarak pazar istikrarını artırmakta ve genel ekonomik verimliliğe katkı sağlamaktadır (Ahmad vd., 2024).

Otomotiv sektöründe, uluslararası piyasalarda fiyat rekabeti önemli bir unsur olmaya devam ederken, günümüzde kalite, ürün çeşitliliği ve geleceğe yönelik yatırımlar rekabetin belirleyici faktörleri haline gelmiştir. Tüketici tercihleri, satışlar üzerinde doğrudan etkili olup, bu bağlamda ürün geliştirme, marka yenilikleri ve model çeşitliliği gibi unsurlar sektördeki rekabet dinamiklerini şekillendirmektedir (Sanayi Bakanlığı). Otomotiv endüstrisi, değişen pazar koşulları ve teknolojik ilerlemeler doğrultusunda araç fiyatlarını doğru bir şekilde tahmin etme gibi karmaşık bir zorlukla karşı karşıyadır. Fiyat tahmin süreçleri; marka, model, üretim yılı, kilometre ve genel durum gibi çeşitli faktörlere dayalı olarak gerçekleştirilmektedir ve bu süreç, alıcılar ve satıcılar için önemli içgörüler sağlamaktadır. Etkili tahmin modelleri, karar alma mekanizmalarını destekleyerek pazar verimliliğini artırmakta ve endüstrinin değişen koşullara uyum yeteneğini güçlendirmektedir (Ahmad vd., 2024).

Otomobil satış tahmini, otomotiv şirketleri için üretim planlaması ve iş stratejilerini optimize etme açısından hayati bir öneme sahiptir. Bu tahmin modelleri, gelecekteki satışları ve tüketici eğilimlerini belirlemek için mevsimsel değişimleri, rakip stratejilerini ve pazar talebini analiz etmektedir. Örneğin, yaz aylarında seyahatlerin artması ve yeni model lansmanları nedeniyle otomobil satışları yükselirken, kış aylarında hava koşullarının etkisiyle düşüş yaşanabilmektedir. Şirketler, bu tür tahminlerden yararlanarak üretim planlarını dinamik bir şekilde ayarlayabilir, aşırı üretimi önleyerek maliyetleri optimize edebilir ve kaynak yönetimini daha verimli hale getirebilirler. Doğru tahmin modelleri, işletmelerin rekabetçi kalmasına ve hızlı değişen piyasa koşullarına uygun stratejiler geliştirmesine olanak tanımaktadır (Harsha vd., 2024).

Yeni model araçların piyasaya sürülmesi, yalnızca sıfır kilometre araç piyasasını değil, aynı zamanda ikinci el otomobil sektörünü de doğrudan etkilemektedir. İkinci el araç piyasası, geniş bir alıcı ve satıcı kitlesini içermesi nedeniyle dinamik bir yapıya sahiptir ve bu pazarda fiyat dengelerinin oluşumu önemli bir araştırma konusu olarak öne çıkmaktadır. İkinci el araç fiyatlandırmasında; teknik özellikler, donanım seviyesi, güvenlik önlemleri, değişen parçalar ve boya geçmişi gibi faktörlerin belirleyici olduğu tespit edilmiştir. Bu unsurlar, tüketici tercihleri ve piyasa koşullarıyla birleşerek fiyatlandırma süreçlerini daha karmaşık hale getirmektedir. Dolayısıyla hem yeni hem de ikinci el otomobil piyasasında doğru fiyatlandırma modellerinin geliştirilmesi, piyasa istikrarını sağlamak ve ekonomik sürdürülebilirliği artırmak adına kritik bir gereklilik olarak karşımıza çıkmaktadır.

Geçmiş verilere dayalı olarak gelecekteki eğilimleri öngörmeye yönelik bir süreç olan tahminleme, çeşitli yöntemler kullanılarak uygulanmaktadır. Bu bağlamda, talep tahmini, otomotiv sektöründeki firmalar için stratejik bir öneme sahiptir. Satış gelirlerini artırmak, üretim süreçlerini optimize etmek ve genel verimliliği yükseltmek gibi hedeflere ulaşmada talep tahmini kritik bir rol oynamaktadır (Efendigil ve Eminler, 2017).

Otomotiv sektöründe fiyat tahmini, yalnızca üreticiler ve satıcılar için değil, aynı zamanda tüketiciler ve politika yapıcılar açısından da stratejik bir öneme sahiptir. Ancak literatürdeki çalışmaların büyük bir kısmı, genellikle sınırlı veri setleri, belirli algoritmalar veya geleneksel yöntemler üzerine odaklanmaktadır. Bridge-Nduwimana vd. (2025), ikinci el araç fiyat tahmininde eksik veri sorununu gidermek amacıyla özel bir özellik seçme algoritması ile HistGradientBoostingRegressor tabanlı bir imputasyon yöntemi entegre eden yenilikçi bir yaklaşım önermektedir. Çalışma, XGBoost ve Gradient Boosting Regressor gibi güçlendirilmiş modellerin fiyat tahmininde üstün performans sergilediğini göstererek, otomotiv sektöründe veri kalitesinin artırılması ve tahmin doğruluğunun iyileştirilmesine önemli katkılar sunmaktadır. Benzer şekilde, Ogundele vd. (2024), ikinci el araç fiyat tahmininde doğruluk ve hata oranlarını optimize etmek için derin öğrenme tabanlı hibrit bir özellik seçme yöntemi geliştirmiştir. Çalışmada, önerilen modelin %96,9 doğruluk oranına ulaştığı ve Karar Ağacı (%87,8) ile Destek Vektör Makineleri (%88,3) gibi geleneksel yöntemlerden daha başarılı olduğu ortaya konulurken, diğer modellerin daha düşük hata oranlarına sahip olduğu belirtilmektedir. Araç fiyat tahmini yalnızca stratejik karar alma süreçlerini desteklemekle kalmayıp, aynı zamanda kaynak planlaması, üretim ve pazarlama stratejilerinin daha etkin bir şekilde geliştirilmesine de katkı sağlamaktadır (Fantazzini ve Toktamsoya, 2015: 97).

Tahmin modelleri, kantitatif yöntemler aracılığıyla piyasa talebini ve fiyat dinamiklerini anlamak için etkili araçlar sunmaktadır. Kantitatif yöntemler, büyük veri setlerinden yararlanarak matematiksel modeller ile belirsizlikleri azaltmakta ve gelecekteki durumları öngörmeyi mümkün kılmaktadır (Karaatlı vd., 2012). Bu bağlamda, makine öğrenmesi algoritmaları, araç fiyat tahminleri için güçlü bir çözüm sunmaktadır. Makine öğrenmesi yöntemleri, büyük ve karmaşık veri setleri üzerinde çalışarak gelecekteki değerleri doğru bir şekilde tahmin etme kapasitesine sahiptir ve bu durum, araç fiyat tahmini gibi çok boyutlu problemlerin çözümünde önemli bir avantaj sağlamaktadır. Ancak literatürdeki birçok çalışma, makine öğrenmesi algoritmalarının performanslarının sistematik bir şekilde karşılaştırılmasına odaklanmamış ve bu nedenle hangi algoritmaların hangi durumlarda daha etkili olduğu sorusu açık kalmıştır.

Bu çalışma, Ocak 2022 ile Aralık 2023 dönemini kapsayan geniş ve çeşitli bir veri setini kullanarak araç fiyat tahmini üzerine odaklanmaktadır. Lineer Regresyon, Ridge Regresyon, Lasso Regresyon, Karar Ağaçları, Rastgele Orman (Random Forest), Gradient Boosting, Destek Vektör Regresyonu (SVR), XGBoost, Multi-Layer Perceptron (MLP) ve Radyal Tabanlı Fonksiyon (RBF) gibi makine öğrenmesi algoritmalarını karşılaştırmalı olarak test eden bu araştırma, literatürdeki şu iki temel boşluğu doldurmayı amaçlamaktadır:

1. Farklı makine öğrenmesi algoritmalarının performanslarının sistematik karşılaştırılması.
2. Araç fiyat tahmini gibi karmaşık bir problemde makine öğrenmesi algoritmalarının etkinliğini artıracak veri ön işleme ve model optimizasyon stratejilerinin analizi.

Araştırmanın sonuçları, fiyat tahmininde kullanılan yöntemlerin etkinliğini artırmaya yönelik uygulamalı öneriler sunmakta ve sektörel stratejilere rehberlik edecek anlamlı bilgiler sağlamaktadır. Böylelikle hem literatüre katkı sağlamakta hem de otomotiv sektöründe daha doğru ve güvenilir tahmin araçlarının geliştirilmesine ışık tutmaktadır.

2. Literatür Taraması

Makine öğrenmesi, veri analizi, tahmin, sınıflandırma ve kümeleme gibi birçok problemin çözümünde etkili bir araç olarak literatürde geniş bir yer edinmiştir. Büyük veri setlerindeki karmaşık ilişkileri anlamlandırma ve gelecekteki değerleri tahmin etme yeteneği, makine öğrenmesini özellikle ticari ve ekonomik analizler için değerli bir yöntem haline getirmiştir. Otomobil sektörü gibi dinamik yapıya sahip bir alanda, bu yöntemlerin kullanımı, satış tahminlerini daha doğru ve güvenilir bir şekilde yapmayı mümkün kılmaktadır. Literatürde, otomobil satışlarının tahminine yönelik birçok çalışma, farklı yöntemler ve değişkenler kullanılarak gerçekleştirilmiştir.

Karaatlı vd. (2012), Türkiye’de yeni otomobil satışlarını tahmin etmek için yapay sinir ağı (YSA) yöntemini kullanmıştır. Çalışmada, Ocak 2007 ile Haziran 2011 arasındaki dönemden elde edilen gayri safi yurtiçi hasıla, reel kesim güven endeksi, yatırım harcamaları, tüketim harcamaları, tüketici güven endeksi, dolar kuru ve zaman gibi değişkenler ele alınmıştır. Araştırma sonucunda, YSA yönteminin otomobil satış tahminlerinde etkili bir yöntem olduğu bulunmuştur.

Fantazzini ve Toktamyssoya (2015) ise 2001 Ocak ile 2014 Haziran dönemini kapsayan, Alman Federal Motorlu Taşıtlar Kurumu’na kayıtlı araçlardan oluşan bir veri setini analiz etmiştir. Çalışmada, otomobil satışlarının piyasa koşulları, ekonomik krizler, petrol fiyatlarındaki artış, vergi avantajları ve faiz oranları gibi dış faktörlerden etkilendiği tespit edilmiştir. Ayrıca, uzun vadeli tahminler için basit tek değişkenli modellerin yetersiz olduğu ve çok değişkenli modellere ihtiyaç duyulduğu sonucuna ulaşılmıştır.

