T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ İLE HAZIR GİYİM PERAKENDE SEKTÖRÜNDE SATIŞ TAHMİNİ

Salih CEYLAN

YÜKSEK LİSANS TEZİ

Matematik Mühendisliği Anabilim Dalı

Matematik Mühendisliği Programı

Danışman

Doç. Dr. Birol ASLANYÜREK

Temmuz, 2024

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ İLE HAZIR GİYİM PERAKENDE SEKTÖRÜNDE SATIŞ TAHMİNİ

Salih CEYLAN tarafından hazırlanan tez çalışması 17.07.2024 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Matematik Mühendisliği Anabilim Dalı, Matematik Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Doç. Dr. Birol ASLANYÜREK
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Doç. Dr. Birol ASLANYÜREK, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Hülya ŞAHİNTÜRK, Üye
Yıldız Teknik Üniversitesi _____

Dr. Öğr. Üye. Mustafa Zahid GÜRBÜZ, Üye
Doğuş Üniversitesi _____

Danışmanım Doç. Dr. Birol ASLANYÜREK sorumluluğunda tarafımca hazırlanan Makine Öğrenmesi ile Hazır Giyim Perakende Sektöründe Satış Tahmini başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Salih CEYLAN

İmza

Aileme

ve

sevgili eşime

Bu çalışma, “Yıldız Teknik Üniversitesi Bilimsel Araştırma Proje Koordinatörlüğü’nün FYL-2023-5651” numaralı projesi ile desteklenmiştir.

TEŞEKKÜR

Bu tezde perakende hazır giyim sektöründe tek sezon için satışa sunulan ya da yeni satışa sunulan ürünlerin mağazalardaki haftalık toplam satış adetinin çeşitli makine öğrenmesi yöntemleriyle tahmin edilmesini sağlamaya çalıştım. Buna göre Türkiye'de faaliyet gösteren çok mağazalı bir hazır giyim firmasının tek sezonluk satış verilerini kullanarak, belli bir ürünün ya da ürün gruplarının mağazalardaki haftalık toplam satışının tahmin edilmesini hedefledim.

Bu çalışmamda öncelikle hiçbir desteğini benden esirgemeyen, beraber defalarca kez beyin fırtınası yaptığım danışmanım Doç. Dr. Birol Aslanyürek'e teşekkürü borç bilirim.

Ürün satış datalarını incelemem için bana yardımcı olan yöneticilerime ve İş Zekâsı ekibine teşekkür ederim.

Eğitim hayatım boyunca hep yanımda olan sevgili aileme ve tez yazma sürecimde bana sabır gösteren eşime teşekkürlerimi sunarım.

Salih CEYLAN

İÇİNDEKİLER

KISALTMA LİSTESİ viii
ŞEKİL LİSTESİ ix
TABLO LİSTESİ xiii
ÖZET xiii
ABSTRACT xv
1 GİRİŞ 1
1.1 Literatür Özeti 1
1.2 Tezin Amacı 4
1.3 Hipotez 4
2 YAPAY ZEKA VE MAKİNE ÖĞRENMESİ 6
2.1 Yapay Zekaya Genel Bakış 6
2.2 Makine Öğrenmesi 7
2.2.1 Gözetimli Öğrenme (Supervised Learning) 7
2.2.2 Gözetimsiz Öğrenme (Unsupervised Learning) 11
2.2.3 Takviye Öğrenme (Reinforcement Learning) 11
3 VERİ ÖN İŞLEME VE VERİLERİN İSTATİSTİKSEL ANALİZİ 12
3.1 Veri Setinin Tanıtılması ve Ön-İşleme 12
3.2 Betimleyici İstatistik Analizi 16
3.3 Çıkarımsal İstatistik Analizi 23
3.3.1 Korelasyon Analizi 23
3.3.2 Hipotez Testleri 24
4 MAKİNE ÖĞRENMESİ MODELLERİNİN ELDE EDİLMESİ VE TAHMİN BAŞARILARININ TEST VERİLERİYLE SINANMASI 30
4.1 Eğitim ve Test Veri Setlerinin Oluşturulması 30
4.2 Makine Öğrenmesi Modellerinin Eğitilmesi 32
4.3 Eğitim ve Test Sonuçları 33
5 SONUÇ 45
KAYNAKÇA 47
TEZDEN ÜRETİLMİŞ YAYINLAR 49

KISALTMA LİSTESİ

AG	Alt Grup
BG	Buyer Grup
BF	(İndirimsiz) Birim Fiyat
BİM	Birim İndirim Miktarı
BK	Brüt Kar
BKM	Brüt Kar Marjı
DDİ	Doğal Dil İşleme
GSRS	Gün Sonu Reyon Stok
GSTS	Gün Sonu Toplam Stok
HN	Hafta No
HÖGS	Haftadaki Özel Gün Sayısı
İBF	İndirimli Birim Fiyat
İLST	İskontolu Satış Tutar
İO	İndirim Oranı
İST	İskontosuz Satış Tutar
KY	Kırıklık Yüzdesi
MK	Model Kod
MN	Mağaza No
MU	Mark-Up
MYG	Marka Yaş Grup
RK	Renk Kod
RÖ	Raf Ömrü
SA	Satış Adeti
SK	Sezon Kod
SMBM	Satılan Malın Birim Maliyeti
ÜK	Ürün Klasman
ÜON	Ürün Opsiyon No
YZ	Yapay Zekâ

ŞEKİL LİSTESİ

Şekil 3.1 *Veri Seti A*'da bulunan *Buyer Grup* özniteliğine ait ortalama satış adetleri 17

Şekil 3.2 *Veri Seti A*'da bulunan *Buyer Grup* özniteliğine ait histogram 17

Şekil 3.3 *Veri Seti A'*da bulunan *Mağaza No* özniteliğine ait histogram 18

Şekil 3.4 *Veri Seti A'*da bulunan *Mağaza No* özniteliğine ait ortalama satış adetleri 18

Şekil 3.5 *Veri Seti A'*da bulunan *Renk Kod* özniteliğine ait histogram 18

Şekil 3.6 *Veri Seti A'*da bulunan *Renk Kod* özniteliğine ait ortalama satış adetleri 18

Şekil 3.7 *Veri Seti A'*da bulunan *Ürün Klasman* özniteliğine ait histogram 18

Şekil 3.8 *Veri Seti A'*da bulunan *Ürün Klasman* özniteliğine ait ortalama satış adetleri 18

Şekil 3.9 *Veri Seti B'de* bulunan *Buyer Grup* özniteliğine ait histogram 19

Şekil 3.10 *Veri Seti B'de* bulunan *Buyer Grup* özniteliğine ait ortalama satış adetleri 19

Şekil 3.11 *Veri Seti B*'de bulunan *Mağaza No* özniteliğine ait histogram 19

Şekil 3.12 *Veri Seti B*'de bulunan *Mağaza No* özniteliğine ait ortalama satış adetleri 19

Şekil 3.13 *Veri Seti B*'de bulunan *Renk Kod* özniteliğine ait histogram 19

Şekil 3.14 *Veri Seti B*'de bulunan *Renk Kod* özniteliğine ait ortalama satış adetleri 19

Şekil 3.15 *Veri Seti B*'de bulunan *Ürün Klasman* özniteliğine ait histogram 20

Şekil 3.16 *Veri Seti B'*de bulunan *Ürün Klasman* özniteliğine ait ortalama satış adetleri 20

Şekil 3.17 *Veri Seti A*'de bulunan *Birim İndirim Miktarı* için kutu diyagramı 20

Şekil 3.18 *Veri Seti A*'de bulunan *Satılan Malın Birim Maliyeti* için kutu diyagramı 20

Şekil 3.19 *Veri Seti A*'de bulunan *Brüt Kar Marjı* için kutu diyagramı 20

Şekil 3.20 *Veri Seti A*'de bulunan *Haftadaki Özel Gün Sayısı* için kutu diyagramı 21

Şekil 3.21 *Veri Seti A*'de bulunan *İndirimli Birim Fiyat* için kutu diyagramı 21

Şekil 3.22 *Veri Seti A*'de bulunan *İndirim Oranı* için kutu diyagramı 21

Şekil 3.23 *Veri Seti A*'de bulunan *İndirimsiz Birim Fiyat* için kutu diyagramı 21

Şekil 3.24 *Veri Seti A*'de bulunan *Kırıklık Yüzdesi* için kutu diyagramı 21

Şekil 3.25 *Veri Seti A*'de bulunan *Mark-Up* için kutu diyagramı 21

Şekil 3.26 *Veri Seti A'*de bulunan *Satış Adeti* için kutu diyagramı 22

Şekil 3.27 *Veri Seti B*'de bulunan *Haftadaki Özel Gün Sayısı* için kutu diyagramı 22

Şekil 3.28 *Veri Seti B'*de bulunan *Kırıklık Yüzdesi* için kutu diyagramı 22

Şekil 3.29 *Veri Seti B'*de bulunan *Satış Adeti* için kutu diyagramı 22

Şekil 3.30 *Veri Seti A*'da bulunan *Hafta No* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması 26

Şekil 3.31 *Veri Seti A*'da bulunan *Mağaza No* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması 26

Şekil 3.32 *Veri Seti A*'da bulunan *Buyer Grup* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması 27

Şekil 3.33 *Veri Seti A*'da bulunan *Ürün Klasman* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması 27

Şekil 3.34 *Veri Seti A*'da bulunan *Renk Kod* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması 27

Şekil 3.35 *Veri Seti B*'de bulunan *Hafta No* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması 28

Şekil 3.36 *Veri Seti B*'de bulunan *Mağaza No* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması 28

Şekil 3.37 *Veri Seti B*'de bulunan *Buyer Grup* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması 28

Şekil 3.38 *Veri Seti B*'de bulunan *Ürün Klasman* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması 29

Şekil 3.39 *Veri Seti B*'de bulunan *Renk Kod* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması 29

Şekil 4.1 *Veri Seti A1* için gerçek ve tahmin değer dağılımı 34

Şekil 4.2 *Veri Seti A2* için gerçek ve tahmin değer dağılımı 35

Şekil 4.3 *Veri Seti B1* için gerçek ve tahmin değer dağılımı 35

Şekil 4.4 *Veri Seti B2* için gerçek ve tahmin değer dağılımı 36

Şekil 4.5 *Veri Seti A2* için tahmincilerin önemi 36

Şekil 4.6 *Veri Seti B2* için tahmincilerin önemi 37

Şekil 4.7 *Veri Seti A2* için *'Takım'* değeri sonuçları 40

Şekil 4.8 *Veri Seti A2* için *'Dokuma Alt'* değeri sonuçları 40

Şekil 4.9 *Veri Seti B2* için *'Takım'* değeri sonuçları 41

Şekil 4.10 *Veri Seti B2* için *'Dokuma Alt'* değeri sonuçları 41

Şekil 4.11 *Veri Seti A2* için belirli ürünleri gerçek ve tahmin edilmiş haftalık satış verileri 42

Şekil 4.12 *Veri Seti B2* için belirli ürünleri gerçek ve tahmin edilmiş haftalık satış verileri 42

Şekil 4.13 *Veri Seti A2* için *Model Kod* eklendiğinde gerçek ve tahmin değer dağılımı 44

TABLO LİSTESİ

Tablo 3.1 Ham veri setinde bulunan öznitelikler 13

Tablo 3.2 *Veri Seti A'da bulunan* sayısal öznitelikler için betimleyici istatistik değerleri 17

Tablo 3.3 *Veri Seti B'*de *bulunan* sayısal öznitelikler için betimleyici istatistik değerleri 17

Tablo 3.4 *Veri Seti A*'da bulunan sayısal öznitelikler için korelasyon katsayıları 23

Tablo 3.5 *Veri Seti A*'da bulunan sayısal öznitelikler için *p* değerleri 24

Tablo 3.6 *Veri Seti B*'da bulunan sayısal öznitelikler için korelasyon katsayıları 24

Tablo 3.7 *Veri Seti B*'da bulunan sayısal öznitelikler için *p* değerleri 24

Tablo 3.8 Kategorik öznitelikler için *p* değerleri 25

Tablo 4.1 Veri Setlerinde Kullanılan Hedef Değişkeni ve Öznitelikler 31

Tablo 4.2 Makine Öğrenmesi Aşamasında Kullanılan Veri Setleri ve Kayıt Sayıları 32

Tablo 4.3 *A1 Veri Seti* için Test ve Eğitim Sonuçları 33

Tablo 4.4 *A2 Veri Seti* için Test ve Eğitim Sonuçları 33

Tablo 4.5 *B1 Veri Seti* için Test ve Eğitim Sonuçları 33

Tablo 4.6 *B2 Veri Seti* için Test ve Eğitim Sonuçları 33

Tablo 4.7 *A2 Veri Seti* için özniteliklerin önem seviyeleri 38

Tablo 4.8 *A2 Veri Seti* için Öznitelik Seçimi Sonrası Test ve Eğitim Sonuçları .. 38

Tablo 4.9 *B2 Veri Seti* için özniteliklerin önem seviyeleri 39

Tablo 4.10 *B2 Veri Seti* için Öznitelik Seçimi Sonrası Test ve Eğitim Sonuçları 39

Tablo 4.11 *A2 Veri Seti* için *Buyer Grup* özniteliğinin değerlerine göre tahmin ölçümleri 39

Tablo 4.12 *B2 Veri Seti* için *Buyer Grup* özniteliğinin değerlerine göre tahmin ölçümleri 41

Tablo 4.13 *A2* Veri Seti için *Model No* Özniteliği Eklendiğinde Test ve Eğitim Sonuçları 43

ÖZET

Makine Öğrenmesi ile Hazır Giyim Perakende Sektöründe Satış Tahmini

Salih CEYLAN

Matematik Mühendisliği Anabilim Dalı

Matematik Mühendisliği Programı

Yüksek Lisans Tezi

Danışman: Doç. Dr. Birol ASLANYÜREK

Son yarım yüzyılda ülkemizde perakende hazır giyim sektörü hızla gelişmiştir ve günümüzde bu sektörde iş yapan çok sayıda ulusal ve uluslararası firma bulunmaktadır. Perakende sektöründe satışların önceden tahmin edilmesi; üretim planlaması, lojistik, reyonlardaki stok yönetimi ve iş gücü planlaması açısından önem arz etmektedir. Ürün karakteristiği, fiyat, hava durumu, özel günler, rakip firmaların durumu, alım gücü gibi birçok etken perakende sektöründe günlük ya da haftalık satış miktarını etkiler. Diğer perakende sektörleri ile karşılaştırıldığında hazır giyim sektörü, müşteri taleplerinin modaya bağlı olarak hızlı bir şekilde değişebildiğinden dolayı dinamik bir yapıya sahiptir.

Perakende hazır giyim sektöründe talep ya da satış tahminine dönük zaman serileri analizine dayanan makine öğrenmesi yöntemleri revaçtadır. Makine öğrenmesi yöntemlerini kullanan çalışmaların çok büyük kısmı geçmiş satış verilerini kullanarak eski ürünlerin satış tahminine odaklanmıştır. Az sayıdaki çalışmada ise, satışa yeni sürülen ve dolayısıyla eski satış verileri olmayan ürünler için, genellikle az sayıda özniteliğe sahip benzer özellikteki ürünlerin verileri kullanılmıştır.

Bu çalışmanın amacı; Türkiye'de faaliyet gösteren bir firmanın 2019 yaz sezonuna ait verilerini kullanarak, perakende hazır giyim sektöründe tek sezon için yeni satışa sunulan ürünlerin mağazalardaki haftalık toplam satışının istatistiksel olarak analiz edilmesi ve çeşitli makine öğrenmesi yöntemleriyle tahmin edilmesidir. İstatistiksel analiz kısmında, ilk olarak veri seti üzerinde temizleme ve dönüştürme gibi veri ön işlemleri yapılmıştır. Sayısal öznitelikler için Pearson korelasyon analizi yapılmıştır. Kategorik öznitelikler için ise Kruskal-Wallis Testi uygulanarak gruplar arasında haftalık satış adeti açısından bir fark olup olmadığı saptanmıştır. 0,05'ten küçük *p* değerleri için sonuçlar anlamlı kabul edilmiştir. Makine öğrenmesi aşamasındaysa, farklı veri setleri için 10 kat çapraz doğrulama ile çeşitli makine öğrenmesi modelleri eğitilmiş ve ardından bu modellerin başarısı test veri seti üzerinden karşılaştırılmıştır. Böylece, geçmiş yıllarda satış verisi bulunmayan, yeni satışa sunulan ürünlerin mağazalardaki haftalık toplam satışının tahmin edilmesine yardımcı olacak modeller geliştirilmiştir.

Anahtar Kelimeler: Makine öğrenmesi, ürün satış tahmini, perakende hazır giyim.

ABSTRACT

Sales Forecasting in Apparel Retail Industry with Machine Learning

Salih CEYLAN

Department of Mathematical Engineering

Master's Thesis

Supervisor: Assoc. Prof. Dr. Birol ASLANYÜREK

The apparel retail industry has developed rapidly in our country in the last half century, and today there are many national and international companies operating in this sector. Forecasting sales in the retail sector is important in terms of production planning, logistics, shelf stock management and workforce planning. Many factors such as product characteristics, price, weather, special days, the status of competitors, purchasing power affect the daily or weekly sales amount in the retail sector. Compared to other retail sectors, the ready-to-wear sector has a dynamic structure because customer demands can change rapidly depending on fashion trends.

Many studies have been conducted in the retail ready-to-wear sector using classical methods such as time series analysis or simple linear regression for demand or sales forecasting. However, these methods use linear models and can often fail to predict complex problems. For this reason, machine learning methods using nonlinear models have become popular in recent years. The vast majority of studies using machine learning methods have focused on sales forecasting of old products using past sales data. In a small number of studies, data from similar products with a small number of attributes were used for products that were newly launched and therefore did not have old sales data.

The aim of this study is to statistically analyze the total weekly sales of newly launched products in stores for a single season in the retail ready-to-wear sector using the data from the 2019 summer season of a company operating in Turkey and to estimate them with various machine learning methods. In the statistical analysis section, first, data pre-processing such as cleaning and transformation was performed on the data set. Pearson correlation analysis was performed for numerical attributes. For categorical attributes, Kruskal-Wallis Test was applied to determine whether there was a difference between the groups in terms of weekly sales. Values where the p value was less than 0.05 were considered significant. In the machine learning phase, various machine learning models were trained with 10-fold cross-validation for different data sets, and then the success of these models was compared on the test data set. Thus, models were developed that would help estimate the total weekly sales of newly launched products in stores for which there was no sales data in previous years.

Keywords: Machine learning, product sales forecasting, retail apparel.

1
GİRİŞ

1.1 Literatür Özeti

Son elli yılda, ülkemizde perakende *hazır* giyim sektörü büyük bir hızla büyümüş ve günümüzde bu alanda faaliyet gösteren birçok ulusal ve uluslararası firma bulunmaktadır. Bu firmalar için ürün satışlarının önceden tahmin edilmesi; üretim planlaması, lojistik, reyonlardaki stok yönetimi ve iş gücü planlaması açısından ciddi kolaylıklar sağlayacaktır. Ürünlerin satış adetinin önceden tahmin edilmesi özellikle raflardaki stok planlaması açısından önemlidir. Raflardaki stokların tükenmesi müşterilerin rakip firmalara gitmesine neden olacak ve firma ekonomik kayba uğrayacaktır. Stokların fazla olması ise stok maliyetlerini artıracaktır, raflardaki ürün çeşitliliğinin azalmasına ve ayrıca ürünlerin sezon sonunda elde kalmasına neden olacaktır.

