PERAKENDE SEKTÖRÜNDE VERİ MADENCİLİĞİ TEKNİKLERİ İLE SATIŞ TAHMİNİ

Pelin Dinçoğlu
191402203

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Yüksek Lisans Programı
Danışman: Dr. Öğr. Üyesi Hüseyin Aygün

İstanbul
T.C. Maltepe Üniversitesi
Lisansüstü Eğitim Enstitüsü
Ocak, 2022

PERAKENDE SEKTÖRÜNDE VERİ MADENCİLİĞİ TEKNİKLERİ İLE SATIŞ TAHMİNİ

Pelin Dinçoğlu
191402203
Orcid: 0000-0002-4259-0382

YÜKSEK LİSANS TEZİ

Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Yüksek Lisans Programı
Danışman: Dr. Öğr. Üyesi Hüseyin Aygün

İstanbul
T.C. Maltepe Üniversitesi
Lisansüstü Eğitim Enstitüsü
Ocak, 2022



JÜRİ VE ENSTİTÜ ONAYI

Bu belge, Yükseköğretim Kurulutarafından 19.01.2021 tarihli *"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"* ile bildirilen 6689 Sayılı Kişisel Verilerin Korunması Kanunu kapsamında gizlenmiştir.

ETİK İLKE VE KURALLARA UYUM BEYANI

Bu belge, Yükseköğretim Kurulutarafından 19.01.2021 tarihli *"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"* ile bildirilen 6689 Sayılı Kişisel Verilerin Korunması Kanunu kapsamında gizlenmiştir.

TEŞEKKÜR

Tez sürecimde her aşamada bana yardımcı olan, bilgi, birikim ve tecrübesiyle beni yönlendiren değerli tez danışmanım Dr. Öğr. Üyesi Hüseyin Aygün'e, eğitimime her zaman destek veren şirketim Migros Tic. A.Ş ailesine, bu süreçte bana bilgisini ve desteğini esirgemeyen sevgili arkadaşım Neslihan Yegen'e ve her zaman arkamda olan eşim ve aileme teşekkürlerimi sunarım.

Pelin Dinçoğlu

Şubat, 2022

ÖZ

VERİ MADENCİLİĞİ TEKNİKLERİ İLE PERAKENDE SEKTÖRÜNDE SATIŞ TAHMİNİ

Pelin Dinçoğlu
Yüksek Lisans Tezi
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Yüksek Lisans Programı
Danışman: Dr. Öğr. Üyesi Hüseyin Aygün
Maltepe Üniversitesi Lisansüstü Eğitim Enstitüsü, 2021

Gelişen teknoloji bize birçok verinin sanal ortamda saklanması ve analizlerde kullanılması ile değerli bilgilere ulaşma imkanı sağlamıştır. Bu veriler içerisinden anlamlı bir bilgiye ulaşmak için veri madenciliği konusunda pek çok gelişim sağlanmıştır. Veri madenciliği uygulamaları arasında şirketlerin satış tahminlerinin belirlenmesi amacıyla elde edilen veriler ile tahminleme konusunda birçok çalışma bulunmaktadır.

Bu çalışmada perakende sektörüne ait satış verileri kullanılarak satış tahminlemesi veri madenciliğinin tahminleme algoritmaları kullanılarak yapılacaktır. İlk olarak veri madenciliği, kullanım alanları ve işlevleri incelenecektir. Sonrasında veri madenciliği detaylandırılarak satış tahminlemesinde kullanılacak tahminleme modelleri tanıtılacaktır.

Anahtar Sözcükler: Veri Madenciliği, Satış Tahminleme, Veri.

ABSTRACT

SALES FORECASTING IN THE RETAIL SECTOR WITH DATA MINING TECHNIQUES

Pelin Dinçoğlu
Master Thesis
Department of Computer Engineering
Computer Engineering Programme
Thesis Advisor: Asst. Prof. Hüseyin Aygün
Maltepe Üniversitesi Master Graduate School, 2021

The developing technology has provided us to reach valuable information by storing many data in a virtual environment and using them in analysis. There has been many improvements have been made to reach the meaningful information through these information. There are many studies on forecasting with the data obtained to determine the sales estimates of companies will being applied in data mining applications.