Kuvvetli vd. (2015), çevresel ve ekonomik parametreleri dikkate alarak farklı segment ve markalar için aylık araç satışlarını tahmin etmeyi amaçlamıştır. 2008-2012 yılları arasındaki dönemi kapsayan çalışmada, segment ve markaya bağlı satış miktarları, taşıt kredisi faiz oranları, gayri safi milli hasıla, CO2 emisyonu, yakıt tüketim miktarları ve fiyatı gibi parametreler, YSA modeli ile analiz edilmiştir. Bu yaklaşım, satış tahminlerinde başarılı sonuçlar elde etmiştir.

Arslankaya ve Öz (2018), talep tahmin yöntemlerini kullanarak otomobil satışlarını tahmin etmiştir. Çalışmada, geçmiş aylara ait sipariş sayıları, tüketici fiyat endeksi, gayri safi yurtiçi hasıla, dolar kuru, tüketici güven endeksi ve reel sektör güven endeksi gibi değişkenlerle YSA yöntemi uygulanmıştır. Model, MAPE değeri 7,44 ile yüksek doğruluk oranı sağlamıştır.

Kurtgeri (2018) tarafından gerçekleştirilen çalışmada, Türkiye’de 2014-2022 yılları arasındaki veriler kullanılarak otomobil talep tahmini gerçekleştirilmiştir. Otomobil talebini etkileyen ekonomik ve sosyal faktörler göz önünde bulundurularak, bağımlı değişken olarak otomobil satış verileri, bağımsız değişkenler olarak ise ithalat miktar endeksi, reel kesim güven endeksi, tüketici güven endeksi, ortalama taşıt kredi faiz oranları ve Türkiye otomobil üretim adetleri modele dahil edilmiştir. Çalışmada, makine öğrenmesi algoritmaları (CatBoost, Gradient Boosting, Random Forest, SVR), yapay sinir ağı (YSA) ve zaman serileri analizi yöntemleri kullanılarak tahmin performansları karşılaştırılmıştır. Veri ön işleme sürecinde aykırı değer analizi uygulanarak model doğruluğu artırılmıştır. Sonuçlar, %85 R² değeri ile en başarılı yöntemin CatBoost algoritması olduğunu ortaya koymuştur. Çalışma, otomotiv sektöründe talep tahmini için güvenilir ve doğru sonuçlar sunan algoritmaların belirlenmesinde önemli bir katkı sağlamaktadır.

Yang ve Zhang (2020), sosyal medya verilerinin otomobil satışlarına etkisini incelemiştir. Audi markasına ait 2011-2019 yıllarını kapsayan aylık satış verileri, Baidu arama dizinleri ve Weibo duygu dizinleri gibi sosyal medya kaynaklarından elde edilen verilerle zenginleştirilmiş ve makine

öğrenmesi modelleri ile analiz edilmiştir. Çalışma, sosyal medya verilerinin tahmin modellerini iyileştirme potansiyeline işaret etmiştir.

Kayapınar Kaya ve Yıldırım (2020), otomobil satış tahminleri için 8 katmanlı Derin Sinir Ağı (DSA) modeli önermiştir. Çalışmada döviz kuru, gayri safi yurtiçi hasıla, tüketici güven endeksi ve tüketici fiyat endeksi gibi değişkenler kullanılmıştır. Araştırma, DSA modelinin çeşitli satış tahmin problemlerinde etkili olduğunu ortaya koymuştur.

Kayakuş vd. (2023) çalışmasında, Türkiye'deki hafif ticari araç satışlarının tahmini için yapay sinir ağı (YSA), çoklu doğrusal regresyon (MLR) ve karar ağacı regresyonu (DTR) yöntemleri kullanılmıştır. Hafif ticari araç talebindeki değişimlerin ekonomik canlılığın bir göstergesi olduğu vurgulanmış ve makroekonomik değişkenlerle oluşturulan modelin tahmin başarısı değerlendirilmiştir. Çalışma sonuçlarına göre, R^2 değerleri YSA için %94,6, MLR için %64,1 ve DTR için %82,2 olarak hesaplanmış ve YSA yönteminin en başarılı sonuçları verdiği belirlenmiştir. Bu çalışma, hafif ticari araç satışlarının ekonomik analizine ve otomotiv sektörüne yönelik tahmin süreçlerine önemli bir katkı sunmuştur.

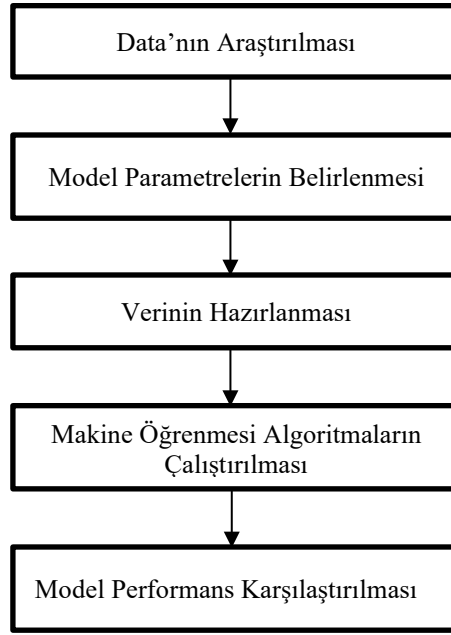
Dere ve Kasapbaşı (2024) tarafından gerçekleştirilen çalışmada, Türkiye'deki ikinci el araç fiyatlarının tahmini için makine öğrenmesi yöntemleri uygulanmıştır. Çalışmada, sahibinden.com platformundan Selenium test aracı kullanılarak 01.03.2023 ile 22.06.2023 tarihleri arasında toplanan 26.000'den fazla veri ve 28 öznitelik içeren bir veri seti kullanılmıştır. Veri setinde araçların markası, modeli, üretim yılı, vites tipi, motor gücü, kilometre bilgisi gibi özellikler yer alırken, fiyat tahmini için etkili olmayan açıklama ve URL gibi öznitelikler çıkarılmıştır. Fiyat tahmininde Doğrusal Regresyon, Rastgele Orman (Random Forest), Gradyan Artırıcı (GBoosted) ve Ağaç Topluluğu (Tree Ensemble) algoritmaları KNIME Analytics Platform 4.7.3 kullanılarak uygulanmıştır. Performans ölçütü olarak R^2 metriği kullanılmış ve en yüksek doğruluk Rastgele Orman algoritmasıyla (%83) elde edilmiştir. Çalışma, ikinci el araç fiyatlarının tahmininde farklı makine öğrenmesi yöntemlerinin performansını karşılaştırarak bu alana önemli katkılar sunmuştur.

Son olarak, Gür vd. (2024), MLP ve RBF yapay sinir ağı modellerini kullanarak Türkiye'nin aylık binek otomobil ihracatını tahmin etmeye odaklanmıştır. Çalışmada bağımlı değişken olarak otomobil ihracatı, bağımsız değişkenler olarak ise ithalat miktarları, dolar kuru, yeni otomobil satış adetleri, motorlu kara taşıtları üretim endeksi ve yurtdışı üretici fiyat endeksi kullanılmıştır. Sonuçlar, önerilen modellerin tahmin performansının yüksek olduğunu ortaya koymuştur.

Bu literatür, makine öğrenmesi yöntemlerinin otomobil satış tahmininde önemli bir potansiyele sahip olduğunu ve farklı değişkenlerle zenginleştirildiğinde doğruluk oranlarının artırılabilirliğini göstermektedir.

3. Materyal ve Yöntem

Bu çalışmada, araç satış miktarının tahmin edilmesi amacıyla çeşitli makine öğrenmesi algoritmaları ve yapay sinir ağı modelleri kullanılmıştır. Kullanılan yöntemler arasında Multi-Layer Perceptron (MLP), Radyal Tabanlı Fonksiyon (RBF), Lineer Regresyon, Ridge Regresyon, Lasso Regresyon, Karar Ağaçları, Rastgele Orman (Random Forest), Destek Vektör Regresyonu (SVR) ve XGBoost yer almaktadır. Bu algoritmaların seçilmesindeki temel neden, farklı veri yapılarına uygunlukları ve regresyon problemlerindeki tahmin performanslarının literatürde sıklıkla kanıtlanmış olmasıdır.



Şekil 1. Araştırma Akış Şeması

Bu çalışma kapsamında öncelikle kaggle platformundan araç satışı için veri araştırması yapılmıştır. Elde edilen veri .csv formatında kayıt edilmiştir. Araştırmada, 23.900 veri noktasından oluşan özgün bir veri seti kullanılmıştır. Veri seti, 11 bağımsız değişken ve bir bağımlı değişkenden oluşmaktadır. Bağımlı değişken, araçların aylık satış miktarını ifade etmekte olup, bağımsız değişkenler araçların ve müşterilerin özelliklerinden oluşmaktadır. Bağımsız değişkenler arasında satış tarihi, aracın marka ve modeli, motor hacmi, vites türü, renk, kasa tipi, araç fiyatı gibi araçla ilgili özelliklerin yanı sıra, aracı alan müşterinin cinsiyeti, aylık geliri ve aracın satıldığı bölge gibi müşteri ve satış koşullarına ilişkin değişkenler yer almaktadır.

Çalışmanın yöntemsel süreci, veri setinin işlenmesi ve analiz edilmesi üzerine yapılandırılmıştır. Veri seti, önce makine öğrenmesi algoritmalarına uygun hale getirilmiş, ardından eğitim ve test aşamalarında kullanılmak üzere %80 eğitim, %20 test verisi olarak ayrılmıştır. Veri ön işleme aşamasında eksik veriler doldurulmuş, kategorik değişkenler One-Hot Encoding yöntemi ile sayısal verilere dönüştürülmüş ve sayısal değişkenler normalize edilerek ölçeklendirilmiştir. Bu işlemler, veri setindeki potansiyel dengesizliklerin modellenme sürecine olumsuz etkisini azaltmayı hedeflemiştir.

Makine öğrenmesi algoritmaların çalıştırılması adımında veri seti üzerinde belirlenen makine öğrenmesi algoritmaları kullanılarak model eğitimi gerçekleştirilmiştir. Eğitim veri seti algoritmaların öğrenme sürecinde kullanılırken, doğrulama veri seti modelin hiperparametre optimizasyonunda değerlendirilmiştir. Test veri seti ise nihai model performansını değerlendirmek amacıyla kullanılmıştır. Çalışmada, Multi-Layer Perceptron (MLP), Radyal Tabanlı Fonksiyon (RBF), Lineer Regresyon, Ridge Regresyon, Lasso Regresyon, Karar Ağaçları, Rastgele Orman (Random Forest), Destek Vektör Regresyonu (SVR) ve XGBoost gibi makine öğrenmesi teknikleri uygulanmış ve her bir algoritma, performans ölçütleri olan MAE, MAPE ve R^2 kullanılarak karşılaştırılmıştır. Algoritmaların çalıştırılmasında Python programlama dili kullanılmıştır.

Bu bölümde, çalışmada kullanılan veri setinin kapsamı, tahmin modelleri ve uygulanan yöntemler detaylı bir şekilde açıklanmıştır. Elde edilen veriler ve kullanılan algoritmalar aracılığıyla araç satış tahminlerinde doğruluğun artırılması ve bu alandaki literatüre anlamlı katkılar sağlanması amaçlanmıştır.