Giyim sektöründe satış tahmini; ürün özellikleri (renk, beden, kumaş, giysi tipi), maliyet, satış fiyatı, müşterilerin alım gücü, hava durumu, coğrafya, özel günler, mevsim özellikleri gibi birçok faktöre bağlı olduğu için zor bir problemdir. Bu nedenle özellikle son 30 yıldır bu problem hakkında çok sayıda çalışma yapılmıştır ve problem halen güncelliğini korumaktadır. Bu konu hakkında yapılan çalışmalarda kullanılan tahmin yöntemleri, kabaca klasik istatistik yöntemleri ve makine öğrenmesi yöntemleri şeklinde ikiye ayrılır.

Hazır giyim sektöründe satış tahmini için kullanılan klasik yöntemlerin büyük kısmı zaman serileri analizine dayanan yöntemlerin türevlerinden (Naive yöntemi, ortalama yöntemleri, üssel düzeltme yöntemleri) oluşmaktadır [1-7]. Zaman serileri analizine dayanan yöntemler dışında kullanılan diğer bir klasik yöntem doğrusal regresyon analizidir [8-9].

Klasik yöntemler doğrusal modeller kullanırlar ve bu nedenle karmaşık problemlerin modellenmesinde başarıları düşebilmektedir. Perakende hazır giyim sektöründe satış miktarları yukarıda bahsedilen birçok farklı etmene bağlı olduğu için nominal verileri de içeren büyük veri setlerini işleyecek ve doğrusal olmayan karmaşık ilişkileri de dikkate alabilecek yapıda yöntemlerin kullanılmasının satış tahmini başarısını artırması beklenir [10]. Günümüzde sıkça kullanılan, yapay zekâ

temelli olan makine öğrenmesi yöntemlerinin önemli bir kısmı bu özellikleri taşımaktadır. Aşağıda perakende hazır giyim sektöründeki satış tahmini için makine öğrenmesi yöntemlerini kullanan çalışmalar ele alınmıştır.

Giyim sektöründe ürünlerin bir kısmı tüm yıl boyunca satılırken, bir kısmı sezonluk satışa sunulmaktadır. Önceki yıllarda veya önceki sezonlarda satışa sunulmuş ürünlerin geçmiş satış verilerine bakarak bu ürünlerin gelecekteki satış adetleri tahmin edilebilir. Bu tip durumlar için, geçmiş satış verilerine bakarak, bir ürünün ya da ürün grubunun gelecekteki satış adetinin tahmini için birçok makine öğrenmesi yöntemi önerilmiştir [1-2, 8, 11-13]. [1]'de sunulan çalışmada; ABD menşeli bir giyim firmasının 1997-2000 yılları arasına ait sadece fiyat, renk, beden, satış tarihi ve kıyafet tipi içeren verileri toplanmış ve yapay sinir ağları ile zaman serileri analizine dayanan iki yöntem aracılığıyla satış tahminleri karşılaştırılmıştır. [2] ile verilen çalışmada, yine ABD menşeli, e-ticaret de yapan bir giyim firmasının satış verileri kullanılarak, uzun dönemli satış tahminleri çeşitli makine öğrenmesi yöntemleriyle karşılaştırılmıştır. Türkiye'de bulunan bir hazır giyim firmasına ait 2014-2018 yılları arasındaki satış verilerini toplayan bir çalışmada ise [11, 12], çok daha geniş bir girdi seti (ürün özellikleri, ürün maliyeti, satış fiyatı, ülkenin ekonomik durumu, özel günler, gün sonu reyon stok vs. içeren veriler) kullanılarak, yapay sinir ağları, destek vektör makineleri ve rastgele orman yöntemleri karşılaştırılmıştır. Ayrıca, Türkiye'de faaliyet gösteren başka bir giyim firmasına ait 2017-2020 yılları arasındaki veriler toplanmış ve çeşitli makine öğrenmesi yöntemleriyle satış tahmini yapılmıştır [8]. Bu çalışmada girdi sayısı geniş olmasına rağmen ürün fiyatı, ürün maliyeti, ürünlerin renk, malzeme ve beden bilgileri kullanılmamıştır. Bunların dışında ulusal pazarda faaliyet gösteren diğer bir firmaya ait 36 aylık promosyon verileri, mağazaya gelen müşteri sayısı, hava sıcaklığı verileri kullanılarak yapay sinir ağlarına dayanan modeller geliştirilmiş ve satış tahmini yapılmıştır [13].

Yukarıda bahsedilen çalışmalar uzun dönemli geçmiş satış verileri kullanılarak gerçekleştirilmiştir. Fakat hazır giyim sektöründe toplumsal ürün beğenileri hızlı bir şekilde değiştiğinden dolayı daha önceki yıllarda satışa sunulmayan birçok yeni ürün pazara sunulmaktadır ve dolayısıyla piyasaya sunulan yeni ürünlerde eski satış kayıtları mevcut değildir. Bu durumda piyasaya yeni sürülen bir ürünün satış tahmini için, bu ürünün belli bir dönem satış istatistikleri kayıt altına alınıp eğitim

verisi olarak kullanılabilir. Eğer sezonluk satışa sunulan bir ürün söz konusuysa, satış istatistiklerinin toplanacağı bu dönem çok daha kısa olacaktır ve bu da tahmin problemini zorlaştıracaktır. Bu kapsamda; Türkiye'de faaliyet gösteren, spor ürünleri satan bir firmanın piyasaya yeni çıkan ürünlerinin ilk üç/altı haftadaki satış kayıt verilerini kullanarak makine öğrenmesi modelleri geliştirilmiştir [18]. Bu çalışmada geniş bir öznitelik sayısı (ürün grubu, ürün tipi, spor kodu, renk, cinsiyet, ürün alt grubu, ürün kırılımı, tema, ürün sıralaması, ürün kataloğu, ürün bedeni, ürün üretim ve gönderim maliyeti, ürün satış fiyatı, vb.) kullanılarak yapay sinir ağları, rastgele orman yöntemi ve kümeleme yöntemleri uygulanmıştır.

Piyasaya yeni sürülen bir ürünün satış istatistiklerinin kısa süreli de olsa kayıt altına alınması satış tahmini için anlamlı bir veri sağlasa da diğer yandan bir ürünün satışa sunulmadan satış adetinin tahmin edilebilmesi satış yapan firmaya üretim planlaması, lojistik ve reyonlardaki stok yönetimi gibi konularda ciddi kolaylıklar sağlayacaktır. Bunun için yeni ürünün geçmiş satış verileri yerine benzer özelliklere sahip eski ürünlerin satış istatistikleri kullanılabilir. Fransa'da [14, 15], İspanya-Portekiz'de [10] ve Hong Kong'da [16,17] faaliyet gösteren firmalara ait satış verileri toplanarak bu tip makine öğrenmesi modelleri kuran araştırmalar yapılmıştır. Fransa'da bulunan firma verilerinden sadece fiyat, satışın başladığı tarih ve ürünün kullanım ömrü girdi olarak alınmış ve karar ağaçlarına [14] ve yapay sinir ağlarına [15] dayanan yöntemlerle satış tahmini gerçekleştirilmiştir. Hong Kong menşeli firmalara ait verilerden ise sadece ürün rengi, ürün bedeni ve fiyat girdi olarak alınmış ve yapay sinir ağlarına dayalı makine öğrenmesi yöntemleri geliştirilerek satış tahmini yapılmıştır [16,17]. İspanya ve Portekiz'de yapılan araştırmada ise ürün tipi, rengi, boyutu, ürünün satıldığı mağaza tipi, ürün fiyatı gibi ürün özellikleri yanında mağazada çalışan uzmanların görüşlerinden elde edilen ürünün tahmini satış seviyesi, modaya uygunluk seviyesi gibi öznel nitelikler de girdi olarak kullanılmış ve çeşitli makine öğrenmesi yöntemleriyle satış tahmini yapılmıştır.

Görüldüğü üzere geçmiş satış istatistikleri olmayan yeni bir ürünün satış tahmini için yapılan çalışmalarda, [10] dışında, çok az girdi ile modeller kurulmuştur. Oysaki yeni bir ürünün satış tahmini için benzer nitelikteki ürünlerin çeşitli bilgileri oldukça önem taşımaktadır ve bu özelliklerin girdi olarak kullanılması makine öğrenmesi modellerinin başarısını artırması beklenir. [10]'da yapılan çalışmada ise

çok sayıda öznitelik kullanmakla birlikte ürünün tahmini satış seviyesi, modaya uygunluk seviyesi gibi önceden toplanması zor olan, profesyonel kişilerin görüşlerine dayalı veri setlerine ihtiyaç vardır.

1.2 Tezin Amacı

Bu çalışmanın amacı; uzmanların profesyonel görüşlerine gerek duymadan, "renk, beden, ürün tipi, ürünün raf ömrü, beden bulunurluğu, ürün fiyatı, ürün maliyeti, takvimdeki özel günler, vb." öznitelikler kullanarak yeni satışa sunulan ürünlerin mağazalardaki haftalık toplam satışının çeşitli makine öğrenmesi yöntemleriyle tahmin edilmesidir. Bunun için Türkiye'de faaliyet gösteren çok mağazalı bir hazır giyim firmasının 2019 yaz sezonuna ait satış verileri kullanılmış ve karar ağaçları, destek vektör makineleri, rastgele orman yöntemi ve yapay sinir ağları gibi modellerle belli bir ürünün tüm mağazalardaki haftalık toplam satışının tahmini gerçekleştirilmiştir.

Çalışmanın ilk aşamasında veriler ön-işlemeye tabi tutulmuş ve ardından betimleyici ve çıkarımsal istatistik yöntemleriyle analiz edilmiştir. İstatistiksel analiz kısmında, ilk olarak veri seti üzerinde temizleme ve dönüştürme gibi veri ön işlemleri yapılmıştır. Sayısal öznitelikler için Pearson korelasyon analizi yapılmıştır. Kategorik öznitelikler için ise Kruskal-Wallis Testi uygulanarak gruplar arasında mağazalardaki haftalık toplam satış adeti açısından bir fark olup olmadığı saptanmıştır. 0,05'ten küçük p değerleri için sonuçlar anlamlı kabul edilmiştir. Makine öğrenmesi aşamasındaysa, 10 kat çapraz doğrulama ile çeşitli makine öğrenmesi modelleri eğitilmiş ve ardından bu modellerin başarısı test veri seti üzerinden karşılaştırılmıştır.

1.3 Hipotez

Perakende hazır giyim sektöründe ürün satışları; ürün karakteristiği (renk, beden, kumaş, ürün tipi, vb.), ürün fiyatı, ürünün maliyeti, hava durumu, özel günler, ürünlerin reyonda bulunurluğu ve sunumu gibi birçok faktöre bağlıdır. Buna göre, satış tahmini için doğrusal olmayan, karmaşık modelleri çözebilecek makine öğrenmesi modelleri kullanılabilir. Ayrıca giyim sektöründe insanların beğenileri hızlı bir şekilde değişebilmekte ve piyasaya sürekli yeni ürünler çıkabilmektedir. Fakat yeni ürünlerin eski satış bilgileri mevcut değildir. Bu kapsamda, bu çalışmada

dayandığımız hipotez; perakende hazır giyim sektöründe yeni satışa çıkarılan ürünlerin satış tahmini için fazla sayıda öznitelik (renk, beden, ürün tipi, ürünün raf ömrü, beden bulunurluğu, ürün fiyatı, ürün maliyeti, brüt kâr marjı, takvimdeki özel günler, vb.) ve benzer özellikteki eski ürünlere ait satış istatistiklerini kullanarak, ürünlerin mağazalardaki haftalık toplam satış tahminini başarılı bir şekilde gerçekleştirebilecek makine öğrenmesi modellerinin kurulabileceğidir.

2
YAPAY ZEKA VE MAKİNE ÖĞRENMESİ

2.1 Yapay Zekaya Genel Bakış

Yapay zekâ, bilgisayar sistemlerine insan benzeri zekâ ve öğrenme yetenekleri kazandırmayı amaçlayan bir bilim dalıdır. YZ, karmaşık problemleri çözmek, örüntüleri tanımak, kararlar vermek, dil anlamını anlamak, sesi tanımak gibi insan zekasına benzer görevleri gerçekleştirmek için kullanılan bir dizi teknik ve algoritma kullanır. YZ'nin temel amacı, bilgisayarların veri ve deneyimlerden öğrenerek, yeni bilgileri işlemesini, kararlar almasını ve problemleri çözmesini sağlamaktır.

YZ, birçok endüstride büyük potansiyele sahip olup, sağlık, ulaşım, eğitim, finans, otomasyon gibi birçok alanda kullanılmaktadır. Örneğin, tıp alanında YZ, teşhis koyma, ilaç geliştirme ve hastaların takibi gibi konularda önemli bir rol oynamaktadır. Otomotiv sektöründe, otonom araçlar YZ tekniklerini kullanarak trafikte sürücüsüz olarak seyahat edebilirler.

YZ hızla gelişen bir alan olup, her geçen gün yeni teknikler, algoritmalar ve uygulamalar geliştirilmektedir. Ancak, YZ'nin kullanımı bazı etik ve sosyal sorunları da beraberinde getirebilir. Örneğin, gizlilik, güvenlik, işsizlik gibi konular YZ'nin etik ve sosyal boyutlarıdır ve dikkate alınması gereken önemli hususlardır.

YZ'nin geliştirilmesinde kullanılan temel yöntemlerden bazıları şunlardır: Makine Öğrenmesi, Derin Öğrenme, Doğal Dil İşleme, Görüntü İşleme, vb.

YZ'nin önemli bir bileşeni olan makine öğrenmesi, bilgisayar sistemlerinin deneyimlerden ve verilerden öğrenmesini sağlayan algoritmalardan oluşur. Makine öğrenmesi, veri setlerini analiz eder, örüntüleri tanır ve gelecekteki olayları tahmin etmek için bu örüntüleri kullanır.

Derin öğrenme, insan beyninin sinir ağlarından ilham alarak çalışan bir makine öğrenme alt dalıdır. Büyük veri setlerini analiz etmek ve karmaşık problemleri çözmek için derin sinir ağları kullanır. Derin öğrenme, görüntü tanıma, doğal dil işleme, otonom sürüş gibi birçok alanda büyük başarılar elde etmiştir.

Doğal dil işleme (DDİ), bilgisayarların insan dilini anlaması ve üretmesi için tasarlanmış bir daldır. Metin tabanlı verileri işlemek, metinleri anlamak, dilbilgisi ve anlam analizi yapmak gibi görevleri gerçekleştirir. DDİ, otomatik metin çevirisi, metin sınıflandırması, metin tabanlı soru cevaplama gibi birçok uygulamada kullanılır.

Görüntü işleme, bilgisayarların görüntüleri analiz etmesi ve anlamasıyla ilgilenen bir alandır. Görüntü işleme, görüntü tanıma, yüz tanıma, nesne algılama, görüntü sınıflandırma gibi uygulamaları içerir. YZ'nin birçok uygulamasında temel bir bileşen olarak yer alır.

2.2 Makine Öğrenmesi

Makine öğrenmesi, bir bilgisayarın belirli bir görevi gerçekleştirmek veya belirli bir problemi çözmek için veri ve deneyimlerden otomatik olarak öğrenmesini sağlayan bir YZ alanıdır. Bu öğrenme süreci, bir algoritmanın belirli bir performans ölçütüne (örneğin, doğruluk oranı veya hata miktarı) dayanarak deneyimleri analiz etmesini, örüntüleri tanımasını ve gelecekteki verilere uygulayabileceği genellemeler yapmasını içerir.

Makine öğrenmesi, genel olarak üç temel kategoriye ayrılır ve bu kategoriler aşağıda kısaca anlatılmaktadır.

2.2.1 Gözetimli Öğrenme (Supervised Learning)

Gözetimli öğrenme, giriş verileri ve bu verilere karşılık gelen hedef çıktıları içeren bir eğitim veri kümesine dayanır. Algoritma, bu eğitim verilerini kullanarak giriş ve çıktı arasındaki ilişkiyi anlamaya çalışır ve sonuçta yeni giriş verileri için doğru çıktıları tahmin edebilir. Örneğin, bir gözetimli öğrenme algoritması, görüntülerin etiketlendiği bir veri kümesini kullanarak, yeni bir görüntüyü tanımlayabilir veya sınıflandırabilir.

Bu tezde; gözetimli öğrenme yöntemlerinden olan karar ağaçları, destek vektör makineleri, rastgele orman yöntemi ve yapay sinir ağları kullanıldığı için, aşağıda bu yöntemlerden kısaca bahsedilmiştir.

2.2.1.1 Karar Ağaçları

Karar ağaçları, veri setindeki özelliklere göre dallara ayrılarak kararlar veren ve sonrasında bu kararların sonucunu tahmin eden bir makine öğrenmesi algoritmasıdır. Hem sınıflandırma hem de regresyon problemlerinde kullanılabilir.

Çalışma Prensibi:

Kök Düğüm (Root Node): Ağaç, kök düğümden başlar ve veri setindeki en önemli özelliğe göre bölünür.

Dallanma (Branching): Her bir düğüm, veri setini alt kümelere ayıran bir karar noktasıdır. Karar noktaları, belirli bir özelliğin belirli bir değeri ile karşılaştırılarak belirlenir.

Yaprak Düğüm (Leaf Node): Ağaçta karar noktalarına ulaşıldığında, bu noktalar sınıflandırma veya tahmin sonucunu temsil eden yaprak düğümlerine dönüşür.

Avantajlar:

- Kolay anlaşılır ve görselleştirilebilir.
- Veri ön işleme gereksinimi düşüktür.
- Hızlı ve etkili sonuçlar sağlar.

Dezavantajlar:

- Overfitting (aşırı uyum) riski yüksektir.
- Büyük veri setlerinde ve çok sayıda özellik olduğunda performansı düşebilir.

2.2.1.2 Destek Vektör Makineleri

Destek vektör makineleri (Support Vector Machines - SVM), verileri sınıflandırmak veya regresyon yapmak için kullanılan güçlü ve esnek bir makine öğrenmesi algoritmasıdır. Destek vektör makineleri, veriyi sınıflandırmak için en iyi ayırma düzlemini (hiper düzlemi) bulmaya çalışır.

Çalışma Prensibi:

Hiper Düzlem (Hyperplane): Destek vektör makineleri, veriyi iki sınıfa ayıran en iyi düzlemi bulur. Bu düzlem, sınıflar arasındaki mesafeyi maksimize edecek şekilde yerleştirilir.

Marjin (Margin): Hiper düzlem ile en yakın veri noktaları arasındaki mesafe. Destek vektör makineleri, bu marjini maksimize etmeye çalışır.

Çekirdek Hileleri (Kernel Tricks): Doğrusal olarak ayrılabilir olmayan veriler için, çekirdek fonksiyonları kullanılarak veri, daha yüksek boyutlu bir uzaya dönüştürülür ve burada ayrılması sağlanır.

Avantajlar:

- Yüksek doğruluk sağlar.
- Overfitting'e karşı dayanıklıdır (özellikle büyük boyutlu veri setlerinde).

Dezavantajlar:

- Eğitim süresi uzundur.
- Parametre ayarı (örneğin çekirdek fonksiyonu seçimi) zordur.

2.2.1.3 Rastgele Orman

Rastgele orman (Random Forest), birden fazla karar ağacının bir araya gelerek oluşturduğu bir topluluk öğrenme algoritmasıdır. Hem sınıflandırma hem de regresyon problemlerinde kullanılır.

Çalışma Prensibi:

Bagging (Bootstrap Aggregating): Veri setinden rastgele örnekler alınarak birçok alt veri seti oluşturulur. Her bir alt veri seti, bir karar ağacı eğitmek için kullanılır.

Random Subspaces: Her bir karar ağacı, belirli bir özellik alt kümesi ile eğitilir.

