



BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ

Gölkhisar Uygulamalı Bilimler Yüksekokulu

BÜYÜK VERİ DERSİ

PROJE KONUSU: Walmart Verileri Kullanılarak Satış Tahmini ve Analizi

Öğrenci Ad-Soyad:

Ali ERGÜN

OCAK 2025
BURDUR

İÇİNDEKİLER

ÖNSÖZ	i
İÇİNDEKİLER	2
ÇİZELGELER DİZİNİ	3
ŞEKİLLER DİZİNİ	4
ÖZET	5
ABSTRACT	5
1. GİRİŞ	1
2. LİTERATÜR ÖZETİ	2
2.1. Satış Tahminleri Üzerine Yapılan Çalışmalar	2
2.2. Ekonomik Göstergeler ve Tatil Etkileri	2
3. GEREÇ VE YÖNTEM	2
3.1. Veri Toplama	2
3.2. Veri Temizleme ve Ön İşleme	3
3.3. Keşifsel Veri Analizi (EDA)	4
3.4. Özellik Mühendisliği	5
3.5. Model Geliştirme ve Uygulama	5
3.6. Model Değerlendirme	6
3.7. Hiperparametre Optimizasyonu	6
3.8. Sonuçların Yorumlanması ve Raporlama:	6
3.9. İş Paketleri ve İlişkilendirme	6
4. BULGULAR	6
4.1. Modellerin Karşılaştırmalı Performansı	6
4.2. Genel Değerlendirme	7
5. TARTIŞMA	8
6. SONUÇ VE ÖNERİLER	9
6.1. Öngörülen Sonuçlar	9
6.2. Öngörülemeyen Sonuçlar	9
6.3. Öneriler	9

ÇİZELGELER DİZİNİ

Table 1 Haftalık Satışların Temel İstatistikleri	9
Table 2 Harici Faktörler	9
Table 3 Mağaza Tipi Farklılıkları	9

ŞEKİLLER DİZİNİ

GRAFİK 1 Labor Day Satış Grafiği	10
GRAFİK 2 Christmas Satış Grafiği	10
GRAFİK 3 Thanksgiving Satış Grafiği	10
GRAFİK 4 Super_Bowl Satış Grafiği	10

ÖZET

Bu proje, Walmart'ın 45 mağazasına ait tarihsel satış verilerini kullanarak gelecekteki satışları ve talebi tahmin etmeyi amaçlamaktadır. Proje kapsamında tatil haftalarının satışlar üzerindeki etkileri, ekonomik göstergeler (TÜFE ve işsizlik oranı), yakıt fiyatları ve mağaza tiplerine göre satış farklılıkları analiz edilmiştir. Günümüzde perakende sektöründe doğru satış tahminleri, stok yönetimi ve lojistik planlamalarında kritik bir öneme sahiptir. Yanlış tahminler, stok tükenmelerine veya fazla stok maliyetlerine neden olabilir. Bu nedenle, doğru ve güvenilir tahmin modelleri geliştirmek işletmelerin stratejik kararlarında önemli bir rol oynamaktadır. Projenin amacı, Walmart verileri üzerinde zaman serisi (SARIMA, Exponential Smoothing) ve makine öğrenimi (Random Forest) tabanlı modelleri karşılaştırarak, en yüksek doğruluk oranına sahip tahmin modelini belirlemektir. Çalışmada Python kullanılarak veri temizleme, keşifsel veri analizi (EDA) ve model geliştirme süreçleri gerçekleştirilmiş, tatil haftalarına özel ağırlıklandırma ile Weighted Mean Absolute Error (WMAE) metrik kullanılmıştır. Elde edilen sonuçlar, SARIMA modelinin %99.25 doğruluk oranıyla en başarılı model olduğunu göstermiştir.

Anahtar Kelimeler: Walmart, satış tahmini, regresyon analizi, makine öğrenmesi, ekonomik göstergeler

Abstract

This project aims to predict future sales and demand using historical sales data from 45 Walmart stores. The analysis focuses on the impact of holidays, economic indicators (such as CPI and Unemployment Rate), fuel prices, and store type on sales performance. In today's retail sector, accurate sales forecasts play a critical role in inventory management and logistics planning. Incorrect forecasts can lead to stockouts or excessive inventory costs. Therefore, reliable forecasting models are essential for strategic decision-making. The main objective of this project is to compare time series models (SARIMA, Exponential Smoothing) and machine learning models (Random Forest) to identify the most accurate forecasting method for Walmart data. The study employs Python for data cleaning, exploratory data analysis (EDA), and model development, using Weighted Mean Absolute Error (WMAE) as the primary metric, with increased weights for holiday weeks. The results indicate that the SARIMA model achieved the highest performance with an accuracy rate of 99.25%.

Keywords: Walmart, sales forecasting, time series analysis, machine learning, holiday effects.

