# *Perakende Sektöründe Makine Öğrenmesi Algoritmalarının Karşılaştırmalı Performans Analizi: Black Friday Satış Tahminlemesi*

## *Comparative Performance Analysis of Machine Learning Algorithms in the Retail Industry: Black Friday Sales Forecasting*

*Vahid SİNAP* *

***ÖZ***

*Büyük perakende zincirlerinin şube ağlarının genişlemesi, müşteri tabanlarının büyümesi ve artan müşteri profili heterojenliği satış tahminleme süreçlerinin karmaşıklığını artırmaktadır. Müşteri çeşitliliği ve bu çeşitliliğin yönetilmesi, perakendeciler için hem stratejik planlama hem de operasyonel düzeyde uygulama açısından önemli bir güçlük oluşturmaktadır. Bu noktada, müşteri segmentasyonu ve kişiselleştirilmiş pazarlama stratejileri geliştirmek, her bir müşteri grubuna özel yaklaşımlar belirlemek ve bu çeşitliliği anlayarak etkili bir şekilde yönetmek önem kazanmaktadır. Gelişen teknolojiler, özellikle makine öğrenmesi yöntemleri söz konusu zorluklarla başa çıkma potansiyeli sunmaktadır. Bu kapsamda araştırmanın amacı, bir perakende firmasının Black Friday günündeki satış veri seti üzerinde Doğrusal Regresyon, Rastgele Orman Regresyonu, K-En Yakın Komşu Regresyonu, XGBoost Regresyonu, Karar Ağacı Regresyonu ve LGBM Regresyonu isimli makine öğrenmesi algoritmaları aracılığıyla satış tahminlemesi gerçekleştirmek ve algoritmaların performanslarını karşılaştırarak en iyi performans gösteren algoritmayı belirlemektir. Ayrıca, GridSearchCV kullanarak hiperparametrelerin ayarlanması ve bu ayarlamaların modellerin performanslarına etkisinin incelenmesi amaçlanmaktadır. Buna ek olarak, veri seti üzerinde Keşifsel Veri Analizleri yürütülerek, perakende sektöründeki işletmelerin ellerinde bulunan verilerden ne tür bilgiler çıkarabileceklerine ve bu bilgileri nasıl değerlendirebileceklerine ilişkin bir örnek oluşturmak araştırmanın diğer önemli bir amacıdır. Araştırmadan elde edilen sonuçlara göre, satışları tahminlemede en başarılı algoritma GridSearchCV ile hiperparametreleri ayarlanmış XGBoost Regresyonu olmuştur. Firma müşterilerinin en çok 26-35 yaş aralığında bireylerden oluştuğu, erkek müşterilerin kadınlara, bekar müşterilerin evlilere göre önemli ölçüde daha yüksek tutarlı alışverişler yaptığı saptanmıştır. Ayrıca, satın alım tutarı ortalaması bağlamında bakıldığında en yüksek harcama ortalamasına sahip yaş grubu 51-55 yaş aralığı olarak tespit edilmiştir.*

***ANAHTAR KELİMELER***

*Satış Tahminlemesi, Makine Öğrenmesi, Regresyon Algoritmaları, Black Friday, Perakende Sektörü*

***ABSTRACT***

*The expansion of branch networks of large retail chains, the growth of their customer base, and the increasing diversity of customer profiles are exacerbating the complexity of sales forecasting processes. Managing this diversity and its implications presents a significant challenge for retailers in terms of both strategic planning and operational implementation. At this point, developing customer segmentation and personalized marketing strategies, determining unique approaches for each customer group, and effectively managing this diversity are becoming increasingly crucial. The emerging technologies, particularly machine learning methods, present the potential to cope with these challenges. In light of this, the main objective of the research is to perform sales forecasting on a retail company's Black Friday sales data using machine learning algorithms named Linear Regression, Random Forest Regression, K-Nearest Neighbors Regression, XGBoost Regression, Decision Tree Regression and LightGBM Regression and determine the best performing algorithm by comparing their performances. It is also aimed to tune the hyperparameters using GridSearchCV and examine the effect of these adjustments on the performance of the models. Additionally, Exploratory Data Analysis will be conducted on the dataset to create a sample example for businesses in the retail sector on how they can extract useful information from their available data and effectively evaluate it. According to the results obtained from the research, the most successful algorithm in predicting sales was the XGBoost Regression with hyperparameters tuned using GridSearchCV. It has been determined that the majority of the company's customers consist of individuals aged 26-35, with male customers making significantly higher purchases compared to females and single customers spending more than married ones. Furthermore, when examining the average amount of purchases made by each age group, it was identified that those within the range of 51-55 years had the highest average spending rate.*

***KEYWORDS***

*Sales Forecasting, Machine Learning, Regression Algorithms, Black Friday, Retail Industry*

* Dr. Öğretim Üyesi, Ufuk Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, vahidsinap@gmail.com, ORCID: 0000-0002-8734-9509

| *Makale Geliş Tarihi / Submission Date*<br>*07.12.2023* | *Makale Kabul Tarihi / Date of Acceptance*<br>*28.02.2024* |
|---|---|
| ***Atıf*** | Sinap, V. (2024). Perakende Sektöründe Makine Öğrenmesi Algoritmalarının Karşılaştırmalı Performans Analizi: Black Friday Satış Tahminlemesi. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 27 (1), 65-90. |

## GİRİŞ

Black Friday, Amerika'da başlayan ancak zamanla dünya genelinde büyük rağbet gören bir alışveriş fenomenidir. Amerika'da Şükran Günü'nden sonraki cuma günü olarak kutlanan bu gün, yüksek indirimler, özel teklifler ve cazip fırsatlarla dolu alışveriş maratonu olmakla bilinmektedir (Alagarsamy ve diğ., 2023). Black Friday çoğu perakendeci açısından değerlendirildiğinde ise yılın en yoğun günü olarak nitelendirilmektedir. Öyle ki Black Friday'den Noel Tatili'ne kadar geçen bir aylık süreçte yaklaşık olarak tüm yıllık perakende satışların %30'unun gerçekleşmesiyle bu dönem birçok işletme açısından yılın en zahmetli ancak buna karşın kârlı dönemi olarak görülmektedir. Bununla birlikte Black Friday günü gerçekleşen satış oranlarında da yıldan yıla artış yaşanmaktadır. Pazaryerleri ve bireysel kullanıcılar için ödeme altyapısı sağlayan İyzico'nun 2022 Black Friday dönemini kapsayan çevrimiçi alışveriş istatistiklerine göre İyzico üzerinden Türkiye'de, 2022 yılında, bir önceki yıla oranla işlem hacminde %145.7, işlem adedinde ise %27.1 artış gerçekleşmiştir. Tek seferde gerçekleştirilen ortalama sipariş harcaması 337.5 TL'den 652 TL'ye ulaşarak %93'lük artış yaşanmıştır (İyzico, 2022). Benzer şekilde, üye işyerlerine tüm ödeme süreçlerini tek bir sistemden yönetme imkânı tanıyan ödeme platformu Craftgate üzerinden 2022 Kasım ayında gerçekleşen ödeme hacmi, Ekim ayına göre %203, 2021'in Kasım ayına oranla ise %867 artış göstermiştir. Ödeme adedi açısından bakıldığında Kasım ayındaki ödeme adedi Ekim ayına kıyasla %178 yükselmiş, Kasım 2022'deki ödeme adedi, bir önceki yıla göre %445'lik bir artış sergilemiştir (Yalçın, 2022).

Tüketiciler Black Friday döneminde daha çok alışveriş yapma eğiliminde olduklarından perakendeciler bu tür alışveriş dönemleri yaklaştıkça ciddi bir hazırlık sürecine girmektedirler (Swilley ve Goldsmith, 2013). Perakendeciler genellikle bu alışveriş etkinliğine hazırlık olarak daha fazla çalışan işe almakta, ürünlerin stoklarını artırmakta, yeni promosyonlar hazırlamakta ve mağazalarını cazip kılacak düzenlemeler gerçekleştirmektedirler. Bunun yanı sıra perakendeciler, müşterileri çevrimiçi veya fiziksel mağazalarına çekmek için çeşitli reklam kampanyaları yürütmektedirler. Söz konusu hazırlıklardan en yüksek verimi alabilmek için bu hazırlıkların müşteri odaklı geliştirilmesi gerekmektedir (Trung ve diğ., 2021). Dolayısı ile müşterilerin alışveriş alışkanlıklarını en iyi şekilde anlamak, perakendecilerin en fazla kârı elde etmelerine yardımcı olacak daha etkili pazarlama stratejileri geliştirmelerine olanak tanımaktadır.

Günümüzde büyük ölçekli işletmelerin müşteri tabanlarının genişlemesiyle müşteri ihtiyaçlarını ve alışkanlıklarını anlamak işletmeler açısından önemli bir zorluk haline gelmiştir. Geçmişte işletmelerin genellikle küçük ölçekli olması müşteri ilişkilerinin daha kişisel bir seviyede kurulabilmesine olanak tanımaktaydı. Özellikle süpermarketler veya büyük mağazalar yerine yerel esnafların ve küçük işletmelerin çoğunlukta olması müşteriyle yakın ilişkiler kurulabilmesine, müşterilerin alışveriş alışkanlıklarını doğrudan gözlemlenebilmesine ve ihtiyaçlarına daha etkili bir şekilde yanıt verilebilmesine imkân vermekteydi (Wu ve diğ., 2018). Ancak, zamanla bu küçük işletmeler büyüyerek ulusal veya uluslararası çapta faaliyet gösteren büyük perakende zincirlerine dönüşmesiyle işletmelerin müşteri sayıları artmış ve müşteri profilleri karmaşık hale gelmiştir. Yüzlerce şubesi olan büyük perakende zincirleri, müşterilerinin kişisel tercihlerini, alışveriş alışkanlıklarını ve ihtiyaçlarını anlamakta zorlanmaya başlamıştır (Meyer ve Schwager, 2007). Bu ihtiyaçları karşılama kapsamında, gelişen teknolojilerle birlikte özellikle makine öğrenmesi yöntemleri önemli bir rol üstlenmektedir.

Makine öğrenmesi, bilgisayar sistemlerinin veri setlerindeki desenleri tanıması ve bu desenleri kullanarak yeni veriler üzerinde tahminler yapabilmesi için kullanılan bir yapay zekâ alt dalıdır (Bi ve diğ., 2019). Temelde, bir makine öğrenmesi modeli, belirli bir görevi yerine getirmek veya belirli bir problemi çözmek için tasarlanmaktadır. Bu model, geçmiş verileri analiz ederek öğrendiği bilgileri, gelecekte benzer görevleri gerçekleştirmek için kullanmaktadır (Alzubi ve diğ., 2018). Makine öğrenmesi pazarlama alanı bağlamında düşünüldüğünde büyük perakende zincirlerinin geniş veri setlerini analiz ederek müşteri davranışlarını anlama ve satış tahminlemesi yapma amacıyla kullanılabilmektedir. Bu teknoloji, geçmiş alışveriş alışkanlıkları, yaş, cinsiyet, gelir düzeyi gibi demografik bilgiler, coğrafi konum, hava durumu gibi dışsal faktörler ve müşteri web sitesindeki gezinme davranışları gibi diğer faktörler üzerinden müşteri segmentasyonu yaparak, her bir müşteri grubunun ihtiyaçlarına özel stratejiler geliştirmeyi mümkün kılabilmektedir (Timoshenko ve Hauser, 2019). Black Friday gibi yoğun alışveriş dönemlerinde, makine öğrenmesi algoritmaları kullanılarak gerçekleştirilen satış tahminlemeleri ile stok yönetimi optimize edilebilmekte ve müşterilere özel indirimler sunularak satışlar artırılabilmektedir. Bu sayede, büyük perakende zincirleri, geçmişteki küçük işletmelerin kişisel yaklaşımlarını modern teknolojilerle birleştirerek sürdürebilmekte ve müşteri memnuniyetini artırabilmektedir (Bohanec ve diğ., 2017). Ayrıca, Black Friday gibi yüksek alışveriş yoğunluğunun yaşanacağı öngörülen dönemlere yönelik satış tahminlemesi gerçekleştirilerek stok durumları, gerekli iş gücü gibi faktörler hiçbir aksaklığa sebep olmayacak şekilde ayarlanabilmektedir. Bu bağlamda, büyük

perakendeciler olan Apple, Amazon ve Walmart gibi firmalar da satışlarını ve müşteri davranışını analiz etmek ve kişiselleştirilmiş iletişimler sunmak için makine öğrenmesi algoritmalarını kullanmaktadırlar (Marr, 2016). Ancak, makine öğrenmesinin önemli getirilerinin yanı sıra bazı zorlukları da bulunmaktadır. Örneğin, modelin veriler üzerindeki performansı, veri kalitesi ve miktarıyla doğrudan ilişkilidir. Eğer yetersiz, yanlış veya eksik verilerle çalışılıyorsa, modelin doğruluğu ve güvenilirliği önemli ölçüde azalabilmektedir. Ayrıca, makine öğrenmesi modellerinin karmaşıklığı ve anlaşılabilirliği de bir zorluk olarak görülmektedir. Çok karmaşık modeller, iç mekanizmalarının anlaşılamaz ve kararlarının açıklanamaz olması sebebiyle "siyah kutu" olarak adlandırılmaktadır (Hassija ve diğ., 2024). Bu durum, modelin güvenilirliğini ve kabul edilebilirliğini azaltabilir.

Veri güvenilirliği ve kalitesi, açıklanabilirlik sorunu, büyük veri ihtiyacı, gizlilik ve etik sorunları bu teknolojinin uygulanmasını kısıtlayan faktörler arasında yer almaktadır (Jagatheesaperumal ve diğ., 2021). Ayrıca, doğru algoritmanın seçimi, doğru bir modelin geliştirilmesinde önemli bir rol oynamaktadır. Doğru algoritmanın seçimi, modelin tahmin yeteneklerini ve doğruluğunu da büyük ölçüde etkilemektedir. Geliştirilen model için hangi algoritmanın daha iyi ve verimli olacağı dikkat edilmesi gereken diğer bir husustur. Her algoritmanın avantajları ve dezavantajları olduğundan, doğru algoritma seçimi yapmak işletmelerin ve müşterilerin çıkarlarını korumak açısından önemli görülmektedir. Doğru algoritmanın seçilmesi önemli olduğu kadar, model hiperparametrelerinin belirlenmesi de önemli bir adımdır. Hiperparametreler, bir modelin öğrenme sürecini kontrol eden ayarlar olarak düşünülebilir ve doğru şekilde ayarlanmaları modelin performansını büyük ölçüde etkileyebilmektedir (Eker ve diğ., 2023). Ancak, bu hiperparametrelerin doğru şekilde belirlenmesi genellikle deneme-yanılma yöntemiyle yapılır ve bu, zaman alabilir ve uzmanlık gerektirebilir. Otomatik makine öğrenmesi (AutoML), bu tür zorlukları aşmak için geliştirilen bir yaklaşımdır. AutoML, makine öğrenmesi sürecini otomatikleştirerek, veri seti üzerinde çalışacak en iyi modeli ve bu model için en iyi hiperparametreleri belirlemeye çalışmaktadır (Selvi ve diğ., 2021). Bu yaklaşım, makine öğrenmesi konusunda uzman olmayan kişiler için model seçimi ve hiperparametre ayarlama gibi karmaşık adımları daha erişilebilir hale getirebilmektedir. AutoML, farklı algoritmaları ve hiperparametre ayarlarını deneyerek en iyi kombinasyonu bulmaya çalışmakta ve bu sayede makine öğrenmesi modellerinin daha verimli ve etkili olmasını sağlamaktadır (Özdemir ve Örslü, 2019). Buna ek olarak, veri ön işleme adımları veri setindeki gürültüyü azaltarak ve eksik verileri düzelterek modelin performansını artırabilmektedir (Garcia ve diğ., 2016). Bu sayede modelin gelecekteki trendleri ve müşteri davranışlarını daha doğru bir şekilde anlaması sağlanabilmektedir. Aynı zamanda veri setinin yapısına uygun görselleştirme teknikleri kullanılarak elde edilen bulgular müşteri davranışlarını ve ihtiyaçlarını daha anlaşılır hale getirerek mağaza sahiplerine müşterilerini daha iyi anlama konusunda yardımcı olabilmektedir. Bu açıdan, perakende sektöründeki işletmelerin etkili pazarlama stratejileri geliştirmeleri ve müşteri ihtiyaçlarına daha iyi yanıt vermeleri için model kurulum aşamaları, seçilen algoritmalar ve veri ön işleme adımları bazı zorluklara sahip olsa da kritik bir önem taşımaktadır.