Oylama (Voting): Tüm ağaçların sonuçları birleştirilerek nihai sonuç elde edilir. Sınıflandırma problemlerinde çoğunluk oylaması, regresyon problemlerinde ise ortalama alınır.

Avantajlar:

- Overfitting riskini azaltır.
- Daha yüksek doğruluk sağlar.
- Eksik verilerle başa çıkabilir.

Dezavantajlar:

- Yorumlanabilirliği düşüktür.
- Eğitim süresi ve hesaplama maliyeti yüksektir.

2.2.1.4 Yapay Sinir Ağları

Yapay sinir ağları (Artificial Neural Networks), insan beyninin çalışma prensiplerinden ilham alınarak geliştirilmiş, çok katmanlı ve birbirine bağlı yapay nöronlardan oluşan bir modeldir. Hem sınıflandırma hem de regresyon problemlerinde kullanılır.

Çalışma Prensibi:

Katmanlar (Layers): Yapay sinir ağları, giriş katmanı (input layer), gizli katmanlar (hidden layers) ve çıkış katmanı (output layer) olmak üzere katmanlardan oluşur.

Nöronlar (Neurons): Her bir katmanda bulunan ve bilgiyi işleyen temel birimlerdir. Nöronlar, girişlerinden aldıkları bilgiyi işleyip aktivasyon fonksiyonu aracılığıyla bir sonraki katmana iletirler.

Ağırlıklar (Weights) ve Önyargılar (Biases): Nöronlar arasındaki bağlantılar, belirli ağırlıklar ve önyargılar ile tanımlanır. Bu ağırlıklar, öğrenme sürecinde güncellenir.

Avantajlar:

- Karmaşık veri ilişkilerini öğrenebilir.
- Büyük veri setleriyle iyi performans gösterir.
- Derin öğrenme (deep learning) ile daha karmaşık yapılar oluşturulabilir.

Dezavantajlar:

- Yüksek hesaplama maliyeti ve uzun eğitim süreleri gerektirir.
- Overfitting riski yüksektir.
- Yorumlanabilirliği düşüktür.

2.2.2 Gözetimsiz Öğrenme (Unsupervised Learning)

Gözetimsiz öğrenme, etiketlenmemiş veri kümesini analiz etmek için kullanılır. Bu durumda, algoritma verilerdeki örüntüleri ve ilişkileri kendiliğinden keşfeder. Gözetimsiz öğrenme, veri kümesindeki benzerlikler ve farklılıklar arasında gruplamalar yapabilir veya verinin temsilini öğrenebilir.

Örneğin, bir gözetimsiz öğrenme algoritması, bir e-ticaret sitesindeki müşteri alışveriş verilerini analiz ederek benzer alışverişçi profillerini gruplayabilir.

2.2.3 Takviye Öğrenme (Reinforcement Learning)

Takviye öğrenme, bir ajanın bir ortamda belirli bir hedefi elde etmek için gerçekleştirdiği aksiyonları öğrendiği bir öğrenme yöntemidir. Ajan, çevresel durumları gözlemleyerek ve belirli aksiyonları gerçekleştirerek çevredeki ödül veya cezaları deneyimler. Amacı, zamanla en yüksek ödülü elde edecek şekilde davranışlarını optimize etmektir.

Örneğin, bir takviye öğrenme algoritması, bir robotun belirli bir hedefe giden en etkili yol ve hareketler kombinasyonunu öğrenebilir.

Makine öğrenmesi, gerçek dünya problemlerinin çözümünde geniş bir kullanım alanına sahiptir. Örnek uygulamalar arasında görüntü ve ses tanıma, doğal dil işleme, pazarlama analitiği, finansal tahminler ve tıbbi teşhisler bulunur. Makine öğrenmesi algoritmaları sürekli olarak geliştirilmekte ve büyük veri setlerinin analizinde ve karmaşık problemlerin çözülmesinde önemli bir rol oynamaktadır.

3
VERİ ÖN İŞLEME VE VERİLERİN İSTATİSTİKSEL ANALİZİ

Firmadan alınan 2019 yaz sezonuna ait veri seti ham halde olduğu için ilk olarak veri seti üzerinde temizleme ve dönüştürme gibi veri ön işlemleri yapılmıştır. Sonrasında ise işlenmiş veri seti üzerinde sırayla betimleyici ve çıkarımsal istatistik analiz yapılmıştır.

Firmadan alınan veri seti ham halde olduğu için eksik ve hatalı veriler içermektedir. Ayrıca gereksiz öznitelikler de içermektedir. Bunun yanında ürünlerin birim satış fiyatı ve maliyeti gibi verileri doğrudan içermemektedir, fakat diğer öznitelikler kullanılarak bu bilgilere erişmek mümkündür. Bu nedenle, ilk olarak veri seti üzerinde temizleme ve dönüştürme gibi veri ön işlemleri yapılmıştır.

İkinci aşamada işlenmiş veri seti üzerinde betimleyici istatistik analizi yapılmıştır. Sonrasında ise çıkarımsal istatistik analizi yapılmıştır.

3.1 Veri Setinin Tanıtılması ve Ön-İşleme

Firmadan alınan veri seti ham halde olduğu için eksik ve hatalı veriler içermektedir. Ayrıca gereksiz öznitelikler de içermektedir. Bunun yanında ürünlerin birim satış fiyatı ve maliyeti gibi verileri doğrudan içermemektedir, fakat diğer öznitelikler kullanılarak bu bilgilere erişmek mümkündür. Aşağıda veri seti tanıtıldıktan sonra veri seti üzerinde temizleme ve dönüştürme gibi veri ön işlemleri yapılmıştır.

Firmadan alınan veri seti ham haliyle 1.953.948 adet kayıt içermektedir. Bu kayıtlar, firmanın 2019 yılında yaptığı haftalık satışlarının bilgisini içermektedir. Firmada veriler belirli öznitelikler ile sezonsal olarak tutulmaktadır. Sezon olarak ifade edilen terim ise aslında mevsim olarak düşünülebilir.

Mevsimsel olarak satış miktarları, satılan ürünlerin özellikleri farklılık göstereceği için bu durum önem arz etmektedir. Bu yüzden tahminleme yapılırken aslında sezonsal veriler baz alınmıştır.

Veri setinde yer alan sütunlar aşağıdaki tabloda verilmiştir.

Tablo 3.1 Ham veri setinde bulunan öznitelikler

Öznitelik Adı	Tipi	Açıklama
Hafta No	Kategorik	2019 yılının kaçıncı haftasının olduğu bilgisidir.
Mağaza No	Kategorik	Mağaza kodudur.
Ürün Opsiyon No	Kategorik	Kullanılmayacaktır.
Marka Yaş Grup	Kategorik	Satılan ürünlerin hitap ettiği yaş grubudur. Kullanılmayacaktır.
Alt Grup	Kategorik	Satılan ürünlerin hitap ettiği yaş grubunun alt kırılımlarıdır. Kullanılmayacaktır.
Buyer Grup	Kategorik	Satılan ürünlerin kumaş ve kesim türünü anlatan kategorik veridir.
Ürün Klasman	Kategorik	Ürünleri yer aldığı klasmanın kodudur.
Model Kod	Kategorik	Ürünün modelini belirten koddur. Her ürün için eşsiz bir koda sahiptir.
Renk Kod	Kategorik	Ürünün rengini belirten koddur.
Sezon Kod	Kategorik	Ürünlerin mevsimsel olarak hangi sezona ait olduğunun bilgisidir. Kullanılmayacaktır.
Satış Adeti	Sayısal	Haftalık satış adedidir.
Gün Sonu Toplam Stok	Sayısal	Kullanılmayacaktır.
Gün Sonu Reyon Stok	Sayısal	Kullanılmayacaktır.
İskontosuz Satış Tutarı	Sayısal	Ürünün bir haftalık iskontosuz toplam satış tutarıdır. Doğrudan kullanılmayacak olup, dönüştürme işleminde kullanılacaktır.
İskontolu Satış Tutarı	Sayısal	Ürünün bir haftalık iskontolu toplam satış tutarıdır. Doğrudan kullanılmayacak olup, dönüştürme işleminde kullanılacaktır.
Cover	Sayısal	Mağaza stok sayısının 7 günlük net satışa bölünerek hesaplanan sayısal karşılığıdır.
Mark-Up	Sayısal	Ürünün satış fiyatının maliyetine oranıdır.
Gün Sonu Yol Stok	Sayısal	Kullanılmayacaktır.
Brüt Kar	Sayısal	Üründen elde edilen haftalık brüt kardır. Doğrudan kullanılmayacak olup, dönüştürme işleminde kullanılacaktır.
Brüt Kar Marjı	Sayısal	Brüt karın KDV'siz satış tutarına oranıdır.
Beden Bulunurluğu	Sayısal	Ürün bedeninin bulunurluğu bilgisidir. Kullanılmayacaktır.
Raf Ömrü	Sayısal	Bir mağazada ürünün satıldığı veya gün sonu stoğunun sıfırdan büyük olduğu gün sayılarının toplamıdır. Çok fazla NULL değer içerdiği için kullanılmayacaktır.
Kırıklık Yüzdesi	Sayısal	Fiili beden sayısının tanımlı beden sayısına oranıdır.

Veri setindeki *Brüt Kar*, *İskontosuz Satış Tutarı*, *İskontolu Satış Tutarı* adlı öznitelikler haftalık satış adetine göre hesaplandığı için haftalık satış adetinin

doğrusal bir fonksiyonudur. Bu nedenle bunların satış adetinin tahmin edilmesi probleminde doğrudan kullanılmaları uygun değildir. Fakat bunları, ayrıntıları aşağıda anlatılan bazı dönüştürme işlemleriyle, birim ürün için dolaylı yoldan kullanmak mümkündür. Bu dönüştürme işlemleri haftalık satış adetinin 0 olmadığı durumlarda yapılabilir. Satış Adetinin 0 olduğu çok fazla sayıda kayıt olduğu için bu kayıtların silinmesi ciddi veri kaybı anlamına gelmektedir. Diğer yandan haftalık satış adetinin 0 olmadığı kayıtlardan dönüştürme yoluyla elde edilecek yeni öznitelikler de ciddi bilgi içermektedir. Bu nedenle bu iki durumu ayrı ayrı incelemek üzere iki ayrı veri seti analiz edilmiştir. Aşağıda adım adım veri setlerine uygulanan işlemler anlatılmaktadır.

- İlk veri seti *Satış Adeti* 0 olan kayıtlar silinerek oluşturulmuştur. Bu işlem sonucunda bu sette 982.994 kayıt kalmıştır ve *Veri Seti A* olarak adlandırılmıştır.
- İkinci veri seti ise tüm kayıtları içermektedir ve 1.953.948 kayıt içermektedir. Bu veri seti *Veri Seti B* olarak adlandırılmıştır.
- *Veri Seti B* için *Satış Adet* değerleri 0 olup, *İskontosuz Satış Tutarı* değeri 0 olmayan ya da *İskontolu Satış Tutarı* değeri 0 olmayan 1.133 kayıt silinmiştir.
- *Veri Seti B* için *Satış Adet* ve *İskontolu Satış Tutarı* değeri 0 olup ve *Brüt Kar* değeri 0 olmayan 3.586 kayıt silinmiştir.
- *Veri Seti A* ve *Veri Seti B* için *İskontolu Satış Tutarı* ve *İskontosuz Satış Tutarı* değerlerinden sadece birinin 0 olduğu 8.256 kayıt silinmiştir.
- *Veri Seti A* ve *Veri Seti B* için *İskontolu Satış Tutarı* ve *İskontosuz Satış Tutarı* değerlerinden sadece birinin negatif olduğu 83 kayıt silinmiştir.
- Temizlenen kayıtlar sonucunda *Veri Seti A* için 974.655 kayıt, *Veri Seti B* için ise 1.940.890 kayıt kalmıştır.
- *Veri Seti A* için dönüştürme işlemi ile aşağıdaki yeni öznitelikler elde edilmiştir:

$$\text{İ}ndirimli\ Birim\ Fiyat\ = \frac{\text{İ}skontosuz\ Satış\ Tutarı}{SA} \tag{3.1}$$

$$Birim\ İndirim\ Miktarı\ = \frac{İskontosuz\ Satış\ Tutarı\ - İskontolu\ Satış\ Tutarı}{SA} \quad (3.2)$$

- *Brüt Kar Marjı* tanım gereği;

$$Brüt\ Kar\ Marjı\ = \frac{Brüt\ Kar}{İskontolu\ Satış\ Tutarı / 1{,}08} \quad (3.3)$$

formülü ile hesaplanabilir. Burada 1,08 sabiti %100 oranına %8 KDV eklenmesiyle elde edilmiştir. *Veri Seti A* için bu formül kullanılarak hesaplanan *Brüt Kar Marjı* ile veri setinde önceden bulunan *Brüt Kar Marjı* değerleri karşılaştırılarak, ikisi arasındaki farkın 0.01'den küçük olduğu 33.197 kayıt silinmiştir.

- *Veri Seti A* için *Brüt Kar Marjı* değeri null olan 56.536 kayıt kullanılarak güncellenmiştir.
- *Veri Seti B* için hesaplanamayan *Mark-Up* ve *Brüt Kar Marjı* alanları veri setinden silinmiştir.
- *Veri Seti A* ve *Veri Seti B* için o haftadaki resmî tatiller incelenmiş ve bununla ilgili yeni bir öznitelik eklenmiştir. Böylece ilgili haftalardaki resmî tatil olan gün sayıları tespit edilerek daha doğru bir analiz yapılabilecektir.
- *Veri Seti A* için *İskontolu Satış Tutarı* ve *İskontosuz Satış Tutarı* kullanılarak *indirim oranı* hesaplanmıştır ve daha iyi analiz yapılabilmesi için yeni bir öznitelik olarak eklenmiştir.
- *Veri Seti A* için, indirimli birim fiyat, birim fiyat, birim indirim miktarı ve satılan malın birim maliyeti (*SMBM*) yeni öznitelikler olarak eklenmiştir.

Kırıklık Yüzdesi raflarda bulunmayan beden oranı hakkında bilgi içerdiği için satış tahmini probleminde önemli bir parametredir. Diğer yandan *Kırıklık Yüzdesi* değerinin NULL olduğu çok fazla sayıda kayıt vardır. *Kırıklık Yüzdesi* özniteliğinin silinmesinin ya da NULL olan kayıtların silinmesinin ayrı ayrı bilgi kayıplarına neden olması beklenir. Hangi seçenekteki bilgi kaybının fazla olacağı net olmadığı için bu iki durum ayrı ayrı analiz edilecektir. Böylece *Veri Seti A'da Kırıklık Yüzdesi* özniteliği çıkarılarak *Veri Seti A1* ve *Kırıklık Yüzdesi* değerinin NULL olduğu kayıtlar silinerek *Veri Seti A2* üretilmiştir. Benzer şekilde *Veri Seti B*

üzerinden Veri Seti B1 ve *Veri Seti B2* üretilmiştir. Bu aşamadan sonra aşağıdaki veri ön-işleme adımları izlenmiştir:

- *Veri Seti A, Veri Seti A1* ve *Veri Seti A2* setlerinden mağazaya göre gruplama yapılarak mağaza bazında kayıt sayısı 200'ün altında olan mağazalar silinmiştir. Bunun sonucunda *Veri Seti A* ve *Veri Seti A1* veri setlerinden 625, *Veri Seti A2* veri setinden 501 kayıt silinmiştir.
- *Veri Seti A, Veri Seti A1* ve *Veri Seti A2* setleri için *Renk Kod* özniteliğine göre gruplama yapılmış ve kayıt sayısı 50'den az olan renk kodları silinmiştir. Böylece *Veri Seti A* ve *Veri Seti A1* veri setlerinden 660, *Veri Seti A2* veri setinden 159 kayıt silinmiştir.
- *Veri Seti B, Veri Seti B1* ve *Veri Seti B2* setlerinden mağazaya göre gruplama yapılarak mağaza bazında kayıt sayısı 200'ün altında olan mağazalar silinmiştir. Bunun sonucunda *Veri Seti B* ve *Veri Seti B1* setlerinden 139, *Veri Seti B2* veri setinden 6 kayıt silinmiştir.
- *Veri Seti B* için *Ürün Klasman* özniteliğine göre gruplama yapılmış ve kayıt sayısı 50'den az olan kayıtlar silinmiştir. Bunun sonucunda *Veri Seti B* ve *Veri Seti B1* veri setlerinden 33 kayıt silinmiştir.
- *Veri Seti B, Veri Seti B1* ve *Veri Seti B2* setleri için *Renk Kod* özniteliğine göre gruplama yapılmış ve kayıt sayısı 50'den az olan renk kodları silinmiştir. Böylece *Veri Seti B* ve *Veri Seti B1* setlerinden 171, *Veri Seti B2* setinden 24 kayıt silinmiştir.

Tüm bu işlemler sonucunda *Veri Seti A* setinde 940.173, *Veri Seti A1* setinde 940.173, *Veri Seti A2* setinde 576.925, *Veri Seti B* setinde 1.940.547, *Veri Seti B1* setinde 1.940.547, *Veri Seti B2* setinde 609.716 adet kayıt kalmıştır.

3.2 Betimleyici İstatistik Analizi

Veri Seti A ve *Veri Seti B*'de bulunan sayısal öznitelikler için hesaplanan medyan, ortalama ve standart sapma değerleri aşağıdaki tablolarda verilmiştir.

Tablo 3.2 *Veri Seti A'da bulunan* sayısal öznitelikler için betimleyici istatistik değerleri

Öznitelik (*)	Ortalama	Standart Sapma	Medyan
SA	3,56	4,38	2,00
MU	1,90	0,44	1,98
BKM	0,39	0,19	0,46
KY	34,92	31,35	30,00
İBF	26,47	11,50	24,99
BF	30,67	11,49	24,99
BİM	4,20	6,04	0,00
SMBM	14,21	5,99	12,33
İO	0,14	0,18	0,00
HÖGS	0,72	1,48	0,00

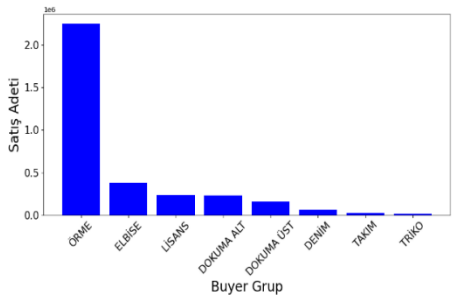
(*) SA: Satış Adeti, MU: Mark-Up, BKM: Brüt Kar Marjı, KY: Kırıklık Yüzdesi, İBF: İndirimli Birim Fiyat, BF: İndirimsiz Birim Fiyat, BİM: Birim İndirim Miktarı, SMBM: Satılan Malın Birim Maliyeti; İO: İndirim Oranı, HÖGS: Haftadaki Özel Gün Sayısı

Tablo 3.3 *Veri Seti B'*de *bulunan* sayısal öznitelikler için betimleyici istatistik değerleri

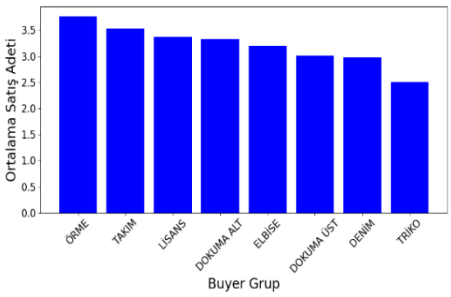
Öznitelik (*)	Ortalama	Standart Sapma	Medyan
SA	1,78	3,56	0,00
KY	35,26	31,42	30,00
HÖGS	0,72	1,47	0,00

(*) SA: Satış Adeti, KY: Kırıklık Yüzdesi, HÖGS: Haftadaki Özel Gün Sayısı

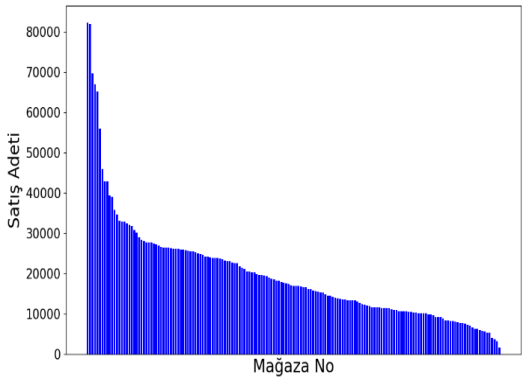
Veri setlerinde bulunan kategorik özniteliklerin her bir grubuna karşı düşen toplam satış adetleri ve ortalama haftalık satış adetleri aşağıdaki şekillerde verilmiştir. Gruplar arasındaki farkların çok olması ilgili özniteliklerin satış adeti tahmin probleminde anlamlı öznitelikler olabileceğini göstermektedir.