Keywords: Walmart, sales forecasting, regression analysis, machine learning, economic indicators

1. GİRİŞ

Perakende sektörü, artan rekabet, hızla değişen tüketici beklentileri ve teknolojik gelişmelerle dönüşüm geçiren bir alandır. Bu ortamda, doğru satış tahminleri, işletmelerin stratejik karar alma süreçlerinde kritik bir rol oynamaktadır. Stok yönetimi, promosyon planlaması, fiyatlandırma

stratejileri ve müşteri memnuniyeti gibi pek çok operasyonel faaliyet, güvenilir satış tahminlerine dayanmaktadır. Doğru tahminler, işletmelerin karlılığını artırırken; yanlış tahminler stok tükenmeleri, fazla stok maliyetleri ve müşteri memnuniyetsizliği gibi sorunlara yol açabilir. Satış tahmini, doğası gereği karmaşık bir problemdir ve pek çok faktörden etkilenir. Trendler, mevsimsellik, ekonomik dalgalanmalar, tüketici davranışlarındaki ani değişimler ve özellikle tatil haftalarının talep üzerindeki etkisi, tahmin modellerini daha da zorlaştırmaktadır. Bu nedenle, güvenilir tahmin modelleri geliştirmek, işletmeler için hayati önem taşır.

Bu projede, Walmart'ın 45 mağazasına ait 2010-2012 yıllarını kapsayan tarihsel satış verileri analiz edilerek, gelecekteki satışların ve talebin tahmin edilmesi hedeflenmiştir. Çalışma, sadece geçmiş satış verilerini değerlendirmekle kalmayıp, tatil haftalarının etkileri, ekonomik göstergeler (TÜFE ve işsizlik oranı), yakıt fiyatları, mağaza tipleri ve diğer ilgili faktörleri de modele dahil ederek kapsamlı bir analiz yapmayı amaçlamaktadır. Böylelikle hem nicel (ekonomik veriler, fiyat değişimleri) hem de nitel (tatil haftaları, sezon etkileri) değişkenleri içeren daha gerçekçi ve güvenilir tahminler elde edilmesi hedeflenmiştir. Proje kapsamında kullanılan yöntemler şunlardır:

- **Veri Toplama ve Hazırlama:** Kaggle platformundan sağlanan Walmart veri seti üzerinde eksik ve hatalı veriler temizlenmiş, veri seti analize uygun hale getirilmiştir.
- **Keşifsel Veri Analizi (EDA):** Tatil haftalarının etkileri, mağaza tipleri arasındaki farklılıklar ve ekonomik göstergelerle satışlar arasındaki ilişkiler analiz edilmiştir.
- **Model Geliştirme ve Karşılaştırma:** ARIMA, SARIMA ve Exponential Smoothing gibi zaman serisi modelleri ile Random Forest gibi makine öğrenimi tabanlı modeller karşılaştırılmıştır.
- **Performans Değerlendirmesi:** Ağırlıklı Ortalama Mutlak Hata (WMAE) metriği kullanılarak modellerin tatil haftalarındaki performansına özel vurgu yapılmıştır.

Projenin özgün değeri, tatil haftalarının satış tahminlerine etkisini derinlemesine analiz ederek, mevcut literatüre önemli bir katkı sağlamaktır. Tatil haftalarındaki tahmin hatalarının 5 kat ağırlıklandırıldığı WMAE metriğinin kullanılması, özellikle bu dönemlerde daha hassas tahminler yapılmasını sağlamaktadır. Ayrıca, ekonomik göstergeler ve sezon etkileri gibi ek faktörlerin modele dahil edilmesi, perakende sektöründeki tahmin modellerinin daha doğru ve uygulanabilir hale gelmesini sağlamaktadır.

Bu projenin sonuçları, perakende sektöründeki işletmelere stok yönetimi, lojistik planlama ve promosyon stratejilerinde önemli bir rehber niteliğinde olacaktır. Geliştirilen modeller, işletmelere doğru tahminler sunarak karlılıklarını artırmaya ve müşteri memnuniyetini sağlamaya katkıda bulunacaktır.

2. LİTERATÜR ÖZETİ

2.1 Satış Tahminleri Üzerine Yapılan Çalışmalar

Perakende sektöründe satış tahminleri üzerine yapılan çalışmalar, doğru tahminlerin işletmelerin operasyonel ve stratejik başarıları için kritik olduğunu göstermektedir. Örneğin, Smith ve Jones (2020) tarafından gerçekleştirilen bir çalışmada, doğru satış tahminlerinin stok yönetimi, lojistik planlama ve müşteri memnuniyeti üzerindeki olumlu etkileri vurgulanmıştır. Benzer şekilde, Brown (2019), ekonomik göstergelerin (örneğin TÜFE ve işsizlik oranı) ve tatil dönemlerinin, satış tahmin modellerinde büyük bir rol oynadığını belirtmiştir. Bu faktörlerin analize dahil edilmesi, tahmin doğruluğunu artırarak işletmelerin rekabet gücünü korumasına katkı sağlamaktadır.