Bu araştırmanın temel amacı, perakende sektöründe makine öğrenmesi algoritmalarının etkinliğini ve uygulanabilirliğini değerlendirmektir. Özellikle büyük perakende zincirlerinin geniş veri setleri üzerinde gerçekleştirilen müşteri segmentasyonu, satış tahminlemesi ve stok yönetimi gibi kritik işlemlerde kullanılan Doğrusal Regresyon (Linear Regression - LR), Rastgele Orman Regresyonu (Random Forest Regressor - RF), K-En Yakın Komşu Regresyonu (K-Nearest Neighbors Regression - KNN), XGBoost Regresyonu (XGBoost Regression - XGB), Karar Ağacı Regresyonu (Decision Tree Regression - DT) ve LGBM Regresyonu (LightGBM Regression - LGBM) isimli makine öğrenmesi algoritmaları detaylı bir şekilde incelenecek ve bu algoritmaların gösterdiği performanslar karşılaştırılacaktır. Buna ek olarak araştırmanın diğer bir önemli amacı GridSearchCV kullanarak hiperparametrelerin ayarlanması ve bu ayarlamaların modellerin performanslarına etkisinin incelenmesidir. Black Friday satış veri seti üzerinden elde edilen sonuçlarla farklı makine öğrenmesi modellerinin farklı hiperparametre ayarlarıyla tahminleme yeteneklerini performans metriklerine bağlı olarak objektif bir şekilde değerlendirmek araştırmanın odak noktasını oluşturmaktadır. Bu çerçevede, her bir algoritmanın avantajları, dezavantajları ve uygulama alanları ele alınarak perakende sektöründeki işletmelerin makine öğrenmesi tekniklerini daha etkili bir şekilde kullanmalarına olanak sağlanacaktır. Ayrıca bu çalışma, perakende sektöründeki işletmelerin pazarlama stratejilerini güçlendirmeleri, müşteri memnuniyetini artırmaları ve yoğun alışveriş dönemlerinden en yüksek faydayı elde edebilmeleri açısından bir rehberlik görevinde bulunmayı hedeflemektedir.

## 1. İLGİLİ ARAŞTIRMALAR

Satış tahminleme ve makine öğrenmesi üzerine yapılan araştırmalar, perakende sektöründe önemli gelişmeler sağlamıştır. Bu bağlamda, alanyazında bazı öne çıkan çalışmalar incelenmiştir.

Nacar ve Erdebilli (2021) gerçekleştirdikleri çalışmada satın alma tahmininde yaşanan zorlukları ve bu zorlukların neden olduğu olumsuz etkileri ele almışlardır. Çok boyutlu verilerin yönetilmesinde karşılaşılan zorluklar vurgulanmış ve bu durumun işletmeler için ne gibi sorunlar doğurabileceği açıklanmıştır. Makine öğrenmesi ve yapay zekâ tekniklerinin, özellikle satın alma tahmininindeki zorluklara çözüm getirebileceği belirtilmiştir. Çalışmanın uygulama bölümünde, LR, Ridge, Lasso, ElasticNet, KNN ve RF isimli makine öğrenmesi algoritmaları kullanılarak bir satın alma tahmin modeli geliştirilmiştir. Bu modelleme çalışmasında en iyi performansı RF algoritması göstermiştir. Dılkı (2020) tarafından yapılan çalışmada sınıflandırma problemleri üzerinde durulmuş ve daha önce piyasaya sürülen benzer ürünlerin verileri kullanılarak ürünün satılma olasılığının tahmin edildiği bir model geliştirilmiştir. Kullanılan denetimli öğrenme algoritmaları arasında KNN, Naive Bayes (NB) ve Destek Vektör Regresyonu (Support Vector Regression – SVR) yer almaktadır. Sonuç olarak, KNN algoritmasının en yüksek doğruluk oranı olan %71 ile en iyi performansı gösterdiği belirlenmiştir. Beştaş (2023) çalışmasında, ilaç sektöründe satışların önemini vurgulamış ve bu satışların gelecekteki trendlerinin tahmin edilmesinin gerekliliğine dikkat çekmiştir. Güneydoğu Anadolu bölgesindeki bir eczanenin üç yıllık satış verileri incelenerek, makine öğrenmesi ve zaman serileri analizleri kullanılarak gelecek 15 güne ait satışların tahmini yapılmıştır. Çalışma sonucunda, en başarılı tahmin yönteminin ARIMA modeli olduğu ve bu modelle elde edilen ortalama hataların karekök değerinin 23 olduğu belirtilmiştir. Ecemiş ve Irmak (2018) tarafından yapılan çalışmada, paslanmaz çelik sektöründe faaliyet gösteren bir firmanın satış tahminleri üzerine odaklanılmıştır. Çalışmada, firmanın Ocak 2008 ile Mart 2016 tarihleri arasındaki günlük satış verileri kullanılarak, sektörlere göre satış tahminleri yapılmıştır. Veri setindeki satış hareketleri müşteri bilgileriyle eşleştirilerek, sektörlere ait satış rakamları belirlenmiştir. SVR ve Yapay Sinir Ağları (Artificial Neural Network – ANN) kullanılarak toplam satış ve sektörlere göre satış tahminleri gerçekleştirilmiştir. Yapılan uygulama sonucunda, SVR yönteminin diğer yöntemlere göre nispeten daha başarılı olduğu tespit edilmiştir. Kayakuş ve diğerleri (2023) tarafından yapılan çalışmada, hafif ticari araç satışlarının tahmin edilmesi üzerine odaklanılmıştır. Çalışmada, hafif ticari araç satış ve ithalatının tahmin edilmesinin genel ekonomik göstergelerin değerlendirilmesine ve otomotiv firmaları için etkin kurumsal kaynak planlaması ve verimli kaynak kullanımı sağlamasına katkı sağlayacağı düşünülmüştür. Tasarlanan tahmin modeli, önceki çalışmaların analizi ve hafif ticari araç satışlarını etkileyebilecek makroekonomik değişkenlerin modele dahil edilmesiyle oluşturulmuştur. Modelin başarısını ölçmek için ANN, çoklu doğrusal regresyon (Multiple Linear Regression - MLR) ve DT gibi makine öğrenmesi yöntemleri kullanılmıştır. Yapılan çalışma sonucunda, YSA yönteminin %94,6 doğruluk ile en başarılı yöntem olduğu tespit edilmiştir. Erol ve İnkaya (2024), satış tahmini için derin öğrenme ve transfer öğrenme yöntemlerinin kullanıldığı bir çalışma gerçekleştirmiştir. Çalışmada, farklı ürünlerin satış tahmin modellerinden elde edilen bilgilerin, gelecekteki tahmin modellerine aktarılması için derin transfer öğrenme yaklaşımı önerilmiştir. Bu yöntem, satış verilerini tek değişkenli zaman serisi olarak ele almakta ve uzun kısa vadeli hafıza (Long Short-Term Memory - LSTM) ağlarını kullanmaktadır. Yapılan deneysel çalışmalar, önerilen yöntemin tahmin doğruluğunu artırdığını ve eğitim süresini azalttığını göstermektedir. Çiçek ve Selçuk (2023) gerçekleştirdikleri çalışmada demografik özellikleri kullanarak çeşitli makine öğrenmesi teknikleri ile bireylerin sanal market kullanımları tahmin edilmiş ve en yüksek doğruluk oranına rasgele orman tekniği ile ulaşılmıştır (%78 doğruluk). Ramachandra ve diğerleri (2021), çalışmalarında geliştirdikleri regresyon modeli ile müşteri davranışları ile ürün tercihleri arasındaki ilişkiyi analiz etmeyi amaçlamışlardır. Model, belirlenen değişkenler arasındaki ilişkiyi %86 oranında doğrulukla tahmin etmiştir. Chen ve diğerleri (2021) Keras ve TensorFlow ile eğiterek geliştirdikleri yapay sinir ağı modelini kullanarak müşteri taleplerini öngörmeye odaklanmışlardır. Araştırma kapsamında geliştirilen modelin müşteri alışkanlıklarını ve ürün tercihlerini belirlemede etkili olduğu sonucuna ulaşılmıştır. Awan ve diğerleri (2021) yaptıkları çalışmada LR ve RF regresyonu algoritmalarını kullanarak ürünlerin satış tahminlemelerini gerçekleştirmişlerdir. Ürün segmentasyonu ve talep analizi konularında çeşitli veri madenciliği yaklaşımlarının nasıl entegre edilebileceğini gösteren çalışmada geliştirilen model, satış tahmini için %89'luk bir doğruluk yakalamıştır. Cheriyan ve diğerleri (2018) satış tahminlemesi için LR kullandıkları araştırmalarında, bu yöntemin organizasyonlar için etkili bir araç olduğunu ortaya koymuşlardır. LR, belirli bir sonucu etkileyen faktörleri anlamak ve değerlendirmek amacıyla, bir bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkiyi modellemek için kullanılmaktadır. Liao ve diğerleri (2020) otomobil ürünlerinin satışlarını tahmin etmek amacıyla yığın modelleme (mass customization) yaklaşımını kullanmışlardır. Çalışma dahilinde müşterilerin otomobil parçalarına olan ilgisini ve satın alma davranışlarını belirlemede etkili bir regresyon modeli geliştirilmiştir. Niu (2020) yürüttüğü araştırmada satış iş planlarını öngörmek için XGB regresyonu ve LR algoritmalarından faydalanmıştır. Çalışmada satış geçmişi ve ürün özellikleri üzerinden gelecekteki satışları tahmin etmeye odaklanarak satış tahminleme modeli geliştirmiştir. Geliştirilen LR modeli, belirli bir

ürünün satış tahminlemesini gerçekleştirmede %78'lik bir doğruluk oranı elde etmiştir. Zeng ve diğerleri (2019), Çin festivallerini inceleyerek, kullanıcıların çevrimiçi tarama ve satın alma davranışlarını değerlendirmişlerdir. Araştırmadan elde edilen sonuçlar, kullanıcıların çevrimiçi davranışlarından elde edilen verilerin, satışları artırmak için öneri sistemleri ve çevrimiçi ürün promosyonları gibi uygulamalarda nasıl kullanılabileceğini göstermektedir. Ma ve Sun (2020) çalışmalarında çevrimiçi reklamların dinamik özelliklerini anlamak ve keşfetmek amacıyla derin öğrenme (deep learning) yöntemleri ile çevrimiçi reklamları incelemişlerdir. Alibaba tarafından sunulan küresel alışveriş festivallerinin çevrimiçi reklamlarının etkileşimlerini analiz ederek, özellikle e-ticaret alanında makine öğrenmesi uygulamalarının nasıl kullanılabileceği konusunda değerli bir bakış açısı sunmuşlardır.

Perakende sektöründe satış tahminleme ve makine öğrenmesi konularında gerçekleştirilen araştırmalar, müşteri davranışlarını analiz etmek, ürün taleplerini öngörmek ve satış tahminlemelerini optimize etmek adına önemli gelişmeler sağlamıştır. Bu çalışmalar, regresyon modelleri, yapay sinir ağları, veri madenciliği teknikleri, LR ve derin öğrenme gibi çeşitli makine öğrenmesi yaklaşımlarını içermekte ve organizasyonlara etkili araçlar sunmaktadır. Buna dayalı olarak perakende sektöründeki aktörler, makine öğrenmesi modellerini kullanarak müşteri segmentasyonu, ürün talep analizi ve satış tahminleme gibi değerlendirmelerden elde ettikleri sonuçlar doğrultusunda stratejik ve operasyonel kararlar alabilmektedirler. Bu çalışmada ise perakende sektöründe önemli bir gün olan Black Friday gününe yönelik gerçekleştirilecek satış tahminleme görevi açısından hangi makine öğrenmesi algoritmasının en iyi performansı sergilediği araştırılmaktadır. Alanyazındaki araştırmalara ek olarak ise GridSearchCV kullanılarak gerçekleştirilen hiperparametre ayarlamalarının varsayılan hiperparametrelere göre model performansına nasıl etki ettiği irdelenmektedir.

## 2. KULLANILAN ALGORİTMALAR

Makine öğrenmesi, bilgisayar sistemlerinin veriler üzerinden öğrenmesini sağlayan bir alan olarak öne çıkmaktadır. Bu alandaki algoritmalar, genellikle denetimli öğrenme, denetimsiz öğrenme ve pekiştirmeli öğrenme olmak üzere üç temel kategoriye ayrılmaktadır.

Denetimli öğrenme, hem girdi (input) hem de çıktı (output) verilerine dayalı olarak bir model geliştirmektedir (Sen ve diğ., 2020). Bu tür algoritmalar, öğrenme sürecinde etiketlenmiş veri kümesini kullanmaktadır. Sınıflandırma ve regresyon algoritmaları, denetimli öğrenmenin iki ana alt dalını oluşturmaktadır. Sınıflandırma, belirli bir girdi için doğru sınıfı tahmin etmeye çalışmaktadır. Örneğin, bir şirket müşteri geri bildirimleriyle ilgili bir anket düzenlediğinde, sınıflandırma algoritmalarını kullanarak bu geri bildirimleri analiz edebilir. Bu bağlamda, algoritma, müşteri yorumlarını olumlu, olumsuz veya nötr olarak sınıflandırabilmektedir. Regresyon ise çıktının sürekli bir sayısal değer olduğu durumlarda kullanılmaktadır. Örneğin, bir şirket reklam harcamaları ile satışlar arasındaki ilişkiyi anlamak amacıyla regresyon analizi kullanabilmektedir. Bu açıdan reklam harcamaları sürekli bir sayısal değer olarak kabul edilirken, satış hacmi regresyon analizi ile tahmin edilmeye çalışılır. Bu sayede reklam bütçesinin artırılması veya azaltılmasının, satışlara olan etkisi daha net bir şekilde anlaşılabilmektedir.

Denetimsiz öğrenme, etiketlenmemiş veri kümesi üzerinde çalışmaktadır. Bu tür algoritmalar, veri içindeki yapıları keşfetmeye çalışarak genellikle gruplamalar (clustering) ve boyut indirgeme (dimensionality reduction) gibi görevlerle ilgilenmektedir. Pekiştirmeli öğrenme ise bir yapay zekâ ajanının çevresiyle etkileşimde bulunarak belirli bir görevi öğrenmeye çalıştığı bir öğrenme yaklaşımını ifade etmektedir. Bu paradigma içinde ajan, çeşitli eylemleri gerçekleştirir ve bu eylemlerin sonuçlarına bağlı olarak ödüller veya cezalar alır. Ajan doğru bir eylem yaptığında ödüllendirilir, yanlış bir eylem yaptığında ise cezalandırılır. Bu şekilde istenen davranışı öğrenmektedir (Sathya ve Abraham, 2013). Son dönemlerde, denetimli ve denetimsiz öğrenmenin bir birleşimi olarak ortaya çıkan yarı denetimli öğrenme algoritmaları belirli problemlerde kullanılmaktadır. Yarı denetimli öğrenme hem etiketlenmiş hem de etiketlenmemiş veri kümesini kullanarak öğrenen algoritmaları ifade etmektedir. Genellikle büyük veri setlerinde, verilerin yalnızca bir kısmının etiketlenmiş olduğu durumlar için kullanılmaktadır (Van Engelen ve Hoos, 2020).