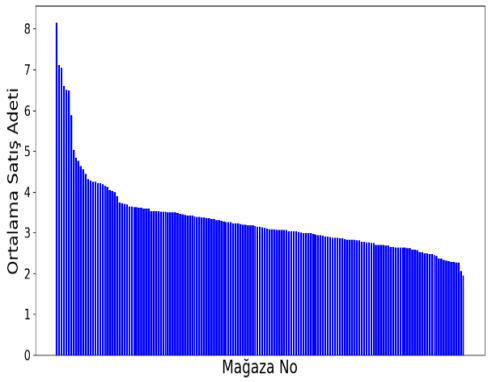
Şekil 3.1 Veri Seti A'da bulunan Buyer Grup özniteliğine ait ortalama satış adetleri



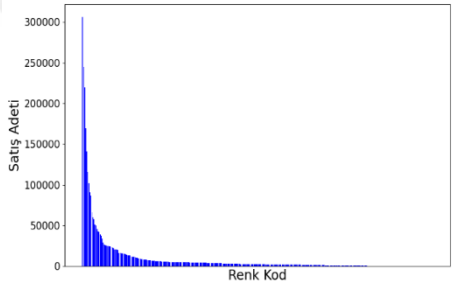
Şekil 3.2 Veri Seti A'da bulunan Buyer Grup özniteliğine ait histogram



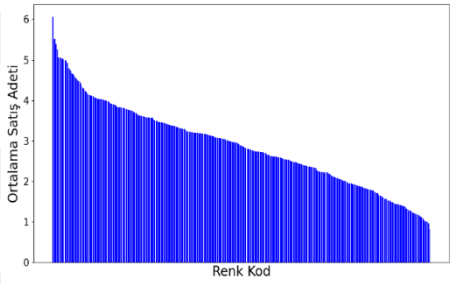
Şekil 3.3 *Veri Seti A'*da bulunan *Mağaza No* özniteliğine ait histogram



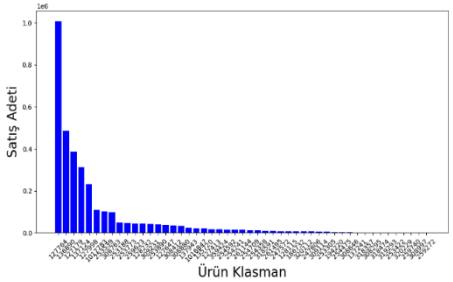
Şekil 3.4 *Veri Seti A'*da bulunan *Mağaza No* özniteliğine ait ortalama satış adetleri



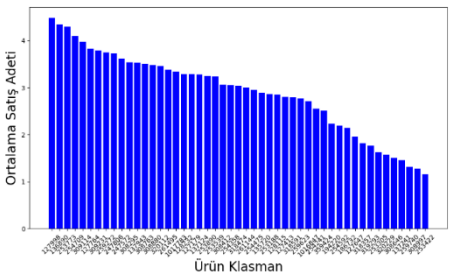
Şekil 3.5 *Veri Seti A'*da bulunan *Renk Kod* özniteliğine ait histogram



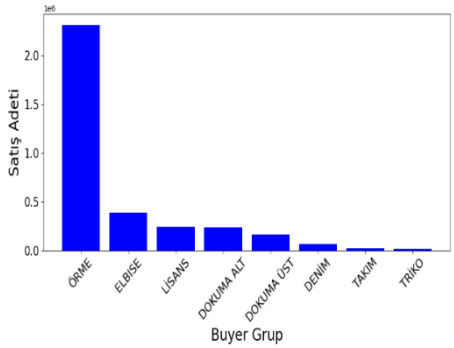
Şekil 3.6 *Veri Seti A'*da bulunan *Renk Kod* özniteliğine ait ortalama satış adetleri



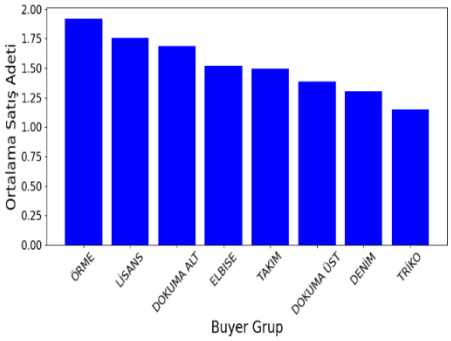
Şekil 3.7 *Veri Seti A'*da bulunan *Ürün Klasman* özniteliğine ait histogram



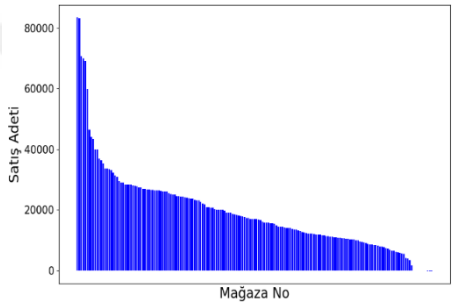
Şekil 3.8 *Veri Seti A'*da bulunan *Ürün Klasman* özniteliğine ait ortalama satış adetleri



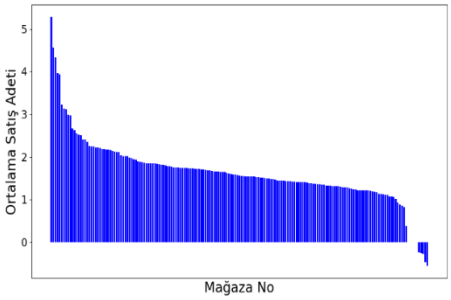
Şekil 3.9 *Veri Seti B'de* bulunan *Buyer Grup* özniteliğine ait histogram



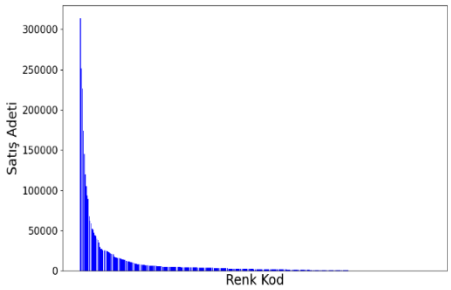
Şekil 3.10 *Veri Seti B'de* bulunan *Buyer Grup* özniteliğine ait ortalama satış adetleri



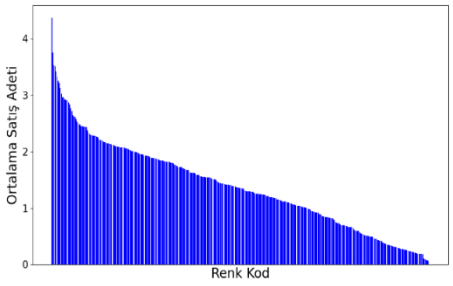
Şekil 3.11 Veri Seti B'de bulunan *Mağaza No* özniteliğine ait histogram



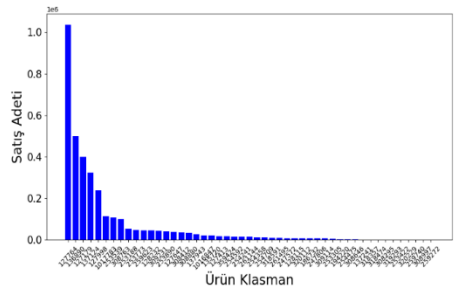
Şekil 3.12 *Veri Seti B*'de bulunan *Mağaza No* özniteliğine ait ortalama satış adetleri



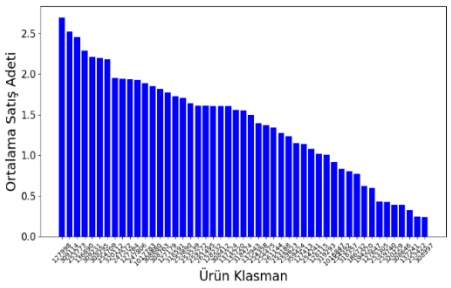
Şekil 3.13 Veri Seti B'de bulunan *Renk Kod* özniteliğine ait histogram



Şekil 3.14 *Veri Seti B*'de bulunan *Renk Kod* özniteliğine ait ortalama satış adetleri

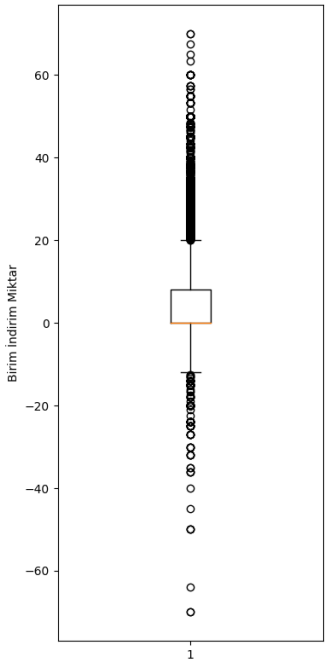


Şekil 3.15 Veri Seti B'de bulunan *Ürün Klasman* özniteliğine ait histogram

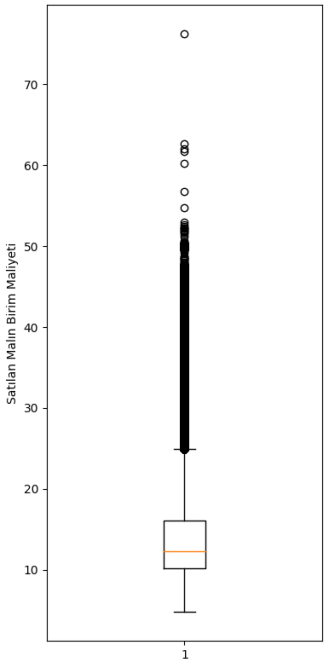


Şekil 3.16 *Veri Seti B'*de bulunan *Ürün Klasman* özniteliğine ait ortalama satış adetleri

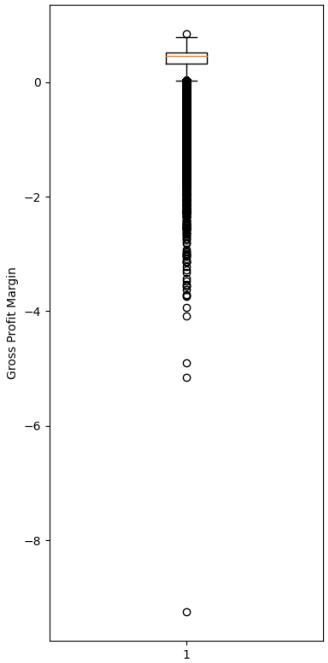
Betimleyici istatistik analizinin son kısmında veri setlerinde bulunan sayısal öznitelikler için kutu diyagramları çizdirilmiştir. Grafiklerde oluşan aykırı değerlerin satış verilerinin doğal bir sonucu olabileceği sonucuna varılmış ve dolayısıyla silinmemeleri gerektiği kanaatine varılmıştır.



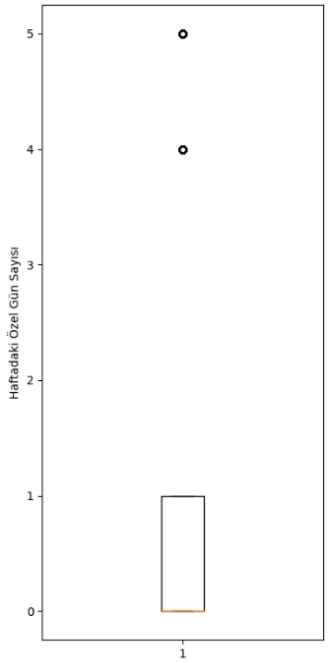
Şekil 3.17 *Veri Seti A*'de bulunan *Birim İndirim Miktarı* için kutu diyagramı



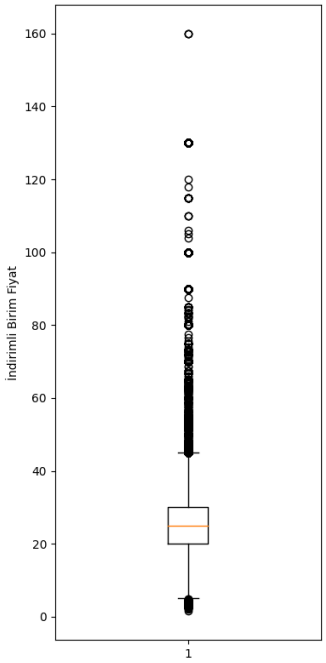
Şekil 3.18 *Veri Seti A*'de bulunan *Satılan Malın Birim Maliyeti* için kutu diyagramı



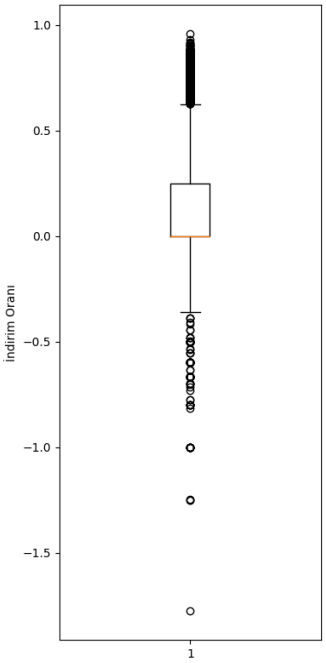
Şekil 3.19 *Veri Seti A*'de bulunan *Brüt Kar Marjı* için kutu diyagramı



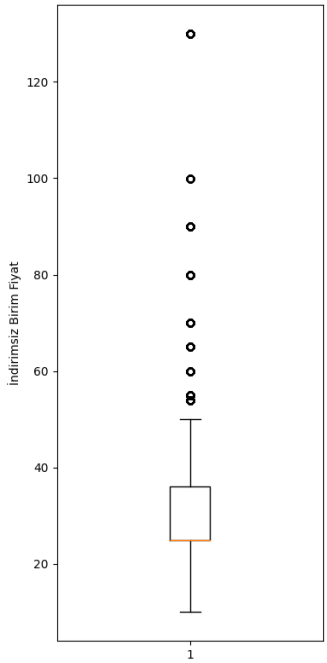
Şekil 3.20 *Veri Seti A*'de bulunan *Haftadaki Özel Gün Sayısı* için kutu diyagramı



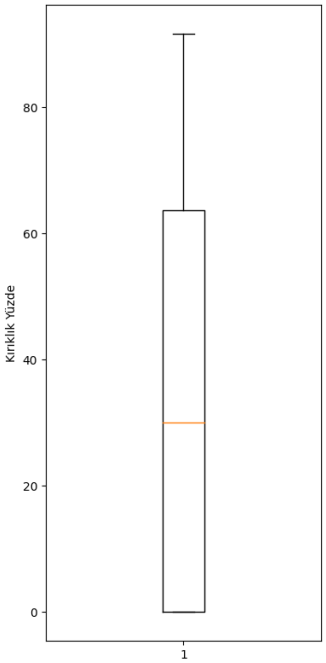
Şekil 3.21 *Veri Seti A*'de bulunan *İndirimli Birim Fiyat* için kutu diyagramı



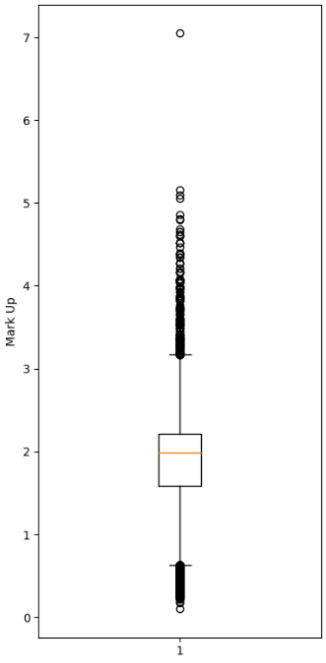
Şekil 3.22 *Veri Seti A*'de bulunan *İndirim Oranı* için kutu diyagramı



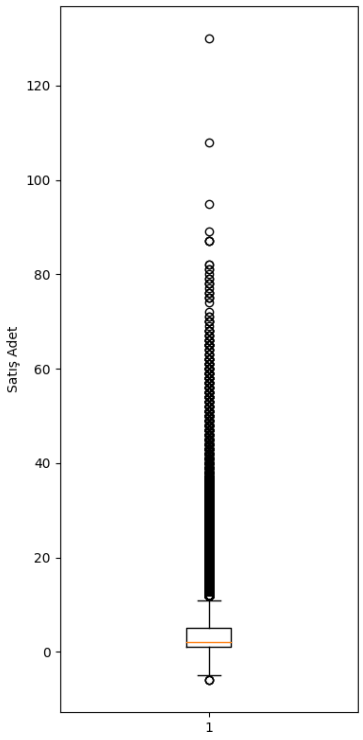
Şekil 3.23 Veri Seti A'de bulunan İndirimsiz Birim Fiyat için kutu diyagramı



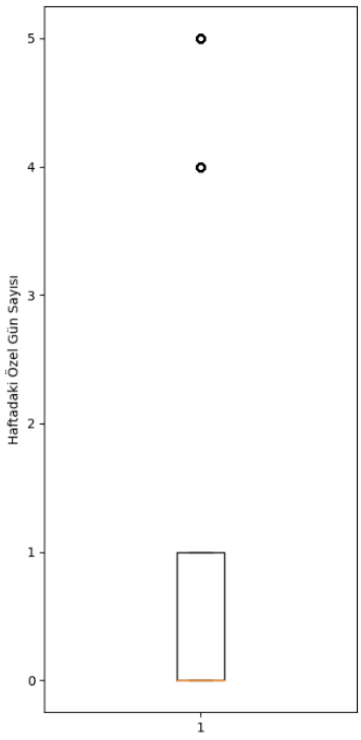
Şekil 3.24 Veri Seti A'de bulunan Kırıklık Yüzdesi için kutu diyagramı



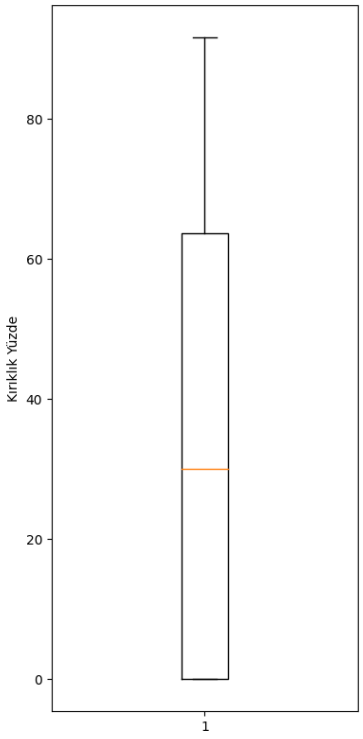
Şekil 3.25 Veri Seti A'de bulunan Mark-Up için kutu diyagramı



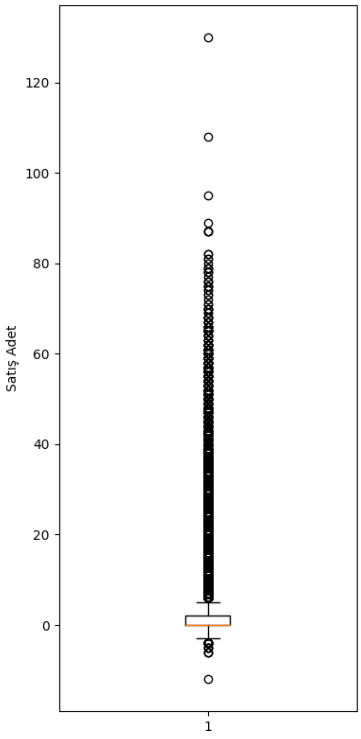
Şekil 3.26 *Veri Seti A'*de bulunan *Satış Adeti* için kutu diyagramı



Şekil 3.27 Veri Seti B'de bulunan Haftadaki Özel Gün Sayısı için kutu diyagramı



Şekil 3.28 *Veri Seti B'*de bulunan *Kırıklık Yüzdesi* için kutu diyagramı



Şekil 3.29 *Veri Seti B'*de bulunan *Satış Adeti* için kutu diyagramı

3.3 Çıkarımsal İstatistik Analizi

Çıkarımsal istatistik analizi iki aşamadan oluşmaktadır. Birinci aşamada sayısal özniteliklerin kendi aralarındaki korelasyon analizi yapılmıştır. İkinci aşamada ise kategorik değişkenlerdeki gruplar arasında satış adeti açısından istatistiksel farkların olup olmadığı test edilmiştir. 0,05'ten küçük p değerleri için sonuçlar anlamlı kabul edilmiştir. Çıkarımsal istatistik analizi için Matlab programı kullanılmıştır.