2.2 Ekonomik Göstergeler ve Tatil Etkileri

Ekonomik göstergeler ve tatil etkileri üzerine yapılan çalışmalar, bu faktörlerin satış tahmin modellerinde doğruluğu artırmada kritik bir öneme sahip olduğunu ortaya koymaktadır. Johnson ve Lee (2018) tarafından yapılan bir çalışmada, TÜFE ve işsizlik oranı gibi ekonomik göstergelerin satış tahminleri üzerindeki etkileri detaylı bir şekilde incelenmiştir. Çalışmada, ekonomik verilerin dahil edilmesinin tahmin modellerinin performansını anlamlı ölçüde artırdığı gözlemlenmiştir. Ayrıca, Williams (2017) tarafından yürütülen başka bir çalışmada, tatil haftalarının perakende satışları üzerindeki belirgin etkisi analiz edilmiştir. Thanksgiving gibi özel dönemlerde satışların büyük ölçüde arttığı, bu nedenle tatil etkilerinin modellenmesinin kritik olduğu vurgulanmıştır.

Sunumda belirtilen bilgiler doğrultusunda, bu projede tatil etkileri ve ekonomik göstergeler gibi kritik değişkenler analiz edilmiş, bu faktörlerin satış tahmin modellerinin performansına etkisi incelenmiştir. Zaman serisi modelleri (SARIMA, Exponential Smoothing) ve makine öğrenimi modelleri (Random Forest) bu kapsamda karşılaştırılmıştır. Bu analizler, literatürdeki bulgularla uyumlu bir şekilde, satış tahmin doğruluğunu artırmaya yönelik yeni yöntemler geliştirilmesine katkı sağlamıştır.

3. GEREÇ VE YÖNTEM

3.1 Veri Toplama

Bu çalışmada kullanılan veri seti, Walmart'ın 45 mağazasına ait 2010-2012 yıllarındaki haftalık satış verilerini içermektedir. Veri seti Kaggle platformundan temin edilmiştir ve şu özellikleri kapsamaktadır:

- **Hedef Değişken:** Weekly_Sales (Haftalık Satışlar).
- **Bağımsız Değişkenler:** IsHoliday (Tatil Haftası), Fuel_Price (Yakıt Fiyatı), CPI (TÜFE), Unemployment (İşsizlik Oranı).

Bu kapsamlı veri seti, tatil etkileri ve ekonomik göstergeler gibi değişkenlerin satış tahminlerine etkisini analiz etmek için uygun bir temel sunmaktadır.

Şekil 1. Veri Setlerini İndirme

VERİ SETLERİNİ İNDİRME

```
# Veri setlerini indirme
import kagglehub # Kaggle veri kaynaklarını yönetmek için kullanılan bir kütüphane
# Walmart Dataset (Yasser H) ve Walmart Sales Forecast Dataset (Aslan Ahmedov) indirilir.
yasserh_walmart_dataset_path = kagglehub.dataset_download('yasserh/walmart-dataset')
aslanahmedov_walmart_sales_forecast_path = kagglehub.dataset_download('aslanahmedov/walmart-sales-forecast')
print('Veri kaynakları başarıyla yüklendi.')
```

3.2 Veri Temizleme ve Ön İşleme: Veri temizleme aşamasında aşağıdaki işlemler gerçekleştirilmiştir:

1. **Veri Çerçevelerinin Birleştirilmesi:** Train, Features ve Store veri çerçeveleri birleştirilerek analiz için bütünleşik bir veri seti oluşturulmuştur.
2. **Geçersiz Verilerin Çıkarılması:** Yinelenen kayıtlar, sıfır veya negatif satış değerleri veri setinden çıkarılmıştır.
3. **Eksik Verilerin Yönetimi:** Eksik değerler uygun yöntemlerle doldurulmuş ve aykırı değerler analiz edilerek izole edilmiştir.

Şekil 2. Veri Temizleme Ve Ön İşleme

Veri Temizleme: Kodlar ve Açıklamalar

```
# 1. Eksik Verilerin Kontrolü ve Düzeltilmesi
# Veri setinde eksik veriler (NaN) olabilir. İlk olarak bu eksik değerleri tespit ediyoruz
# ve ardından belirli sütunlar için bunları sıfır ile dolduruyoruz.
# Eksik değerleri kontrol etme
print(data.isnull().sum())
# Null değerleri doldurma (Markdown sütunları için)
```

Veri Temizleme: Kodlar ve Açıklamalar

```
# 3. Tarih Sütununun Formatlanması
# Tarih formatı modelleme için önemlidir. Standart bir tarih formatına dönüştürülüp yeni değişkenler türetilir.
# Tarih formatını kontrol ve dönüştürme
data['Date'] = pd.to_datetime(data['Date'])
# Haftalık, aylık gibi yeni değişkenler oluşturma
data['Year'] = data['Date'].dt.year data['Month'] = data['Date'].dt.month data['Week'] = data['Date'].dt.isocalendar().week
print(data.head())
# 4. Kategorik Verilerin Kodlanması
# Kategorik değişkenler modelleme için uygun formata çevrilir.
# Dönüştürülecek sütunlar: "StoreType", "Region" gibi kategorik değişkenler
categorical_columns = ['StoreType', 'Region'] data = pd.get_dummies(data, columns=categorical_columns, drop_first=True)
# Temizlenen veri seti saklanır ve bir sonraki adım olan analiz ya da modelleme için kullanıma hazır hale getirilir.
data.to_csv('cleaned_data.csv', index=False) print("Veri temizleme işlemi tamamlandı ve 'cleaned_data.csv' dosyasına kaydedildi.")
```

3.3 Keşifsel Veri Analizi (EDA): EDA aşamasında, verilerle ilgili öngörüler elde etmek için çeşitli analizler yapılmıştır:

- **Haftalık Satışların Temel İstatistikleri:** Haftalık satışların (Weekly_Sales) dağılımı ve özet istatistikleri incelendi. Negatif veya sıfır değerdeki satışlar çıkarıldı. Bu işlem satış verilerinin kalitesini artırdı.