Makine öğrenmesi algoritmaları, problem türüne ve kullanılacak veri kümesine bağlı olarak seçilmektedir. Bu araştırmada, bir perakende firmasının Black Friday gününde yaptığı satışları tahminlemek amacıyla LR, RF, KNN, XGB, DT ve LGBM olmak üzere toplam altı adet denetimli regresyon algoritması kullanılmıştır.

### 2.1. Doğrusal Regresyon

LR, istatistik ve makine öğrenmesinde kullanılan temel bir regresyon analiz yöntemidir. Bu yöntem, bir bağımsız değişkenin, bir veya birden fazla bağımlı değişkeni, tahmin etmesi için kullanılmaktadır (Kim ve diğ., 2022). LR, bu bağımlılık ilişkisini birinci dereceden bir polinomla ifade eder ve genellikle Eşitlik 1'deki formül ile temsil edilir:

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (1)$$

Bu formülde $Y$, bağımlı değişkeni; $X$, bağımsız değişkeni temsil etmektedir. $\beta_0$, LR'nin kesme noktasını ifade eden sabit bir katsayıdır. $\beta_1$, bağımsız değişkenin yani $X$ 'in katsayısını belirtmektedir. $\varepsilon$ ise hata terimidir ve gözlemlenen ve tahmin edilen değer arasındaki farkı ifade etmektedir. LR, gözlemler arasındaki ilişkiyi simgeleyen en uygun doğruyu (regresyon çizgisini) bulmaya çalışmaktadır. Bu, genellikle gözlemler arasındaki hataların karelerinin toplamını minimize eden en küçük kareler yöntemi ile gerçekleştirilmektedir. LR analizi, bağımsız değişkenin etkisinin istatistiksel olarak anlamlı olup olmadığını belirlemek ve gelecekteki değerleri tahmin etmek için yaygın olarak kullanılmaktadır.

### 2.2. Rastgele Orman Regresyonu

RF tekniği, birçok karar ağacının bir araya getirilmesi ile oluşturulan bir topluluk öğrenmesi (ensemble) modelidir. RF, ağaçların rastgele örnekler ve özellikler üzerinde eğitildiği bir yöntemdir. Ağaçlar arasındaki çeşitliliği artırmak ve aşırı öğrenmeyi (overfitting) önlemek için kullanılmaktadır (Jain ve diğ., 2022). RF, ekseriyetle Eşitlik 2'deki formül ile ifade edilmektedir.

$$Y = \beta_0 + \sum_{j=1}^{M} f_j(X) + \varepsilon \quad (2)$$

Söz konusu formülde $Y$ bağımlı değişkeni temsil etmektedir. $\beta_0$ modelin sabit bir terimini gösteren bir katsayıdır. $M$ kullanılan ağaç sayısını simgelemektedir. $f_j(X)$ bir ağacın bağımlı değişkeni tahmin etmek için kullandığı fonksiyonu belirtmektedir. $\varepsilon$ hata terimidir ve gözlemlenen ve tahmin edilen değer arasındaki farkı ifade etmektedir. RF, her ağacın rastgele bir alt kümesi üzerinde eğitildiği ve her biri farklı özellikleri dikkate aldığı için geniş bir uygulama yelpazesine sahiptir. Topluluk öğrenmesi modeli, birbirinden farklı ağaçların tahminlerini birleştirerek genel bir tahmin oluşturur ve bu genellikle daha güçlü ve kararlı bir regresyon modeli sağlamaktadır. RF, büyük boyutlu veri setleriyle başa çıkabilmekte, gürültülü veya eksik verilerle etkili bir şekilde çalışabilmektedir. Veri setinin boyutuna ve ölçeğine duyarlılığı azdır. Ayrıca, özellik seçimi konusunda da etkili olmakla birlikte aşırı uyum sorununu önleyerek daha iyi genelleştirme sağlamaktadır (Eker ve diğ., 2023). Bu yöntem özellikle büyük ve karmaşık veri setlerinde etkili olmaktadır.

### 2.3. K-En Yakın Komşu Regresyonu

KNN, bir gözlemin tahmin edilmesinde, komşu gözlemlerin ortalamasının kullanıldığı bir regresyon yöntemidir. Bu yöntem, bir örneği tahmin etmek için en yakın k komşusunun etkisini kullanır (Kohli ve diğ., 2020). KNN, Eşitlik 3'teki formül ile ifade edilmektedir.

$$\hat{Y}(\pi) = \frac{1}{k} \sum_{i=1}^{k} y_i \quad (3)$$

Eşitlikteki formülde $\hat{Y}(\pi)$, $\pi$ noktasının tahmini değerini temsil etmektedir. $k$ komşu sayısını göstermektedir. $y_i$, $\pi$ noktasına en yakın $k$ komşunun bağımlı değişken değerlerini simgelemektedir. KNN, bir örneği tahmin etmek için, önce bu örnek ile diğer tüm örnekler arasındaki uzaklıkları hesaplamaktadır. Daha sonra, en yakın $k$ komşuyu seçer ve bu komşuların bağımlı değişken değerlerinin ortalamasını kullanarak tahmin yapar. Uzaklık hesaplama genellikle Öklidyen Mesafe veya Manhattan Mesafe gibi metriklerle gerçekleştirilir. KNN, özellikle veri setinin düzensiz dağıldığı ve karmaşık yapıları olan durumlarda kullanışlıdır.

### 2.4. XGBoost Regresyonu

XGB, bir dizi zayıf öğrenici (weak learner) modelin bir araya getirilmesiyle güçlü bir model oluşturma hedefindedir. XGB, özellikle regresyon problemlerinde yüksek performans sergileyen bir gradyan artırma (gradient boosting) algoritmasıdır. Bu yöntem, zayıf tahmincileri (genellikle karar ağaçları) birleştirerek güçlü bir model oluşturmaktadır (Zhu ve diğ., 2021). Gradyan artırma algoritmaları, önceki tahmincilerin hatalarını düzeltmeye odaklanarak iteratif bir şekilde öğrenmektedir. Gradyan artırmanın temel prensibi, bir hata fonksiyonunun (loss function) gradyanına göre bir sonraki tahmincinin parametrelerini güncellemektir. Bu süreç, kayıp fonksiyonunu minimize etmek üzere gerçekleştirilir (Wang ve diğ., 2022). XGB, Eşitlik 4'te yer alan formülle temsil edilmektedir.

$$\hat{Y}_i = \phi(x_i) = \sum_{k=1}^{K} f_k(x_i), \text{where } f_k \in F \quad (4)$$

Yukarıdaki formülde $\hat{Y}_i$, $i$-inci gözlemin tahmini değerini; $x_i$, $i$-inci gözlemin özellik vektörünü; $f_k$, $k$-inci ağaç modelini temsil etmektedir. $F$ ise tüm ağaç modellerinin kümesidir. XGB, çeşitli zayıf öğrenicileri birleştirmek için gradyan artırma yöntemini kullanır. Her bir ağaç modeli, önceki ağaçlardan kaynaklanan hataları düzeltmeye odaklanır. Her ağaç, önceki ağaçların öğrenemediği hataları ele alır ve bu sayede genel hatayı azaltır ve optimizasyon gerçekleştirilir. Optimizasyonun amacı, ağırlıklı bir şekilde hata terimine dayalı olan bir kayıp fonksiyonunu minimize etmektir. Bu amaçla, birinci ve ikinci türevler kullanılarak iteratif bir optimizasyon süreci uygulanmaktadır.

### 2.5. Karar Ağacı Regresyonu

DT bir veri kümesindeki ilişkileri belirlemek ve gelecekteki değerleri tahmin etmek için kullanılmaktadır (Mahendra ve Roopashree, 2023). DT, veri kümesindeki değişkenlerin değerlerine göre veri noktalarını bölerek bir ağaç yapısı oluşturur. Veri noktaları her dalda belirli bir kurala göre bölünür ve bu bölünmeler sonucunda elde edilen alt küme veri noktalarının ortalama değeri, o bölgenin tahmini değeri olarak kullanılır. Bu yöntem, doğrusal olmayan ilişkileri tanımlama yeteneği sayesinde lineer regresyon gibi yöntemlerden daha esnek bir yapı sunmaktadır (Thomas ve diğ., 2020). Ayrıca, karar ağaçları, sonuçları açıkça yorumlanabilir yapıda olduğundan modelin nasıl kararlar aldığının anlaşılmasının önemli olduğu alanlar açısından sık tercih edilmektedir (Gilmore ve diğ., 2021). DT'nin bir avantajı da aşırı uyuma eğilimini azaltabilmesidir. Ağaç yapısının derinliği, bölünmelerin sayısı ve diğer hiperparametreler ayarlanmasıyla modelin genelleme yeteneği artırılabilmektedir. Ancak, çok derin ağaçların oluşturulması modelin aşırı uyuma olan direncini kırabilmektedir. Bu nedenle hiperparametrelerin dikkatli bir şekilde belirlenmesi önemli bir unsurdur.

### 2.6. LightGBM Regresyonu

LGBM, topluluk öğrenme tekniklerinden yararlanarak, büyük veri setleri üzerinde etkili performans sergileyen bir makine öğrenimi algoritması olarak öne çıkmaktadır. Bu algoritma, karar ağaçlarının bir araya gelmesiyle oluşturulan bir modelleme yaklaşımı benimseyerek, karmaşık ilişkileri başarılı bir şekilde modelleyebilmektedir (Talkhi ve diğ., 2023). LGBM, düşük bellek kullanımıyla birlikte yüksek hızda çalışabilmektedir. Bu özelliğiyle büyük ölçekli veri setlerindeki analizler için tercih sebebidir (Zhang ve Gong, 2020). Ayrıca, LGBM, değişkenler arasındaki etkileşimleri ve doğrusal olmayan ilişkileri daha iyi yakalayabilme yeteneğiyle dikkat çekmektedir. LGBM, değişkenler arasındaki etkileşimleri ve doğrusal olmayan ilişkileri yakalayabilme yeteneğini, karar ağaçları üzerindeki belirli özellikleri optimize ederek elde etmektedir. LGBM'in yapısı, karar ağaçlarının dallara bölünmesi sırasında, daha önce bölünmüş olan dalların göz önünde bulundurulmasını sağlar. Bu sayede, karar ağaçları daha az derinleştirilerek daha geniş bir alanı kapsayacak şekilde optimize edilir. LGBM bu özellikleri sayesinde, özellikle büyük ve karmaşık yapıdaki veri setleri üzerinde etkili bir şekilde kullanılabilecek önemli bir makine öğrenimi aracı olarak kabul edilmektedir (Chen ve diğ., 2019).

## 3. YÖNTEM

### 3.1. Performans Metrikleri

Regresyon modelleri oluşturulduktan sonra en iyi tahminlemeyi yapan algoritmanın belirlenmesinde modelin ve problemin yapısına uygun bazı performans değerlendirme metriklerinden faydalanılmaktadır. Araştırmada kullanılan metrikler Ortalama Mutlak Hata (Mean Absolute Error – MAE), Ortalama Kare Hata (Mean Squared Error – MSE), Kök Ortalama Kare Hata (Root Mean Squared Error – RMSE) ve R-kare ($R^2$) skoru şeklindedir.

MAE, bir regresyon modelinin tahmin ettiği değerlerle gerçek değerler arasındaki farkları ölçmektedir. Ancak, bu farkları değerlendirmek için her bir farkın mutlak değeri alınır, yani negatif değerler pozitif hale getirilir. Bu mutlak değerlerin ortalaması alındığında, her bir tahminin gerçek değerden ne kadar uzak olduğunu temsil eden bir sayı elde edilmektedir (Chai ve Draxler, 2014). MAE'nin formülü Eşitlik 5'te verilmiştir.

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

MSE, regresyon modelinin tahminlerinin gerçek değerlere olan karesel uzaklığını ölçmektedir. Sonrasında, bu kare hata değerlerinin ortalaması alınır. Bu sayede, modelin genel olarak ne kadar hata yaptığını ölçen bir sayıya dönüştürülür. Hatalar karesi alındığı için büyük hataların etkisi büyümektedir ve MSE büyük hatalara daha fazla vurgu yapmaktadır. Ancak, kare alma işlemi, aykırı (outlier) değerlere daha hassas olma eğilimindedir (Wang ve Bovik, 2009). MSE'nin formülü Eşitlik 6'da verilmiştir.

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

RMSE, MSE'nin kareköküdür ve MSE'nin birimine geri dönmektedir. Yani bu metrik, tahmin hatalarının standart sapmasını ifade etmektedir. RMSE, MSE ile aynı avantajlara sahiptir ancak orijinal ölçü birimine döndüğü için tahmin hatalarının standart sapmasını yorumlamak daha kolaylaşmaktadır (Chai ve Draxler, 2014). RMSE'nin formülü Eşitlik 7'de verilmiştir.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \check{y}_i)^2} \quad (7)$$

$R^2$ skoru, regresyon modellerinin ne kadar iyi bir şekilde veriyi açıkladığını ölçen istatistiksel bir ölçüdür. Bu skor, bağımlı değişkenin varyasyonunun bağımsız değişkenler tarafından açıklanan yüzdesini temsil etmektedir (Chicco ve diğ., 2021). $R^2$ skoru, 0 ile 1 arasında bir değer almaktadır. Skor, 1'e ne kadar yakınsa, modelin veriyi o kadar iyi açıkladığı anlamına gelir. 0'a ne kadar yakınsa, modelin veriyi açıklama gücü o kadar düşüktür. Formülü Eşitlik 8'de yer almaktadır.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \check{y}_i)^2}{\sum_{i=1}^{n}(y_i - \grave{y}_i)^2} \quad (8)$$

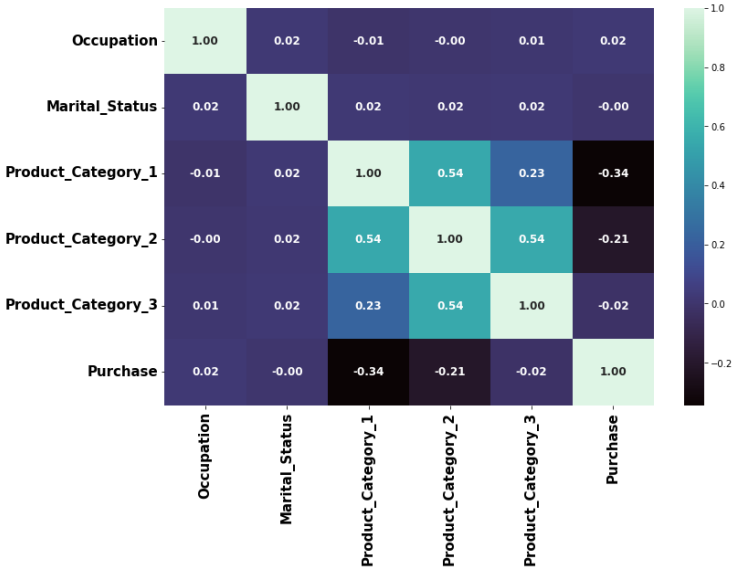
### 3.2. Veri Seti

Araştırmada kullanılan veri seti "Analytics Vidhya Black Friday Sales Dataset" isminde olup, Kaggle platformunda da bulunabilen popüler bir açık kaynaklı veri setidir (Analytics Vidhya, 2016). Veri setinde bulunan veriler Amerika Birleşik Devletleri'nde faaliyet gösteren, ismi "ABC Private Limited" şeklinde gizlenmiş bir perakende firmasına aittir. Bu veri seti, çeşitli makine öğrenmesi modellerini eğitmek, Black Friday satışları sırasında gerçekleştirilen müşteri davranışlarını belirlemek ve satış tutarlarını tahmin etmek için kullanılmaktadır. Satın alma tahminleri gerçekleştirilerek müşterilerin yoğunlukla tercih ettiği ürünler ortaya çıkarılmakta ve bu sayede perakendecilerin müşterilerine sundukları teklifleri incelemeleri ve uyarlamaları mümkün olmaktadır. Veri seti iki kısımdan oluşmaktadır. Eğitim seti 550,069 kayıt ve 12 öznitelik içermektedir. Test seti ise 233,600 kayıttan oluşmakta ve eğitim setinden farklı olarak satın alma tutarı özniteliğini içermediği için 11 özniteliğe sahiptir. Veri gizliliğini sağlamak amacıyla müşterilerin meslekleri ve satın alınan ürünlerin kategorileri gibi bazı öznitelikler veri setinin yayımcısı tarafından maskelenmiştir. Tablo 1'de, veri setine ait özniteliklere ve açıklamalara, Şekil 1'de veri setinin ısı haritasına yer verilmiştir.