3.3.1 Korelasyon Analizi

Veri Seti A'da bulunan sayısal öznitelikler için Pearson korelasyon analizi yapılmıştır ve elde edilen regresyon katsayıları ve p değerleri sırasıyla Tablo 1 ve Tablo 2'de verilmiştir. *Kırıklık Yüzdesi* dışında *Satış Adeti* ile korelasyon halinde bulunan değişken saptanamamıştır. *Kırıklık Yüzdesi* ile *Satış Adeti* arasında anlamlı orta seviyede negatif korelasyon vardır ($r = -0{,}47, p = 0{,}00$). *Brüt Kar Marjı* ile *Mark-Up* arasında *İndirimsiz Birim Fiyat* ile *Satılan Malın Birim Maliyeti* arasında ve *Birim İndirim Miktarı* ile *İndirim Oranı* arasında anlamlı çok yüksek pozitif korelasyon vardır. Bunların dışında, Tablo 1'de mavi ile işaretli hücreler anlamlı güçlü korelasyonları, turuncu ile işaretli hücreler anlamlı zayıf korelasyonları göstermektedir.

Tablo 3.4: *Veri Seti A*'da bulunan sayısal öznitelikler için korelasyon katsayıları

(*)	SA	MU	BKM	KY	İBF	BF	BİM	SMBM	İO	HÖGS
SA	1,00									
MU	0,15	1,00								
BKM	0,13	0,92	1,00							
KY	-0,47	-0,24	-0,21	1,00						
İBF	0,02	0,33	0,35	-0,08	1,00					
BF	-0,07	-0,09	-0,08	0,06	0,86	1,00				
BİM	-0,16	-0,80	-0,81	0,25	-0,27	0,26	1,00			
SMBM	-0,06	-0,20	-0,16	0,05	0,83	0,95	0,23	1,00		
İO	-0,16	-0,86	-0,86	0,26	-0,46	0,02	0,91	-0,01	1,00	
HÖGS	-0,03	0,07	0,06	-0,01	0,03	-0,00	-0,07	-0,00	-0,08	1,00

(*) SA: Satış Adeti, MU: Mark-Up, BKM: Brüt Kar Marjı, KY: Kırıklık Yüzdesi, İBF: İndirimli Birim Fiyat, BF: İndirimsiz Birim Fiyat, BİM: Birim İndirim Miktarı, SMBM: Satılan Malın Birim Maliyeti; İO: İndirim Oranı, HÖGS: Haftadaki Özel Gün Sayısı

Tablo 3.5: *Veri Seti A*'da bulunan sayısal öznitelikler için *p* değerleri

	SA	MU	BKM	KY	İBF	BF	BİM	SMBM	İO	HÖGS
SA	1,00									
MU	0,00	1,00								
BKM	0,00	0,00	1,00							
KY	0,00	0,00	0,00	1,00						
İBF	0,00	0,00	0,00	0,00	1,00					
BF	0,00	0,00	0,00	0,00	0,00	1,00				
BİM	0,00	0,00	0,00	0,00	0,00	0,00	1,00			
SMBM	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00		
İO	0,00	0,00	0,00	0,00	0,00	0,006	0,00	0,00	1,00	
HÖGS	0,00	0,00	0,00	0,00	0,00	0,12	0,00	0,13	0,00	1,00

(*) SA: Satış Adeti, MU: Mark-Up, BKM: Brüt Kar Marjı, KY: Kırıklık Yüzdesi, İBF: İndirimli Birim Fiyat, BF: İndirimsiz Birim Fiyat, BİM: Birim İndirim Miktarı, SMBM: Satılan Malın Birim Maliyeti; İO: İndirim Oranı, HÖGS: Haftadaki Özel Gün Sayısı

Pearson korelasyon analizi *Veri Seti B*'de bulunan sayısal özniteliklere uygulandığında benzer sonuçlar elde edilmiştir (Tablo 3 ve Tablo 4'e bakınız).

Tablo 3.6 *Veri Seti B*'da bulunan sayısal öznitelikler için korelasyon katsayıları

	SA	KY	HÖGS
SA	1,00		
KY	-0,47	1,00	
HÖGS	-0,02	-0,01	1,00

(*) SA: Satış Adeti, KY: Kırıklık Yüzdesi, HÖGS: Haftadaki Özel Gün Sayısı

Tablo 3.7 *Veri Seti B*'da bulunan sayısal öznitelikler için *p* değerleri

	SA	KY	HÖGS
SA	1,00		
KY	0,00	1,00	
HÖGS	0,00	0,00	1,00

(*) SA: Satış Adeti, KY: Kırıklık Yüzdesi, HÖGS: Haftadaki Özel Gün Sayısı

3.3.2 Hipotez Testleri

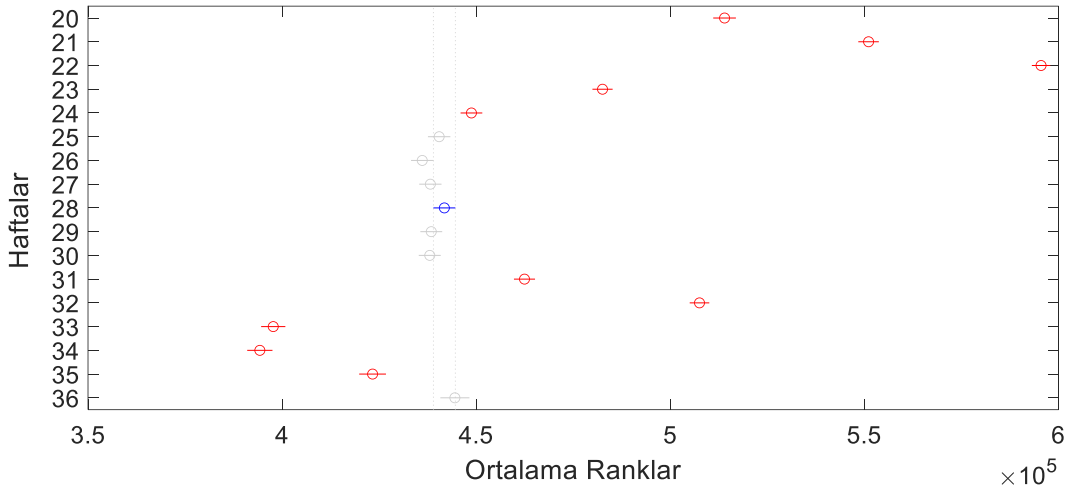
Veri Seti A ve *Veri Seti B*'de bulunan kategorik tipteki verilerin (*Hafta No, Mağaza No, Buyer Grup, Ürün Klasman, Renk Kod*) her grubuna karşılık düşen haftalık satış adetlerinin normal dağılıma uygunluğu Kolmogorov-Smirnov testi ile sınanmıştır. *Mağaza No* adlı öznitelikte çok fazla sayıda grup (2716) olduğu için

hipotez testi uygulanmamıştır. *Veri Seti A*'da bulunan *Ürün Klasman* adlı özniteliğin bir grubu dışında, diğer tüm gruplara karşılık düşen satış adetlerinin normal dağılmadığı tespit edilmiştir. Bu nedenle gruplar arasındaki istatistiksel farkların olup olmadığı parametrik olmayan Kruskal-Wallis Testi uygulanarak belirlenmiştir. Burada Mann-Whitney U testi en fazla iki grubun karşılaştırılmasında kullanıldığı ve veri setindeki öznitelikler ikiden fazla grup içerdiği için, Mann-Whitney U testi yerine onun genelleştirilmiş hali olan Kruskal-Wallis Testi kullanılmıştır.

Tablo 3.8 Kategorik öznitelikler için p değerleri

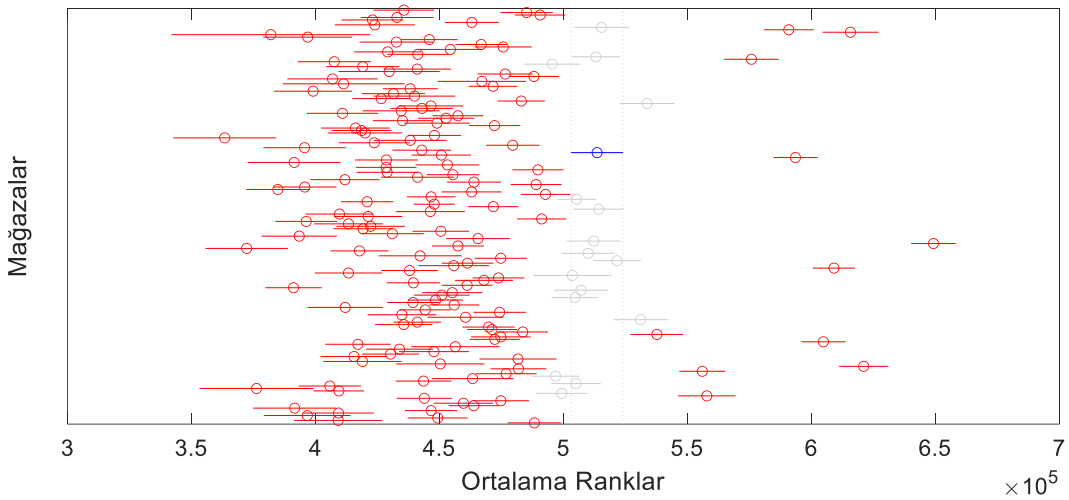
	Hafta No	***Mağaza No***	***Buyer Grup***	***Ürün Klasman***	***Renk Kod***
Veri Seti A	0,00	0,00	0,00	0,00	0,00
Veri Seti B	0,00	0,00	0,00	0,00	0,00

Kruskal-Wallis Testi sonucu elde edilen p değerleri Tablo 3.5'te verilmiştir. Tabloda görüldüğü üzere tüm p değerleri 0 olarak hesaplanmıştır. Dolayısıyla verilen özniteliklere ait her gruba karşılık düşen satış adetleri arasında istatistiksel olarak bir fark olmadığı hipotezi reddedilmektedir. Örneğin *Hafta No* adlı özniteliği düşünürsek; farklı haftalarda gerçekleşen (farklı gruplara karşılık düşen) satış adetleri arasında istatistiksel olarak fark olduğunu söyleyebiliriz. Hangi haftalara ait satış adetlerinin farklılaştığını daha ayrıntılı bir şekilde analiz etmek için her grubu ikili olarak karşılaştırmak gerekir. Bunu yapmak için Dunn-Sidak testi uygulandığında her bir öznitelik için elde edilen sonuçlar aşağıdaki şekillerde verilmiştir. Verilen şekillerde her grup bir yatay çizgi (aralık) ile temsil edilmektedir. Yatay çizgilerin yatay eksendeki izdüşümleri kesişiyorsa, bu çizgilere karşılık düşen gruplar arasında istatiksel açıdan önemli bir fark olmadığı söylenir. Eğer kesişme yoksa gruplar arasında önemli bir fark olduğunu söyleriz.

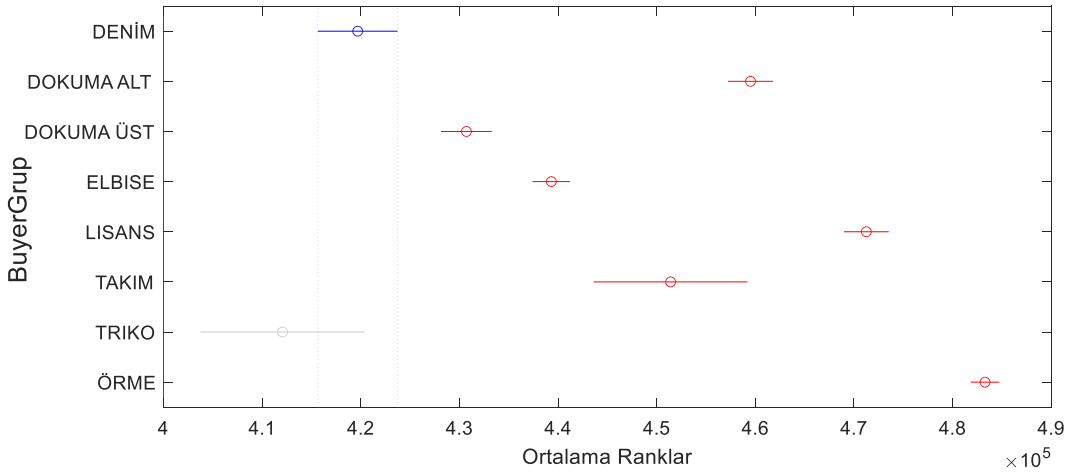


Şekil 3.30 Veri Seti A'da bulunan Hafta No adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması

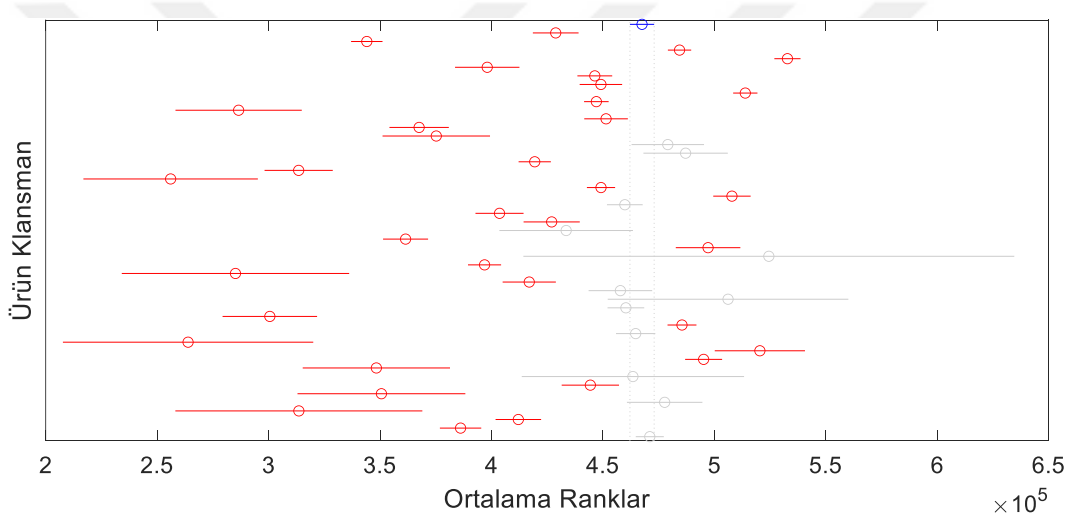
Şekil 3.30'da 28.haftaya karşılık düşen aralık mavi ile gösterilmiştir. Gri ile gösterilen aralıkların yataydaki izdüşümü mavi aralık ile kesiştiği için, gri aralıklara karşılık düşen haftalarla 28.hafta arasındaki satış adeti arasında önemli bir fark yoktur. Kırmızı aralıklara karşılık gelen haftalar ile 28.hafta arasında ise satış adetleri arasında önemli bir fark vardır. Diğer öznitelikler için benzer analizler aşağıdaki şekillere göre yapılabilir.



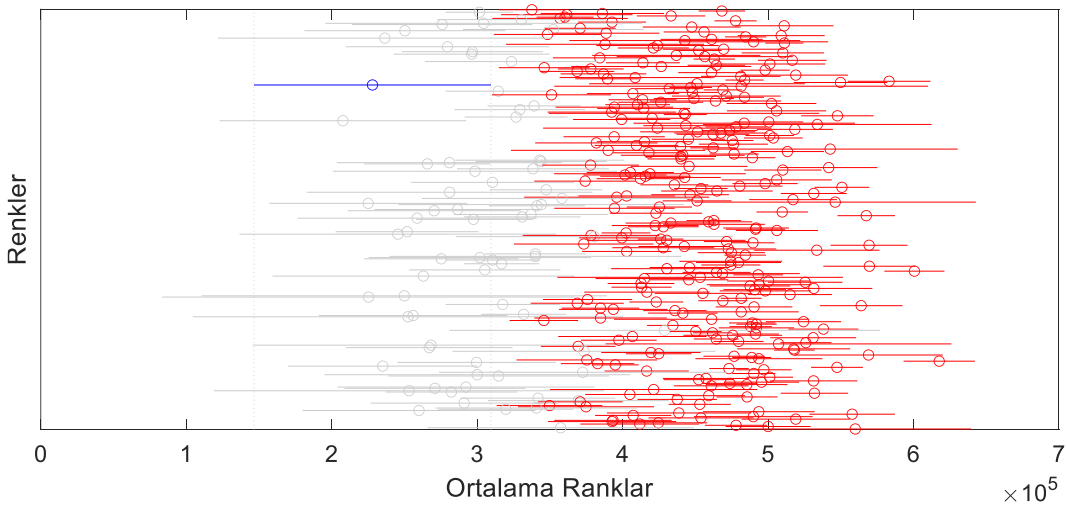
Şekil 3.31 Veri Seti A'da bulunan Mağaza No adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması



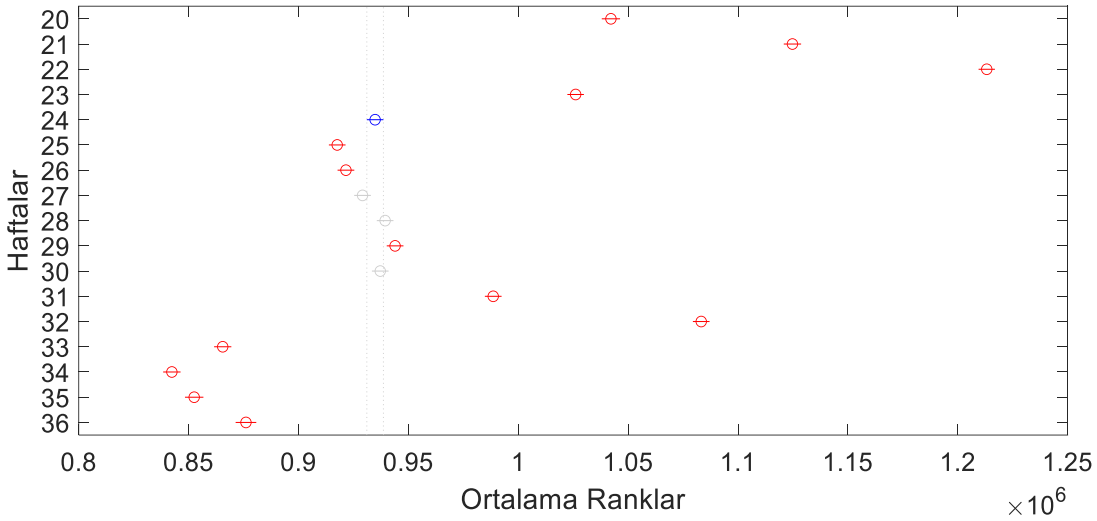
Şekil 3.32 *Veri Seti A*'da bulunan *Buyer Grup* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması



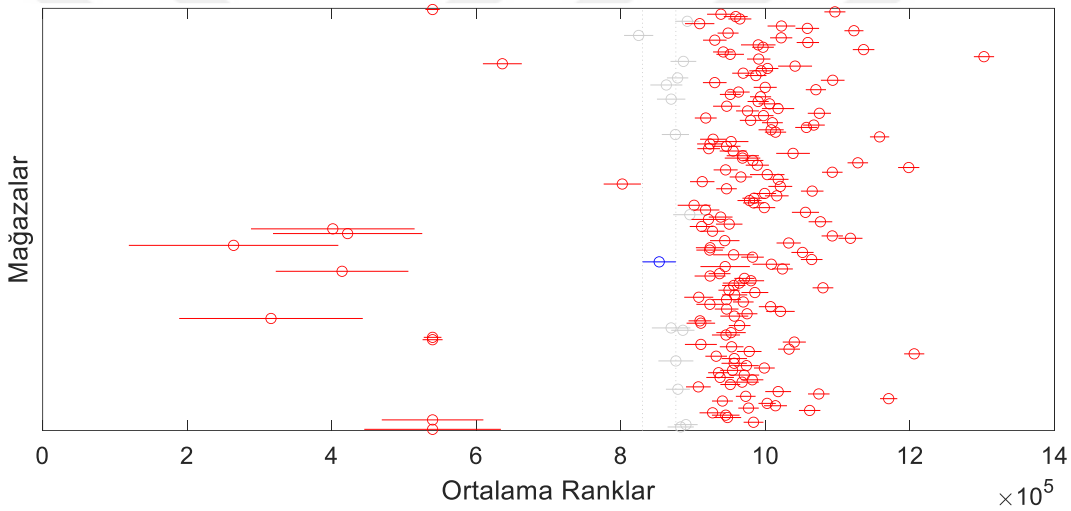
Şekil 3.33 *Veri Seti A*'da bulunan *Ürün Klasman* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması



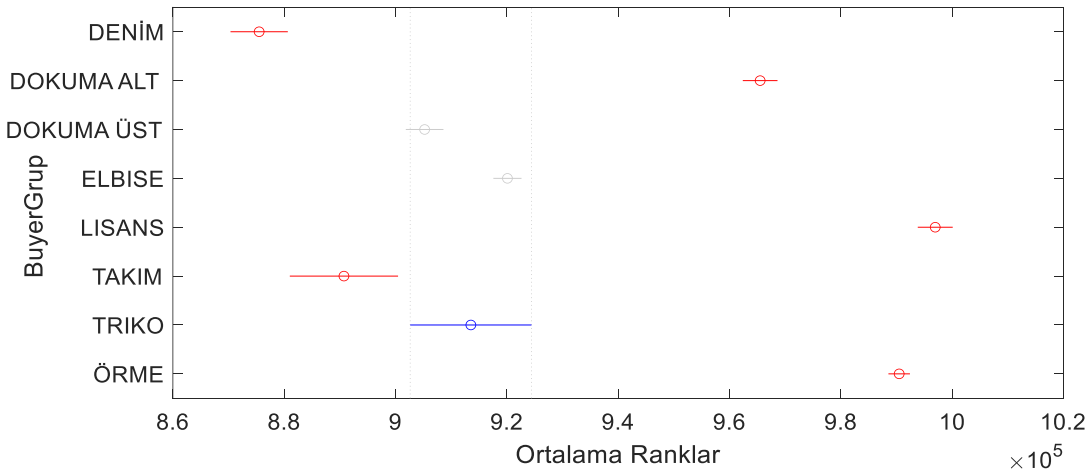
Şekil 3.34 *Veri Seti A*'da bulunan *Renk Kod* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması



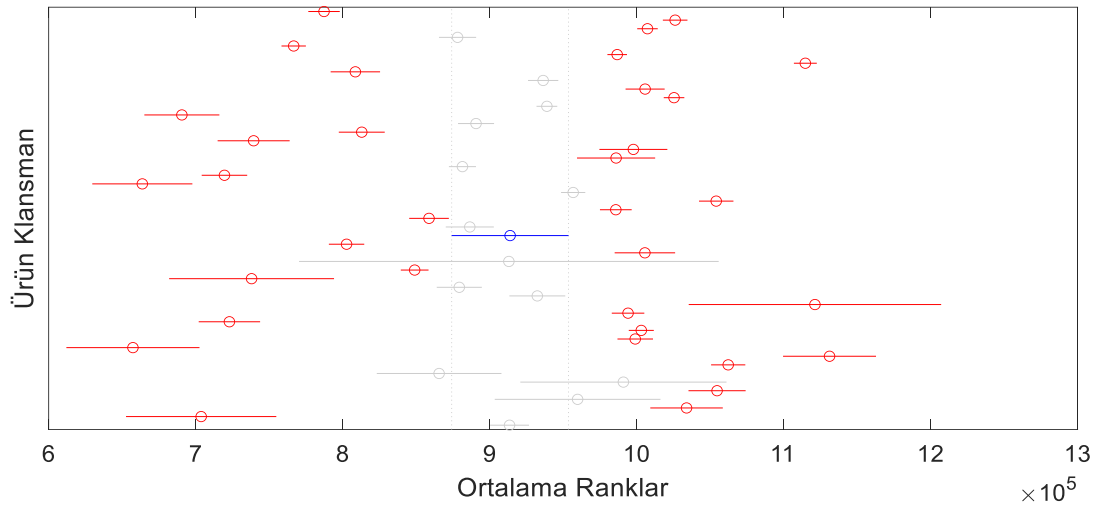
Şekil 3.35 *Veri Seti B*'de bulunan *Hafta No* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması



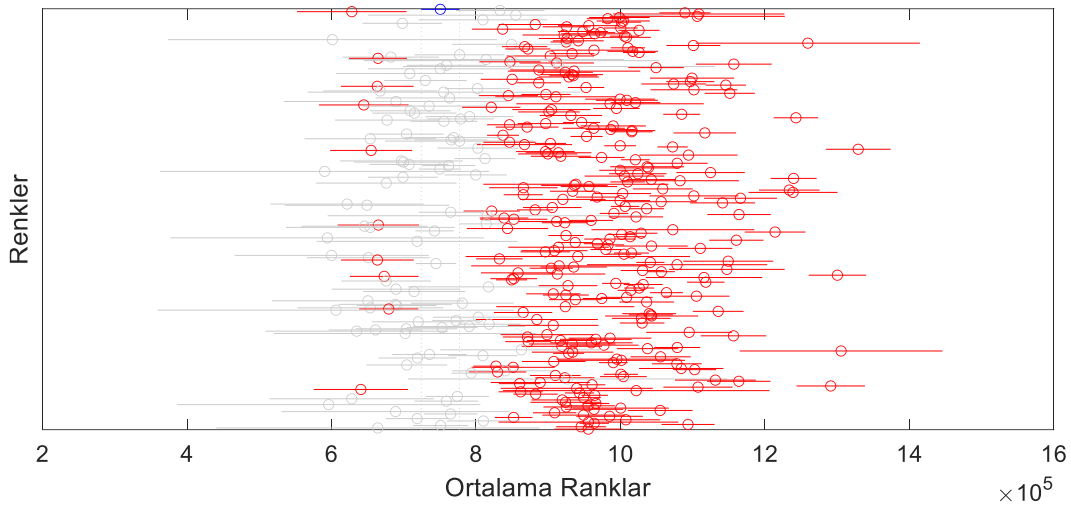
Şekil 3.36 *Veri Seti B*'de bulunan *Mağaza No* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması



Şekil 3.37 *Veri Seti B*'de bulunan *Buyer Grup* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması



Şekil 3.38 *Veri Seti B*'de bulunan *Ürün Klasman* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması



Şekil 3.39 *Veri Seti B*'de bulunan *Renk Kod* adlı özniteliğin gruplarının Dunn-Sidak testi ile karşılaştırılması

4
MAKİNE ÖĞRENMESİ MODELLERİNİN ELDE EDİLMESİ VE TAHMİN BAŞARILARININ TEST VERİLERİYLE SINANMASI

Makine öğrenmesi sürecinde kullanılan veri setleri, ön-işlenmiş ve istatistiksel olarak analiz edilmiş olan aşağıdaki 4 adet veri setinden oluşmaktadır:

- Veri Seti A1
- Veri Seti A2
- Veri Seti B1
- Veri Seti B2

Bu veri setlerinin her biri öncelikle eğitim ve test veri setleri olarak ikiye ayrılmıştır. Sonrasında ise eğitim veri seti üzerinden makine öğrenmesi yöntemleri kullanılarak modeller eğitilmiştir. Eğitim aşamasında 10 kat çapraz doğrulama (10-cross validation) tekniği kullanılmıştır. Son aşamada ise eğitilen modeller test veri seti kullanılarak test edilmiştir. Bu sürecin ayrıntıları (öznitelik seçimi, aykırı verilerin silinmesi, vb.) ve elde edilen sonuçlar aşağıda verilmektedir.

4.1 Eğitim ve Test Veri Setlerinin Oluşturulması

Veri setinde bulunan *Model Kod* adlı öznitelik her bir ürün için eşsiz bir koda sahiptir ve her bir değer, renk ve beden özellikleri hariç ayrı bir ürünü tanımlar. Örneğin, *Model Kod* değeri '375' olan ürün farklı renklere veya farklı bedenlere sahip olabilir ama bu özellikler dışında spesifik bir ürünü temsil eder. Bu nedenle yeni satışa sunulan bir ürünün mağazalardaki haftalık toplam satış adetini tahmin etmek için veri setini eğitim ve test seti olarak ikiye ayırma işlemi *Model Kod* üzerinden yapılmıştır. Böylece test veri setinde bulunan *Model Kod'lara ait kayıtlar* eğitim veri setine katılmamış olacaktır. Bu durumda test veri setindeki kayıtlar yeni ürünleri temsil ederken, eğitim veri setindeki kayıtlar ise eskiden satışa sunulmuş ürünleri temsil edecektir.

Diğer yandan veri setinde bulunan belli bir ürünün belli bir mağazadaki haftalık satış adetinin düşük sayıda olması ve görece fazla sapmaya sahip olması nedeniyle, öncelikle bir ürünün tüm mağazalardaki toplam haftalık satış adetinin tahmin

edilmesi amaçlanmıştır. Bu amaçla; *Model Kod, Renk Kod* ve *Hafta No* değerleri aynı olan kayıtlar farklı *Mağaza No* değerleri için birleştirilip tek bir kayıt şekline dönüştürülmüştür. Bu dönüştürme işleminde Mağaza No adlı öznitelik kaldırılıp Satış Adeti toplanırken, Mark-Up, Brüt Kar Marjı, Kırıklık Yüzdesi, İndirimli Birim Fiyat, İndirimsiz Birim Fiyat, Satılan Malın Birim Maliyeti, İndirim Oranı gibi özniteliklerin ortalaması alınmıştır. Belli bir Hafta No değeri için Haftadaki Özel Gün Sayısı değişmediğinden dolayı sabit tutulmuştur. Ayrıca belli bir haftada belli bir ürünün satışının yapıldığı mağaza sayısı veri setinden hesaplanabileceği için Mağaza Sayısı adlı bir öznitelik veri setlerine eklenmiştir. Son olarak bir ürünün ardışık haftalar boyunca satış adetinin değişimini analiz etme açısından Hafta No adlı öznitelik kategorik türden sayısal bir türe çevrilmiştir. Farklı veri setleri için kullanılan öznitelikler ve hedef değişkeni aşağıdaki tabloda verilmiştir.

Tablo 4.1 Veri Setlerinde Kullanılan Hedef Değişkeni ve Öznitelikler

Veri Seti	**Hedef Değişkeni**	***Öznitelikler***
Veri Seti A1	Satış Adeti	Hafta No, Buyer Grup, Ürün Klasman, Renk Kod, Mark-Up, Brüt Kar Marjı, İndirimli Birim Fiyat, Birim Fiyat, SMBM, Birim İndirim Miktarı, İndirim Oranı, Haftadaki Özel Gün Sayısı, Mağaza Sayısı
Veri Seti A2	Satış Adeti	Hafta No, Buyer Grup, Ürün Klasman, Renk Kod, Mark-Up, Brüt Kar Marjı, İndirimli Birim Fiyat, Birim Fiyat, SMBM, Birim İndirim Miktarı, İndirim Oranı, Haftadaki Özel Gün Sayısı, Mağaza Sayısı, Kırıklık Yüzdesi
Veri Seti B1	Satış Adeti	Hafta No, Buyer Grup, Ürün Klasman, Renk Kod, Haftadaki Özel Gün Sayısı, Mağaza Sayısı
Veri Seti B2	Satış Adeti	Hafta No, Buyer Grup, Ürün Klasman, Renk Kod, Haftadaki Özel Gün Sayısı, Mağaza Sayısı, Kırıklık Yüzdesi

Makine öğrenmesinde eğitim ve test için veri setleri %90 eğitim, %10 test verisi olarak ayrıştırılmıştır. %95 eğitim, %5 test ve %80 eğitim, %20 test olarak eğitimler de yapılmıştır ama en iyi sonucun, veri setlerinin %90 eğitim, %10 test olarak ayrıştırıldıktan sonra alındığı görülünce bu seçim üzerinden devam edilmiştir.

Eğitim veri setindeki herhangi sayısal bir özniteliğin ortalaması μ ve standart sapması σ olmak üzere, z dağılımına göre değerleri sırasıyla $[\mu - 2\sigma, \mu + 2\sigma]$, $[\mu - 3\sigma, \mu + 3\sigma]$ ve $[\mu - 4\sigma, \mu + 4\sigma]$ aralıklarının dışında kalan özniteliklere ait kayıtlar silinmiştir. Benzer şekilde, test veri setleri için de yukarıda verilen sınır değerlerini aşan özniteliklere ait kayıtlar silinip tahminlemeler yapılmıştır. En iyi sonucun $[\mu - 4\sigma, \mu + 4\sigma]$ aralığı için elde edildiği görüldüğünden dolayı aşağıdaki sonuç ve analizler sadece bu seçim üzerinden verilmiştir.

Bu işlemlere göre, her bir veri seti için kullanılan eğitim ve test veri setlerindeki kayıt sayıları Tablo 4.2'de verilmiştir.

Tablo 4.2 Makine Öğrenmesi Aşamasında Kullanılan Veri Setleri ve Kayıt Sayıları

		Kayıt Sayısı
Veri Seti A1	Eğitim Veri Seti	28.157
	Test Veri Seti	3133
Veri Seti A2	Eğitim Veri Seti	17.345
	Test Veri Seti	1905
Veri Seti B1	Eğitim Veri Seti	40651
	Test Veri Seti	4666
Veri Seti B2	Eğitim Veri Seti	17.438
	Test Veri Seti	2203

4.2 Makine Öğrenmesi Modellerinin Eğitilmesi

Mağazalardaki haftalık toplam satış adetinin tahmini için Matlab programında bulunan '*Regression Learner*' adlı araç kutusu kullanılmıştır.

Eğitim aşamasında k katlamalı çapraz doğrulama yönteminde $k = 5$, $k = 10$ ve $k = 20$ olacak şekilde makine öğrenmesi modelleri eğitilmiş olup, en iyi sonucu veren $k = 10$ parametresine göre sonuçlar aşağıda verilmiştir. Verilerin normalizasyonu ile elde edilen sonuçların normalize edilmeden elde edilen sonuçlara göre anlamlı bir başarı farkı olmadığı için veriler normalize edilmeden kullanılmıştır.

A1, A2, B1, B2 veri setleri için başarısının diğer yöntemlere göre daha fazla olduğu gözlemlenen beş farklı regresyon öğrenme yöntemi kullanılmıştır. Ağaç toplulukları (ensembles of trees) yöntemleri olarak; *'Bagged Trees' (*minimum yaprak sayısı: 8, öğrenici sayısı: 30*)* ve *'Boosted Trees' (*minimum yaprak sayısı: 8, öğrenici sayısı: 30*),* karar ağacı olarak *'Coarse Tree' (minimum yaprak sayısı: 36)* kullanılmıştır. Destek Vektör Makineleri (Support Vector Machines) yöntemi olarak *'Fine Gaussian SVM'* (Kernel scale: 0.97) kullanılmıştır. Sinir Ağları (Neural Network) yöntemi olarak ise *'Wide Neural Network'* (bağlı katman sayısı: 1, birinci katman boyutu: 100, iterasyon sınırı: 1000) yöntemi kullanılmıştır.

4.3 Eğitim ve Test Sonuçları

Bir önceki bölümde verilen veri setleri ve öğrenme yöntemleri kullanılarak çeşitli makine öğrenmesi modelleri eğitilmiş ve test edilmiştir. Yöntemlerin başarısı için ortalama mutlak hata (Mean Absolute Error – MAE) ve belirleme katsayısı (R^2) değerleri hesaplanmıştır. Buna göre eğitim ve test veri setleri için elde edilen sonuçlar aşağıdaki tablolarda verilmiştir.