In this study, sales forecasting will be done by using the estimation algorithms of data mining by using sales data of the retail industry. First, data mining, usage areas and functions will be examined. Later, data mining will be detailed and forecasting models to be used in sales forecasting will be introduced.

Keywords: Data Mining, Sales Forecasting, Data.

İÇİNDEKİLER

JÜRİ VE ENSTİTÜ ONAYI ii

ETİK İLKE VE KURALLARA UYUM BEYANI iii

TEŞEKKÜR iv

ÖZ v

ABSTRACT vi

İÇİNDEKİLER vii

TABLOLAR LİSTESİ ix

ŞEKİLLER LİSTESİ xi

ÖZGEÇMİŞ xiii

BÖLÜM 1. GİRİŞ 1

BÖLÜM 2. LİTERATÜR TARAMASI 2

2.1. Veri Madenciliğine Giriş 2

2.2. Veri Madenciliği Tanımı 3

2.3. Kullanım Alanları 4

2.4. Bilgi Keşfi 5

2.5. Veri Ön İşleme 6

2.5.1. Veri Temizleme 6

2.5.2. Veri Birleştirme 7

2.5.3. Veri Dönüştürme 7

2.5.4. Veri İndirgeme 7

2.6. Veri Madenciliği Modelleri 8

2.6.1. Tanımlayıcı Modeller 8

2.6.1.1. Kümeleme 8

2.6.1.2. Birliktelik Kuralı 8

2.6.2. Tahmin Edici Modeller 9

2.6.2.1. Karar Ağaçları 9

2.6.2.2. Regresyon 10

2.6.2.3. Rastgele Orman 12
2.6.2.4. Destek Vektör Makineleri 13
2.6.2.5. K-En Yakın Komşu 14
2.6.2.6. Zaman Serileri 15

BÖLÜM 3. VERİ VE YÖNTEM 19

3.1. Veri Kümesinin Analizi 19
3.2. Verilerin Hazırlanması 21
3.3. Yöntem 28

BÖLÜM 4. BULGULAR VE YORUMLAR 29

4.1. Regresyon İle Satış Tahmini Uygulaması 29
4.1.1. 2019 Yılı Verisi İle Regresyon Modeli 41
4.2. Zaman Serileri İle Satış Tahmini Uygulaması 48
4.2.1. 2019 Yılı Verisi İle Zaman Serisi Modeli 57
4.3. Elde Edilen Bulgular 62