3.1. Veri Seti

Bu çalışmada, Ocak 2022 ile Aralık 2023 dönemini kapsayan aylık araç satış verileri kullanılmıştır. İlgili veri seti 23.900 gözlem içermekte ve ABD pazarındaki araç satışlarına yönelik bilgiler sunmaktadır. Veri seti, Kaggle platformundan sağlanan açık kaynaklı bir veri setinin düzenlenmesiyle elde edilmiştir (Kaynak: Kaggle Car Sales Report). Veri kalitesini değerlendirmek adına eksik veriler analiz edilmiş, uygun dönüşümler sağlanmış ve aykırı değerler tespit edilerek işlenmiştir. Veri setinin gerçek dünya verilerini ne ölçüde temsil ettiğini anlamak için dağılım analizleri yapılmış ve örneklem istatistikleri incelenmiştir. Fakat veri setinin tüm ABD otomotiv pazarını eksiksiz şekilde temsil edip etmediği doğrudan test edilememiştir. Tahmin modellerinde bağımlı değişken olarak dönemsel satış adetleri kullanılmış olup, bağımsız değişkenler ise araçlara ve müşterilere ait çeşitli özelliklerden oluşmaktadır. Veri setindeki değişkenler, modelin doğruluğunu artırmak ve gereksiz gürültüyü önlemek amacıyla dikkatlice seçilmiştir. Car_id, Date, Customer Name, Dealer_Name, Dealer_No, Phone gibi değişkenler, modelin tahmin performansına doğrudan katkıda bulunmayacağı için dışlanmıştır. Ayrıca müşteri isimleri veya telefon numaraları gibi kişisel bilgiler tahmin için anlamlı olmadığı gibi, modelde veri sızıntısına (data leakage) yol açabileceği için kullanılmamıştır. Satış tarihi, müşterinin cinsiyeti, aylık geliri, araç özellikleri ve satış fiyatı gibi değişkenler ise satış adetlerini etkileyebilecek önemli faktörler olduğu için modele dahil edilmiştir. Veri setinde yer alan tüm değişkenler Tablo 1'de detaylı şekilde sunulmaktadır.

Tablo 1. Bağımlı ve Bağımsız Değişkenler

Bağımlı Değişken	Bağımsız Değişkenler
Satış Adedi	Satış Tarihi
	Müşterinin Cinsiyeti
	Müşterinin Aylık Geliri
	Araç Markası
	Araç Modeli
	Araç Motoru
	Vites Türü
	Araç Rengi
	Araç Fiyatı
	Araç Kasa Tipi
	Galeri Bölgesi

Çalışma kapsamında modellerin hangi bağımsız değişkenlere daha fazla önem verdiğini analiz etmek amacıyla özellik önem sıralaması (feature importance) analizi yapılmıştır. Bu analiz sonucunda, araç fiyatı, müşterinin aylık geliri ve araç markası gibi değişkenlerin model tahminlerinde en belirleyici faktörler olduğu gözlemlenmiştir. Aykırı değerlerin tespiti için ise z-skoru ve IQR (Interquartile Range) yöntemleri kullanılmıştır. Aykırı değer olduğu tespit edilen veriler modelin performansını düşürebileceğinden, bu gözlemler ya veri setinden çıkarılmış ya da uç değerleri sınırlayan yöntemler uygulanmıştır.

Çalışmada, Yapay Sinir Ağları modellerinden Multi-Layer Perceptron (MLP) ve Radyal Tabanlı Fonksiyon (RBF) modelleri ile araç satış tahmini gerçekleştirilmiştir. Modellerin uygulanmasında, Python programlama dili ve Google Colab platformu kullanılmıştır. Veri dönüşümü ve modelleme süreçleri bu platformda Python kütüphaneleri yardımıyla gerçekleştirilmiştir.

Çalışmada kullanılan veri seti ABD pazarına ait olduğundan, enflasyon oranı görece düşük seviyede (%3 civarında) olup, kısa vadede araç fiyatları üzerindeki etkisi sınırlıdır. Çalışma kapsamında model bireysel araç fiyat tahmini üzerine odaklandığından, makroekonomik göstergelerin (döviz kuru, faiz oranları vb.) doğrudan dahil edilmesi metodolojik olarak uygun olmayabilir.

3.1.1. Veri Ön İşleme

Veri setinin analiz için uygun hale getirilmesi amacıyla çeşitli ön işleme adımları uygulanmıştır:

1. **Veri Formatı ve Yükleme:** Kaggle'dan indirilen veri seti, tarih ve satış adetleri gibi bilgileri içeren bir Excel dosyası olarak Colab ortamına yüklenmiştir.
2. **Kategorik Değişkenlerin Kodlanması:** Veri setinde yer alan kategorik değişkenler (örneğin, müşterinin cinsiyeti, araç markası, kasa tipi) One-Hot Encoding yöntemi ile sayısal verilere dönüştürülmüştür. Bu yöntem, kategorik verilerin modellemede kullanılabilir hale getirilmesini sağlamıştır.
3. **Tarihsel Verilerin İşlenmesi:** Satış tarihi bilgisi, yıl ve ay formatına dönüştürülerek gruplandırılmıştır. Bu işlem, datetime kütüphanesi kullanılarak gerçekleştirilmiş ve tarih verileri tahmin modelleri için hazır hale getirilmiştir.
4. **Veri Setinin Ayrıştırılması:** Dönüştürülen veriler, eğitim ve test verisi olarak ikiye ayrılmıştır. Veri setinin %80'i eğitim, %20'si ise test için kullanılmıştır. Bu ayırım model performansının nesnel bir şekilde değerlendirilmesi için yapılmıştır.
5. **Sayısal Verilerin Standardizasyonu:** Sayısal değişkenler, farklı ölçeklerin etkisini minimize etmek amacıyla standardizasyon tekniklerine tabi tutulmuştur.
6. **Verilerin Saklanması:** Dönüşümden geçen veriler farklı dosya isimleriyle kaydedilerek gelecekteki işlemler için korunmuştur.

3.1.2. Modelleme Süreci

Veri seti hazırlandıktan sonra, farklı tahmin modelleri döngüsel olarak çalıştırılmış ve elde edilen sonuçlar görsel ve metrik tabanlı olarak raporlanmıştır. Modellerin bağımlı değişkeni aylık satış adetleri olup, bağımsız değişkenler arasında müşterinin cinsiyeti, aylık geliri, aracın marka ve modeli, vites türü, kasa tipi, motor hacmi, renk ve satıldığı bölge gibi faktörler yer almıştır.

Makine öğrenmesi modellerinin doğruluğunu artırmak için hiperparametre optimizasyonu sürecinde Lasso ve Ridge regresyonu gibi modeller için grid search yöntemi kullanılarak en uygun alpha değeri belirlenmiştir. XGBoost ve SVR modelleri için ise Grid Search CV (Grid Search Cross-Validation) ve Randomized Search CV yöntemleri uygulanmıştır. XGBoost modelinde learning_rate, max_depth, n_estimators ve subsample gibi kritik hiperparametreler optimize edilmiştir. SVR modeli için ise C, epsilon ve gamma parametreleri üzerinde ince ayarlamalar yapılmış, RBF çekirdek fonksiyonuna ek olarak doğrusal (linear) ve polinom (poly) çekirdek fonksiyonları da test edilmiştir.

Bu süreçler sayesinde, veri seti üzerindeki karmaşık ilişkiler anlaşılmış ve araç satış tahminleri için etkin modeller oluşturulmuştur. Elde edilen sonuçlar, sektörel analizler ve stratejik planlamalar için değerli bilgiler sunmaktadır.

3.2. Multi - Layer Perceptron (MLP)

Multi-Layer Perceptron (MLP), biyolojik sinir hücrelerini taklit ederek oluşturulmuş bir yapay sinir ağı modeli olup, 1986 yılında Rumelhart vd. tarafından geliştirilmiştir. Bu model, hata yayma ya da geri yayılım modeli olarak da bilinmekte ve temel amacı, ağın beklenen çıktısı ile ürettiği çıktı arasındaki hatayı en aza indirmektir (Kıvrak, 2016). MLP, giriş katmanı, bir veya daha fazla gizli katman ve bir çıkış katmanından oluşan çok katmanlı bir sinir ağı yapısına sahiptir. Her bir katmanda, bir önceki katmandan gelen sinyaller alınarak bir aktivasyon fonksiyonu yardımıyla hesaplama yapılır. Hesaplanan sonuçlar, bir sonraki katmana iletilerek ileri yayılım gerçekleştirilir. MLP, karmaşık problemleri modelleme ve çok boyutlu veri setlerinden tahmin yapma amacıyla yaygın olarak kullanılmaktadır (Gür vd., 2024: 25).

MLP'nin ilk katmanında, veri setindeki özellik sayısına eşit sayıda nöron bulunmaktadır. Bu giriş katmanını takiben gelen gizli katmanlar, ağı daha karmaşık problemleri çözmesine olanak tanır. Gizli katmanlardaki nöron sayısının artırılması veya daha fazla gizli katman eklenmesi, ağı daha karmaşık ilişkileri öğrenmesini sağlamaktadır (Bhushan vd., 2012: 135). Bununla birlikte, literatürde gizli katmandaki nöron sayısının belirlenmesi için kesin bir kural bulunmamaktadır ve bu sayı genellikle deneme-yanılma veya hiperparametre optimizasyonu yöntemleriyle belirlenmektedir (Aslay ve Özen, 2013: 142).

Çıkış katmanı ise, gizli katmandan aldığı girdilerle nihai hesaplamayı yaparak tahmin sonucunu üretir. Literatürde, çıkış katmanındaki nöron sayısının genellikle giriş katmanındaki nöron sayısına eşit olması gerektiği vurgulanmaktadır (Hamid ve Iqbal, 2004: 1118). MLP'nin bu esnek ve güçlü yapısı, karmaşık ilişkilerin öğrenilmesi ve doğrusal olmayan problemler için uygun bir çözüm sunmasını mümkün kılmaktadır.

MLP'nin bu temel özellikleri, çalışmamızda kullanılan araç satış tahmin modelinin geliştirilmesinde önemli bir rol oynamıştır. Aktivasyon fonksiyonları ve geri yayılım algoritması aracılığıyla, model veri setindeki karmaşık ilişkileri öğrenerek tahmin doğruluğunu artırmayı hedeflemiştir.

3.3. Radial Basis Function (RBF)

Radial Basis Function (RBF), ileri beslemeli bir yapay sinir ağı modeli olup, girdi verileri ile çıkışlar arasındaki ilişkileri modellemek için gizli katmanlarında radyal tabanlı fonksiyonlar kullanan bir mimariye sahiptir. RBF ağı, genellikle bir giriş katmanı, bir gizli katman ve bir çıkış katmanından oluşan üç katmanlı bir yapıya sahiptir (Gür vd., 2024: 25). Bu mimari, basit yapısı sayesinde diğer yayılım algoritmalarına kıyasla daha hızlı bir şekilde eğitilebilmektedir (Kıvrak, 2016: 59).