Tablo 4.3 A1 Veri Seti için Test ve Eğitim Sonuçları

	MAE		R^2	
	Eğitim	***Test***	***Eğitim***	***Test***
Coarse Tree	28,77	30,81	0,83	0,8
Boosted Trees	26,68	**27,92**	0,85	**0,83**
Bagged Trees	25,04	**28,56**	0,87	**0,83**
Fine Gaussian SVM	45,87	63,54	0,55	0,33
Wide Neural Network	28,74	43,43	0,87	0,67

Tablo 4.4 A2 Veri Seti için Test ve Eğitim Sonuçları

	MAE		R^2	
	Eğitim	***Test***	***Eğitim***	***Test***
Coarse Tree	28,83	29,56	0,83	0,8
Boosted Trees	25,04	**25,22**	0,86	**0,86**
Bagged Trees	24,12	**27,02**	0,87	**0,84**
Fine Gaussian SVM	50,37	65,22	0,53	0,37
Wide Neural Network	25,72	44,99	0,89	0,65

Tablo 4.5 B1 Veri Seti için Test ve Eğitim Sonuçları

	MAE		R^2	
	Eğitim	***Test***	***Eğitim***	***Test***
Coarse Tree	31,69	45,17	0,65	0,52
Boosted Trees	32,04	**41,63**	0,66	**0,59**
Bagged Trees	27,33	**42,97**	0,72	**0,57**
Fine Gaussian SVM	39,65	57,61	0,34	0,13
Wide Neural Network	32,1	55,74	0,7	0,29

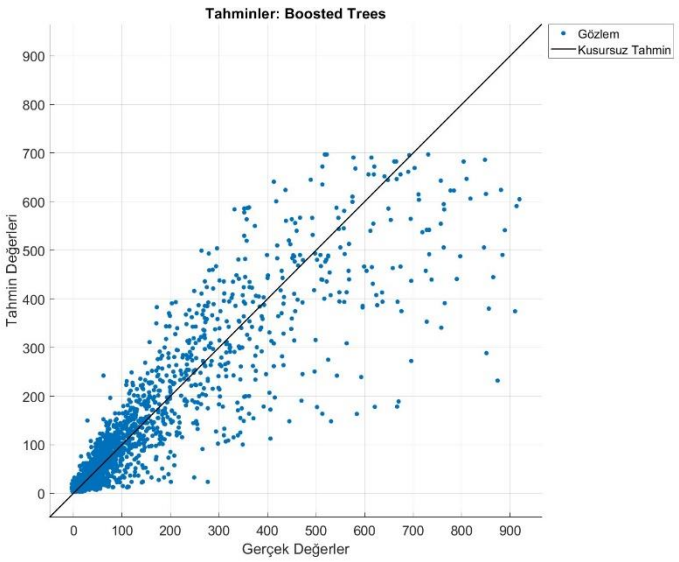
Tablo 4.6 B2 Veri Seti için Test ve Eğitim Sonuçları

	MAE		R^2	
	Eğitim	***Test***	***Eğitim***	***Test***
Coarse Tree	29,24	33,55	0,82	0,79
Boosted Trees	25,79	**28,52**	0,86	**0,85**
Bagged Trees	25,11	**30,31**	0,87	**0,83**
Fine Gaussian SVM	45,16	76,16	0,55	0,26
Wide Neural Network	26,95	52,62	0,87	0,47

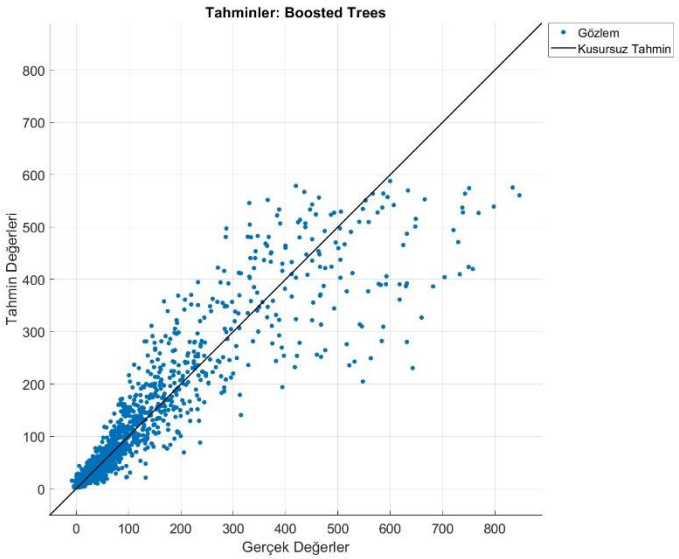
Ortalama mutlak hata verinin ortalama büyüklüğüne bağlı olup benzer dağılımların karşılaştırılması için uygun bir ölçüdür. R^2 katsayısı ise verinin ortalama büyüklüğünden fazla etkilenmez ve daha evrensel bir karşılaştırma niteliğine

sahiptir. Tahmin başarısı arttıkça R^2 katsayısı 1 değerine yaklaşır, başarı düştükçe bu katsayı 0 değerine yaklaşır.

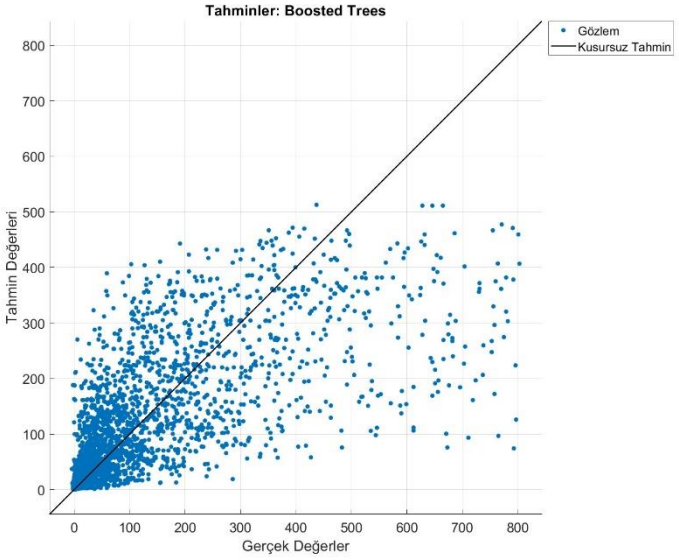
Yukarıda verilen tablolarda görüldüğü üzere; tüm veri setleri için test verisi üzerinden R^2 katsayısına göre en başarılı yöntemler sırasıyla *Boosted Trees*, *Bagged Trees* ve *Coarse Tree* olmuştur. *Boosted Trees* yöntemiyle eğitilen modeller için test veri setleri üzerinde yapılan tahminlemede elde edilen dağılımlar Şekil 4.1 - Şekil 4.4'de verilmiştir. Şekillerdeki siyah çapraz çizgi ideal çözüm bölgesini göstermektedir. Görüldüğü üzere R^2 katsayısı arttıkça mavi ile gösterilen tahmin sonuçlarının ideal çizginin etrafında toplanma oranı artmaktadır.



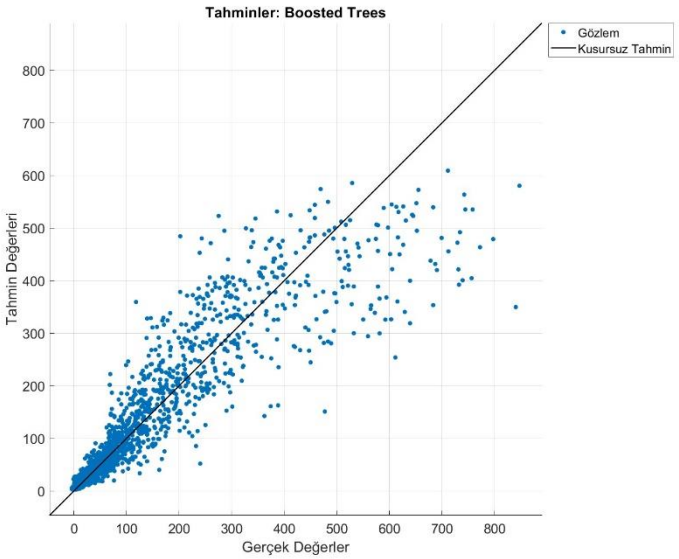
Şekil 4.1 *Veri Seti A1* için gerçek ve tahmin değer dağılımı



Şekil 4.2 *Veri Seti A2* için gerçek ve tahmin değer dağılımı

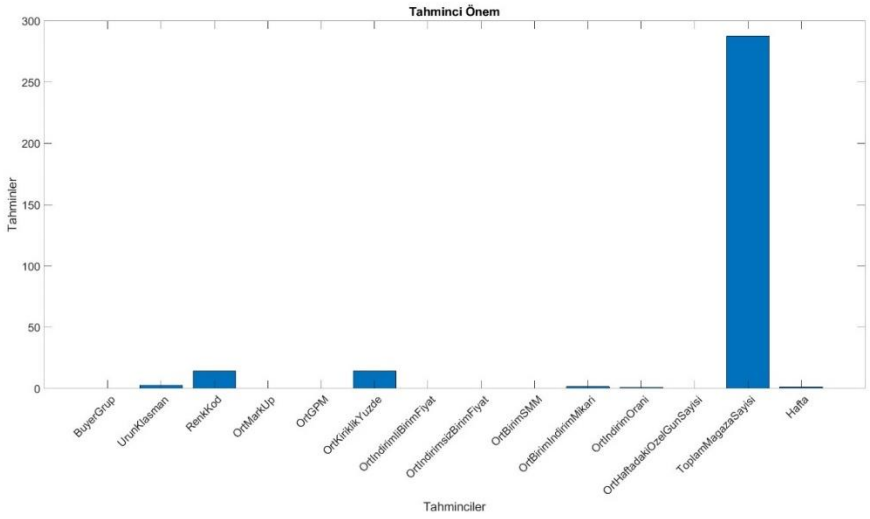


Şekil 4.3: *Veri Seti B1* için gerçek ve tahmin değer dağılımı



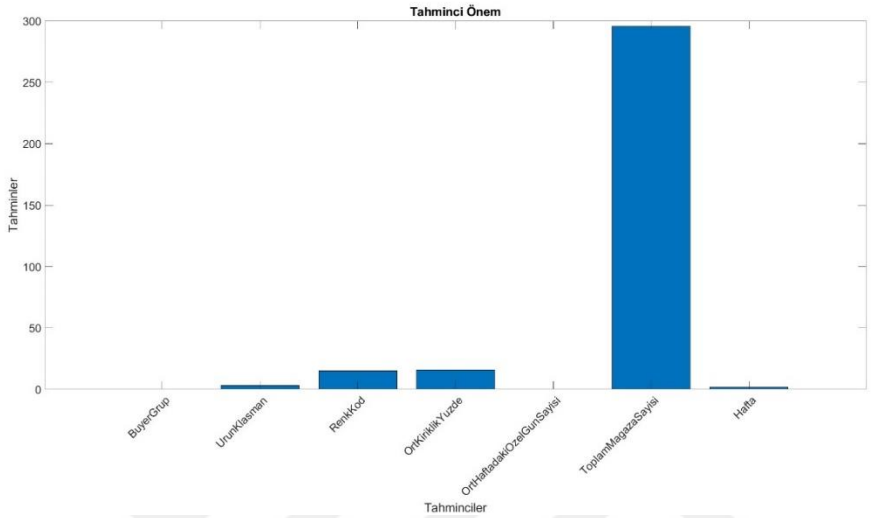
Şekil 4.4: *Veri Seti B2* için gerçek ve tahmin değer dağılımı

Tablo 4.3 – Tablo 4.6'da görüldüğü üzere en başarılı sonuçlar *A2* ve *B2* veri setleri için elde edilmiştir. Bu iki veri setinin özelliği, diğer veri setlerinde olmayan *Kırıklık Yüzdesi* özniteliğinin dahil edilmiş olmasıdır. Kalın olarak işaretlenmiş değerler en başarılı değerlerdir. En başarılı sonuçlar *A2* ve *B2* veri setleri için alındığı için alt analizler bu veri setleri üzerinden yapılacaktır. Bu iki veri seti için en iyi sonuçları veren *'Boosted Tree'* yöntemine ait aşağıda en önemli tahmincilerin analizi verilmiştir.



Şekil 4.5 *Veri Seti A2* için tahmincilerin önemi

Şekil 4.5'te görüldüğü üzere *A2 veri seti* için tahminleme yaparken *Mağaza Sayısı* özniteliği çok baskın olmuştur. Onun dışında *Kırıklık Yüzdesi, Renk Kod, Ürün Klasman* öznitelikleri de etkili olmuştur. Ancak diğer özniteliklerin etkilerinin minimal olduğu gözlemlenmiştir.



Şekil 4.6 *Veri Seti B2* için tahmincilerin önemi

Şekil 4.6'da görüldüğü üzere *B2* veri seti için tahminleme yaparken *Mağaza Sayısı* özniteliği çok baskın olmuştur. Onun dışında *Kırıklık Yüzdesi* ve *Renk Kod* öznitelikleri de etkili olmuştur. Ancak diğer özniteliklerin etkilerinin minimal olduğu gözlemlenmiştir.

En başarılı sonuçları aldığımız *A2* ve *B2* veri setleri için öznitelik seçimi (feature selection) uygulanmıştır. Öznitelik seçimi için MRMR algoritması uygulanacaktır. MRMR algoritmasından kısaca bahsetmek gerekirse; MRMR, Maximum Relevance Minimum Redundancy (Maksimum İlgililik Minimum Yedeklilik) anlamına gelir. MRMR sadece girdi olarak kullanılan özniteliklerin minimum korelasyon sağlamalarını dikkate almaz. Aynı zamanda çıktı değişkeniyle de özniteliklerin yüksek korelasyon sağladığından emin olur [19]. MRMR tekniği *Veri Seti A2* üzerine uygulandığında elde edilen özniteliklerin önem seviyeleri aşağıdaki tabloda verilmiştir.

Tablo 4.7 A2 Veri Seti için özniteliklerin önem seviyeleri

Öznitelikler	**MRMR**
Mağaza Sayısı	1,24
Buyer Grup	0,68
Hafta No	0,48
Kırıklık Yüzdesi	0,48
Brüt Kar Marjı	0,18
İndirimli Birim Fiyat	0,16
Renk Kod	0,16
Birim İndirim Miktarı	0,12
İndirim Oranı	0,10
Ürün Klasman	0,07
Mark-Up	0,06
İndirimsiz Birim Fiyat	0,03
Ortalama Haftadaki Özel Gün Sayısı	0,03
SMBM	0,02

En az etkili olan dört öznitelik, *Mark-Up, İndirimsiz Birim Fiyat, Haftadaki Özel Gün Sayısı* ve *SMBM* çıkartılarak tekrar makine öğrenmesi modelleri eğitilmiştir. Ancak sonuçlar etkisiz özniteliklerin çıkarılmadığı hali ile neredeyse birebir aynı çıkmıştır (bak Tablo 4.8).

Tablo 4.8 *A2 Veri Seti* için Öznitelik Seçimi Sonrası Test ve Eğitim Sonuçları

	MAE		R^2	
	Eğitim	***Test***	***Eğitim***	***Test***
Coarse Tree	28,8	29,55	0,83	0,80
Boosted Trees	25,10	**25,21**	0,86	**0,86**
Bagged Trees	24,14	**27,6**	0,87	**0,83**

Benzer şekilde *B2 Veri Seti* için MRMR algoritmasıyla özniteliklerin önemi seviyeleri elde edilmiştir (bak Tablo 4.9). En az etkili olan öznitelik, *Haftadaki Özel*

Gün Sayısı çıkartılarak tekrar öğrenme yöntemleri uygulanmıştır. Ancak sonuçlar etkisiz özniteliğin çıkarılmadığı hali ile neredeyse birebir aynı çıkmıştır. Sonuçlar Tablo 4.10'da verilmiştir.

Tablo 4.9 B2 Veri Seti için özniteliklerin önem seviyeleri

Öznitelikler	**MRMR**
Mağaza Sayısı	1,24
Buyer Grup	0,49
Hafta No	0,34
Kırıklık Yüzdesi	0,35
Renk Kod	0,31
Ürün Klasman	0,12
Haftadaki Özel Gün Sayısı	0,04

Tablo 4.10 *B2 Veri Seti* için Öznitelik Seçimi Sonrası Test ve Eğitim Sonuçları

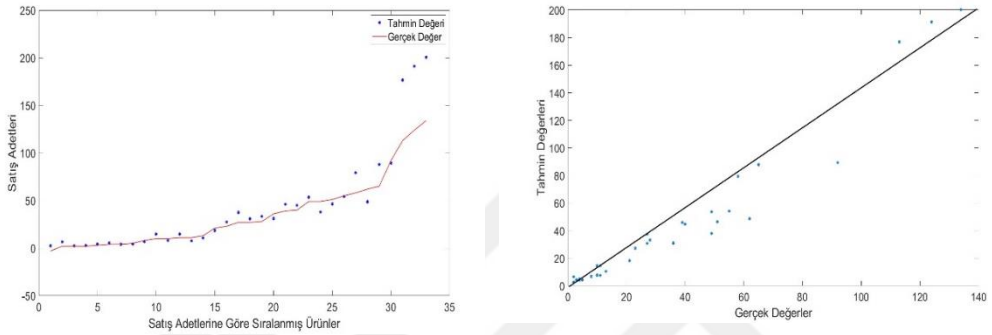
	MAE		R^2	
	Eğitim	***Test***	***Eğitim***	***Test***
Coarse Tree	29,2	33,55	0,82	0,79
Boosted Trees	25,85	**28,52**	0,86	**0,85**
Bagged Trees	25,18	**30,58**	0,86	**0,83**

A2 ve *B2* veri setleri için *Buyer Grup* özniteliği üzerinden alt analizler yapılmıştır. Buna göre test verisinde bulunan *Buyer Grup* özniteliğindeki her bir değerin ayrı ayrı başarıları hesaplanmıştır. *A2* veri seti için elde edilen sonuçlar Tablo 4.11'de verilmiştir. *Buyer Grup* öznitelik değeri *'TRİKO'* olan verilerin sayısı çok az olduğu için bu değere ait alt analiz yapılmamıştır.

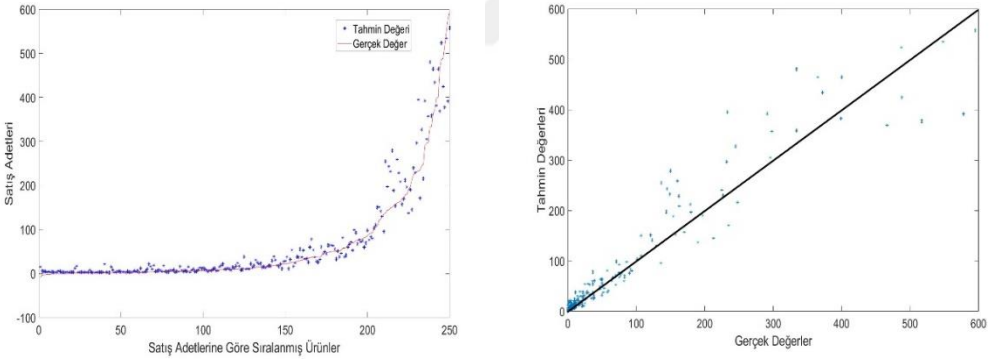
Tablo 4.11 *A2 Veri Seti* için Buyer Grup özniteliğinin değerlerine göre tahmin ölçümleri

	Tüm Veri	Örme	Takım	Dokuma Üst	Dokuma Alt	Elbise	Lisans
MAE	25,22	30,46	10,73	16,00	15,80	24,80	20,46
R^2	0,86	0,84	0,66	0,91	0,92	0,85	0,83

Yukarıdaki tabloda görüldüğü gibi en kötü sonuç *'takım'* verisi ile alınırken en iyi sonuç *'Dokuma Alt'* verisi ile alınmıştır. *'Takım'* verisi ve *'Dokuma Alt'* verisi için elde edilen tahminler sırasıyla Şekil 4.7 ve Şekil 4.8'de gösterilmektedir. Solda verilen grafiklerde kırmızı eğri küçükten büyüğe sıralanmış gerçek satış adetlerini gösterirken, mavi yıldızlar karşılık düşen tahmin değerlerini göstermektedir. Soldaki grafikler ise mavi yıldızlarla belirtilen tahmin değerlerinin siyah çapraz çizgi ile gösterilen ideal sonuçların etrafında toplanmasını göstermektedir.