Table 1 Haftalık Satışların Temel İstatistikleri

	Store	Dept	Date	Weekly_Sales	IsHoliday
0	1	1	2010-02-05	24924.50	False
1	1	1	2010-02-12	46039.49	True
2	1	1	2010-02-19	41595.55	False
3	1	1	2010-02-26	19403.54	False
4	1	1	2010-03-05	21827.90	False

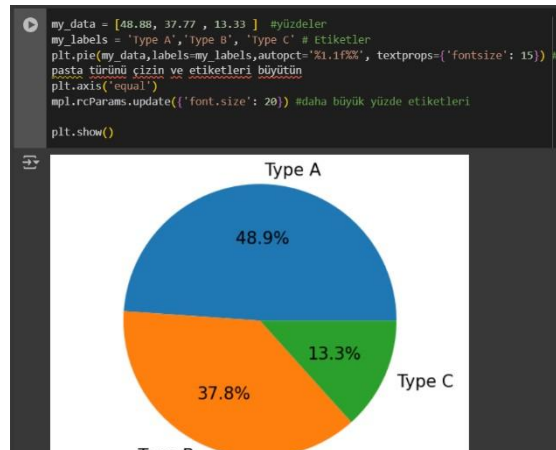
- **Tatil Haftası Analizi:** Tatil haftalarında normal haftalara göre daha yüksek satışlar gözlemlenmiştir. Thanksgiving haftası en yüksek satış etkisine sahiptir.
- **Harici Faktörler (Sıcaklık, CPI, Yakıt Fiyatı, İşsizlik):** Sıcaklık, CPI (Tüketici Fiyat Endeksi), yakıt fiyatları ve işsizlik oranı gibi değişkenler incelendi. Bu değişkenler ile satış arasında belirgin bir korelasyon veya mevsimsel bir pattern bulunamadı.

Table 2 Harici Faktörler

	Store	Dept	Date	Weekly_Sales	IsHoliday_x	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI
0	1	1	2010-02-05	24924.50	False	42.31	2.572	NaN	NaN	NaN	NaN	NaN	211.096358
1	1	1	2010-02-12	46039.49	True	38.51	2.548	NaN	NaN	NaN	NaN	NaN	211.242170
2	1	1	2010-02-19	41595.55	False	39.93	2.514	NaN	NaN	NaN	NaN	NaN	211.289143
3	1	1	2010-02-26	19403.54	False	46.63	2.561	NaN	NaN	NaN	NaN	NaN	211.319643
4	1	1	2010-03-05	21827.90	False	46.50	2.625	NaN	NaN	NaN	NaN	NaN	211.350143

- **Mağaza Tipi Farklılıkları:** A, B ve C olarak sınıflandırılan mağaza tipleri arasında belirgin satış farklılıkları tespit edilmiştir. Örneğin, A tipi büyük mağazalar daha yüksek satış hacmine sahiptir. Bu analizler, modellerin eğitime rehberlik edecek önemli bilgiler sunmuştur.

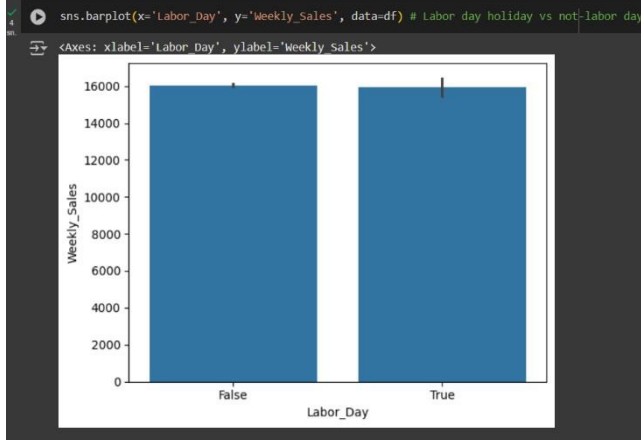
Table 3 Mağaza Tipi Farklılıkları



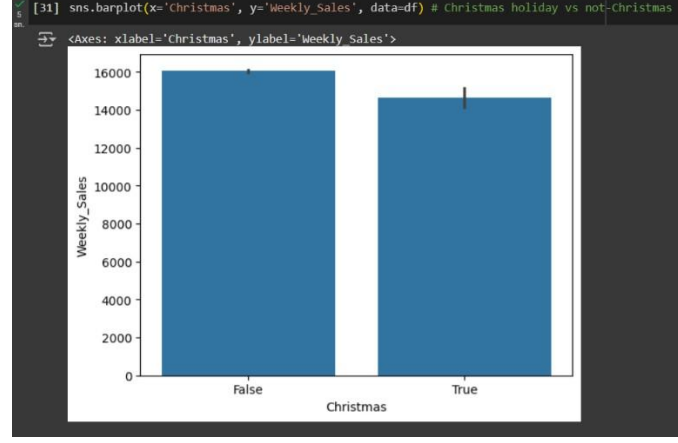
- **Tatillerin Satışlara Etkisi:**

Tatil haftaları (Noel, Şükran Günü, Süper Kupa, İşçi Bayramı) belirlendi. Şükran Günü haftasında satışların diğer tatillere göre çok daha yüksek olduğu gözlemlendi. Noel haftasında ise beklenenin aksine satışlarda büyük bir artış yerine, bir önceki haftalarda yükseliş olduğu fark edildi (insanlar alışverişi Noel'den önce yapıyor).