BÖLÜM 5. SONUÇ 65

KAYNAKÇA 66

TABLOLAR LİSTESİ

Tablo 3.1. Veri Sayıları 19
Tablo 3.2. Veri Kümesi 20
Tablo 3.3. Veri Kümesi Son Hali 27
Tablo 4.1. 1096 Numaralı Mağaza İçin Modelin Girdilerinin Gün Bazında Ortalamaları (2018-2019) 33
Tablo 4.2. 2018-2019 Verileri Kullanılarak Hesaplanan 1096 Numaralı Mağazanın 2020 Satış Verileri 34
Tablo 4.3. 5144 Numaralı Mağaza İçin Modelin Girdilerinin Gün Bazında Ortalamaları (2018-2019) 35
Tablo 4.4. 2018-2019 Verileri Kullanılarak Hesaplanan 5144 Numaralı Mağazanın 2020 Satış Verileri 36
Tablo 4.5. Genel MAPE Oranları (2018-2019) 36
Tablo 4.6. Mağaza ve Gün Bazında Regresyon MAPE Değerleri 37
Tablo 4.7. Mağaza Bazında Regresyon Model Doğrulukları 38
Tablo 4.8. 1096 Numaralı Mağaza İçin Modelin Girdilerinin Gün Bazında Ortalamaları (2019) 42
Tablo 4.9. 2019 Verileri Kullanılarak Hesaplanan 1096 Numaralı Mağazanın 2020 Satış Verileri 43
Tablo 4.10. 5144 Numaralı Mağaza İçin Modelin Girdilerinin Gün Bazında Ortalamaları (2019) 43
Tablo 4.11. 2019 Verileri Kullanılarak Hesaplanan 5144 Numaralı Mağazanın 2020 Satış Verileri 44
Tablo 4.12. Genel MAPE Oranları (2019) 45
Tablo 4.13. Mağaza ve Gün Bazında Regresyon MAPE Değerleri (2019) 45
Tablo 4.14. 1096 ve 5144 Numaralı Mağazaların 2020 Satış Tahminleri (2018-2019). 52
Tablo 4.15. Mağaza Bazında Genel MAPE Oranları 52
Tablo 4.16. Mağaza ve Gün Bazında Time Series MAPE Değerleri 53
Tablo 4.17. 1096 ve 5144 Numaralı Mağazaların 2020 Satış Tahminleri (2019) 58
Tablo 4.18. Zaman Serisi Genel MAPE Değerleri (2019) 58
Tablo 4.19. Mağaza ve Gün Bazında Zaman Serisi MAPE Değerleri (2019) 58

Tablo 4.20. Mağaza, Model ve Yıl Bazında MAPE Değerleri 62
Tablo 4.21. 1096 Numaralı Mağaza için 2020 Satış Tahmini 63
Tablo 4.22. 5144 Numaralı Mağaza için 2020 Satış Tahmini 64

ŞEKİLLER LİSTESİ

Şekil 2.1. Veri Madenciliğinin İlişkili Disiplinleri 2
Şekil 2.2. Veri, Enformasyon, Bilgi 3
Şekil 2.3. Bilgi Keşfi Adımları 5
Şekil 3.1. 5144 Numaralı Mağazanın Satış Verilerinin Birleştirilmesi 23
Şekil 3.2. 1096 Numaralı Mağazanın Satış Verilerinin Birleştirilmesi 23
Şekil 3.3. IBM SPSS Modeler ile Auto Data Prep Kullanımı 24
Şekil 3.4. Aykırı Değerler Temizlenmiş Veriler 25
Şekil 3.5. IBM SPSS Modeler ile Filler Komutu Kullanımı 25
Şekil 3.6. IBM SPSS Modeler ile Derive Komutu Kullanımı 26
Şekil 4.1. IBM SPSS Modeler ile Aggregate Komutu Kullanımı 29
Şekil 4.2. IBM SPSS Modeler ile Type Komutu Kullanımı 29
Şekil 4.3. IBM SPSS Modeler ile Partition Komutu Kullanımı 30
Şekil 4.4. 1096 Numaralı Mağaza Regresyon Model Katsayıları 31
Şekil 4.5. 5144 Numaralı Mağaza Regresyon Model Katsayıları 31
Şekil 4.6. IBM SPSS Modeler ile Regresyon Modelinin Kullanımı 32
Şekil 4.7. 1096 Numaralı Mağaza Regresyon Grafiği (2018-2019) 37
Şekil 4.8. 5144 Numaralı Mağaza Regresyon Grafiği (2018-2019) 38
Şekil 4.9. 1096 Numaralı Mağaza Regresyon Modeli Girdi Önem Sıralaması 39
Şekil 4.10. 5144 Numaralı Mağaza Regresyon Modeli Girdi Önem Sıralaması 39
Şekil 4.11. 1096 numaralı Mağazanın Model Çıktıları (2018-2019) 40
Şekil 4.12. 5144 numaralı Mağazanın Model Çıktıları (2018-2019) 40
Şekil 4.13. 1096 Numaralı Mağaza Regresyon Model Katsayıları 41
Şekil 4.14. 5144 Numaralı Mağaza Regresyon Model Katsayıları 41
Şekil 4.15. 1096 Numaralı Mağaza Regresyon Grafiği (2019) 46
Şekil 4.16. 5144 Numaralı Mağaza Regresyon Grafiği (2019) 46
Şekil 4.17. 1096 numaralı Mağazanın Model Çıktıları (2019) 47
Şekil 4.18. 5144 numaralı Mağazanın Model Çıktıları (2019) 48