**Tablo 1. Veri Setinde Bulunan Öznitelikler ve Açıklamaları**

| Öznitelik | Açıklama |
|---|---|
| User_ID | Müşterinin eşsiz kimliği |
| Product_ID | Ürünün eşsiz kimliği |
| Gender | Müşterinin cinsiyeti |
| Age | Müşterinin yaşı |
| Occupation | Müşterinin mesleği (maskelenmiş) |
| City_Category | Müşterinin yaşadığı şehrin kategorisi |
| Stay_In_Current_City | Müşterinin bulunduğu şehirde ne kadar süredir yaşadığı |
| Maritial_Status | Müşterinin medeni durumu |
| Product_Category_1 | Ürün kategorisi (maskelenmiş) |
| Product_Category_2 | Ürün kategorisi (maskelenmiş) |
| Product_Category_3 | Ürün kategorisi (maskelenmiş) |
| Purchase | Müşterinin satın alım miktarı |

**Şekil 1. Veri Setine Ait Isı Haritası**




### 3.3. Verilerin Hazırlanması

Regresyon algoritmalarının uygulanmasına başlanmadan önce veri üzerinde bazı ön işlemler gerçekleştirilmiştir. İlk olarak kategorik veriler, analiz sürecinde kullanılabilir hale gelebilmesi için sayısal değerlere dönüştürülmüştür. "Yaş (age)" ve "bulunulan şehirde kalma süreleri (stay_in_current_city)" gibi belirli aralıkları temsil eden kategorik değişkenler tam sayı veya sürekli bir değer olarak sayısal bir formata dönüştürülerek istatistiksel analizlere hazır hale getirilmiştir. Ürün kategorisi 2 ve 3'teki eksik değerlere karşı çözüm olarak "Ürün Kategorisi 0" adında yeni bir kategori oluşturulmuştur. Daha sonra, bu yeni kategoriye mod değeri atanarak eksik verilerin işlenmesi sağlanmıştır. Mod değeri, bir veri kümesinde en sık tekrarlanan değeri temsil etmektedir (Huang, 1998). Bu şekilde, eksik veriler genel eğilimlere daha yakın bir şekilde doldurularak veri seti üzerindeki boşluklar giderilmektedir. Ayrıca, test setinin satın alım tutarlarına sahip olmaması durumunda, eğitim setinden alınan harcama tutarları kullanılarak test ve geliştirme verilerine atama (imputation) yapılmıştır. Bahsedilen ön işlemler, veri setini regresyon algoritmalarına daha uygun hale getirmek ve model performansını artırmak için gerçekleştirilmiştir.

Çalışma kapsamında veri analizi ve model testleri için Python programlama dili kullanılmıştır. Veri analizi sürecinde pandas ve NumPy gibi temel veri işleme kütüphaneleri işe koşulmuş, model oluşturma ve test aşamalarında ise scikit-learn kütüphanesi tercih edilmiştir. Sonuçlar, matplotlib ve seaborn gibi görselleştirme kütüphaneleriyle analiz edilmiş ve bulgular grafiklerle sunulmuştur. Bu işlemler Jupyter Notebook geliştirme ortamında gerçekleştirilmiştir. Bununla birlikte, veri seti hacminin büyüklüğünden dolayı bazı analizlerin ve model testlerinin gerçekleştirilmesinde Google Colab gibi bulut tabanlı hizmetlerden de faydalanılmıştır. Bu hizmetler, daha fazla işlem gücü ve bellek kapasitesi sağlayarak analiz sürecini hızlandırmış ve bu büyüklükteki bir veri setiyle çalışmayı mümkün kılmıştır.

## 4. DENEYSEL ÇALIŞMA VE BULGULAR

Bir perakende mağazasının Black Friday günündeki satışlarını tahminlemeyi amaçlayan bu çalışmada, denetimli regresyon algoritmalarından LR, RF, KNN, XGB, DT ve LGBM algoritmaları kullanılmıştır. Modeller oluşturulurken, veri seti %75 eğitim ve %25 test olmak üzere iki parçaya bölünmüştür. Araştırmanın deneysel kısmında öncelikle Keşifsel Veri Analizi (Exploratory Data Analysis – EDA) yürütülerek veri seti üzerinde bazı çıkarımlarda bulunulmuştur. Sonrasında algoritmalara ait performans ölçümleri gerçekleştirilerek karşılaştırmalar yapılmıştır. Modeller için en iyi hiperparametre ayarları, GridSearchCV yöntemi kullanılarak belirlenmiştir. Bu yöntem, belirtilen hiperparametre aralıkları içinden farklı kombinasyonlar deneyerek en iyi performansı veren hiperparametreleri seçmektedir. Örneğin, bir RF modeli için GridSearchCV kullanarak hiperparametrelerin ayarlanması istenildiğinde, GridSearchCV belirlenen hiperparametre aralıkları içinden her bir kombinasyonu deneyerek birçok farklı RF modeli oluşturmaktadır. Daha sonra, bu modelleri belirli bir metrik (genellikle çapraz doğrulama) kullanarak değerlendirmekte ve en iyi performansı veren hiperparametre kombinasyonunu seçmektedir (Ranjan ve diğ., 2019). Tablo 2'de, farklı

makine öğrenmesi modelleri için GridSearchCV kullanılarak belirlenen en iyi hiperparametre ayarları listelenmiştir. LR hiperparametre almayan bir model olduğu için herhangi bir ayar girilmemiştir.

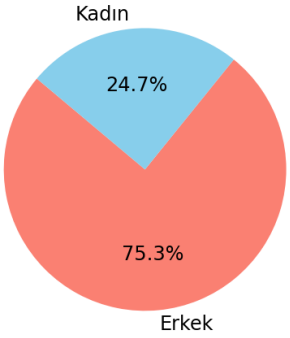
**Tablo 2. Modeller için En İyi Hiperparametre Ayarları**

| Model | Hiperparametreler | Ayarlar |
|---|---|---|
| LGBM | n_estimators | 1000 |
| | learning_rate | 0.05 |
| | max_depth | 5 |
| | num_leaves | 31 |
| | min_child_samples | 20 |
| | subsample | 0.8 |
| | colsample_bytree | 0.8 |
| | reg_alpha | 0.1 |
| | reg_lambda | 0.1 |
| | metric | 'rmse' |
| DT | max_depth | 10 |
| | min_samples_split | 5 |
| | min_samples_leaf | 2 |
| | max_features | 'auto' |
| XGB | objective | 'reg:linear' |
| | colsample_bytree | 0.3 |
| | learning_rate | 0.05 |
| | max_depth | 10 |
| | alpha | 10 |
| | n_estimators | 1000 |
| KNN | n_neighbors | 3 |
| | weights | 'distance' |
| | algorithm | 'auto' |
| | leaf_size | 30 |
| RF | n_estimators | 1200 |
| | min_samples_split | 5 |
| | min_samples_leaf | 1 |
| | max_features | 'sqrt' |
| | max_depth | 25 |
| | bootstrap | True |
| LR | - | - |

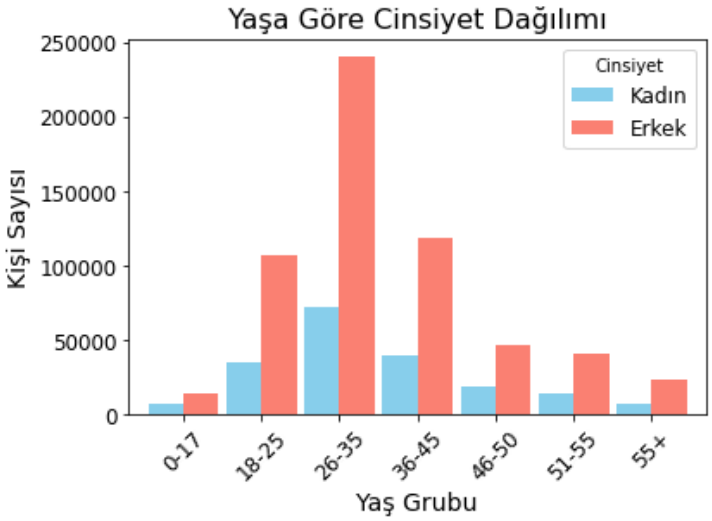
### 4.1. Keşifsel Veri Analizi

EDA, bir veri setinin temel özelliklerini ve yapılarını anlamak, potansiyel desenleri keşfetmek ve önemli bilgileri ortaya çıkarmak amacıyla uygulanan bir veri analizi yöntemidir (Milo ve Somech, 2020). Veri setinin istatistiksel özetleri ve grafiksel gösterimleri kullanılarak gerçekleştirilen bu süreç, verilerdeki eğilimleri, dağılımları ve ilişkileri belirleyerek veri setinin içsel yapısını ortaya koymayı amaçlamaktadır. EDA, veri madenciliği projelerinde ve veri bilimi çalışmalarında genellikle ilk aşama olarak kullanılmaktadır. Bu yöntem, derinlemesine analizlerin ve modelleme çalışmalarının temelini oluşturarak veri tabanlı kararlar alınmasına yardımcı olmaktadır. Şekil 2'de müşterilerin cinsiyet dağılımı verilmiştir. Şekil 3'te ise müşteri cinsiyetlerinin yaş gruplarına göre dağılımını gösteren grafik bulunmaktadır.

**Şekil 2. Cinsiyet Dağılımı**

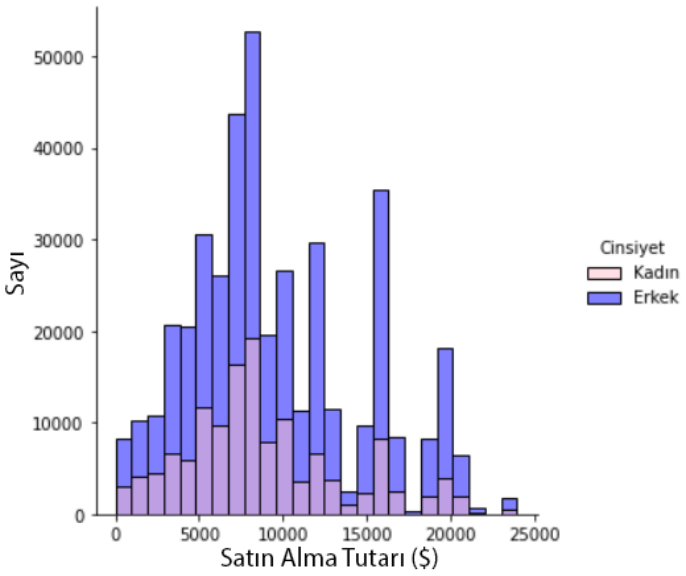



**Şekil 3. Cinsiyetlerin Yaş Gruplarına Göre Dağılımı**



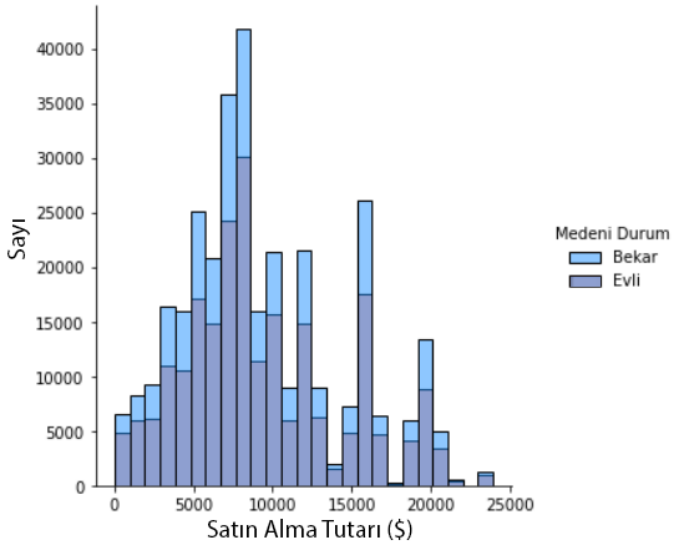

Şekil 2 incelendiğinde müşterilerin %75,3'ünün erkek, %24,7'sinin kadın olduğu görülmektedir. Şekil 3'teki verilere göre ise müşterilerin en çok 26-35 yaş aralığında bireylerden oluştuğu, 17 yaşın altındaki müşterilerin diğer yaş gruplarına kıyasla daha az oranda bulunduğu anlaşılmaktadır. Şekil 4'te cinsiyete göre satın alma tutarlarının dağılımı ifade edilmiştir.

**Şekil 4. Cinsiyete Göre Satın Alma Tutarlarının Dağılımı**



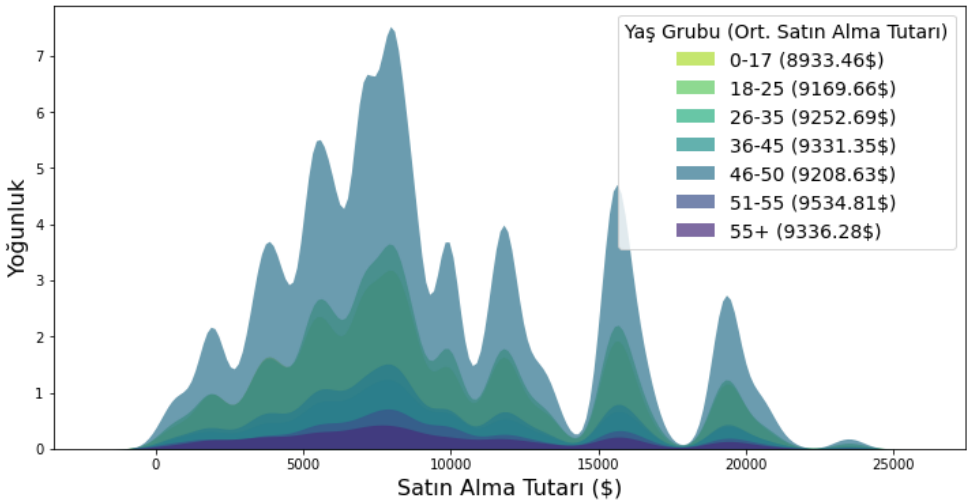
Şekil 4'e göre, erkek müşterilerin harcamalarının kadın müşterilerden belirgin şekilde daha yüksek olduğu gözlemlenmiştir. Ayrıca, genel alışveriş harcamalarının çoğunluğunun 5.000-10.000$ arasında yoğunlaştığı saptanmıştır. Şekil 5'te, müşterilerin medeni durumuna göre satın alım tutarlarının dağılımına yer verilmiştir.