RBF ağındaki her bir gizli katman nöronunun aktivasyon değeri, girdi verisi ile önceden belirlenmiş bir merkez arasındaki mesafeye dayalı olarak hesaplanır. Bu mesafe genellikle Öklid mesafesi ile ölçülür ve elde edilen değerler, radyal tabanlı bir fonksiyon (örneğin, Gauss fonksiyonu) aracılığıyla işlenir. Gizli katmandaki nöronlardan elde edilen aktivasyon değerleri, çıkış katmanına ağırlıklandırılarak iletilir ve nihai çıktı, gizli katman nöronlarının ağırlıklı toplamı olarak hesaplanır (Gür vd., 2024: 25).

RBF ağlarının basit yapısı, karmaşık veri ilişkilerini modelleme kabiliyetine ek olarak hızlı eğitim süresiyle de öne çıkmaktadır. Bu özellik, özellikle büyük veri setleriyle çalışırken avantaj sağlamaktadır. RBF ağlarının güçlü tahmin performansı hem regresyon hem de sınıflandırma problemlerinde etkili bir şekilde kullanılmasını mümkün kılmaktadır.

Çalışmamızda, RBF ağı, araç satış miktarlarının tahmin edilmesinde kullanılan bir diğer model olarak değerlendirilmiştir. RBF'nin aktivasyon fonksiyonları ve hızlı öğrenme yetenekleri, modelin doğruluk ve verimlilik açısından katkı sağlamasına olanak tanımıştır.

3.4. Lineer Regresyon

Lineer Regresyon, bir bağımlı değişken (genellikle y ile gösterilir) ile bir veya daha fazla bağımsız değişken ($x_1, x_2, \dots, x_p$) arasındaki doğrusal ilişkiyi modellemek için kullanılan temel bir istatistiksel yöntemdir. Bu model, bağımlı değişkeni tahmin etmek amacıyla doğrusal bir denklem oluşturur ve bu denklemin katsayıları, bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini açıklar. Bağımlı değişken genellikle açıklanan değişken, tahmin edilen değer veya yanıt değişkeni olarak adlandırılırken, bağımsız değişkenler açıklayıcı değişkenler, kontrol değişkenleri ya da regresörler olarak tanımlanır (Yan ve Su, 2009). Lineer Regresyon, özellikle veri analizi ve tahmin problemlerinde kullanımı kolay, hızlı ve etkili bir yöntem olarak öne çıkar.

Basit doğrusal regresyon, bir bağımlı ve bir bağımsız değişken arasındaki doğrusal ilişkiyi modellemek için kullanılan bir varyasyonudur. Örneğin, kas gücü (y) ile yağsız vücut kütlesi (x) arasındaki ilişkiyi modellemek için kullanılabilir. Bunun yanı sıra, çoklu doğrusal regresyon, birden fazla bağımsız değişkenin bağımlı değişken üzerindeki etkisini analiz etmeye olanak tanır. Model, tahmin hatasını en aza indirmek için genellikle En Küçük Kareler Yöntemi (Ordinary Least Squares) ile eğitilir (Montgomery vd., 2012). Ancak modelin doğruluğu, bağımsız değişkenler arasındaki çoklu bağlantı (multicollinearity) ve bağımlı değişken ile bağımsız değişkenler arasındaki doğrusal varsayımlara bağlıdır. Lineer Regresyon, özellikle büyük veri setlerinde hızlı tahmin yapabilme kapasitesi ve genelleştirme yeteneği ile çeşitli sektörlerde sıklıkla kullanılan temel bir yöntemdir.

3.5. Ridge Regresyon

Ridge Regresyon, çoklu doğrusal bağlantı (multicollinearity) problemlerini çözmek ve standart doğrusal regresyonun tahmin performansını iyileştirmek için kullanılan bir regresyon tekniğidir. Çoklu doğrusal bağlantı, bağımsız değişkenler arasında yüksek korelasyon olduğunda ortaya çıkar ve katsayıların kararsız hale gelmesine, dolayısıyla model performansının düşmesine yol açabilir. Ridge Regresyon, doğrusal regresyonun maliyet fonksiyonuna bir L2 normu ceza terimi ekleyerek bu sorunu giderir. Bu ceza terimi, katsayıların büyüklüğünü kontrol altına alır, böylece modelin genelleme yeteneği artırılır ve aşırı öğrenme (overfitting) riski azaltılır (Hoerl ve Kennard, 1970).

Ridge Regresyonun arkasındaki temel fikir, modelin eğitim verilerine tam olarak uyum sağlamasından ziyade, test verilerinde daha iyi tahmin performansı sergilemesini sağlamaktır. Bu nedenle, Ridge Regresyon modeli, eğitim verilerine daha az uyum sağlasa da, varyansı önemli ölçüde azaltarak uzun vadeli tahminlerde doğruluğu artırır (Arseven ve Çınar, 2023). Bu süreçte, modele eklenen ceza terimi (λ), eğitim verilerine uyan yeni bir doğrusal denklem oluşturmaya yardımcı olur ve bu denklemin optimal parametreleri genellikle çapraz doğrulama (cross-validation) ile belirlenir (Friedman vd., 2001). Ridge Regresyon, özellikle yüksek boyutlu veri setleri ve çok sayıda açıklayıcı değişkenin bulunduğu durumlarda etkili bir yöntemdir ve finansal analizden genetik veri analizine kadar geniş bir uygulama alanına sahiptir.

3.6. Lasso Regresyon

LASSO Regresyon, bağımsız değişkenler arasındaki yüksek korelasyonu yönetmek ve modelde değişken seçimi yapmak için kullanılan etkili bir düzenleme tekniğidir. Bu yöntem, doğrusal regresyonun maliyet fonksiyonuna bir L1 normu ceza terimi ekler ve bu ceza mekanizması, gereksiz veya önemsiz değişkenlerin regresyon katsayılarını sıfıra indirerek modelden çıkarılmasını sağlar (Tibshirani, 1996). Böylece LASSO hem regresyon analizi yapar hem de öznetelik seçimini otomatik olarak gerçekleştirir. Bu özelliği sayesinde, özellikle yüksek boyutlu veri setleri ve çok sayıda açıklayıcı değişkenin bulunduğu durumlarda etkili bir çözüm sunar. Tahmin hatasını en aza indirirken, modeli oluşturan değişkenleri ve bu değişkenlere karşılık gelen regresyon katsayılarını optimize etmeyi amaçlar.

LASSO Regresyon, aşırı risk öngörme ve düşük performans sergileme eğiliminde olan standart regresyon yöntemlerine kıyasla bazı durumlarda daha iyi performans gösterebilmektedir. Bununla birlikte, bu modelin odak noktası, bağımsız değişkenlerin tek tek katkılarını doğru bir şekilde yorumlamaktan ziyade, değişkenlerin en iyi kombinasyonunu bulmaya yöneliktir. Bu nedenle, regresyon katsayılarının bağımsız risk faktörleri açısından yorumlanabilirliği sınırlı olabilir (Ranstam ve Cook, 2018). Ayrıca, LASSO'nun ceza terimini belirleyen hiperparametresi (λ) genellikle çapraz doğrulama yöntemiyle optimize edilerek modelin performansı artırılır (Hastie vd., 2009). LASSO Regresyon, genetik analizden finansal modellemeye kadar birçok farklı alanda yaygın olarak kullanılmakta ve özellikle karmaşık veri yapılarına sahip problemlerde tercih edilmektedir.

3.7. Karar Ağaçları (Decision Trees)

Karar Ağaçları hem sınıflandırma hem de regresyon problemleri için kullanılan, veri odaklı bir modelleme yaklaşımıdır. Regresyon işlemleri bağlamında, Karar Ağaçları bağımlı değişkeni tahmin etmek için veriyi ardışık olarak bölerek doğrusal olmayan ilişkileri modelleyebilir. Model, bağımsız değişkenler üzerindeki eşik değerlerini kullanarak veriyi dallara ayırır ve bu dalların uç noktalarında (yaprak düğümlerinde) tahminler yapar. Her yaprak düğüm, veri alt kümelerinin ortalamasını veya ağırlıklı değerlerini tahmin etmek için kullanılır. Bu özelliği, Karar Ağaçlarını doğrusal varsayımlara ihtiyaç duymadan karmaşık ilişkileri anlamlandırmada etkili bir yöntem haline getirir (Breiman vd., 1984). Ancak, Karar Ağaçları aşırı öğrenme (overfitting) eğiliminde olabilir; bu sorunu azaltmak için genellikle ağaç budama teknikleri veya belirli derinlik sınırları uygulanır. Karar Ağaçları, görsel olarak yorumlanabilir yapısı ve kolay uygulanabilirliği nedeniyle regresyon problemlerinde sıklıkla tercih edilmektedir.

3.8. Rastgele Orman (Random Forest)

Rastgele Orman (Random Forest), regresyon ve sınıflandırma problemlerinde yüksek doğruluk sunan, karar ağaçlarına dayalı bir topluluk öğrenme (ensemble learning) yöntemidir. Regresyon bağlamında, Rastgele Orman, birden fazla karar ağacını rastgele seçilen veri alt kümeleri ve özelliklerle eğiterek oluşturur. Model, her bir karar ağacının tahminlerini birleştirerek (genellikle ortalamasını alarak) nihai tahmini üretir. Bu yapı, modelin tek bir karar ağacına kıyasla aşırı öğrenme (overfitting) eğilimini azaltmasını ve daha genelleştirilebilir tahminler sunmasını sağlar. Rastgele Orman, doğrusal olmayan ilişkileri modelleyebilme yeteneği ve eksik verilerle başa çıkabilmesi nedeniyle regresyon problemlerinde sıklıkla tercih edilmektedir. Ancak, modelin karmaşıklığı ve büyük veri setleri üzerindeki hesaplama maliyeti, hiperparametrelerin dikkatlice optimize edilmesini gerektirir (Breiman, 2001). Özellikle yüksek boyutlu veri setlerinde, Rastgele Orman'ın performansı genellikle diğer geleneksel regresyon yöntemlerini aşmaktadır.

3.9. Destek Vektör Regresyonu (SVR)

Destek Vektör Regresyonu (SVR), doğrusal ve doğrusal olmayan ilişkileri modelleyebilme yeteneğiyle regresyon problemlerinde etkili bir yöntemdir. SVR, temelinde destek vektör makineleri (SVM) algoritmasını kullanır ve tahmin doğruluğunu artırmak için bir hata toleransı (ϵ) tanımlayarak, tahmin edilen değer ile gerçek değer arasındaki hatayı bu aralık içinde tutmayı hedefler. Model, bu hata toleransı dışında kalan veriler için en iyi uyumu sağlayacak hiper düzlemi oluşturur. Çekirdek (kernel) fonksiyonları, SVR'nin doğrusal olmayan ilişkileri modellemesine olanak tanır; yaygın olarak kullanılan çekirdekler arasında doğrusal, polinomal ve radyal tabanlı fonksiyonlar (RBF) bulunur. SVR, büyük veri setlerinde yüksek doğruluk sağlarken, az sayıda destek vektörü kullandığı için bellek verimliliği de sunar (Smola ve Schölkopf, 2004). Ancak, hiperparametrelerin dikkatlice optimize edilmesi gerekmektedir, çünkü bu parametreler modelin genelleştirme yeteneği üzerinde doğrudan etkilidir.