Şekil 4.19. Zaman Serisi Modeli Girdi Çıktı Değerleri 49
Şekil 4.20. Modele Girecek Veri Tiplerinin Nihai Hali 50
Şekil 4.21. IBM SPSS Modeler ile Zaman Serisi Modelinin Kullanımı 51
Şekil 4.22. 1096 Numaralı Mağaza Zaman Serisi Grafiği (2018-2019)..... 53
Şekil 4.23. 5144 Numaralı Mağaza Zaman Serisi Grafiği (2018-2019)..... 54
Şekil 4.24. 1096 Numaralı Mağaza Zaman Serisi Modeli Girdi Önem Sıralaması..... 55
Şekil 4.25. 5144 Numaralı Mağaza Zaman Serisi Modeli Girdi Önem Sıralaması..... 55
Şekil 4.26. 1096 Numaralı Mağazanın Zaman Serisi Model Bilgileri(2018-2019) 56
Şekil 4.27. 5144 Numaralı Mağazanın Zaman Serisi Model Bilgileri(2018-2019) 57
Şekil 4.28. 1096 Numaralı Mağaza Zaman Serisi Grafiği (2019) 59
Şekil 4.29. 5144 Numaralı Mağaza Zaman Serisi Grafiği (2019) 60
Şekil 4.30. 1096 Numaralı Mağazanın Zaman Serisi Model Bilgileri (2019)..... 61
Şekil 4.31. 5144 Numaralı Mağazanın Zaman Serisi Model Bilgileri (2019)..... 61

ÖZGEÇMİŞ

Pelin Dinçoğlu

Bilgisayar Mühendisliği Anabilim Dalı

Eğitim

Derece	*Yıl*	*Üniversite, Enstitü, Anabilim/Anasanat Dalı*
Y.Ls.	2022	Maltepe Üniversitesi, Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı
Ls.	2015	Kadir Has Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi Enformasyon Teknolojileri
Lise	2010	Mevlana Anadolu Lisesi

İş/İstihdam

Yıl	*Görev*
2017 -	Senior Back-End Developer / Migros Tic. A.Ş.
2015- 17	Özel Yazılım Geliştirme ve Raporlama Uzmanı / Nebim Yazılım

BÖLÜM 1. GİRİŞ

Gelişen teknoloji ile başlayan ve hızla ilerleyen verilerin saklanması süreci birçok verinin saklanabilmesine imkan sağlamaktadır. Bilgi şirketler için en önemli güçlerden biri halindedir. Saklanan bu veriler büyüyerek analizi zor bir veri kümesi oluşturmak ile beraber anlamlı sonuçlara ve istenilen bilgiye ulaşmak için bir yönteme ihtiyaç duyulmuştur. Bu sebep ile veri madenciliği metotları geliştirilmiştir. Daha hızlı veriye ve sonuçlara erişim sağlanmıştır. Bu sebeple veri madenciliğinin önemi hızla artmıştır.

Veri madenciliğinin değerinin yüksek olduğu sektörlerin başında perakende gelmektedir. Firmalar, genel satış verileri ve sadakat kartı ile yapılan satış verileri ile tanımladıkları müşterilerin verileriyle davranışlarını inceleyerek özel tahminler yapabilmektedirler. Müşterinin ihtiyacının belirlenmesi, alınabilecek ürünlerin tespiti ile büyük bir rekabet avantajı kazanılabilmektedir. Şirketler bu sebepler ile veri madenciliği çalışmalarına yönelmişlerdir.