**Şekil 5. Medeni Duruma Göre Satın Alma Tutarlarının Dağılımı**



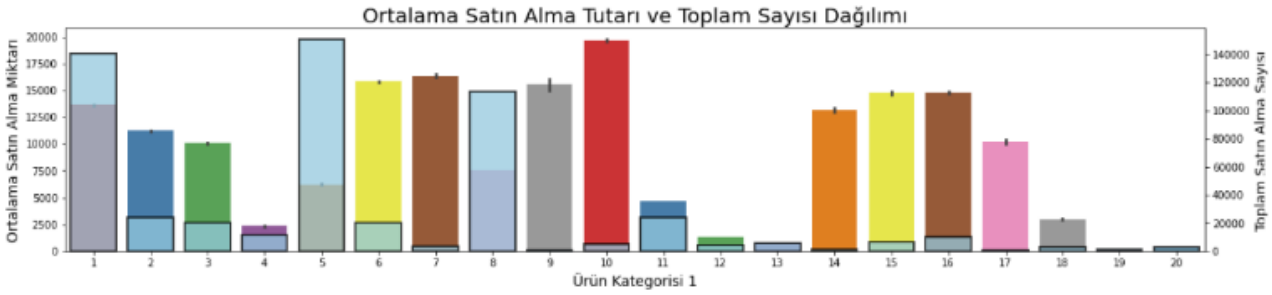

Şekil 5 üzerinde yapılan inceleme, bekar müşterilerin evli müşterilere göre gözle görülür bir biçimde daha fazla harcama yaptığını göstermektedir. Bu durum, bekar bireylerin genellikle daha fazla kişisel harcamaya eğilimli oldukları, belirli ürün veya hizmet kategorilerine daha fazla ilgi gösterdikleri şeklinde geniş bir yoruma açıktır. Şekil 6'da müşterilerin yaş gruplarına göre satın alım tutarlarının ortalamaları gösterilmektedir.

**Şekil 6. Yaş Gruplarına Göre Satın Alma Tutarları Ortalamalarının Dağılımı**

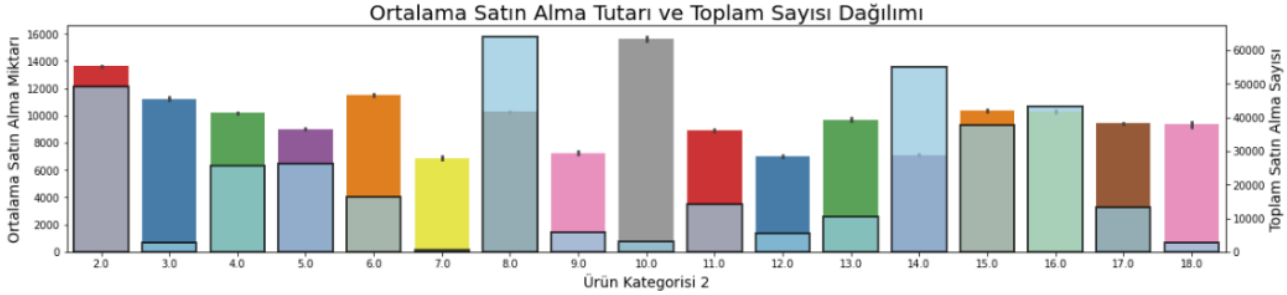



Şekil 6'ya bakıldığında, alışveriş tutarlarının en yüksek ortalamaya sahip olduğu yaş grubu 51-55 yaş aralığıdır. En düşük ortalamaya sahip grup ise 0-17 yaş arasındaki bireylerden oluşmaktadır. Satın alım tutarı ortalamaları 9.200$ bandında yoğunlaşmaktadır. Bu durum, genel müşteri kitlesinin çoğunluğunun benzer bir harcama seviyesine sahip olduğunu ve bu tutarın bir tür merkezi eğilim noktasına tekabül ettiğini göstermektedir. Şekil 7, 8 ve 9'da ürün kategorilerine göre ortalama satın alma tutarlarının ve toplam satın alma sayılarının dağılımları verilmiştir.

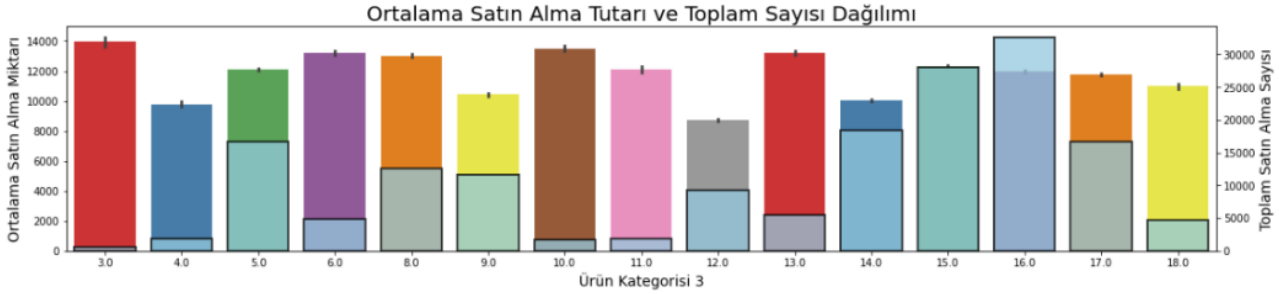
**Şekil 7. Ürün Kategorisi 1'e Göre Ortalama Satın Alma Tutarları ile Toplam Satın Alma Sayılarının Dağılımı**




**Şekil 8. Ürün Kategorisi 2'ye Göre Ortalama Satın Alma Tutarları ile Toplam Satın Alma Sayılarının Dağılımı**




**Şekil 9. Ürün Kategorisi 3'e Göre Ortalama Satın Alma Tutarları ile Toplam Satın Alma Sayılarının Dağılımı**




Şekil 7, 8 ve 9 incelendiğinde, farklı ürün kategorilerindeki belirli ürün numaralarının yüksek değerli (premium) ürünleri temsil ettiğini göstermektedir. Birinci ürün kategorisinde, özellikle 9, 10 ve 14 numaralı ürünlerin "yüksek değerli" olduğu gözlemlenmektedir. İkinci ürün kategorisinde, 10 numaralı ürünün "seçkin" bir ürün olduğu anlaşılmaktadır. Üçüncü ürün kategorisinde ise 3 ve 10 numaralı ürünlerin "yüksek değerli" kategoride yer aldığı görülmektedir. Bu durum, belirli ürün numaralarının sadece kendi kategorilerinde değil, genel olarak yüksek değerli ürünler olduğunu ifade etmektedir.

### 4.2. Algoritmaların Performans Ölçümleri

Araştırmanın bu bölümünde, bir perakende firmasının Black Friday gününde yaptığı satışları tahminlemek amacıyla kullanılan LR, RF, KNN, XGB, DT ve LGBM algoritmalarına ilişkin performans ölçümleri değerlendirilmiştir. Her bir algoritmanın performansını değerlendirmek, regresyon modelinin ne kadar iyi tahmin yaptığını ölçmek amacıyla MAE, MSE, RMSE ve $R^2$ metrikleri kullanılmıştır. Tablo 3'te, varsayılan hiperparametrelerle, Tablo 4'te ise GridSearchCV ile ayarlanmış hiperparametrelerle algoritmaların tahminleme yeteneklerine dair performans ölçümlerine yer verilmiştir. LR hiperparametre almayan bir model olduğu için iki tabloda da aynı performans değerlerine sahiptir. EK 1'de gerçek satış değerleri ile algoritmalar tarafından tahmin edilen değerlerin karşılaştırması bulunmaktadır.

**Tablo 3. Varsayılan Hiperparametrelerle Modellerin Performansı**

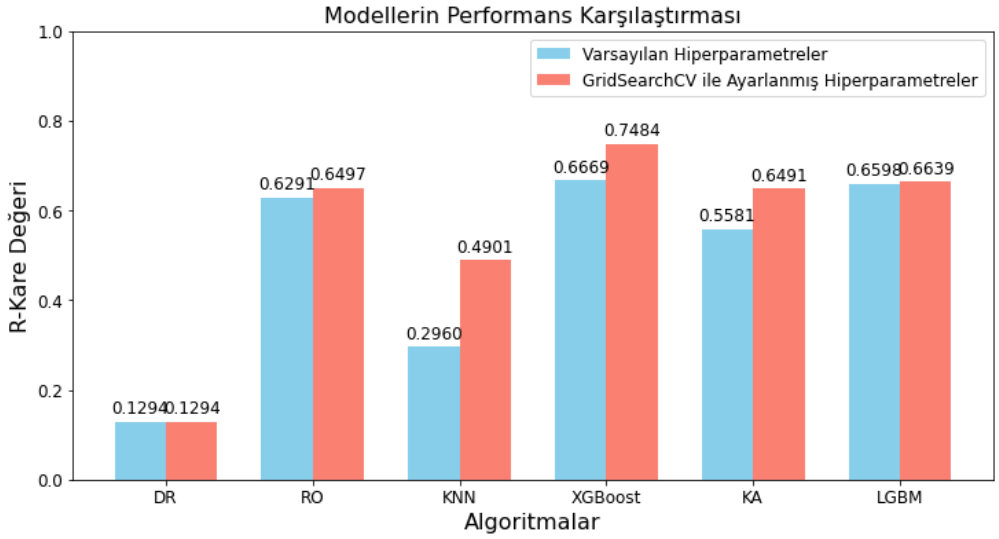
| Algoritma | MAE | MSE | RMSE | $R^2$ |
|---|---|---|---|---|
| LR | 3577.49 | 2193.91 | 4683.92 | 0.1294 |
| RF | 2226.98 | 9346.89 | 3057.26 | 0.6291 |
| KNN | 2925.58 | 1774.11 | 4212.02 | 0.2960 |
| XGB | 2166.31 | 8393.51 | 2897.15 | 0.6669 |
| DT | 2362.04 | 1113.50 | 3336.92 | 0.5581 |
| LGBM | 2201.78 | 8571.63 | 2927.73 | 0.6598 |

**Tablo 4. GridSearchCV ile Ayarlanmış Hiperparametrelerle Modellerin Performansı**

| Algoritma | MAE | MSE | RMSE | $R^2$ |
|---|---|---|---|---|
| LR | 3577.49 | 2193.91 | 4683.92 | 0.1294 |
| RF | 2193.39 | 8826.80 | 2970.99 | 0.6497 |
| KNN | 2561.70 | 1284.94 | 3584.60 | 0.4901 |
| XGB | 1872.67 | 6352.18 | 2520.35 | 0.7484 |
| DT | 2231.30 | 8842.38 | 2973.61 | 0.6491 |
| LGBM | 2179.68 | 8469.63 | 2910.26 | 0.6639 |

MAE değerinin düşük olması, modelin tahminlerinin genellikle gerçek değerlere yakın olduğunu göstermektedir. MSE değeri ne kadar düşükse modelin tahminleri o kadar iyi olmaktadır. RMSE'nin düşük olması modelin genel performansının gücünü ifade etmektedir. $R^2$ değeri ise 1'e yaklaştıkça modelin veriyi iyi açıkladığı anlamına gelmektedir. Tablo 3 ve Tablo 4 incelendiğinde Black Friday satış tahminlemesinde en iyi sonuç veren algoritma GridSearchCV ile ayarlanmış hiperparametrelerle XGB olarak karşımıza çıkmaktadır (RMSE = 2520.35, $R^2$ = 0.7484). En düşük performans değerlerine sahip algoritma ise LR'dir (RMSE = 4683.92, $R^2$ = 0.1294). Bunun yanı sıra hiperparametrelerin GridSearchCV ile ayarlanmasının modellerin performansını genellikle artırdığı anlaşılmaktadır. Şekil 10'da algoritmaların tahminleme doğruluklarını daha iyi karşılaştırabilmek için bir grafik sunulmuştur.

**Şekil 10. Algoritmaların Doğruluk Karşılaştırmaları**




## TARTIŞMA

Araştırma bulgularına dayalı olarak, perakende sektöründe makine öğrenmesi uygulamalarının etkinliğinin ve uygulanabilirliğinin tespit edilmesinin yanı sıra hangi algoritmanın en yüksek performansı sergilediği hakkında önemli bilgiler elde edilmiştir. Araştırmadan elde edilen bulgular, alandaki benzer çalışmalardan alınan bulgularla karşılaştırılarak araştırma sonuçlarının değerlendirilmesi ve bu sonuçların alanyazınla bir bağlantısının kurulması hedeflenmiştir.

Abhinav ve Prasad (2023), Wu ve diğerleri (2018) ile Trung ve diğerleri (2021) tarafından yapılan çalışmalarda, LR, DT, RF, Ekstra Ağaç Regresyonu (Extra Tree Regression – ET) ve XGB modellerinin perakende satış tahminleme görevi açısından performans karşılaştırılmaları gerçekleştirilmiştir. Bu

araştırmanın bulgularına benzer şekilde XGB yüksek tahmin doğruluğuyla öne çıkmıştır. Bu benzerlik, XGB'nin perakende sektöründe satış tahminleme açısından genel olarak etkili bir tahmin aracı olarak kullanılabileceği çıkarımını desteklemektedir. XGB'nin bu etkinliği, büyük veri setlerinde karmaşık ilişkileri modelleme yeteneği ve doğrusal olmayan yapıları ele alma kabiliyeti sayesinde elde edilebilmektedir. XGB'nin karar ağacı tabanlı yapısı veri setindeki değişkenler arasındaki etkileşimleri yakalayabilmekte ve bu da daha kesin tahminler yapılmasını sağlamaktadır. Bu yapının temel mantığı, veri setindeki değişkenler arasındaki karmaşık ilişkileri belirlemek ve öğrenmek için karar ağaçları oluşturmaktır (Amjad ve diğ., 2022). Karar ağaçları, bir sorunun çözümü için bir dizi karar kuralını takip eden ağaç benzeri bir yapıdır. Her iç düğüm (node) bir karar noktasını, her kenar (edge) bir karar kuralını ve her yaprak (leaf) sonuç tahminini temsil etmektedir (Shouval ve diğ., 2021). XGB'nin karar ağacı tabanlı yapısı, bu karar ağaçlarını bir araya getirerek güçlü bir tahmin modeli oluşturulmasını sağlamaktadır.

Alanyazında XGB ile satış tahminleme görevinde düşük performans elde eden çalışmalar da bulunmaktadır (Kalra ve diğ., 2020; Xia ve diğ., 2020). XGB'den en yüksek performansı elde edebilmek için doğru parametre ve model ayarlamalarının yapılması önemlidir. Parametrelerin belirlenmesi ve optimizasyonu, modelin performansını önemli ölçüde etkilemekte ve bu süreç deneyimli bir veri bilimcisinin müdahalesini gerektirmektedir (de Oliveira ve diğ., 2021). XGB'de kullanılan parametrelerin (örneğin, learning rate, max_depth, min_child_weight, subsample gibi) doğru bir şekilde ayarlanması, modelin aşırı uyum veya eksik öğrenme yapmasını önlemek için kritiktir. Aşırı uyum problemlerini çözmek için özel önlemler alınması gerekmektedir. Bu süreç, veri bilimi alanında deneyimi olmayan kişiler için karmaşık ve zorlayıcı olmaktadır. Bu nedenle, alanyazında, model ve parametre ayarlama süreçlerindeki zorluklardan ve bu süreçlerin başarılı bir şekilde yönetilmesi için gereken uzmanlık düzeyinden sıkça bahsedilmektedir (Bergstra ve diğ., 2011; Herodotou ve diğ., 2022; Huang ve diğ., 2019).

Model ve parametre ayarlama süreçlerinde kullanılmak üzere öne çıkan ve bu araştırmada da kullanılan yöntemlerden biri GridSearchCV'dir. Bu yöntem, belirli bir model için en iyi performansı veren parametre kombinasyonunu bulmayı amaçlamaktadır. GridSearchCV'nin çalışma mantığına göre parametre araştırma alanı bir ızgara gibi düşünülmektedir. Her bir ızgara noktası farklı bir parametre kombinasyonunu temsil eder. GridSearchCV, bu ızgaradaki her noktayı modeli eğitmek ve doğrulama setinde test etmek için kullanır. Sonuç olarak, en iyi performansı veren parametre kombinasyonu seçilir ve model bu kombinasyonla eğitilir (Belete ve Huchaiah, 2022). Bu şekilde, GridSearchCV modelin performansını artırmak için en uygun hiperparametreleri belirlemesine yardımcı olmaktadır.