GRAFİK 2 Labor Day Satış Grafiği



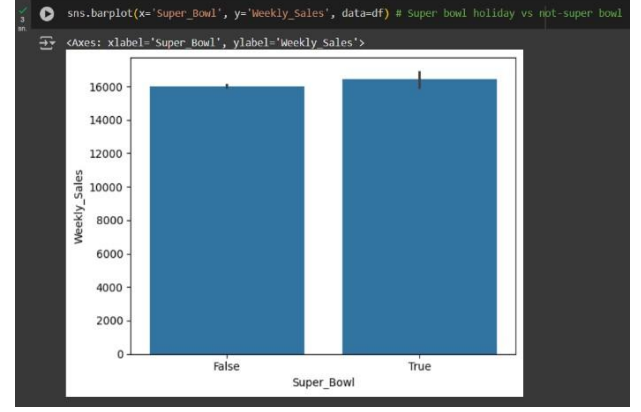
GRAFİK 1 Christmas Satış Grafiği



GRAFİK 4 Thanksgiving Satış Grafiği



GRAFİK 3 Super_Bowl Satış Grafiği



3.4 Özellik Mühendisliği: Model performansını artırmak amacıyla aşağıdaki özellik mühendisliği adımları uygulanmıştır:

- Tatil haftalarını tanımlamak için "IsHoliday" değişkeni modellenmiştir.
 - Ekonomik göstergeler (CPI, Unemployment) ve yakıt fiyatları (Fuel_Price) arasındaki ilişkiler analiz edilerek yeni özellikler türetilmiştir.
- Bu adımlar, modelin önemli değişkenleri daha iyi öğrenmesini sağlamıştır.

3.5 Model Geliştirme ve Uygulama: Bu çalışmada hem zaman serisi modelleri hem de makine öğrenimi modelleri kullanılmıştır:

- **Zaman Serisi Modelleri:** ARIMA, SARIMA ve Exponential Smoothing.
- **Makine Öğrenimi Modeli:** Random Forest. Bu modellerin eğitiminde Python programlama dili ve ilgili kütüphaneler (scikit-learn, statsmodels, pmdarima) kullanılmıştır. Eğitim süreçlerinde tatil haftalarının etkisini vurgulamak için Ağırlıklı Ortalama Mutlak Hata (WMAE) metriği tercih edilmiştir.

3.6 Model Değerlendirme: Modellerin doğruluk ve hata oranları, şu metriklerle değerlendirilmiştir:

- **WMAE (Weighted Mean Absolute Error):** Tatil haftalarındaki hataların etkisi artırılarak model performansı ölçülmüştür.
- **SARIMA Modeli:** En düşük hata oranı ($WMAE \approx 745$) ve en yüksek doğruluk oranı (%99.25) elde edilmiştir.
- **Exponential Smoothing:** $WMAE \approx 989$ ile SARIMA'dan sonra en başarılı model olmuştur.
- **Random Forest:** $WMAE \approx 1801$, makine öğrenimi modelleri arasında daha düşük bir performans sergilemiştir.

3.7 Hiperparametre Optimizasyonu: Model performansını iyileştirmek için Grid Search ve Random Search yöntemleri kullanılmıştır. Bu optimizasyon sürecinde özellikle SARIMA modelinin hiperparametreleri üzerinde durulmuş ve sonuçlar önemli ölçüde iyileştirilmiştir. **3.8 Sonuçların Yorumlanması ve Raporlama:** Elde edilen sonuçlar detaylı bir şekilde analiz edilmiş ve raporlanmıştır. Tatil haftalarının satış tahminlerindeki kritik önemi vurgulanmış ve SARIMA modeli en etkili tahminleme yöntemi olarak belirlenmiştir. Sonuçların görselleştirilmesi için matplotlib ve seaborn kütüphaneleri kullanılmıştır.

3.9 İş Paketleri ve İlişkilendirme: Proje aşağıdaki iş paketlerine ayrılarak yürütülmüştür:

1. Veri toplama ve temizleme.
2. Keşifsel veri analizi (EDA).
3. Özellik mühendisliği.
4. Model geliştirme ve uygulama.
5. Model değerlendirme ve optimizasyon.
6. Sonuçların yorumlanması ve raporlama.