Perakende sektöründe veri madenciliği uygulamaları ile satış tahminleme, sipariş tahminleme, birliktelik analizi ve kampanya yönetimi konularında birçok çalışma bulunmaktadır. Bu çalışmalar tahminlemeler yapılarak maliyetin düşürülmesi ve müşteriye özel sunulabilecek kampanyalar ile satışın arttırılması amacıyla yapılmaktadır.

Bu çalışmada, perakende sektöründe geçmiş alışveriş verileri üzerinden satış tahminlemesi yapılacaktır. Bu tahminlemede birçok veri madenciliği algoritmaları kullanılacaktır.

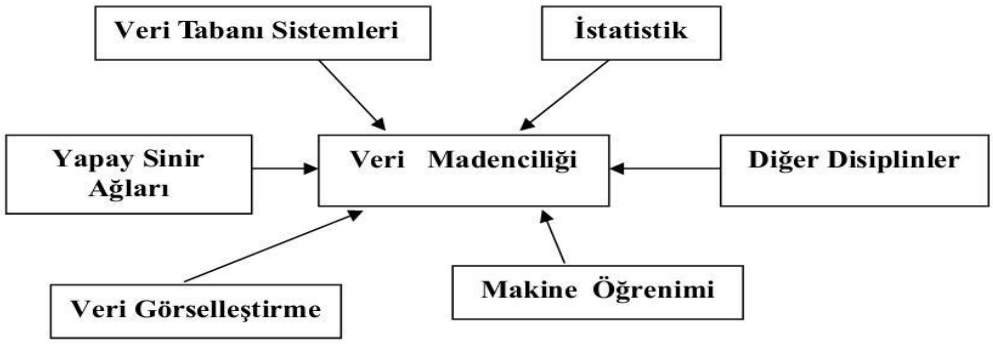
BÖLÜM 2. LİTERATÜR TARAMASI

2.1. Veri Madenciliğine Giriş

Verinin saklanması süreci bilgisayar teknolojisi ile hızla gelişerek çok büyük ve çok çeşitli verinin tutulmasına olanak sağlamaktadır. Günlük hayatımızda kullandığımız birçok elektronik cihaz ile verileri depolamaktayız. Bu veriler market alışverişi, sağlık ve uyku, kredi kartı ve bankacılık işlemleri, sosyal medya gibi çok çeşitli olarak saklanmaktadır. Fazla büyük olan veriler veri tabanlarında saklanarak yönetilmektedir.

Hayatımızda gelişen teknolojinin katkısı olarak elde edilen veriler elektronik ortamda saklanmaktadır. Teknolojinin artması ile veri saklamak daha kolay ve maliyeti düşük bir hal aldığı için veri miktarı özellikle son 20 yılda çok hızlı büyüyerek artmaktadır (DURMAZ & KOCAMIŞ, 2008).

Verinin boyutunun artması ile veriyi okumak daha karmaşık hale gelmiştir. Veri içerisinden anlamlı bir bilgiye ulaşmak zorlaştığı için veri madenciliği yöntemleri geliştirilmiştir. Veri madenciliği, istatistik, algoritmalar ve veri yapıları, bilgisayar mimarisi gibi bilgisayar bilimlerinin birleşiminden oluşmaktadır. Bu birleşim ile veri analiz yöntemleri ortaya çıkmıştır (GEMİCİ, 2012) .



Şekil 2.1. Veri Madenciliğinin İlişkili Disiplinleri

2.2. Veri Madenciliği Tanımı

BÜNCHER, ANAND, & HUGHES tarafından "büyük veri setlerinden belirli bir önemi olan, daha önce bilinmeyen ve anlaşılır ve faydalı örüntülerin keşfi" olarak tanımlanmıştır (BÜCHNER, ANAND, & HUGHES, 1997).