Mayer ve diğerleri (2020) tarafından yapılan araştırma, LR'nin satış tahminlemesi konusunda daha yüksek şeffaflık sağladığını ve anlaşılır bir model gerektiren durumlarda tercih edilebileceğini vurgulamaktadır. Ancak, bu araştırmada LR algoritması satış tahminlemesi konusunda oldukça düşük bir performans göstermiştir. LR modelinin satış tahminleme görevinde en düşük performans değerlerine sahip olması doğrusal olmayan ilişkilere sahip veri setlerinde LR'nin yetersiz kalabileceğini ve daha karmaşık modellerin tercih edilmesi gerektiğini göstermektedir. LR, bağımsız değişkenler ile bağımlı değişken arasındaki ilişkiyi açıklamak için kullanılan basit ve temel bir regresyon modelidir. Bu modelde, bağımlı değişkenin tahmini, bağımsız değişkenler arasında doğrusal bir ilişki kullanılarak yapılır (Uyanık ve Güler, 2013). Yani, bağımlı değişken ile bağımsız değişkenler arasındaki ilişki doğrusal bir fonksiyon ile ifade edilir. Daha karmaşık modeller ise genellikle doğrusal olmayan ilişkileri ifade etmek için kullanılan modellerdir. Bu modeller, LR'den farklı olarak, bağımsız değişkenler ile bağımlı değişken arasındaki ilişkiyi doğrusal olmayan bir şekilde ifade edebilir. Bu bağlamda, LR'nin satış tahmini gibi karmaşık ve çok boyutlu veri setlerinde yetersiz kalabileceği ve daha karmaşık modellerin tercih edilmesi gerekmektedir.

Alanyazında, bu araştırmada kullanılan veri seti üzerinde satış tahminlemesinde bulunan bazı önemli çalışmalar bulunmaktadır. Kalra ve diğerleri (2020) gerçekleştirdikleri çalışmada ET (RMSE = 3252.04), Aher ve diğerleri (2021) RF (MSE = 3062.72), Patil ve diğerleri (2023) RF (RMSE = 3047.12, $R^2$ = 0.6265) ve Abhinav ve Prasad (2023) XGB Regresyonu (RMSE = 2529.36, $R^2$ = 0.7470) algoritmalarını en iyi performans gösteren algoritmalar olarak saptamışlardır. Bu araştırmada ise XGB Regresyonu, RMSE = 2520.35 ve $R^2$ = 0.7484 skorları ile alanyazında aynı veri seti üzerinde satış tahminlemesi gerçekleştiren araştırmalardan daha iyi performans elde etmiştir. Bu araştırmanın diğer araştırmalardan daha yüksek doğruluk elde etmesinin temel sebepleri arasında veri ön işleme adımlarının doğru ve titiz bir şekilde gerçekleştirilmesi yer almaktadır. Veri ön işleme adımlarının başarılı bir şekilde uygulanması, eksik verilerin doğru şekilde doldurulması ve kategorik verilerin uygun şekilde dönüştürülmesi, modelin daha güçlü ve tutarlı tahminler yapmasını sağlamış olabilir. Ayrıca, kullanılan XGB modelinin parametrelerinin GridSearchCV tekniğiyle optimize edilmesi, modelin performansını artıran en önemli etkenlerden biridir. Trung ve diğerleri (2021) de aynı veri setinde GridSearchCV tekniği ile XGB Regresyonunun en yüksek performans verdiği sonucuna ulaşsa da (RMSE =

2647) bu araştırmada daha iyi bir satış tahmin performansı elde edilmiştir. Bunun bir sebebi olarak ise veri seti üzerinde özellik mühendisliği tekniklerinin başarılı bir şekilde uygulanmasının bu farkın oluşmasına katkıda bulunduğu çıkarımı yapılabilir.

## SONUÇ

Araştırma kapsamında elde edilen bulgular, perakende sektöründeki işletmelerin makine öğrenmesi tekniklerini hangi amaçlarla kullanabilecekleri ve makine öğrenmesi uygulamaları yoluyla elde ettikleri bilgileri stratejik ve operasyonel kararlar alma süreçlerinde nasıl işe koşacakları konusunda yardımcı olacak önemli bilgiler sunmaktadır. Özellikle farklı makine öğrenmesi modellerine ait performansların detaylı bir şekilde değerlendirilmesi ve karşılaştırılması, işletmelerin belirli hedeflere ulaşmak için en uygun algoritmaları seçmelerine olanak tanıyacağı düşünülmektedir. Bulgular, XGB'nin, LR, RF, KNN, DT ve LGBM algoritmalarına kıyasla daha yüksek tahmin doğruluğuna sahip olduğunu ortaya koymaktadır. XGB'nin bu yüksek performansı, perakende işletmelerinin satış tahminlemesi, stok yönetimi ve müşteri segmentasyonu gibi kritik süreçlerde daha güvenilir ve etkili sonuçlar elde etmelerine imkân tanıyacağı öngörülmektedir. Diğer taraftan, LR modelinin satış tahminleme görevinde en düşük performans değerlerine sahip olduğu ve satış tahminlenmesi açısından özellikle karmaşık ve büyük veri setlerinde yetersiz kaldığı görülmüştür. Ayrıca, hiperparametrelerin GridSearchCV kullanılarak ayarlanması modellerin performansını belirgin şekilde artırmaktadır. Bunlara ek olarak, araştırma kapsamında gerçekleştirilen EDA ile işletmelere, müşteri davranışlarını daha iyi anlama, talebi öngörme, satış stratejilerini belirleme ve stokları daha etkili bir şekilde yönetme konularında makine öğrenmesi tekniklerini nasıl kullanabilecekleri açısından araştırma bir rehber görevi görmektedir.

Sonuç olarak, bu araştırma perakende sektöründeki işletmelere, veri odaklı kararlar almalarına ve makine öğrenmesi tekniklerini stratejik bir şekilde kullanmalarına yönelik bir kılavuz sunmaktadır. Elde edilen bulguların doğrudan uygulanabilir ve işletmelerin rekabet avantajını artırıcı nitelikte olması, perakende sektöründeki dijital dönüşüm süreçlerine yönelik değerli bir katkı sağlamaktadır.

## KAYNAKÇA


Abhinav, T., & Prasad, P. K. (2023). Black friday sales prediction using machine learning. UGC Care Group I Listed Journal, 13(11), 9-14.

Aher, A., Rajeswari, K., & Vispute, S. (2021). Data Analysis and price prediction of black friday sales using machine learning technique. IJERT, 10(7), 621-627.

Alagarsamy, S., Varma, K. G., Harshitha, K., Hareesh, K., & Varshini, K. (2023, January). Predictive Analytics for Black Friday Sales using Machine Learning Technique. In 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT) (pp. 389-393). IEEE.

Alzubi, J., Nayyar, A., & Kumar, A. (2018, November). Machine learning from theory to algorithms: An overview. In Journal of Physics: Conference Series (Vol. 1142, p. 012012). IOP Publishing.

Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kamiński, P., & Amjad, U. (2022). Prediction of pile bearing capacity using XGBoost algorithm: modeling and performance evaluation. Applied Sciences, 12(4), 2126.

Analytics Vidhya. (2016, July). Black friday sales prediction. https://datahack.analyticsvidhya.com/contest/black-friday

Awan, M. J., Mohd Rahim, M. S., Nobanee, H., Yasin, A., & Khalaf, O. I. (2021). A big data approach to black friday sales. Intelligent Automation & Soft Computing, 27(3), 785-797.

Belete, D. M., & Huchaiah, M. D. (2022). Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. International Journal of Computers and Applications, 44(9), 875-886.

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for hyper-parameter optimization. Advances in Neural Information Processing Systems, 24.

Beştaş, M. (2023). Keşifçi veri analizi ile eczane satış analizi ve satış tahmini. Third Sector Social Economic Review, 58(1), 765-782.

Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. American Journal of Epidemiology, 188(12), 2222-2239.

Bohanec, M., Borštnar, M. K., & Robnik-Šikonja, M. (2017). Explaining machine learning models in sales predictions. Expert Systems with Applications, 71, 416-428.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3), 1247-1250.

Chen, C., Zhang, Q., Ma, Q., & Yu, B. (2019). LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. Chemometrics and Intelligent Laboratory Systems, 191, 54-64.

Chen, J., Koju, W., Xu, S., & Liu, Z. (2021, March). Sales forecasting using deep neural network and SHAP techniques. In 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) (pp. 135-138). IEEE.

Cheriyan, S., Ibrahim, S., Mohanan, S., & Treesa, S. (2018, August). Intelligent sales prediction using machine learning techniques. In 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE) (pp. 53-58). IEEE.

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623.

Çiçek, C. T., & Selçuk, G. D. (2023). Sanal market sektöründe hedef müşteri kitlesinin tanımlanması ve makine öğrenmesi ile tüketim eğilimlerinin tahmini. Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 16(1), 24-35.

de Oliveira, D., Porto, F., Boeres, C., & de Oliveira, D. (2021). Towards optimizing the execution of spark scientific workflows using machine learning-based parameter tuning. Concurrency and Computation: Practice and Experience, 33(5), e5972.

Dılkı, G. (2020). Makine öğrenmesi algoritmalarının sınıflama problemleri üzerinden karşılaştırılması: Satış tahmini. PressAcademia Procedia, 12(1), 82-83.

Ecemiş, O., & Irmak, S. (2018). Paslanmaz çelik sektörü satış tahmininde veri madenciliği yöntemlerinin karşılaştırılması. Kilis 7 Aralık Üniversitesi Sosyal Bilimler Dergisi, 8(15), 148-169.

Eker, R., Alkiş, K. C., Uçar, Z., Aydın, A. (2023). Ormancılıkta makine öğrenmesi kullanımı. Turkish Journal of Forestry, 24(2), 150-177. doi:10.18182/tjf.1282768

Erol, B., & İnkaya, T. (2024). Satış tahmini için uzun kısa-süreli bellek ağı tabanlı derin transfer öğrenme yaklaşımı. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, 39(1), 191-202.

García, S., Luengo, J., & Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. Knowledge-Based Systems, 98, 1-29.

Gilmore, E., Estivill-Castro, V., & Hexel, R. (2021). More interpretable decision trees. In Hybrid Artificial Intelligent Systems: 16th International Conference, HAIS 2021, Bilbao, Spain, September 22–24, 2021, Proceedings 16 (pp. 280-292). Springer International Publishing.

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., ... & Hussain, A. (2024). Interpreting black-box models: a review on explainable artificial intelligence. Cognitive Computation, 16(1), 45-74.

Herodotou, H., Odysseos, L., Chen, Y., & Lu, J. (2022, May). Automatic performance tuning for distributed data stream processing systems. In 2022 IEEE 38th International Conference on Data Engineering (ICDE) (pp. 3194-3197). IEEE.

Huang, C., Li, Y., & Yao, X. (2019). A survey of automatic parameter tuning methods for metaheuristics. IEEE Transactions on Evolutionary Computation, 24(2), 201-216.

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Mining and Knowledge Discovery, 2(3), 283-304.

İyzico. (2022). 2022 iyzico Black Friday Karnesi.

Jagatheesaperumal, S. K., Rahouti, M., Ahmad, K., Al-Fuqaha, A., & Guizani, M. (2021). The duo of artificial intelligence and big data for industry 4.0: Applications, techniques, challenges, and future research directions. IEEE Internet of Things Journal, 9(15), 12861-12885.

Jain, P., Choudhury, A., Dutta, P., Kalita, K., & Barsocchi, P. (2021). Random forest regression-based machine learning model for accurate estimation of fluid flow in curved pipes. Processes, 9(11), 2095.

Kalra, S., Perumal, B., Yadav, S., & Narayanan, S. J. (2020, February). Analysing and predicting the purchases done on the day of Black Friday. In 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE) (pp. 1-8). IEEE.

Kim, S. J., Bae, S. J., & Jang, M. W. (2022). Linear regression machine learning algorithms for estimating reference evapotranspiration using limited climate data. Sustainability, 14(18), 11674.

Kohli, S., Godwin, G. T., & Urolagin, S. (2020). Sales prediction using linear and KNN regression. In Advances in Machine Learning and Computational Intelligence: Proceedings of ICMLCI 2019 (pp. 321-329). Singapore: Springer Singapore.

Liao, W., Ye, G., Yin, Y., Yan, W., Ma, Y., & Zuo, D. (2020, November). Auto Parts Sales Prediction Based on Machine Learning for small data and a long replacement cycle. In 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA) (pp. 1-5). IEEE.

Ma, L., & Sun, B. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. International Journal of Research in Marketing, 37(3), 481-504.

Mahendra, G., & Roopashree, H. R. (2023, February). Prediction of road accidents in the different states of India using machine learning algorithms. In 2023 IEEE International Conference on Integrated Circuits and Communication Systems (ICICACS) (pp. 1-6). IEEE.

Marr, B. (2016). Big data in practice: how 45 successful companies used big data analytics to deliver extraordinary results. John Wiley & Sons.

Mayer, J. H., Meinecke, M., Quick, R., Kusterer, F., & Kessler, P. (2022, December). Applying predictive analytics algorithms to support sales volume forecasting. In European, Mediterranean, and Middle Eastern Conference on Information Systems (pp. 63-76). Cham: Springer Nature Switzerland.

Meyer, C., & Schwager, A. (2007). Understanding customer experience. Harvard Business Review, 85(2), 116.

Milo, T., & Somech, A. (2020, June). Automating exploratory data analysis via machine learning: An overview. In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (pp. 2617-2622).

Nacar, E. N., & Erdebilli, B. (2021). Makine öğrenmesi algoritmaları ile satış tahmini. Endüstri Mühendisliği, 32(2), 307-320.

Niu, Y. (2020, October). Walmart sales forecasting using xgboost algorithm and feature engineering. In 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE) (pp. 458-461). IEEE.

Özdemir, Ş., & Örslü, S. (2019). Makine öğrenmesinde yeni bir bakış açısı: Otomatik makine öğrenmesi (AutoML). Journal of Information Systems and Management Research, 1(1), 23-30.

Patil, S., Nankar, O., Agrawal, R., Sharma, K., Awasthi, S., & Jha, N. (2023, January). Black Friday sales prediction using supervised machine learning. In 2023 International Conference on Artificial Intelligence and Smart Communication (AISC) (pp. 1006-1012). IEEE.

Ramachandra, H. V., Balaraju, G., Rajashekar, A., & Patil, H. (2021, March). Machine learning application for black friday sales prediction framework. In 2021 International Conference on Emerging Smart Computing and Informatics (ESCI) (pp. 57-61). IEEE.

Ranjan, G. S. K., Verma, A. K., & Radhika, S. (2019, March). K-nearest neighbors and grid search cv based real time fault monitoring system for industries. In 2019 IEEE 5th international conference for convergence in technology (I2CT) (pp. 1-5). IEEE.

Sathya, R., & Abraham, A. (2013). Comparison of supervised and unsupervised learning algorithms for pattern classification. International Journal of Advanced Research in Artificial Intelligence, 2(2), 34-38.

Selvi, G., Dag, G., Dirican, E. G., Aktay, T., Aksu, S. M., Özdem, K., ... Akcayol, M. A. (2021). Automated machine learning platform otomatik makine öğrenmesi platformu. 6th International Conference on Computer Science and Engineering, UBMK 2021 (ss.769-774). Ankara, Türkiye.

Sen, P. C., Hajra, M., & Ghosh, M. (2020). Supervised classification algorithms in machine learning: A survey and review. In Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018 (pp. 99-111). Springer Singapore.

Shouval, R., Fein, J. A., Savani, B., Mohty, M., & Nagler, A. (2021). Machine learning and artificial intelligence in haematology. British Journal of Haematology, 192(2), 239-250.

Swilley, E., & Goldsmith, R. E. (2013). Black Friday and Cyber Monday: Understanding consumer intentions on two major shopping days. Journal of Retailing and Consumer Services, 20(1), 43-50.

Talkhi, N., Nooghabi, M. J., Esmaily, H., Maleki, S., Hajipoor, M., Ferns, G. A., & Ghayour-Mobarhan, M. (2023). Prediction of serum anti-HSP27 antibody titers changes using a light gradient boosting machine (LightGBM) technique. Scientific Reports, 13(1), 12775.