3.10. XGBoost

XGBoost, Gradient Boosting algoritmasına dayanan ve yüksek doğruluk sunan güçlü bir makine öğrenmesi modelidir. Regresyon işlemleri bağlamında, XGBoost, doğrusal olmayan ilişkileri modellemek için ardışık karar ağaçlarını birleştirerek çalışır ve her bir ardışık model, önceki modelin yaptığı hataları öğrenerek tahmin doğruluğunu artırır. Modelin temelinde yer alan optimizasyon teknikleri, hesaplama hızını artırırken, overfitting'i önlemek için düzenleme (regularization) terimlerini de içerir. Özellikle L1 ve L2 ceza terimlerinin kullanımı, XGBoost'u diğer gradient boosting modellerine kıyasla daha dayanıklı ve esnek hale getirir. Hızlı hesaplama yeteneği ve büyük

veri setleri üzerindeki yüksek performansı nedeniyle, XGBoost, regresyon problemlerinde sıklıkla tercih edilmektedir (Chen ve Guestrin, 2016). Ayrıca, modelin hiperparametrelerinin geniş bir yelpazede ayarlanabilir olması, farklı veri yapılarına ve problemlere uyarlanmasını sağlar.

3.11. Model Performans Değerlendirmesi

Çalışma sonuçlarını analiz etmek ve yorumlamak amacıyla, model performansının değerlendirilmesinde Ortalama Karesel Hata (Mean Square Error - MSE) ve Ortalama Mutlak Hata Oranı (Mean Absolute Percentage Error - MAPE) metrikleri kullanılmıştır. MAPE, hataların mutlak değerlerinin gerçek değerlere oranlanarak yüzdesel olarak ifade edilmesiyle hesaplanır ve özellikle tahmin sonuçlarının gerçek değerlerin altında veya üstünde olma durumunu anlamak için etkili bir yöntemdir. Bu metrik, model doğruluğunu değerlendirmede önemli bir referans sağlamaktadır.

Literatürde MAPE değerlerine ilişkin çeşitli sınıflandırma kriterleri bulunmaktadır. Witt ve Witt, MAPE değeri %10'un altında olan modelleri "yüksek doğruluk" derecesine sahip olarak nitelendirirken, %10 ile %20 arasındaki modelleri "doğru tahmin" kategorisinde değerlendirmiştir. Benzer şekilde, Lewis (Kıvrak, 2016), MAPE değerlerini şu şekilde sınıflandırmıştır:

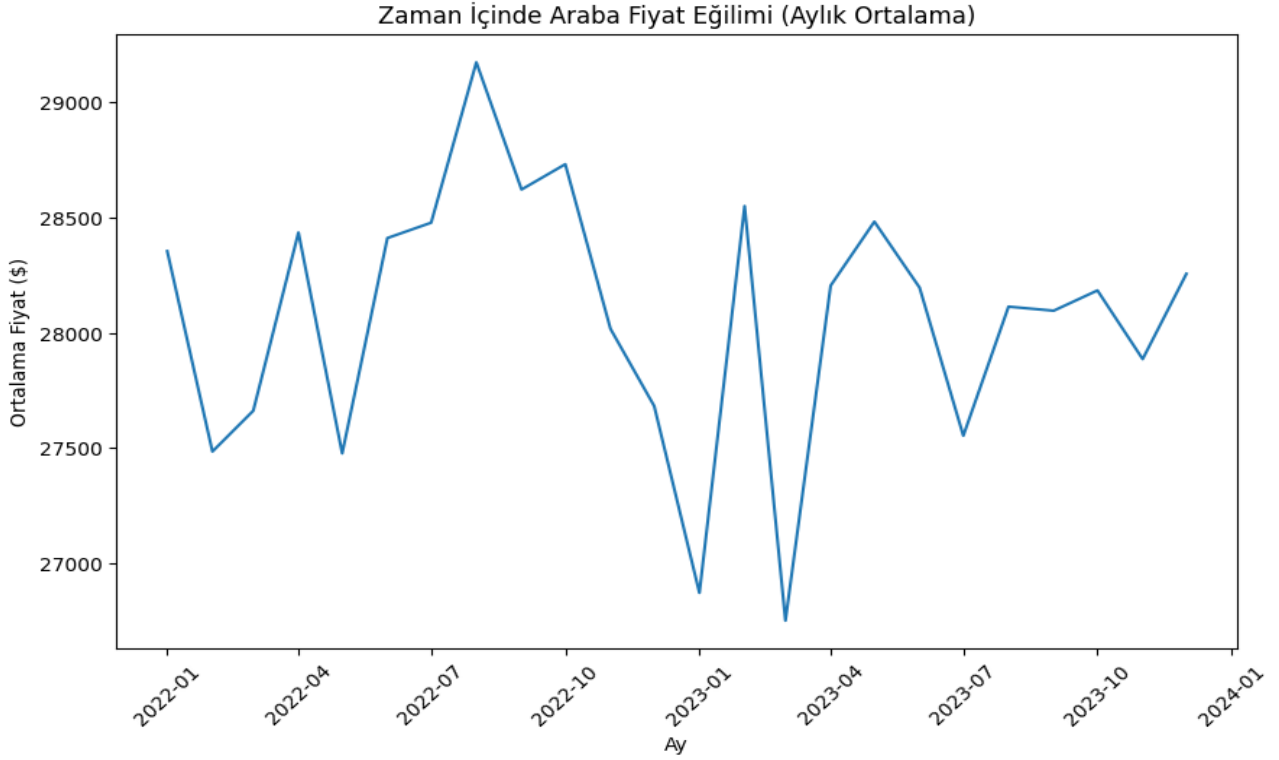
- %10'un altında: Son derece doğru tahmin
- %10 - %20: İyi tahmin
- %20 - %50: Makul tahmin
- %50'nin üzerinde: Yanlış tahmin

Bu sınıflandırma, modellerin doğruluk düzeylerinin değerlendirilmesi ve karşılaştırılmasında temel bir rehber olarak kullanılmaktadır. MAPE ve MSE gibi metriklerin bir arada kullanılması, modellerin yalnızca doğruluk oranlarını değil, aynı zamanda hataların boyutlarını ve dağılımlarını da detaylı bir şekilde anlamaya olanak tanımaktadır. Bu yaklaşımla, çalışmada kullanılan tahmin modellerinin etkinliği kapsamlı bir şekilde değerlendirilmiştir.

4. Bulgular

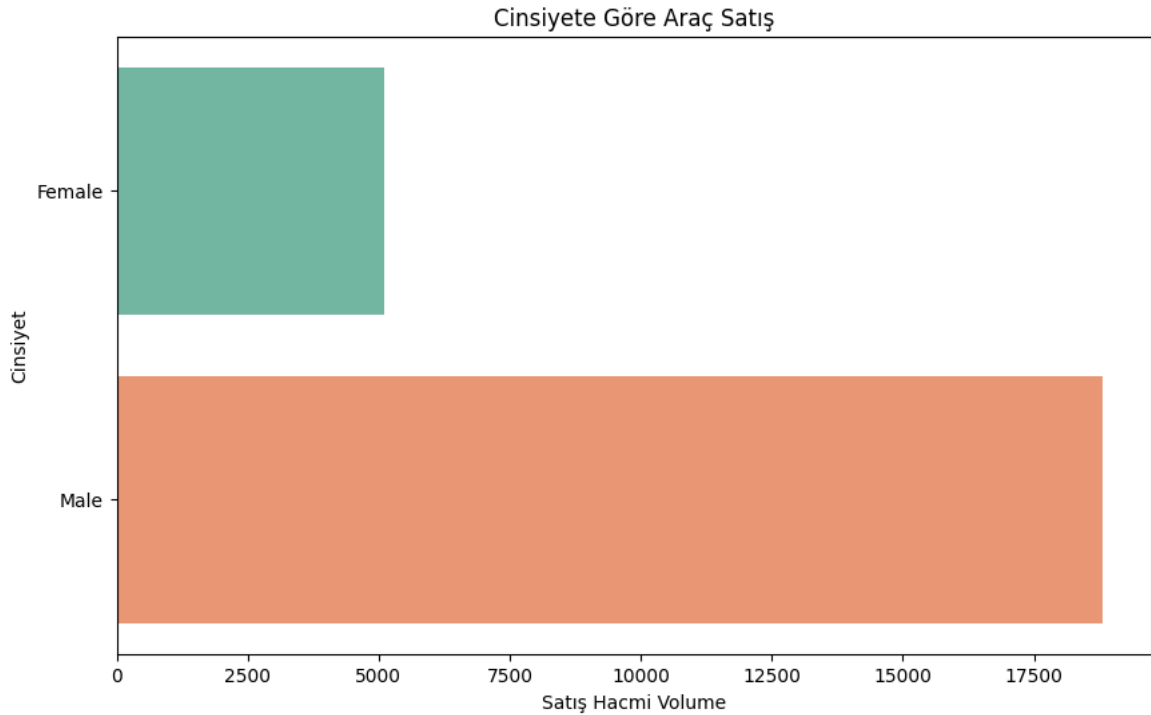
Şekil 2'de, 2022 yılının başından 2024 yılının başına kadar olan dönemdeki aylık ortalama araç fiyatlarının değişimi gösterilmektedir. Grafikte, zaman içinde fiyatlarda gözlemlenen dalgalanmalar dikkat çekmektedir. Araç fiyatları, genel olarak inişli çıkışlı bir eğilim sergilemekte ve belirli dönemlerde ani artışlar ya da düşüşler yaşanmaktadır.

2022 yılının ortalarında, özellikle Temmuz ve Ağustos aylarında, araç fiyatlarının zirveye ulaştığı gözlemlenmiştir. Satışların yaz aylarında artış göstermesi, tatil sezonunda araç talebinin yükselmesi, bireysel seyahat planlarının yoğunlaşması ve yeni model lansmanlarının bu dönemde daha sık gerçekleşmesiyle açıklanabilir. Bununla birlikte, yıl sonuna doğru artan kampanyalar ve filo yenileme süreçleri, satışlardaki mevsimsel dalgalanmaların önemli etkenleri arasında yer almaktadır. Bu dönemdeki fiyat artışı, yüksek talep, ekonomik belirsizlikler ve mevsimsel etkilerden kaynaklanabilir. Ancak, bu yükselişin ardından fiyatlar hızlı bir şekilde düşüş göstermiş ve 2022 yılının sonunda en düşük seviyelerine gerilemiştir. Bu dalgalanma, piyasadaki talep değişimleri ve makroekonomik faktörlerin etkisiyle şekillenmiş olabilir.