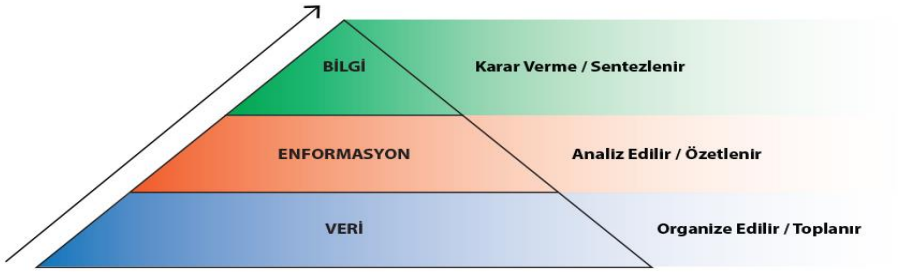
Veri madenciliği, istatistikten farklı olarak büyük veriler üzerinde uygulanmaktadır. İstatiksel uygulamalarda daha küçük yüz adet gibi veri sayısı ile ilgilenilirken veri madenciliği uygulamaları milyarlar seviyesinde veri üzerinde çalışabilme imkanı vermektedir (BİLEKDEMİR, 2010). Veri madenciliği altında istatistik disiplinleri de yer almaktadır.

Veri, enformasyon, bilgi veri madenciliğinin en değerli kavramlarıdır. Elde edilen veri özetlenir ve enformasyon haline dönüşür. Daha sonrasında yapılan analizler sonucunda bilgiye ulaşılır.

Veri: Tek başına herhangi bir anlam ifade etmeyen ham bilgilere veri denir. Ürünler, müşteriler, satış miktarları ayrı ayrı birer veridir. Fakat tek başlarına herhangi bir anlam ifade etmezler (AYDEMİR, 2018).

Enformasyon: Veri üzerinde bazı işlemler yapılarak daha anlamlı bir biçime getirilmiş haline enformasyon denir.

Bilgi: Enformasyonun işlenmesi ile kişisel bir sonuç elde edilmiş haline bilgi denir. Büyük veriler dönüştürülerek daha küçük enformasyon elde edilir. Daha sonrasında bu enformasyon kullanılarak daha küçülerek veri elde edilmiş olur.



Şekil 2.2. Veri, Enformasyon, Bilgi

2.3. Kullanım Alanları

Veri miktarının çoğalması ve bilginin değeri göz önüne alındığı zaman veri madenciliği çok geniş bir kullanım alanına sahiptir.

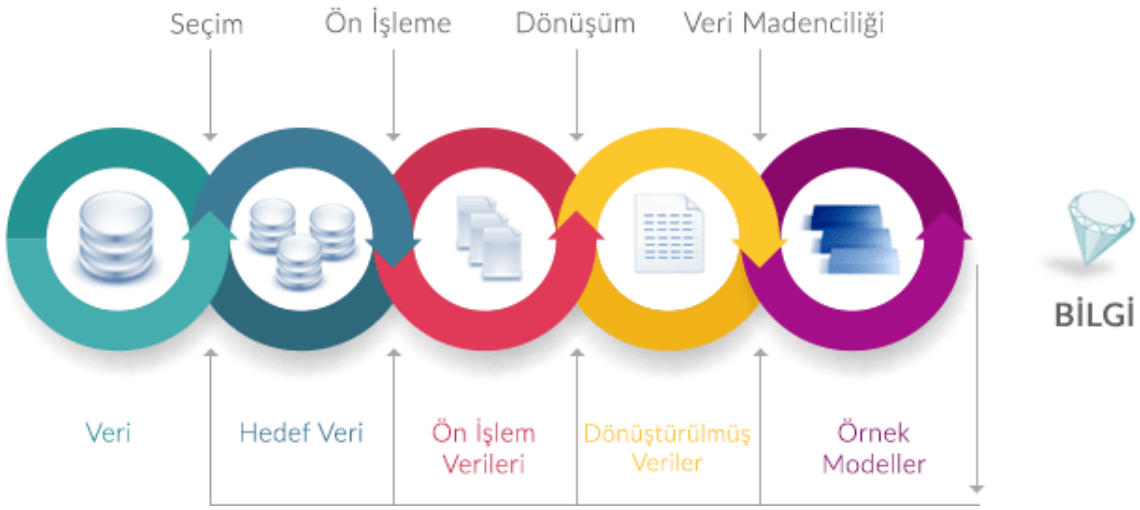
Farklı amaçlar doğrultusunda birçok sektörde veri madenciliği uygulamaları yapılmaktadır. Bunlardan bazıları aşağıdaki gibidir.