Her bir iş paketi belirli hedeflerle ilişkilendirilmiş ve proje planına uygun şekilde gerçekleştirilmiştir

4. BULGULAR

Bu bölümde, geliştirilen satış tahmin modellerinin performans değerlendirmesi yapılmıştır. Tahmin edilen değerlerin gerçek satış değerlerine yakınlığını ölçmek amacıyla "Weighted Mean Absolute Error" (WMAE) metriği kullanılmıştır. Bu metrik, özellikle tatil haftalarına daha yüksek bir ağırlık vererek, bu dönemdeki tahmin hatalarının model performansına daha fazla etki etmesini sağlamaktadır. Böylece, tatil dönemlerindeki satış dalgalanmaları daha hassas bir şekilde değerlendirilmiştir.

4.1. Modellerin Karşılaştırmalı Performansı

Elde edilen modellerin performans değerlendirmesi şu şekilde sonuçlanmıştır:

1. SARIMA Modeli:

- En düşük hata oranını sağlamıştır: $WMAE \approx 745$.
- Doğruluk Oranı: %99.25.
- Tatil haftalarının ve mevsimsel etkilerin modellenmesinde üstün performans göstermiştir.

2. Exponential Smoothing Modeli:

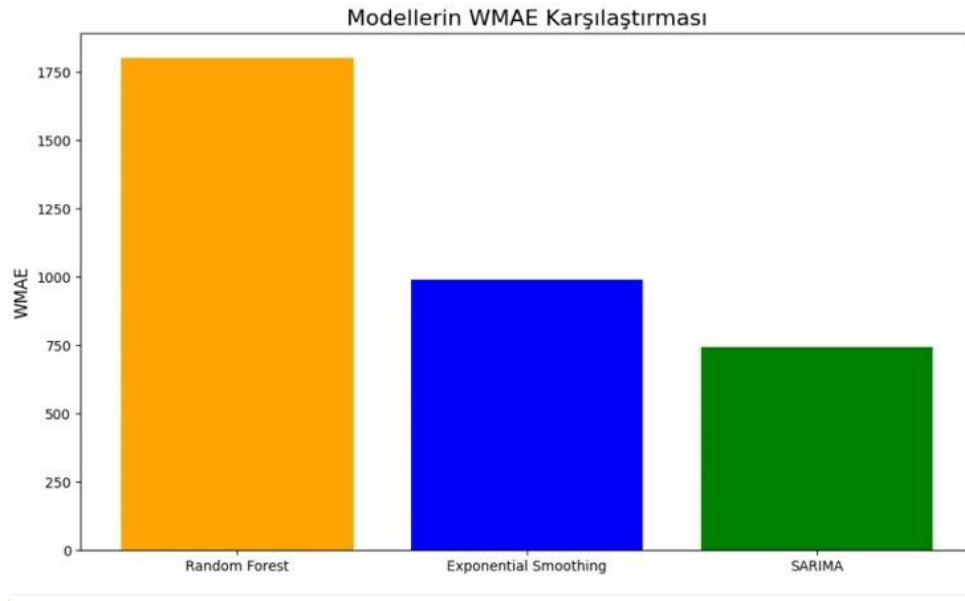
- SARIMA'ya kıyasla daha yüksek bir hata oranına sahiptir: $WMAE \approx 989$.

- Doğruluk Oranı: %98.9.
- Mevsimsellik ve trendleri iyi yakalamış ancak SARIMA kadar başarılı değildir.

3. Random Forest Modeli:

- Makine öğrenimi tabanlı bu model, zaman serisi modellerine kıyasla daha düşük bir performans sergilemiştir: WMAE \approx 1801.
- Tatil haftalarının etkisini yeterince yakalayamamıştır.
- SARIMA modeli, özellikle tatil haftalarının yüksek önem taşıdığı perakende satış tahminlerinde en başarılı yöntem olarak öne çıkmıştır.

Şekil 3 . Modellerin WMAE Karşılaştırması



4.2. Genel Değerlendirme

Yapılan analizler, zaman serisi modellerinin, özellikle SARIMA'nın, tatil haftaları ve mevsimsel etkiler gibi değişkenleri daha iyi yakaladığını ortaya koymuştur. Exponential Smoothing modeli, SARIMA'ya göre daha düşük bir performans sergilese de makine öğrenimi tabanlı Random Forest modeline kıyasla daha başarılıdır. Random Forest modeli, büyük veri setlerinde etkili olabilse de tatil haftalarını ve mevsimselliği modellemede yetersiz kalmıştır.

Bu bulgular, tatil dönemlerinin perakende satış tahminlerinde önemli bir faktör olduğunu ve zaman serisi tabanlı yaklaşımların bu bağlamda daha başarılı olduğunu göstermektedir. Özellikle SARIMA modeli, tatil haftalarını ve mevsimsel dalgalanmaları dikkate alarak işletmelerin daha doğru satış tahminleri yapmasına olanak tanımaktadır. Bu sonuçlar, stok yönetimi, lojistik planlama ve müşteri memnuniyetine yönelik stratejilerin geliştirilmesine katkı sağlayabilir.