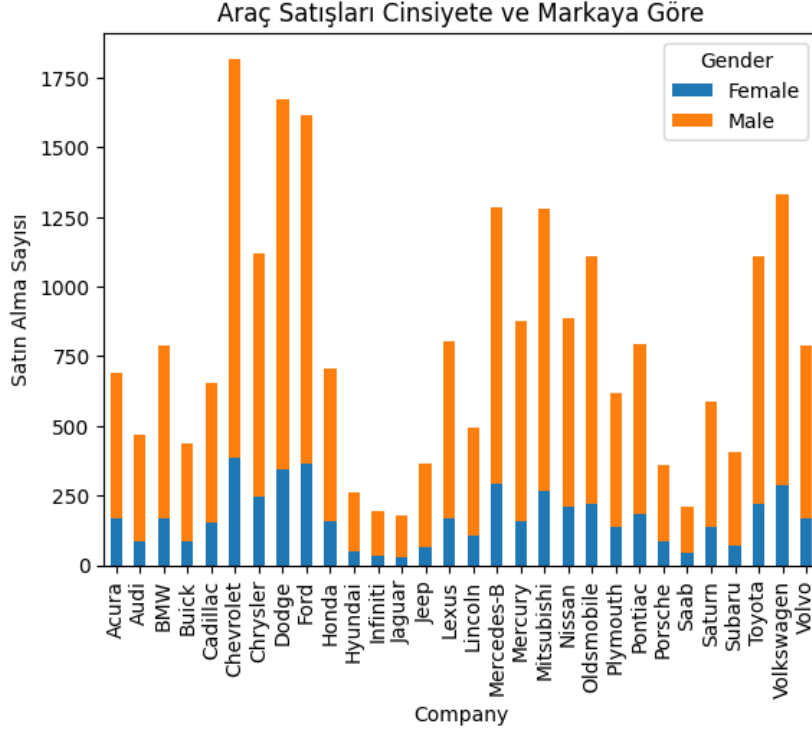
Şekil 2. Zaman içinde Araba Fiyat Eğilimi

2023 yılına gelindiğinde, fiyatlarda tekrar toparlanma eğilimi gözlemlenmektedir. İlk çeyrek itibarıyla fiyatlar yeniden artış gösterirken, yıl ortasında bir dalgalanma yaşanmış ve belirli dönemlerde fiyatlar tekrar düşüşe geçmiştir. Yılın sonlarına doğru ise fiyatlarda nispeten istikrarlı bir seyir görülmektedir. Bu dönemlerde fiyatların, ekonomik dengelenme veya sektörün talebe uygun şekilde yeniden yapılandırılması gibi faktörlerden etkilenmiş olabileceği değerlendirilmektedir.



Şekil 3. Cinsiyete Göre Araç Satış

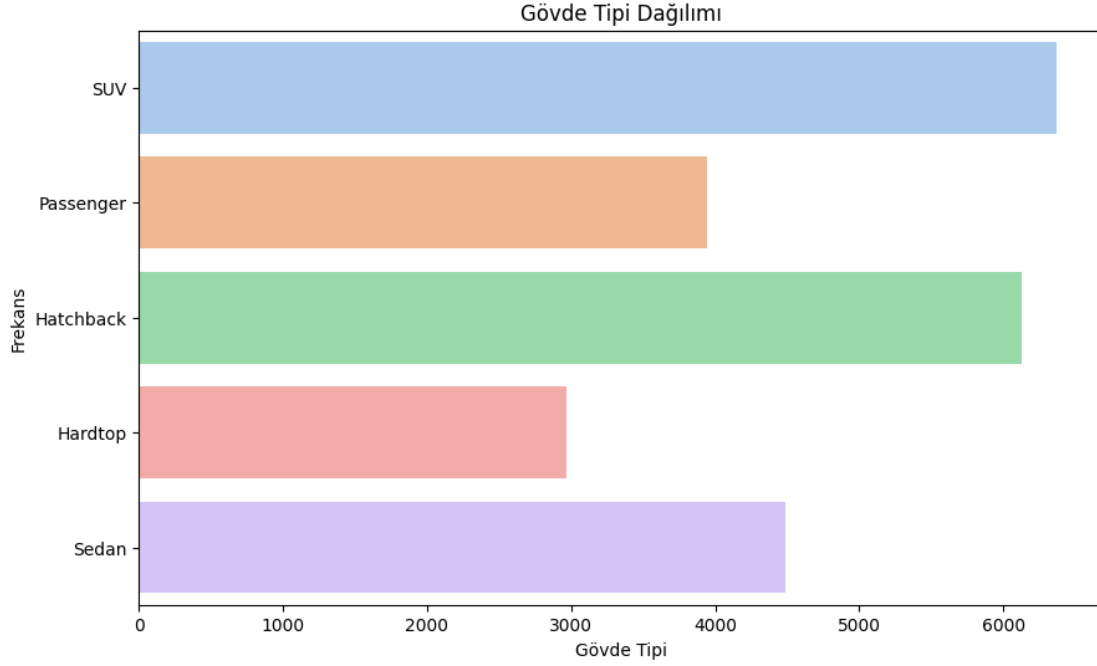
Şekil 3, otomobil satışlarının demografik analize göre 4.891 kadın alıcıya karşı 17.674 erkek alıcının olması önemli bir cinsiyet eşitsizliğini göstermektedir. Bu grafiğe göre otomotiv sektörü erkek odaklı olduğunu göstermekte ve kadın alıcılara yönelik hedefli pazarlama otomobil satışlarındaki temsiliyetin artırılması için potansiyel bir alan olduğunu göstermektedir.



Şekil 4. Araç satışları Cinsiyete ve Markaya Göre

Şekil 4'teki analiz sonuçlarına göre, belirli otomobil markaları erkek tüketiciler tarafından daha fazla tercih edilmektedir. Özellikle Chevrolet ve Dodge, pazarda güçlü bir konumda yer alırken, Ford ve Mercedes-Benz de yüksek satış rakamlarına ulaşmıştır. Buna karşılık, Infiniti ve Jaguar'ın satışlarının daha düşük seviyede olduğu gözlemlenmiş, bu durum ise söz konusu markaların araştırma yapılan tüketici tabanında daha az popüler olması ya da niş bir pazara hitap etmesiyle açıklanabilir. Elde edilen veriler, otomobil üreticilerinin değişen demografik tercihlere duyarlı olmaları gerektiğini göstermektedir.

Pazarlama stratejilerinin, tüketici segmentlerine özgü beklentileri dikkate alacak şekilde şekillendirilmesi büyük önem taşımaktadır. Özellikle kadın tüketicilere yönelik güvenlik, çevre dostu özellikler ve gelişmiş teknolojiye sahip araçlar gibi unsurları öne çıkaran pazarlama kampanyaları, kadın alıcıların otomobil sektöründeki temsiliyetini artırmada etkili olabilir. Analiz bulguları, erkek tüketicilerin genellikle yüksek motor gücüne ve geniş araçlara yöneldiğini, kadın tüketicilerin ise güvenlik, yakıt verimliliği ve teknolojik donanım gibi faktörlere daha fazla önem verdiğini ortaya koymaktadır. Bu farklılıklar, otomotiv sektöründe hedef kitleye yönelik özelleştirilmiş pazarlama stratejilerinin geliştirilmesi gerektiğini vurgulamaktadır.



Şekil 5. Gövde Tipi Dağılımı

Şekil 5'e göre, gövde stili tercihleri, popülerlikte önde gelen SUV'lerin ve onları yakından takip eden Hatchback'lerin açık bir şekilde baskın olduğunu ortaya koymaktadır. Sedanlar ve Binek araçlar daha düşük talep görürken, Hardtop'lar en az tercih edilen seçenek. Bu eğilim, tüketicilerin daha fazla alan ve çok yönlülük sunan daha büyük araçlara olan tercihini göstermektedir.

Özellikle Eylül ve Kasım aylarında olmak üzere belirli aylarda belirgin satış zirveleriyle verilerdeki önemli mevsimselliği vurgulamaktadır. Chevrolet, Ford ve Dodge tüm bölgelerde baskın otomobil markaları olarak ortaya çıkarken Hyundai, Saab ve Jaguar popülerlikte geride kalmaktadır. Bu da daha az popüler markalar tarafından stratejik bir yeniden değerlendirmeye ihtiyaç duyulabileceğini göstermektedir. SUV'ler ve Hatchback'ler en çok tercih edilen gövde stilleri olurken, Sedan'lar geride kalmaktadır. Bu durum ise daha büyük, daha çok yönlü araçlara yönelik bir pazar tercihinin ortaya koymaktadır.

Sonuç olarak, araç fiyatlarının zaman içerisindeki değişimi, piyasa dinamiklerinden, ekonomik faktörlerden ve mevsimsel etkilerden yoğun bir şekilde etkilenmiştir. Şekil 1'deki dalgalanmalar, otomotiv sektöründeki fiyat hareketlerinin karmaşıklığını ve fiyat tahmini yaparken dikkate alınması gereken çok boyutlu faktörlerin varlığını ortaya koymaktadır. Bu eğilimler, doğru tahmin modellerinin geliştirilmesi ve piyasa davranışlarının daha iyi anlaşılması için önemli veriler sunmaktadır.

Tablo 2. Model Performanslarının Karşılaştırılması

Tahmin Yöntemi	MAE	MAPE	R ²
Lineer Regresyon	9,657	0,95	0,99
Ridge Regresyon	9,678	1,11	0,99
Lasso Regresyon	6,074	0,94	0,99
Karar Ağacı	313,2	25,03	0,16
Random Forest	56,65	10,84	0,97
Gradient Boosting	68,54	10,39	0,97
SVR	490,81	60,28	-0,38
XGBoost	331,46	23,55	0,12
MLP	55,36	8,97	0,97
RBF	139,0	11,13	0,88

Tablo 2, farklı tahmin yöntemlerinin performans metriklerini karşılaştırarak araç fiyat tahminindeki başarılarını değerlendirmektedir. Performans değerlendirmesi için Ortalama Mutlak Hata (MAE), Ortalama Mutlak Yüzde Hata (MAPE) ve Determinasyon Katsayısı (R^2) metrikleri kullanılmıştır. Bulgular, modellerin doğruluğu ve genelleme kabiliyetleri açısından önemli farklılıklar gösterdiğini ortaya koymaktadır.