Şekil 4.7 *Veri Seti A2* için *'Takım'* değeri sonuçları

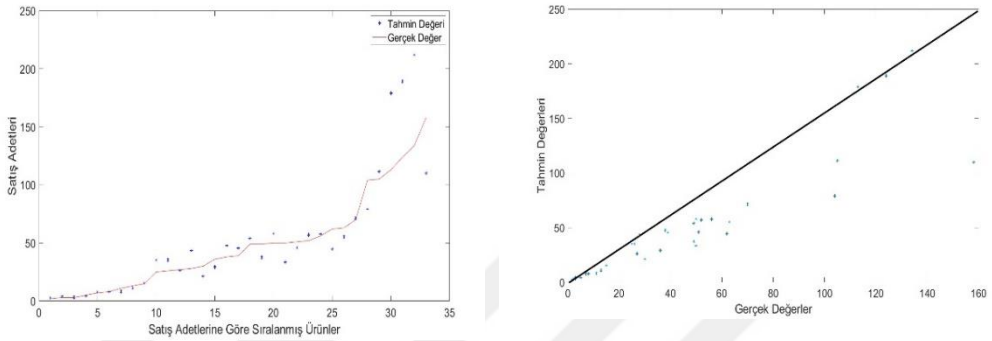


Şekil 4.8 *Veri Seti A2* için *'Dokuma Alt'* değeri sonuçları

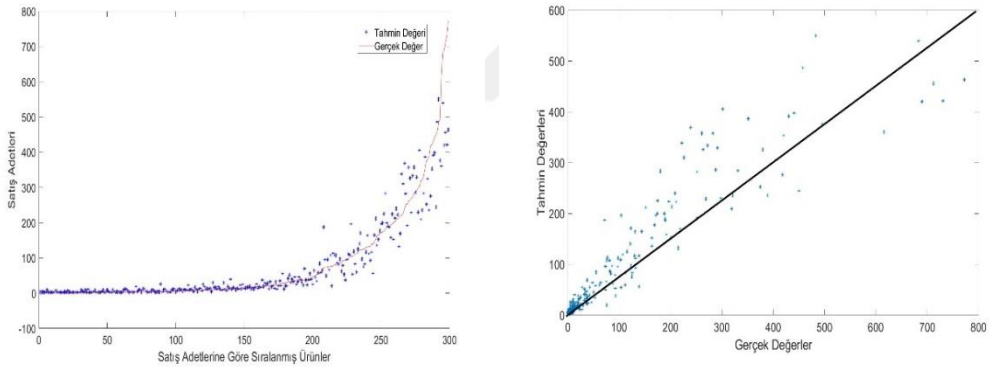
Buyer Grup özniteliğindeki her bir değer için *B2* veri setinden elde edilen sonuçlar Tablo 4.12'de verilmiştir. Benzer şekilde en kötü sonuç *'takım'* verisi ile alınırken en iyi sonuç *'Dokuma Alt'* verisi ile alınmıştır. *'Takım'* verisi ve *'Dokuma Alt'* verisi için elde edilen tahminler sırasıyla Şekil 4.9 ve Şekil 4.10'da gösterilmektedir.

Tablo 4.12 *B2 Veri Seti* için Buyer Grup özniteliğinin değerlerine göre tahmin ölçümleri

	Tüm Veri	Örme	Takım	Dokuma Üst	Dokuma Alt	Elbise	Lisans
MAE	28,52	30,54	13,44	19,74	22,58	31,25	30,40
R^2	0,85	0,85	0,65	0,86	0,87	0,84	0,83

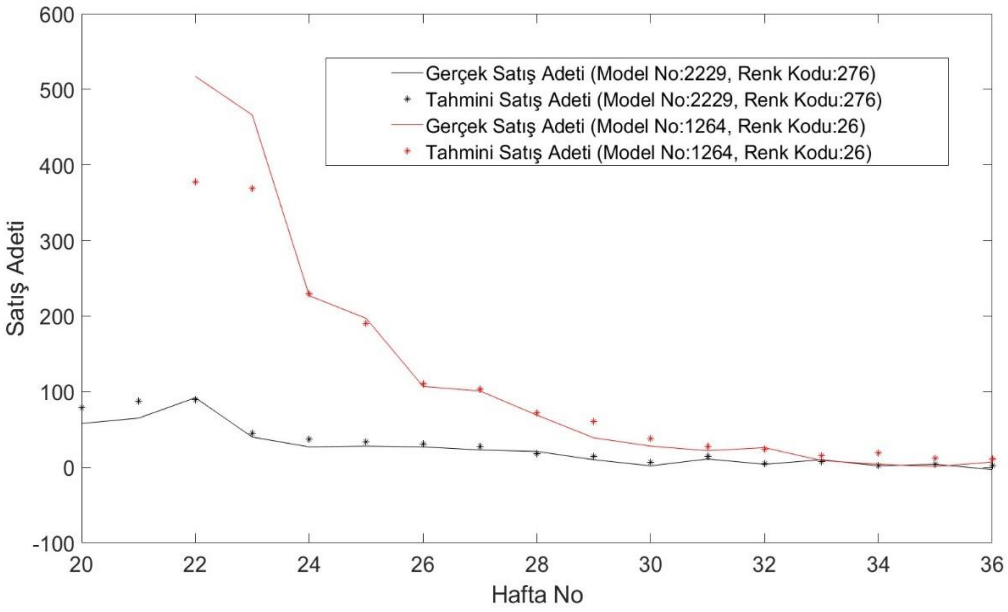


Şekil 4.9 *Veri Seti B2* için *'Takım'* değeri sonuçları



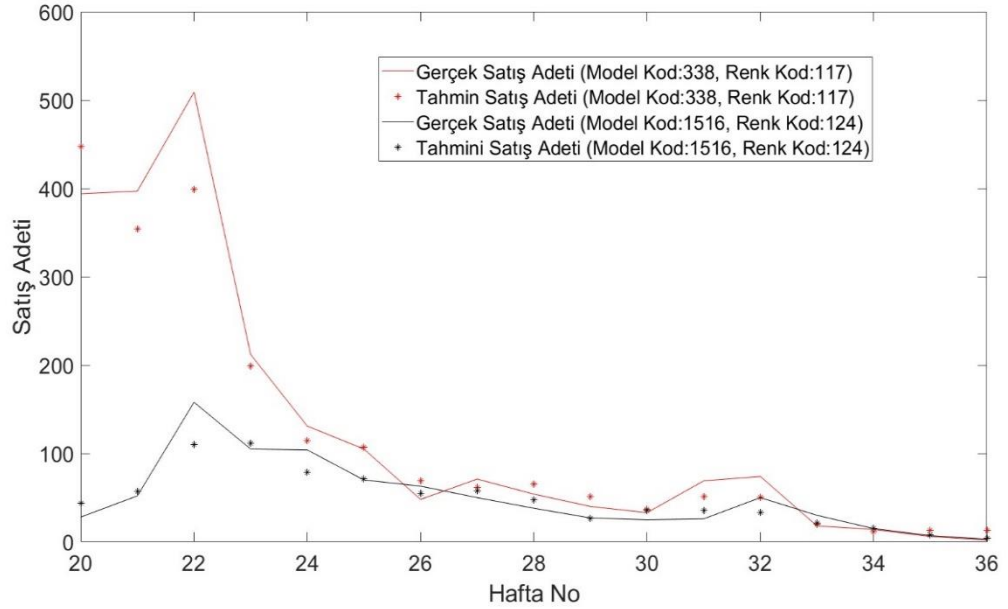
Şekil 4.10 *Veri Seti B2* için *'Dokuma Alt'* değeri sonuçları

A2 veri seti üzerinde *'Boosted Tree'* yöntemiyle eğitilen model kullanılarak, *Model No* değeri 2229 ve *Renk Kod* değeri 276 olan ürünle ve *Model No* değeri 1264 ve *Renk Kod* değeri 26 olan ürünün tüm haftalar boyunca satış tahminleri incelenmiş ve aşağıdaki grafikler elde edilmiştir. Eğriler haftalara göre gerçek haftalık satış adetlerini gösterirken yıldızlar tahmin değerlerini göstermektedir.



Şekil 4.11 Veri Seti A2 için belirli ürünleri gerçek ve tahmin edilmiş haftalık satış verileri

Benzer bir analiz *B2* veri seti üzerinde *Model No* değeri 338 ve *Renk Kod* değeri 117 olan ürünle ve *Model No* değeri 1516 ve *Renk Kod* değeri 124 olan ürün için yapılmıştır. Elde edilen sonuçlar aşağıdaki şekilde verilmiştir.



Şekil 4.12 Veri Seti *B2* için belirli ürünleri gerçek ve tahmin edilmiş haftalık satış verileri

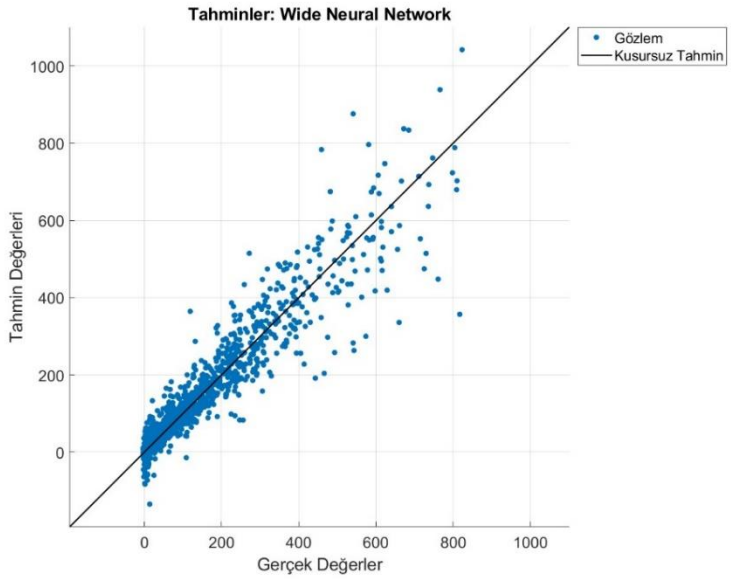
Daha önce de ifade edildiği üzere; *Model No* özniteliği belli bir ürünü tanımlayan eşsiz bir değere sahiptir ve yeni satışa sunulan ürünlerin satış tahminini yapmayı

amaçladığımız için eğitim ve test sürecinde öznitelik olarak kullanılmamıştır. Spesifik bir ürünü temsil eden bu özniteliğin kullanılmamasının tahmin başarısını düşürmesi beklenir. Bu *Model No'*yu öznitelik olarak eğitim modelinde kullanmak yeni satışa sunulan değil, eskiden beri satışta olan bir ürünün satış adetini tahmin etmek anlamına gelir. Bu da görece daha kolay bir problemdir. Bu öngörünün doğrulanması için *Model No* özniteliği dahil edilerek A2 veri seti için eğitim ve test süreci tekrarlanmıştır. Sonuçlar aşağıdaki tabloda verilmiştir.

Tablo 4.13 *A2* Veri Seti için *Model No* Özniteliği Eklendiğinde Test ve Eğitim Sonuçları

	MAE		R^2	
	Eğitim	***Test***	***Eğitim***	***Test***
Coarse Tree	28,08	29,64	0,83	0,83
Boosted Trees	22,86	23,75	0,88	0,88
Bagged Trees	24,28	25,02	0,87	0,87
Wide Neural Network	22,15	**23,54**	0,90	**0,91**

Tablo 4.13'de görüldüğü gibi *Model No'*yu özniteliği dahil edildiğinde başarı oranı artmıştır. Daha önceki analizlerde R^2 sonucu en iyi 0,86 olarak hesaplanmışken, *Model No* özniteliği hesaplandığında Wide Neural Network yönteminde başarı 0,90 olacak şekilde artmıştır. Bu sonuca göre; "%4'lük kaybı yeni model tahmin etmek istediğimiz için yaşamaktayız" denebilir. Wide Neural Network yöntemine göre elde edilen tahmin sonuçlarının grafiği aşağıda verilmiştir.



Şekil 4.13 Veri Seti A2 için Model Kod eklendiğinde gerçek ve tahmin değer dağılımı

5
SONUÇ

Perakende sektöründe, özellikle rakip firmaların fazla olduğu bir ortamda tahminleme büyük önem arz etmektedir. Stok miktarı, stok maliyeti, lojistik gibi bir firmanın ekonomik gelir-gider kalemlerini ciddi şekilde etkileyecek belirlemeleri, günümüzde büyüyerek çoğalan yapay zekâ gücünü kullanmadan yapmak oldukça zordur. Hatta mağazalardaki depoların büyüklüğü, çalışan sayıları gibi parametreler de yapay zekanın gücü kullanılarak yapılan tahminlemelerden etkilenecektir.

Bu çalışmada bu amaçlara hizmet etmek adına, hazır giyim perakende sektöründe büyük bir hacme sahip olan bir firmanın 2019 Yaz mevsimine ait verileri kullanılmıştır. Haftalık bazda satış verileri ile geleceğe yönelik bir tahminleme yapılmıştır.

Tahmin değeri olarak *'Satış Adet'* özniteliği belirlendiği için veri seti ikiye bölünmüştür. *Veri Seti A*'da *'Satış Adet'* değeri 0 olan veriler silinmiştir. Bunun amacı *'Satış Adet'* değeri 0 iken hesaplanamayan özniteliklerin *Veri Seti A'*ya eklenmesidir. Bu sayede *Veri Seti A*'ya *'Birim İndirim Miktarı', 'İndirimli Birim Fiyat', 'İndirimsiz Birim Fiyat', 'Satılan Malın Birim Maliyeti'* gibi yeni öznitelikler dahil edilmiştir. Böylece tahminleme yapılırken başarının artması hedeflenmiştir.

'Kırıklık Yüzdesi' özniteliği raflarda bulunmayan beden oranı hakkında bilgi içerdiği için satış tahmini probleminde önemli bir parametredir. *'Kırıklık Yüzdesi'* parametresinin NULL olduğu çok fazla kayıt tespit edilmiştir. Bu yüzden veri setleri tekrar ikiye bölünerek toplamda dört veri seti olması sağlanmıştır. Böylece *Veri Seti A'da* Kırıklık Yüzdesi özniteliği çıkarılarak *Veri Seti A1* ve Kırıklık Yüzdesi değerinin NULL olduğu kayıtlar silinerek *Veri Seti A2* üretilmiştir. Benzer şekilde *Veri Seti B* üzerinden *Veri Seti B1* ve *Veri Seti B2* üretilmiştir.

Korelasyon analizinde görülmüştür ki; satış adetiyle en yüksek korelasyona sahip öznitelik *"Kırıklık Yüzdesi"* adlı özniteliktir. Bu da bir bedenin bulunurluğu ne kadar yüksekse satılma ihtimalinin de o derece yüksek olacağını göstermiştir.

Matlab programı ile regresyon öğrenme yöntemleri kullanılarak eğitimler ve testler yapılmıştır. Buna göre en iyi sonuçlar *Veri Seti A2 ve Veri Seti B2* veri setleriyle

yapılan eğitimler ve testler ile elde edilmiştir. *Veri Seti A2* için elde edilen tahminleme sonuçlarında R^2 katsayısı %86'ya kadar çıkarken, *Veri Seti B2* için bu oran %85'e kadar çıkmıştır.

Bu çalışmada piyasaya çıkacak yeni ürünlerin tahmini üzerinden ilerlenmiştir. Yani *'Model Kod'* özniteliği değerleri eğitim ve test aşamasında birbirinden ayrıştırılmıştır. Eğer bu ayrıştırılma yapılmasaydı, yani önceden satışa sunulan bir ürünün satış tahmini yapılmak istenseydi, eğitim ve test aşamasında R^2 katsayısı yüzde 91'e kadar çıkacaktı.

Bu çalışmada; 2019 yılının yaz verilerine ait veriler kullanılmıştır ve satışa yeni sunulacak ürünlerin satışını tahmin ederken sonraki yaz dönemlerinde de Türkiye'deki ekonomik ve sosyolojik durumların benzer olacağı varsayılmıştır. Bir savaş, ekonomik kriz veya afet durumunda eğitilen modellerin başarısının düşmesi beklenir. Benzer şekilde yıllara göre değişen moda eğilimi düşünüldüğünde farklı yıllara ait sezonluk satış verileri de kullanılarak makine öğrenmesi modelleri geliştirilebilir.

KAYNAKÇA

[1] C. Frank, A. Garg, L. Sztandera, ve A. Raheja, "Forecasting women's apparel sales using mathematical modeling", *International Journal Of Clothing Science And Technology*, vol. 15, no. 2, pp. 107–125, 2003.

[2] C. Catal, K. Ece, B. Arslan ve A. Akbulut, "Benchmarking of Regression Algorithms and Time Series Analysis Techniques for Sales Forecasting", *Balkan Journal of Electrical and Computer Engineering*, vol. 7, no. 1, pp. 20-26, 2019.

[3] P. Sevgi, "Hazır Giyim Perakendeciliği Yapan Bir Firmada Yapay Sinir Ağları ile Satış Tahmini", Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, 2012.

[4] N. Kılınç ve M.Ç. Aydın, "Hazır Giyim İşletmelerinde Beden Numaralarına Göre Oluşacak Talebin Belirlenmesi: Örnek Bir Uygulama ve Karşılaştırma", *KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi*, vol. 18, no. 31, pp. 30-35, 2016.

[5] M.Ç. Aydın, "Giyim Endüstrisinde Talep Tahmin Yöntemlerinin Uygulanması: Örnek Bir Uygulama", Yüksek Lisans Tezi, Selçuk Üniversitesi, 2017.

[6] C. Li ve A. Lim, "A greedy aggregation–decomposition method for intermittent demand forecasting in fashion retailing", *European Journal Of Operational Research*, vol. 269, no. 3, pp. 860–869, 2018.

[7] M. Xia ve W. K. Wong, "A seasonal discrete grey forecasting model for fashion retailing", *Knowledge-Based Systems*, vol. 57, pp. 119–126, 2014.

[8] H. Işık, "Sales Forecasting in Fashion Retail Industry with Classical and Machine Learning Methods", Yüksek Lisans Tezi, İTÜ, 2020.

[9] Ö. İ. Kalaoğlu, E. S. Akyuz, S. Ecemiş, S. H. Eryürük, H. Sümen, ve F. Kalaoğlu, "Retail demand forecasting in clothing industry", *Tekstil Ve Konfeksiyon*, vol. 25, no. 2, pp. 172–178, 2015.

[10] A.L.D. Loureiro, V.L. Miguéis, ve L.F.M. da Silva, "Exploring the use of deep neural networks for sales forecasting in fashion retail", *Decision Support Systems*, vol. 114, pp. 81–93, 2018.

[11] İ. Güven, "Perakende Hazır Giyim Endüstrisinde Yapay Zeka Yöntemleri İle Talep Tahmini", Yüksek Lisans Tezi, Karabük Üniversitesi, 2020.

[12] İ. Güven ve F. Şimşir, "Demand forecasting with color parameter in retail apparel industry using artificial neural networks (ANN) and support vector machines (SVM) methods", *Computers & Industrial Engineering*, vol. 147, pp. 106678, 2020.

[13] S. I. Satoğlu ve U. Eren, "Perakende Giyim Sektöründe Yapay Sinir Ağları ile Talep Tahmini", *Endüstri-İşletme Kurultayı*, İstanbul, 2017.

[14] S. Thomassey ve A. Fiordaliso, "A hybrid sales forecasting system based on clustering and decision trees", *Decision Support Systems*, vol. 42, no. 1, pp. 408–421, 2006.

[15] S. Thomassey ve M. Happiette, "A neural clustering and classification system for sales forecasting of new apparel items", *Applied Soft Computing*, vol. 7, no. 4, pp. 1177-1187, 2007.

[16] Z.-L. Sun, T.-M. Choi, K.-F. Au, ve Y. Yu, "Sales forecasting using extreme learning machine with applications in fashion retailing", *Decision Support Systems*, vol. 46, no. 1, pp. 411–419, 2008.

[17] Y. Yu, T.-M. Choi, ve C.-L. Hui, "An intelligent fast sales forecasting model for fashion products", *Expert Systems with Applications*, vol. 38, pp. 7373–7379, 2011.

[18] S. Öztürk, "Hızlı Moda Sektöründe Makine Öğrenmesi Yöntemleri ile Satış Miktarlarının Tahmin Edilmesi", Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, 2020.

[19] G. Gülgezen, "Kararlı ve Başarımı Yüksek Öznitelik Seçimi", Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, 2009.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirileri

1. S. Ceylan, B. Aslanyürek, Makine Öğrenmesi ile Hazır Giyim Perakende Sektöründe Satış Tahmini, Atılım Üniversitesi Mühendislik Bilimleri ve Araştırmaları Öğrenci Kongresi Mübak 2024, 16-17 Mayıs, Ankara, Türkiye.

Projeler

1. Makine Öğrenmesi ile Hazır Giyim Perakende Sektöründe Satış Tahmini, 2024, YTÜ BAP, Proje Kodu: FYL-2023-5651.