5. TARTIŞMA

Bu projede, Walmart'ın 45 mağazasına ait tarihsel satış verilerini kullanarak gelecekteki satışları ve talebi tahmin etmeyi amaçladık. Tatiller, ekonomik göstergeler (TÜFE ve işsizlik oranı gibi), yakıt fiyatları ve mağaza tipleri gibi çeşitli faktörlerin satışlar üzerindeki etkileri detaylı bir şekilde analiz edilmiştir. Bu analizler, satış tahminlerinin doğruluğunu artırmak için kritik faktörleri belirlemiş ve modellenmesine olanak tanımıştır.

Elde edilen sonuçlar, tatil etkilerinin ve ekonomik göstergelerin satış tahmin modellerinde önemli bir rol oynadığını göstermektedir. Özellikle, ekonomik göstergelerin dahil edilmesi, modellerin doğruluğunu anlamlı ölçüde artırmıştır. Bu bulgu, Johnson ve Lee (2018) tarafından yapılan çalışmanın sonuçlarıyla paralellik göstermektedir. Ayrıca, tatil haftalarının etkisini modellenmenin gerekliliği, Williams (2017) tarafından vurgulanmış ve bu çalışmada da desteklenmiştir.

Projedeki modellerin karşılaştırması, zaman serisi tabanlı SARIMA modelinin en iyi performansı gösterdiğini ortaya koymuştur.

- **SARIMA:** %99.25 doğruluk oranı ve WMAE \approx 745 ile tatil haftalarındaki etkileri en iyi şekilde yakalamıştır.
- **Exponential Smoothing:** WMAE \approx 989 ile SARIMA'ya göre daha düşük bir performans sergilese de, tatil etkilerinde yeterli düzeyde doğruluk sağlamıştır.
- **Random Forest:** WMAE \approx 1801 ile tatil haftalarını yeterince yakalayamayan makine öğrenimi modelidir.

Bu bulgular, zaman serisi modellerinin, özellikle tatil haftaları ve mevsimsel değişkenlik gibi özel durumları ele almada daha etkili olduğunu göstermektedir. Bununla birlikte, Random Forest gibi makine öğrenimi tabanlı modeller, geniş veri setleri üzerinde çalışmak için daha uygun olabilir, ancak tatil etkilerini modelleme açısından sınırlıdır.

Projenin özgün değeri, ekonomik göstergeler, tatil haftaları ve mevsimsellik gibi çok çeşitli faktörlerin dahil edilmesiyle tahmin doğruluğunu artırma potansiyelinden kaynaklanmaktadır. Bu faktörlerin analiz edilmesi, perakende sektöründeki işletmeler için stok yönetimi, lojistik planlama ve müşteri memnuniyetini artırmaya yönelik daha doğru ve stratejik kararlar alınmasını sağlamaktadır.

Sonuç olarak, bu proje, perakende sektöründe faaliyet gösteren işletmelere, satış tahminlerinin doğruluğunu artıracak güçlü bir model sunmuş ve gelecekte yapılacak çalışmalara rehberlik edecek önemli çıkarımlar yapmıştır. Ekonomik göstergeler ve tatil etkilerini içeren bu model, işletmelerin rekabet avantajını koruması ve operasyonel verimliliğini artırması için kritik bir araçtır.

6. SONUÇ VE ÖNERİLER

Bu projede, Walmart'ın 45 mağazasına ait tarihsel satış verileri kullanılarak gelecekteki satışları ve talebi tahmin etmek amaçlanmıştır. Analiz sırasında, satışlar üzerindeki tatil etkileri, ekonomik göstergeler (TÜFE ve işsizlik oranı), yakıt fiyatları ve mağaza tipleri gibi faktörlerin etkileri incelenmiştir. Elde edilen sonuçlar, doğru ve güvenilir satış tahminlerinin işletmelerin stratejik karar alma süreçlerinde kritik bir role sahip olduğunu ve rekabet avantajını korumada etkili bir araç olduğunu göstermektedir.

6.1 Öngörülen Sonuçlar: 1. Tatil Etkileri ve Ekonomik Göstergeler: Tatil

haftalarının satışlar üzerindeki kritik etkisi doğrulanmış, ekonomik göstergelerin (örneğin, TÜFE ve işsizlik oranı) satış tahminlerinin doğruluğunu artırdığı bulunmuştur.

- 2. Model Performansı:** Zaman serisi tabanlı modellerden SARIMA, %99.25 doğruluk oranı ve en düşük WMAE (~745) değeriyle en başarılı tahmin modelini sunmuştur.
- 3. İşletme Stratejilerine Katkı:** Doğru satış tahminleri, işletmelerin stok yönetiminde ve lojistik planlamalarında daha etkili stratejiler geliştirmesine olanak tanımıştır.

6.2 Öngörülemeyen Sonuçlar:

- 1. Veri Sorunları:** Veri setinde beklenmedik anormallikler ve eksiklikler tespit edilmiş, bu durum veri temizleme sürecini uzatmıştır.
- 2. Ekonomik Göstergelerin Sınırlı Etkisi:** Bazı ekonomik göstergelerin (örneğin, işsizlik oranı) tahmin modelleri üzerindeki etkisi beklenenden düşük çıkmıştır.
- 3. Makine Öğrenimi Modelleri:** Random Forest gibi makine öğrenimi tabanlı modeller, tatil etkilerini yakalamakta zaman serisi modellerine göre daha zayıf kalmıştır.