Lineer Regresyon, Ridge Regresyon ve Lasso Regresyon yöntemleri, oldukça düşük MAE ve MAPE değerleri ile öne çıkmış ve R^2 değerleri 0,99 ile en yüksek doğruluğu sağlamıştır. Bu sonuçlar, bu modellerin tahmin performanslarının oldukça başarılı olduğunu göstermektedir. Özellikle Lasso Regresyon, 6,074 MAE ve 0,94 MAPE değerleri ile bu üç model arasında en iyi performansı sunmuştur. Ridge Regresyon ve Lineer Regresyon modelleri benzer başarı oranları sergilemekle birlikte, Lasso Regresyon'un daha düşük hata değerlerine sahip olması dikkat çekicidir. Lasso'nun L1 düzenleme etkisiyle bazı gereksiz veya düşük etkili değişkenlerin katsayılarını sıfıra indirerek daha sade ve genelleştirilebilir bir model oluşturmaktan kaynaklanmaktadır. Böylece aşırı öğrenme (overfitting) riski azalırken, modelin tahmin doğruluğu artmıştır.

Karar Ağacı modeli, 313,2 MAE ve 25,03 MAPE ile oldukça yüksek hata değerleri sunmuş ve R^2 değeri yalnızca 0,16 olarak hesaplanmıştır. Bu durum, Karar Ağacı modelinin, araç fiyat tahmini için yetersiz kaldığını göstermektedir. Karar Ağaçları veri setine aşırı uyum sağlama eğiliminde olabilmekte, özellikle karmaşık ve sürekli değişkenli veri setlerinde yüksek varyans gösterebilmektedir. Çalıştırılan bu model belirli örnekler üzerine aşırı öğrenme yaparak genelleme performansını düşürebilmektedir. Bu durum çalışmada kullanılan veri setinin yapısıyla da ilişkilidir. Veri noktalarının geniş bir dağılım göstermesi ve değişkenler arasındaki ilişkilerin doğrusal olmaması, Karar Ağaçları'nın başarısını sınırlamış olabilir. Benzer şekilde, XGBoost algoritması da yüksek hata oranlarına (331,46 MAE ve 23,55 MAPE) ve düşük bir R^2 değerine (0,12) sahip olup, bu problemin çözümünde etkili bir yöntem olarak değerlendirilememektedir.

Random Forest, Gradient Boosting ve MLP modelleri ise oldukça benzer sonuçlar vermiştir. Random Forest, 56,65 MAE ve 10,84 MAPE ile tahmin doğruluğu açısından iyi bir performans sergilerken, Gradient Boosting ve MLP modelleri sırasıyla 68,54 MAE ve 55,36 MAE değerleriyle Random Forest'a yakın sonuçlar elde etmiştir. Üç modelin R^2 değerleri de 0,97 olarak hesaplanmış ve yüksek doğruluk sağlamıştır. Bu modeller, genelleme kabiliyetleri açısından güçlü bir alternatif olarak öne çıkmaktadır.

SVR ve RBF modelleri ise diğer yöntemlerle kıyaslandığında daha düşük performans sergilemiştir. Özellikle SVR, 490,81 MAE, 60,28 MAPE ve -0,38 R^2 değeriyle en düşük doğruluğa sahip yöntem olmuştur. Bu durum SVR'nin mevcut veri seti ve problem yapısı için uygun bir model olmadığını göstermektedir. SVR modeli özellikle büyük veri setlerinde ve yüksek boyutlu problemlerde hiperparametre hassasiyetine bağlı olarak değişken performans gösterebilmektedir. Çekirdek (kernel) seçimi ve C ile epsilon gibi hiperparametreler modelin genelleme yeteneğini doğrudan etkileyebilmektedir. Bu çalışmada varsayılan RBF çekirdeği kullanılmış olup, farklı çekirdek fonksiyonlarının (örneğin, polynomial veya linear kernel) denemesi modelin başarısını artırabilir. RBF modeli ise 139,0 MAE ve 11,13 MAPE değerleriyle daha makul bir performans göstermiş, ancak diğer güçlü modellerle kıyaslandığında yeterli bir doğruluğa ulaşamamıştır ($R^2 = 0,88$).

Genel olarak, analiz sonuçları, Lineer Regresyon, Ridge Regresyon, Lasso Regresyon, Random Forest, Gradient Boosting ve MLP modellerinin araç fiyat tahmininde yüksek doğruluk sunduğunu göstermektedir. Özellikle Lasso Regresyon, düşük hata oranlarıyla dikkat çekmekte; Random Forest, Gradient Boosting ve MLP ise genelleme kabiliyeti açısından etkili modeller olarak öne çıkmaktadır. Karar Ağacı, XGBoost ve SVR modelleri ise mevcut problem kapsamında yetersiz performans sergilemiştir. Bu bulgular, makine öğrenmesi ve regresyon yöntemlerinin tahmin performanslarının problemin yapısına ve veri özelliklerine bağlı olarak önemli ölçüde değişebileceğini göstermektedir.

5. Sonuç ve Öneriler

Bu çalışma, araç fiyat tahmini problemine yönelik farklı makine öğrenmesi algoritmalarının performanslarını detaylı bir şekilde analiz etmiş ve literatüre önemli katkılar sunmuştur. Çalışmada kullanılan modeller arasında Lineer Regresyon, Ridge Regresyon, Lasso Regresyon, Karar Ağaçları, Rastgele Orman (Random Forest), Gradient Boosting, Destek Vektör Regresyonu (SVR), XGBoost, Multi-Layer Perceptron (MLP) ve Radyal Tabanlı Fonksiyon (RBF) yer almaktadır. Elde edilen sonuçlar, modellerin doğruluk oranları ve genelleme kapasiteleri arasında belirgin farklılıklar olduğunu göstermiştir. Lasso Regresyon modeli, %99 R^2 değeri ve düşük hata oranları ile en başarılı yöntem olarak öne çıkarken, Rastgele Orman, Gradient Boosting ve MLP algoritmaları da yüksek doğruluk oranları ile güçlü alternatifler arasında yer almıştır. Ancak, Karar Ağaçları, SVR ve XGBoost algoritmalarının daha düşük performans sergilediği belirlenmiş, bu durum model başarısının veri setinin özelliklerine ve problem yapısına bağlı olarak önemli ölçüde değişebileceğini ortaya koymuştur.

Elde edilen bulgular, otomotiv sektöründe fiyat tahminleme süreçlerinin daha sistematik hale getirilmesine yönelik önemli katkılar sunmaktadır. Üreticiler açısından doğru fiyat tahmini, envanter yönetiminin optimize edilmesine, üretim planlamasının daha verimli hale getirilmesine ve rekabet avantajı sağlanmasına olanak tanımaktadır. Tüketiciler açısından ise güvenilir fiyat tahmini, piyasa değerinin belirlenmesini kolaylaştırarak bilinçli satın alma kararları alınmasını desteklemektedir. Bunun yanı sıra, sigorta ve finans sektörleri, fiyat tahmini modellerini kullanarak risk değerlendirme süreçlerini iyileştirebilir ve kredi politikalarını daha sağlam temellere dayandırabilir. Kadın ve erkek tüketiciler arasındaki farklı tercihler doğrultusunda, otomobil üreticilerinin pazarlama stratejilerini cinsiyete özgü tüketici eğilimlerine göre şekillendirmesi büyük önem taşımaktadır. Erkek tüketicilerin daha çok güçlü motorlara ve geniş araçlara yönelmesi, kadın tüketicilerin ise güvenlik, yakıt verimliliği ve teknolojik donanım gibi faktörlere daha fazla önem vermesi, sektörde daha spesifik hedefleme stratejilerinin geliştirilmesini gerektirmektedir. Kadın tüketicilere yönelik güvenlik ve çevre dostu özelliklerin vurgulandığı kampanyaların artırılması, bu kesimin otomotiv sektöründeki temsiliyetini güçlendirebilir.

Çalışma, belirli bir zaman dilimini ve coğrafi bölgeyi kapsayan veri setiyle gerçekleştirildiğinden, genelleme kabiliyeti açısından sınırlılıklar içermektedir. Modelin farklı pazarlara uygulanabilirliğini artırmak için daha geniş, çok merkezli ve uzun dönemli veri setlerinin kullanılması gerekmektedir. Ayrıca, çalışmada XGBoost ve SVR için varsayılan hiperparametreler kullanılmış olup, bu modellerin performansının artırılabilmesi için kapsamlı bir hiperparametre optimizasyonuna ihtiyaç duyulmaktadır. SVR modelinde farklı çekirdek fonksiyonlarının (linear, polynomial) denenmesi ve C, epsilon, gamma gibi hiperparametrelerin optimize edilmesi, modelin hata oranlarını düşürerek daha yüksek doğruluk sağlamasına katkıda bulunabilir. Eksik verilerin doldurulması ve aykırı değerlerin yönetilmesi konusunda farklı tekniklerin karşılaştırmalı olarak incelenmesi de model başarısını artıracak unsurlar arasında yer almaktadır.

Araç fiyatlarının yalnızca teknik özellikler ve pazar dinamikleri ile değil, aynı zamanda makroekonomik faktörlerle de şekillendiği göz önünde bulundurulmalıdır. Döviz kuru dalgalanmaları, faiz oranları, enflasyon, petrol fiyatları ve tüketici güven endeksi gibi makroekonomik değişkenlerin modele entegrasyonu, tahmin doğruluğunu artırabilir ve fiyat dalgalanmalarının arkasındaki nedenleri daha kapsamlı bir şekilde açıklayabilir. Özellikle küresel ekonomik krizler, regülasyon değişiklikleri ve pandemi gibi dışsal faktörlerin otomotiv sektörü üzerindeki etkisi dikkate alındığında, gelecekte bu tür değişkenlerin modele dahil edilmesi önem arz etmektedir. Literatürde de makroekonomik göstergelerin fiyat tahmin modellerine entegrasyonunun model performansını artırdığı gösterilmiştir. Örneğin, Fantazzini ve Toktamyssoya (2015), ekonomik krizler, vergi oranları ve petrol fiyatlarının otomobil talebi üzerindeki etkisini analiz ederek, çok değişkenli modellerin uzun vadeli tahminlerde kritik bir rol oynadığını vurgulamıştır. Benzer şekilde, Kayakuş vd. (2023), Türkiye'deki hafif ticari araç satışlarını tahmin etmek için Yapay Sinir Ağları

(YSA), Çoklu Doğrusal Regresyon (MLR) ve Karar Ağacı Regresyonu (DTR) yöntemlerini kullanmış ve YSA'nın %94,6 R² değeri ile en başarılı yöntem olduğunu göstermiştir.

Çalışmamız, önceki araştırmalarla karşılaştırıldığında, farklı makine öğrenmesi algoritmalarının ayrıntılı performans analizini sunarak literatüre katkı sağlamaktadır. Karaatlı vd. (2012), yapay sinir ağlarının otomobil satış tahmininde etkili bir yöntem olduğunu vurgulamış ve ekonomik göstergelerin modele entegrasyonu ile tahmin performansının artırılabilceğini göstermiştir. Kurtgeri (2018) ve Dere ve Kasapbaşı (2024) ise veri ön işleme süreçlerinin ve model seçiminin tahmin başarısındaki kritik rolünü ortaya koymuştur. Dere ve Kasapbaşı'nın çalışmasında, Rastgele Orman algoritmasının Türkiye'deki ikinci el araç fiyat tahmini problemlerinde güvenilir bir yöntem olduğu belirtilmiştir. Gür vd. (2024) ise MLP ve RBF modellerinin otomobil ihracat tahmini üzerindeki olumlu etkilerini inceleyerek, bu modellerin sektörel planlamalarda kullanılabileceğini göstermiştir.