Thomas, T., P. Vijayaraghavan, A., Emmanuel, S., Thomas, T., P. Vijayaraghavan, A., & Emmanuel, S. (2020). Applications of decision trees. Machine Learning Approaches in Cyber Security Analytics, 157-184.

Timoshenko, A., & Hauser, J. R. (2019). Identifying customer needs from user-generated content. Marketing Science, 38(1), 1-20.

Trung, N. D., Thien, T. D., Luu, T. D., & Huynh, H. X. (2021, July). Black Friday sale prediction via extreme gradient boosted trees. In Proceedings of the 12th National Conference on Basic and Applied Research in Information Technology (FAIR) (pp. 49-57). Acesso em.

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. Procedia-Social and Behavioral Sciences, 106, 234-240.

Wang, R., Wang, L., Zhang, J., He, M., & Xu, J. (2022). XGBoost machine learning algorism performed better than regression models in predicting mortality of moderate-to-severe traumatic brain injury. World Neurosurgery, 163, e617-e622.

Wang, Z., & Bovik, A. C. (2009). Mean squared error: Love it or leave it? A new look at signal fidelity measures. IEEE Signal Processing Magazine, 26(1), 98-117.

Wu, C. S. M., Patil, P., & Gunaseelan, S. (2018, November). Comparison of different machine learning algorithms for multiple regression on black friday sales data. In 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS) (pp. 16-20). IEEE.

Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. Machine Learning, 109(2), 373-440.

Wu, C. S. M., Patil, P., & Gunaseelan, S. (2018, November). Comparison of different machine learning algorithms for multiple regression on black friday sales data. In 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS) (pp. 16-20). IEEE.

Xia, Z., Xue, S., Wu, L., Sun, J., Chen, Y., & Zhang, R. (2020). ForeXGBoost: Passenger car sales prediction based on XGBoost. Distributed and Parallel Databases, 38, 713-738.

Yalçın, F. G. (2022). Craftgate, Kasım Ayı İndirimlerine İlişkin Online Alışveriş Verilerini Açıkladı.

Zeng, M., Cao, H., Chen, M., & Li, Y. (2019). User behaviour modeling, recommendations, and purchase prediction during shopping festivals. Electronic Markets, 29, 263-274.

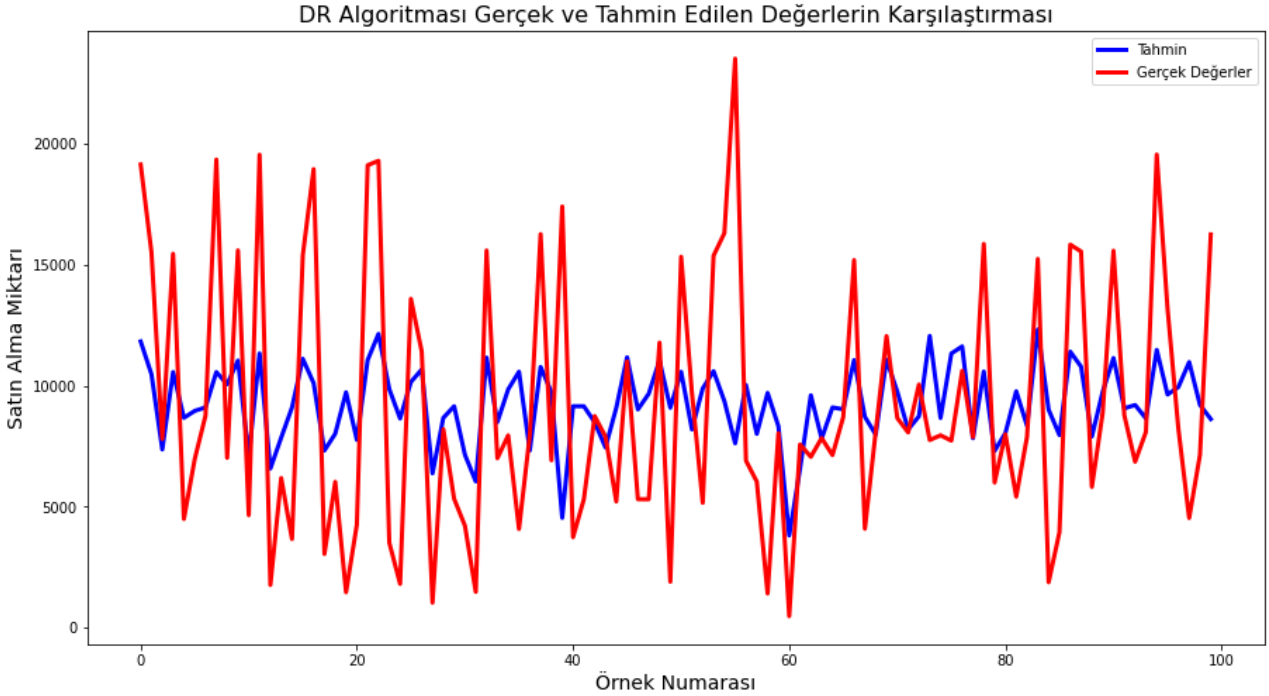
Zhang, D., & Gong, Y. (2020). The comparison of LightGBM and XGBoost coupling factor analysis and prediagnosis of acute liver failure. IEEE Access, 8, 220990-221003.

Zhu, X., Chu, J., Wang, K., Wu, S., Yan, W., & Chiam, K. (2021). Prediction of rockhead using a hybrid N-XGBoost machine learning framework. Journal of Rock Mechanics and Geotechnical Engineering, 13(6), 1231-1245.

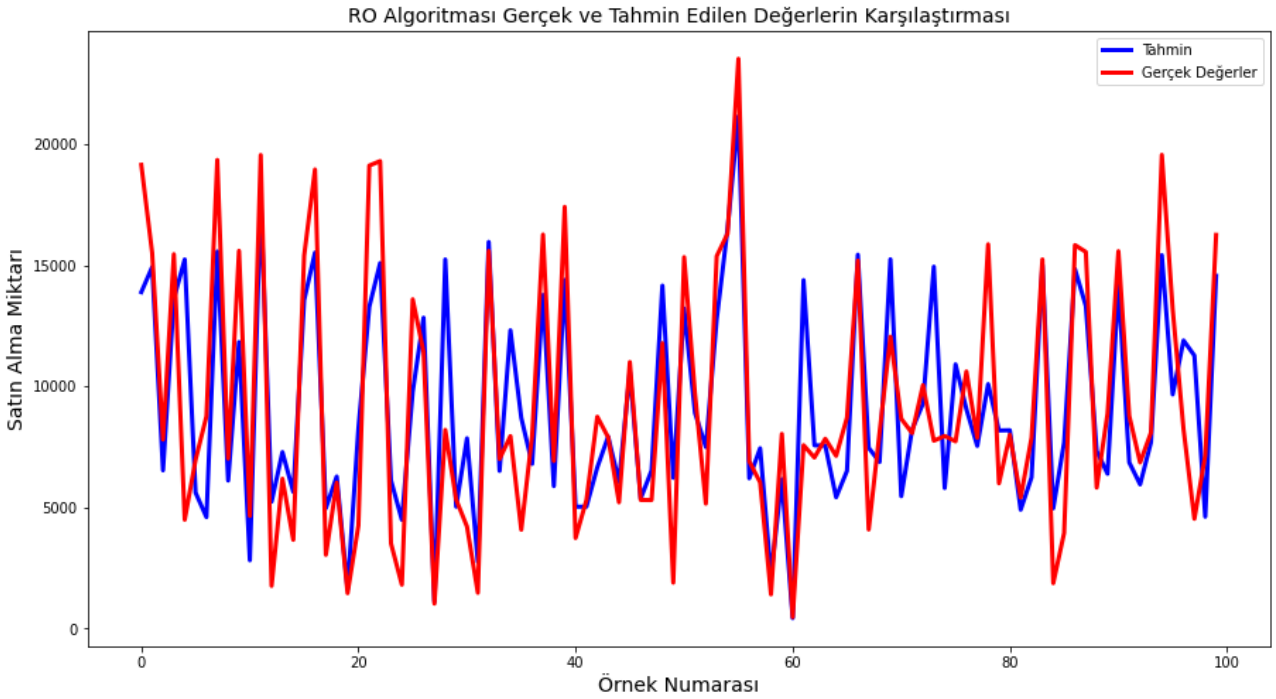
## EKLER

### EK 1- Algoritmalar ve Hiperparametreler Bazında Gerçek Satış Değerleri ile Tahmin Edilen Değerlerin Karşılaştırılması

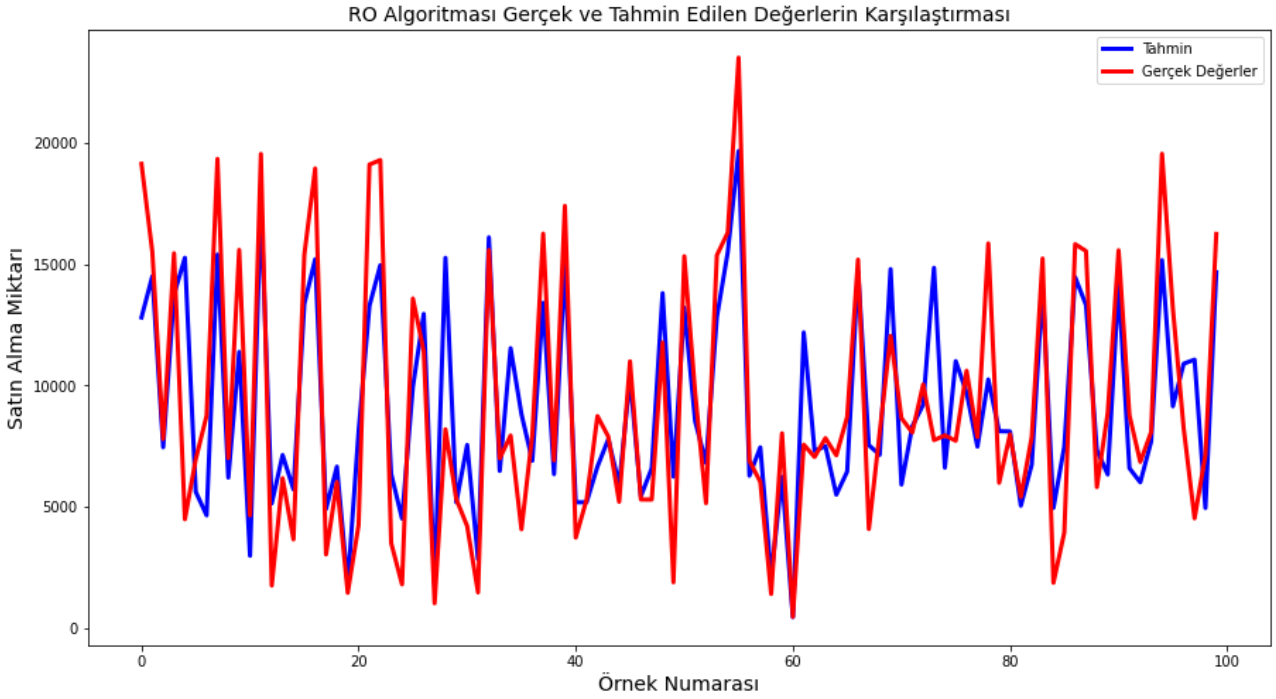
#### a) DR Algoritması Varsayılan Hiperparametre



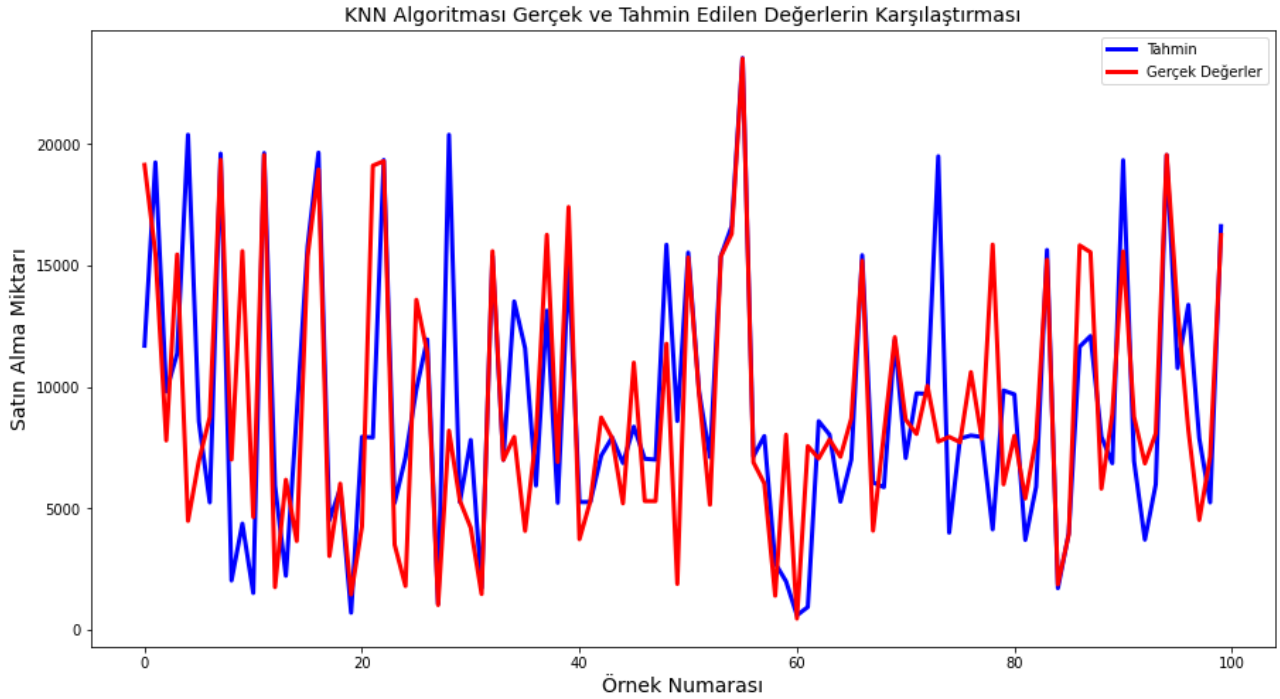

#### b) RO Algoritması Varsayılan Hiperparametre



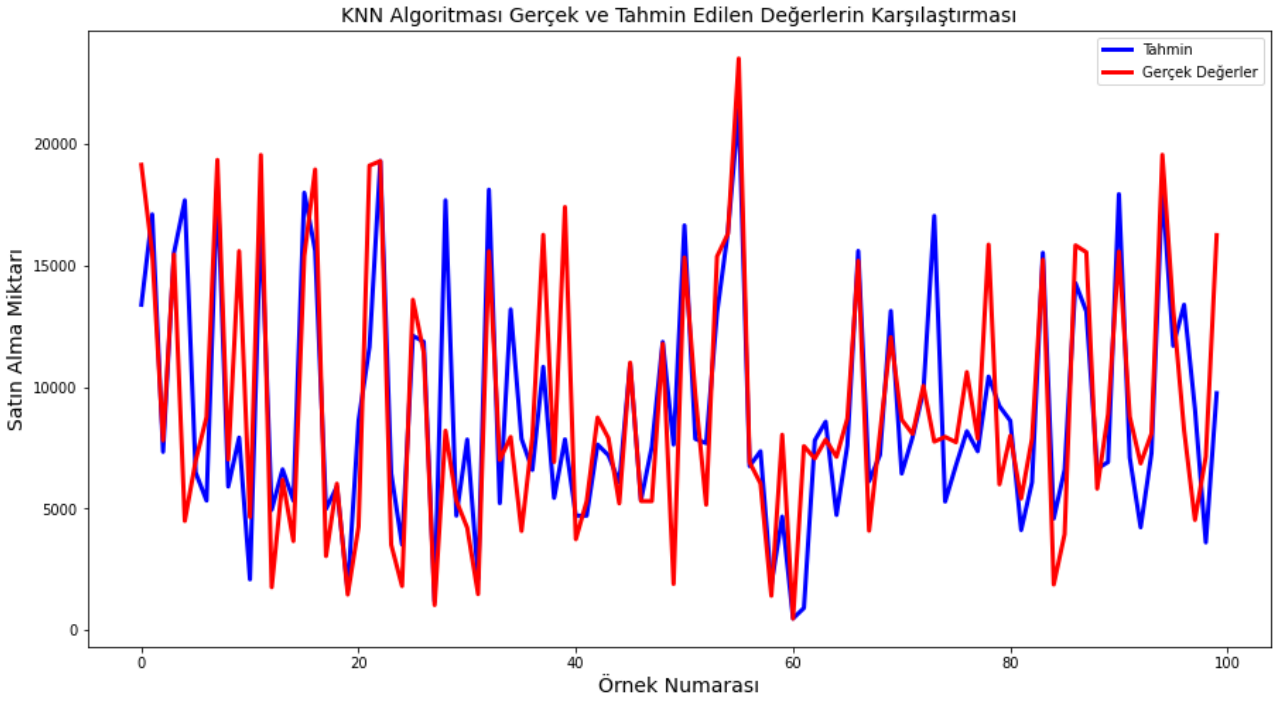
### c) RO Algoritması GridSearchCV Hiperparametre