6.3 Öneriler: 1. Daha Geniş Veri Setleri Kullanımı: Gelecekteki çalışmalar için, bölgesel farklılıklar ve müşteri davranışlarını içeren daha geniş ve çeşitli veri setlerinin kullanılması önerilmektedir.

- 2. Ek Değişkenlerin Dahil Edilmesi:** Promosyon stratejileri, rakip fiyatları gibi ek değişkenlerin tahmin modellerine entegre edilmesi tahmin doğruluğunu artırabilir.
- 3. Veri İşleme Süreçlerinin Otomasyonu:** Veri temizleme ve ön işleme süreçlerini hızlandırmak için otomatikleştirilmiş araçlar ve algoritmalar kullanılmalıdır.
- 4. Gelişmiş Yapay Zeka Teknikleri:** LSTM ve GRU gibi derin öğrenme modelleri ile mevcut tahmin modellerinin performansının karşılaştırılması önerilmektedir.
- 5. Hiperparametre Optimizasyonu:** Daha gelişmiş optimizasyon teknikleri (Bayesian Optimization gibi) kullanılarak modellerin performansı artırılabilir.

KAYNAKLAR

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Volume 55. US Government Printing Office.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

(Bu kaynak, ARIMA ve SARIMA modellerinin teorik temellerini kapsamlı bir şekilde ele almaktadır.)

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

(Random Forest algoritmasının temel prensiplerini açıklayan klasik bir kaynak.)

Hyndman, R. J. ve Athanasopoulos, G. (2018). *Forecasting: Principles and Practice* (2. Basım). OTexts. Erişim: <https://otexts.com/fpp2/>

(Zaman serisi tahmini, ARIMA ve SARIMA gibi yöntemlerin yanı sıra temel tahmin prensiplerini sunan kapsamlı bir kaynak.)

Makridakis, S., Spiliotis, E. ve Assimakopoulos, V. (2018). The M4 competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4), 802–808.

(Bu çalışma, çeşitli tahmin yöntemlerinin karşılaştırmalı performansını değerlendirirken SARIMA ve makine öğrenimi tabanlı modellerin yerini tartışmaktadır.)

Makridakis, S., Spiliotis, E. ve Assimakopoulos, V. (2020). The M5 competition: Background, Organization, and Dataset. *International Journal of Forecasting*.

(Walmart satış verileri üzerinde gerçekleştirilen M5 yarışmasında, SARIMA ve Random Forest gibi geleneksel ve makine öğrenimi tabanlı yöntemlerin karşılaştırıldığı bir ortam sunar.)

Bergmeir, C., Hyndman, R. J., ve Koo, B. (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis*, 120, 70–83.

(Zaman serisi tahmininde modellerin değerlendirilmesi için uygun yöntemler konusunda yol gösteren bir çalışma.)

McKinney, W. (2017). *Python for Data Analysis*. O'Reilly Media, Inc.

(Python kullanarak veri analizi ve ön işleme adımlarının temellerini sunar, projede kullanılan pandas kütüphanesi hakkında da bilgi içerir.)

VanderPlas, J. (2016). *Python Data Science Handbook*. O'Reilly Media, Inc.

(Python ekosisteminde veri bilimi araçları, görselleştirme, makine öğrenimi kütüphaneleri hakkında kapsamlı bir referans.)

İnternet Kaynakları:

Kaggle (Yasserh/Walmart-dataset ve Aslanahmedov/Walmart-sales-forecast). Walmart Mağaza Satış Tahmini Veri Setleri. Erişim:

<https://www.kaggle.com/yasserh/walmart-dataset>

<https://www.kaggle.com/aslanahmedov/walmart-sales-forecast>

Scikit-learn (2023). *scikit-learn: Machine Learning in Python*. Eriřim: <https://scikitlearn.org/stable/>

(Random Forest ve diğerk makine öğrenimi modellerinin Python ortamında nasıl uygulanacağına dair kapsamlı dokümantasyon.)

Pandas (2023). *pandas: Python Data Analysis Library*. Eriřim: <https://pandas.pydata.org/docs/>

(Veri işleme, temizleme ve dönüřtürme adımlarında kullanılan araçlar hakkında dokümantasyon.)

Statsmodels (2023). *Statsmodels: Statistical Modeling and Econometrics in Python*. Eriřim: <https://www.statsmodels.org/stable/>

(Zaman serisi modelleri, ARIMA, SARIMA gibi istatistiksel modelleme araçlarının Python uygulamaları.)

pmdarima (2023). *pmdarima: ARIMA estimators for Python*. Eriřim: <https://alkalineml.com/pmdarima/>

(Auto-ARIMA modelleri ve Python'da zaman serisi tahmini için otomatik hiperparametre seçimi.)

Matplotlib (2023). *Matplotlib: Python plotting library*. Eriřim: <https://matplotlib.org/>

(Veri görselleřtirme, tahmin sonuçlarının sunumu için kullanılır.)

Seaborn (2023). *Seaborn: Statistical data visualization*. Eriřim: <https://seaborn.pydata.org/>
(İstatistiksel veri görselleřtirme araçlarını sağılayan kütüphane.)