Gelecekte yapılacak araştırmalarda, makroekonomik değişkenlerin modele entegrasyonunun yanı sıra hiperparametre optimizasyonu, derin öğrenme algoritmalarının uygulanması ve büyük veri analitiğinin kullanımı gibi yöntemlerin araştırılması, model performansını daha da artırabilir. Mevsimsel eğilimlerin incelenmesi, pazarlama kampanyalarının ve promosyon stratejilerinin optimize edilmesine olanak tanırken, sosyal medya verilerinin müşteri duyarlılıklarını ve pazar taleplerini analiz etmek amacıyla kullanılması, tahmin modellerinin doğruluğunu ve piyasa davranışlarını anlamaya yönelik içgörülerini önemli ölçüde geliştirebilir.

Sonuç olarak, bu çalışma, araç fiyat tahminine yönelik makine öğrenmesi algoritmalarının performanslarını değerlendirerek sektöre yönelik uygulanabilir çözümler sunmuştur. Özellikle Lasso Regresyon, Rastgele Orman ve Gradient Boosting gibi modellerin yüksek doğruluk oranları ile başarılı tahminler sağladığı belirlenmiştir. Bununla birlikte, veri ön işleme süreçlerinin titizlikle uygulanmasının model başarısında kritik bir rol oynadığı ortaya konulmuştur. Gelecekte yapılacak çalışmaların, model doğruluğunu artırmak amacıyla makroekonomik değişkenlerin entegrasyonu, hiperparametre optimizasyonu, derin öğrenme tekniklerinin kullanımı ve büyük veri kaynaklarının analiz edilmesi gibi yöntemlere odaklanması, araç fiyat tahmin süreçlerinin daha kapsamlı ve güvenilir hale gelmesini sağlayacaktır.

Kaynaklar

- Ahmad, M., Farooq, M. A., Hussain, M. Z., Hasan, M. Z., Mustafa, M., Khalid, A. ve Javaid, A. (2024). Car price prediction using machine learning. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)* içinde (s. 1-5). Pune, India. <https://doi.org/10.1109/I2CT61223.2024.10544124>
- Arseven, B. ve Çınar, S. M. (2023). Dünya dışı ışınımlarla iyileştirilmiş ARIMA, ridge regresyon ve lasso regresyon yöntemlerinin saatlik ışınım tahmininde kullanılması. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 12(3), 733-742. <https://doi.org/10.28948/ngumuh.1249503>
- Arslankaya, S. ve Öz, V. (2018). Time series analysis of sales quantity in an automotive company and estimation by Artificial Neural Networks. *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 22(5), 1482-1492. <https://doi.org/10.16984/aufenbilder.456518>
- Aslay, F. ve Özen, Ü. (2013). Meteorolojik parametreler kullanılarak yapay sinir ağları ile toprak sıcaklığının tahmini. *Politeknik Dergisi*, 16(4), 139-145. <https://dergipark.org.tr/tr/pub/politeknik/issue/33068/367992>
- Bhushan, B. ve Singh, M. ve Hage, Y. (2012). Identification and control using MLP, Elman, NARXSP and radial basis function networks: a comparative analysis. *Artificial Intelligence Review*, 37, 133-156. <https://doi.org/10.1007/s10462-011-9223-0>

- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J., Olshen, R. A. ve Stone, C. J. (1984). Classification and regression trees. (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781315139470>
- Bridge-Nduwimana, C. B., El Ouazizi, A., ve Benyakhlef, M. (2025). A new data imputation technique for efficient used car price forecasting. *International Journal of Electrical and Computer Engineering (IJECE)*, 15(2), 2364-2371. <http://doi.org/10.11591/ijece.v15i2.pp2364-2371>
- Chen, T. ve Guestrin, C. (2016, 13-17 Ağustos). Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* içinde (s. 785-794). San Francisco, ABD. <https://doi.org/10.48550/arXiv.1603.02754>
- Dere, E. ve Kasapbaşı, M. C. (2024). İkinci el araç değerlemesinde makine öğrenme tekniklerinin uygulanması. *İstanbul Ticaret Üniversitesi Teknoloji ve Uygulamalı Bilimler Dergisi*, 6(2), 35-51. <https://doi.org/10.56809/icujtas.1329496>
- Ecer, F. (2013). Türkiye’de 2. El otomobil fiyatlarının tahmini ve fiyat belirleyicilerinin tespiti. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 13(4), 101-112. <https://dergipark.org.tr/tr/download/journal-file/20335>
- Efendigil, T. ve Eminler, Ö. E. (2017). Havacılık sektöründe talep tahminin önemi: Yolcu talebi üzerine bir tahmin modeli. *Yaşar Üniversitesi E-Dergisi*, 12(Tarım ve Gıda Değer Zincirlerinde Yöneylem Araştırmaları ve Endüstri Mühendisliği Özel Sayısı), 14-30. <https://dergipark.org.tr/tr/pub/jyasar/issue/45368/568599>
- Eşidir, K. A., Gür, Y. E., Yoğunlu, V. ve Çubuk, M. (2022). Yapay sinir ağları (YSA) ve ARIMA modelleri ile Türkiye’de aylık sıfır km otomobil satış adetlerinin tahmin edilmesi. *Pamukkale Üniversitesi İşletme Araştırmaları Dergisi*, 9(2), 260-277. <https://doi.org/10.47097/piar.1132101>
- Fantazzini, D. ve Toktamysova, Z. (2015). Forecasting German car sales using Google data and multivariate models. *International Journal of Production Economics*, 170(Part A), 97-135. <https://doi.org/10.1016/j.ijpe.2015.09.010>
- Friedman, J., Hastie, T., ve Tibshirani, R. (2001). The elements of statistical learning: Data mining, Inference, and Prediction. Springer. <https://tinyurl.com/4cpndh32>
- Gür, Y. E., Eşidir, K. A. ve Kaldırımcı, Ş. G. (2024). Makine öğrenimi ile binek otomobil ihracat tahmini: MLP ve RBF modeli kullanımı. *Afyon Kocatepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 26(Özel Sayı), 17-34. <https://doi.org/10.33707/akuiibfd.1426338>
- Hamid, S., A. Iqbal, Z. (2004). Using neural networks for forecasting volatility of svep 500 index futures prices, *Journal of Business Research*, 57, 1116-1125. [http://dx.doi.org/10.1016/S0148-2963\(03\)00043-2](http://dx.doi.org/10.1016/S0148-2963(03)00043-2)
- Harsha, C. S. S., Sai, V. A., Kaif, S. M., Dinesh, G. ve Shareefunnisa, S. (2024). Prediction of Car Sales Price Trends using Ensembling Models. *2024 8th International Conference on Inventive Systems and Control (ICISC)* içinde (s. 234-239). IEEE. <https://doi.org/10.1109/ICISC62624.2024.00047>
- Hoerl, A. E. ve Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67. <https://tinyurl.com/5h3c9eh9>
- Karaatlı, M., Helvacıoğlu, Ö. C., Ömürbek, N. ve Tokgöz, G. (2012). Yapay sinir ağları yöntemi ile otomobil satış tahmini. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, 8(17), 87-100. <https://doi.org/10.11122/ijmeb.2012.8.17.290>
- Kayakuş, M., Terzioğlu, M., Yağmur, A. ve Erdoğan, D. (2023). Türkiye’deki hafif ticari araç satışlarının makine öğrenmesi yöntemleriyle tahmin edilmesi. *Gazi Mühendislik Bilimleri Dergisi*, 9(4-ICAİAME 2023), 100-112. <https://dergipark.org.tr/en/pub/gmbd>
- Kayapınar Kaya, S. K. ve Yıldırım, Ö. (2020). A prediction model for automobile sales in Turkey Using Deep Neural Networks. *Endüstri Mühendisliği*, 31(1), 57-74. <https://dergipark.org.tr/tr/pub/endustrimuhendisligi>

- Kıvrak, O. (2016). *Müşteri yaşam boyu değerinin yapay zekâ algoritmaları ile modellenmesi* (Yayımlanmamış Doktora Tezi). Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü, Balıkesir.
- Kurtgeri, B. (2018). *Yapay sinir ağları ve makine öğrenmesi ile otomobil satış tahmininin yapılması ve zaman serileri analizi ile karşılaştırılması* (Yayımlanmamış Yüksek Lisans Tezi). Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Sakarya.
- Kuvvetli, Y., Dağsuyu, C. ve Oturakci, M. (2015). Türkiye'deki araç satışları için ekonomik ve çevresel faktörleri göz önüne alan yapay sinir ağı tabanlı bir tahmin yaklaşımı. *Endüstri Mühendisliği*, 26(3), 23-31. <https://dergipark.org.tr/pub/endustrimuhendisligi>
- Llopis-Albert, C., Rubio, F. ve Valero, F. (2021). Impact of digital transformation on the automotive industry. *Technological Forecasting and Social Change*, 162, 120343. <https://doi.org/10.1016/j.techfore.2020.120343>
- Montgomery, D. C., Peck, E. A. ve Vining, G. G. (2021). Introduction to linear regression analysis. John Wiley ve Sons.
- Ogundele, L. A., Ejidokun, A. O., Ayo, F. E. ve Taiwo, A. I. (2024). Car price prediction using artificial neural networks: A data-driven approach. *Indonesian Journal on Computing*, 9(3), 198-207. <https://tinyurl.com/5n8fec2e>
- Otomotiv Sanayi Derneği. (2024). Otomotiv sanayii 2023 yılı küresel değerlendirme raporu, <https://tinyurl.com/29kadxe5>
- Ranstam, J. ve Cook, J. A. (2018). Lasso regression. *The British Journal of Surgery*, 105(10), 1348-1348). <https://doi.org/10.1002/bjs.10895>
- Smola, A. J. ve Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14, 199-222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
- T.C. Sanayi ve Teknoloji Bakanlığı. (2024, 26 Kasım). *Otomotiv sektör raporu*. <https://www.sanayi.gov.tr/assets/pdf/planprogram/OtomotivSektorRaporu2020.pdf>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288. <https://www.jstor.org/stable/2346178>
- Yan, X. ve Su, X. (2009). Linear regression analysis: Theory and computing. World Scientific. <http://www.manalhelal.com/Books/geo/LinearRegressionAnalysisTheoryandComputing.pdf>
- Yang, Z. ve Zhang, C. (2020, July). Automobile sales forecast based on web search and social network data. *ICEME '20: Proceedings of the 2020 11th International Conference on E-business, Management and Economics* içinde (s. 37-41). <https://doi.org/10.1145/3414752.3414759>
- Yurtsever, M. (2022). LSTM yöntemi ile ekonomik göstergeler kullanılarak otomobil satış tahmini. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, 12(3), 1481-1492. <https://doi.org/10.30783/nevsosbilen.987093>