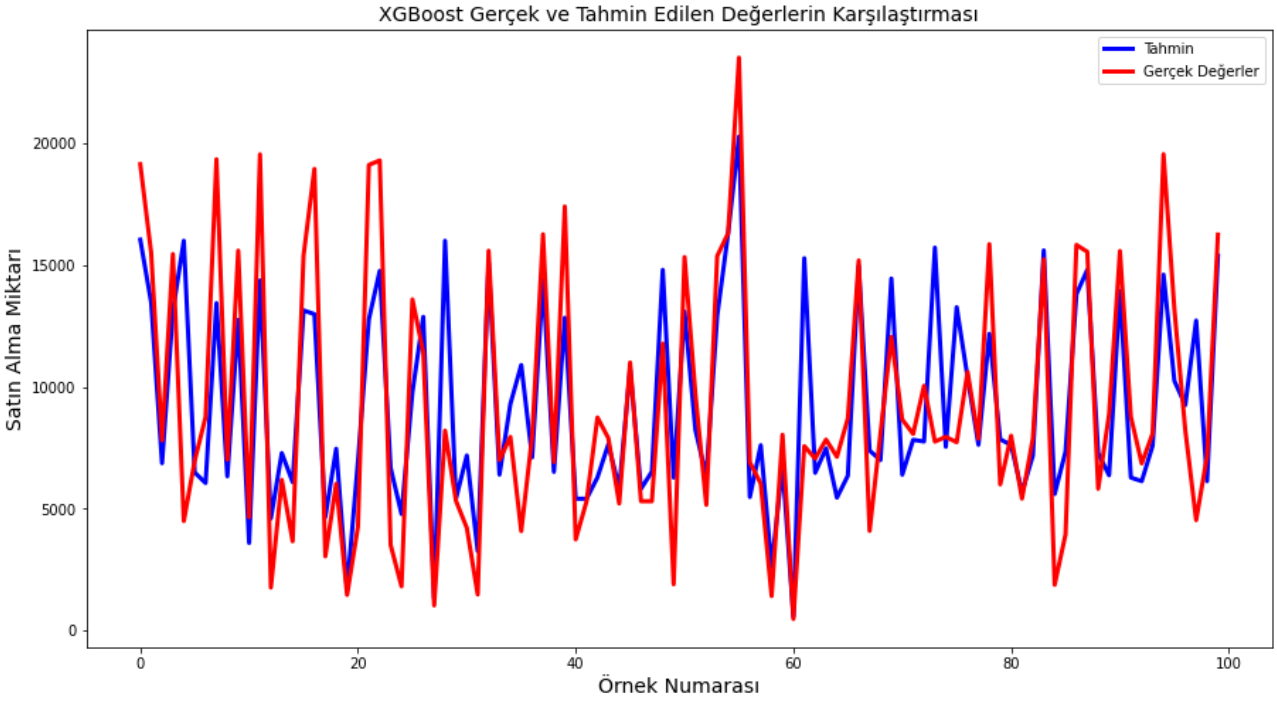

### d) KNN Algoritması Varsayılan Hiperparametre



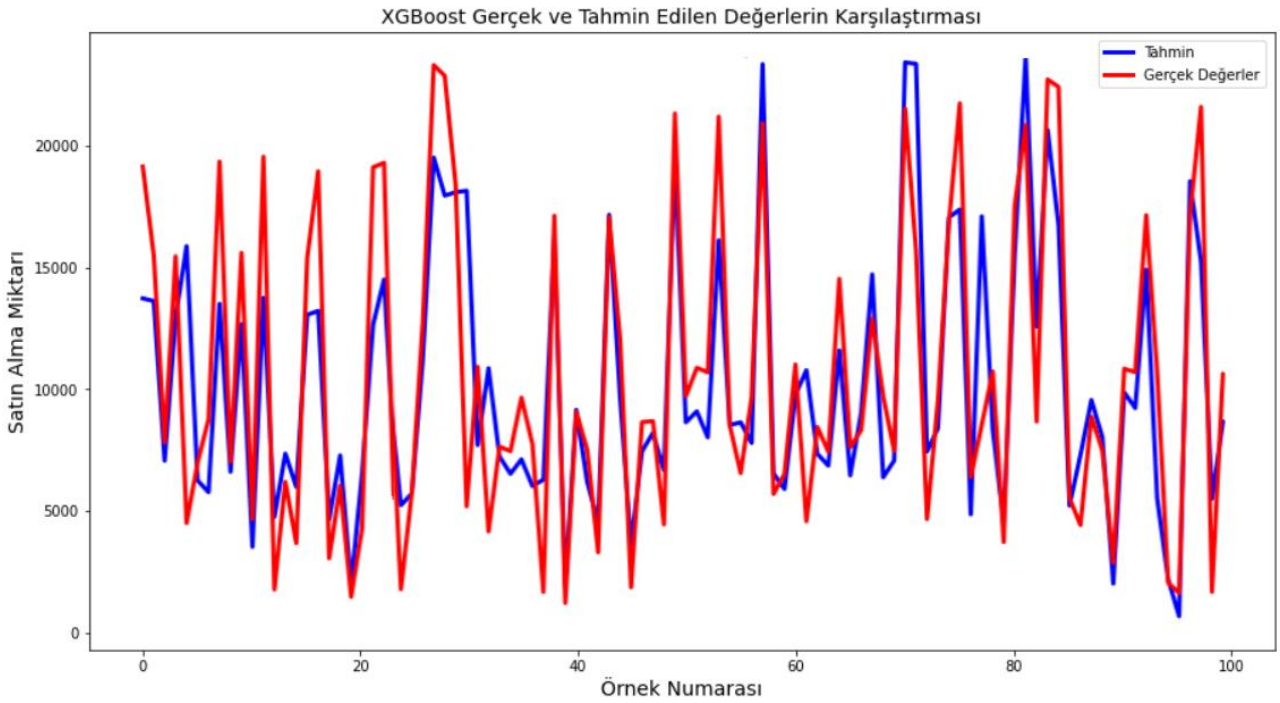
### e) KNN Algoritması GridSearchCV Hiperparametre



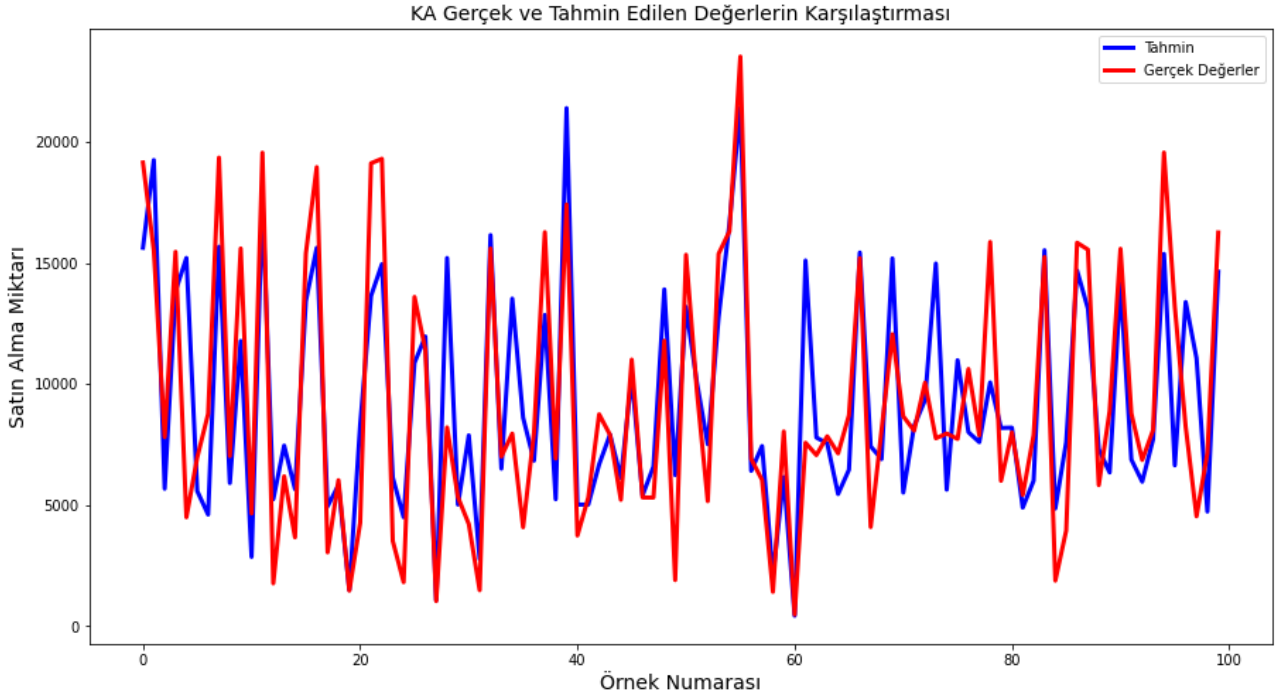

### f) XGB Algoritması Varsayılan Hiperparametre



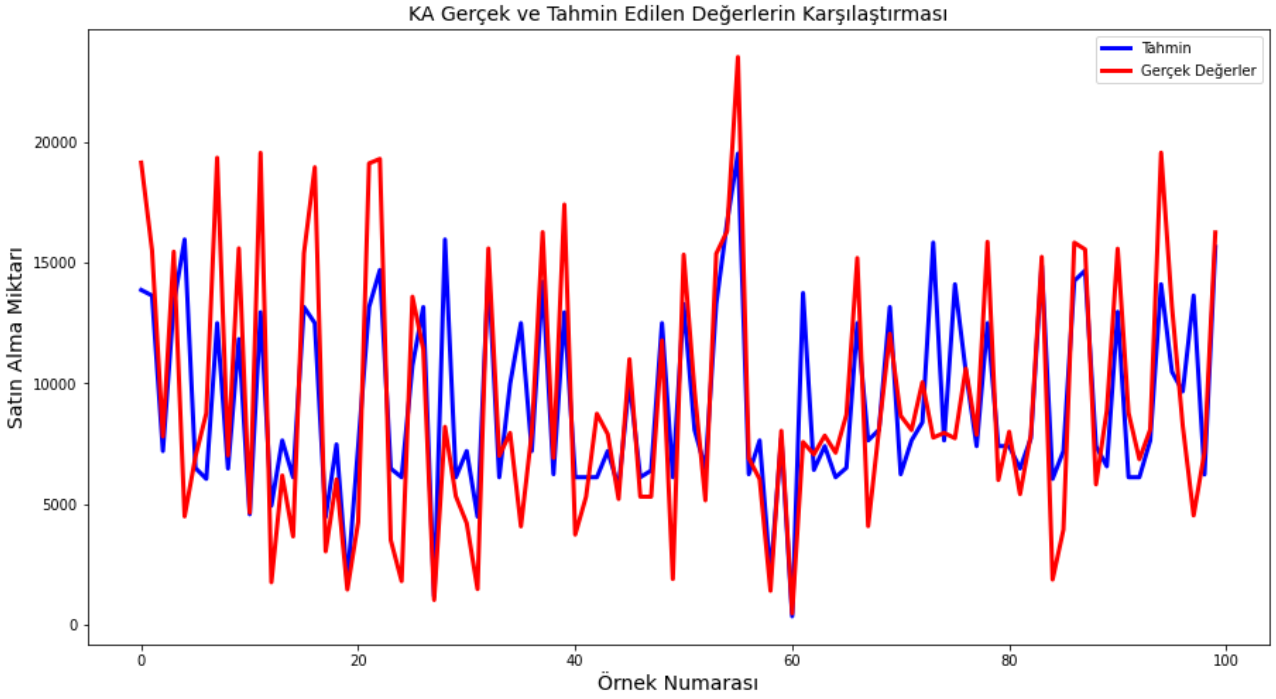
### g) XGB Algoritması GridSearchCV Hiperparametre



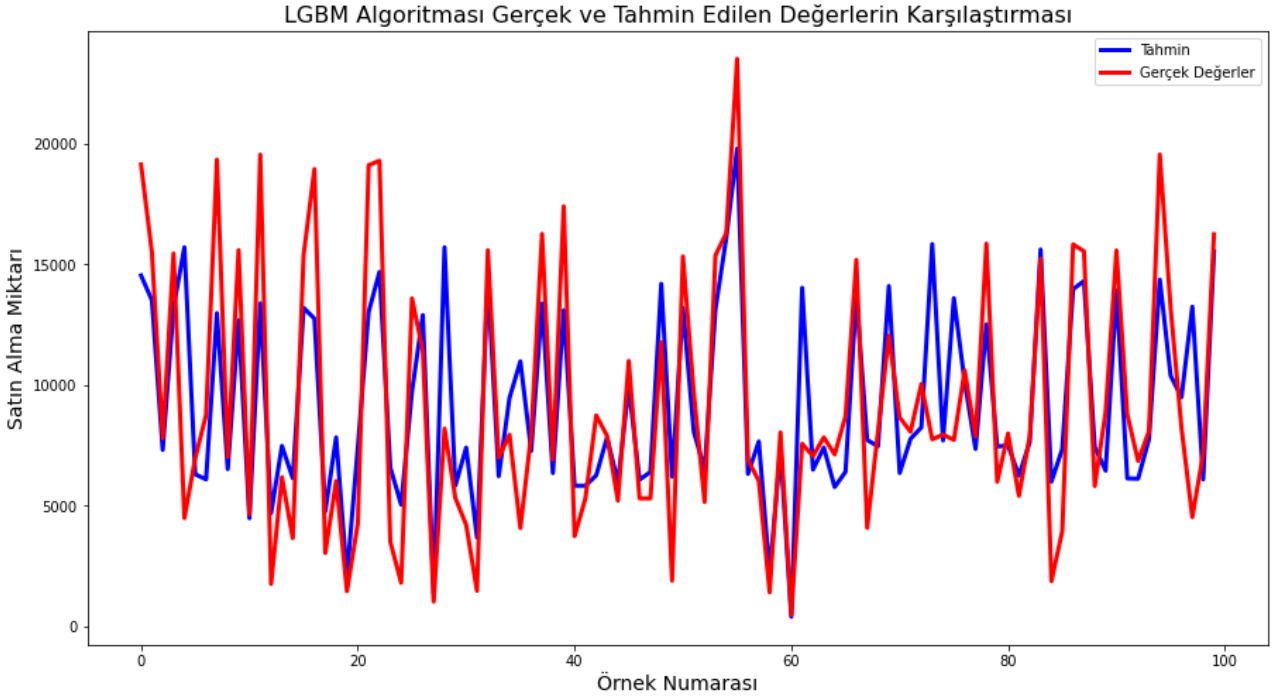

### h) KA Algoritması Varsayılan Hiperparametre



### i) KA Algoritması GridSearchCV Hiperparametre




### j) LGBM Algoritması Varsayılan Hiperparametre



**k) LGBM Algoritması GridSearchCV Hiperparametre**