1) Pazarlama / Perakende
    - Satış tahminlerinin yapılması
    - Müşteri profilinin oluşturulması
    - Müşteriye özel kampanyaların belirlenmesi
    - Siparişlerin tahminlenmesi
    - Sepet analizleri
    - Yeni ürün için pazar araştırmalarının yapılması
2) Sağlık Hizmetleri
    - Teşhislerin tahminlenmesi
    - Tıbbi ürün geliştirilmesi
    - İlaç kullanımlarının etkilerinin analiz edilmesi
    - Gen analizlerinde
3) Bankacılık / Sigortacılık
    - Harcama bilgileri ile müşteri profilinin oluşturulması
    - Kredi ihtiyacının belirlenmesi
    - Harcama tahminlerinin yapılması
    - Müşteriye özel kampanyaların oluşturulması
    - Riskli müşterilerin tespit edilme
4) Eğitim
    - Öğrenci profilinin belirlenmesi ve başarı oranlarının tahminlenmesi
    - Kullanılacak yeni eğitim sistemlerinin başarı oran tahmini

Yukarı belirtilen sektörlerde farklı amaçlarda birçok çalışma bulunmaktadır. Hepsi elde edilen veriler doğrultusunda sektöründe kullanılabilecek bilgiye ulaşmayı amaçlamaktadır.

2.4. Bilgi Keşfi

Büyük veriler içerisinden faydalı bilginin keşfidir. Henüz bilinmeyen verilerin tahminlenmesidir. Veriye, veri seçimi, ön işleme, dönüştürme, veri madenciliği ve değerlendirme adımlarını uygulayarak bilgi elde edilir.



Şekil 2.3. Bilgi Keşfi Adımları

Veri seçimi: Kullanılacak verinin amaca uygun olarak belirlenmesidir. Yanlış ve eksik verilerin kullanılması ile yanlış modeller oluşturulabileceği için model oluşturmanın temeli olan bu adım çok büyük önem taşımaktadır.

Ön İşleme: Kayıp verilerin düzenlenmesi, gürültünün yok edilmesi işlemleri, gereksiz değişkenlerin yok edilmesi bu adımda yapılır. Tüm bu işlemlerin veri biliminde en çok zamanı aldığı kabul edilmektedir.

Dönüştürme: Verilerin bir sonraki adım olan veri madenciliğine uygun hale getirilmesi işlemidir. Düzeltme, birleştirme, genelleştirme ve normalleştirme olarak adlandırılan dönüştürme türlerinin bir veya birkaçının kullanılmasıdır.

Veri Madenciliği: Son adım olan veri madenciliği, uygun tekniğin seçilmesi, algoritmanın belirlenmesi ve bu algoritmanın uygulanması adımı ile üç basamaktan oluşmaktadır.

2.5. Veri Ön İşleme

Veri madenciliği çalışmalarında en önemli adımlardan birisi veri ön işleme basamağıdır. Verilerin eksik olması, veri seti içerisinde anlamlı verilerin olmaması, veriler arasında örüntü oluşturulamayacak verilerin olması ve sapma değerlerine sahip verilerin olması çalışmaların performansını ve doğruluğunu etkilemektedir. Verideki bu aksaklıkların giderilmesi aşamaları veri ön işleme adımında yapılmaktadır. Bu adımda elimizdeki veri, kullanıma uygun ve daha anlamlı olacak şekilde hazırlanmaktadır (GEMİCİ, 2012).

Veri ön işleme ile verinin kalitesinin yükselmesi sonucunda çıktıların hatalı olması oranı azalacaktır. Yapısı çözülen veriler modellerin başarısını arttıracaktır.

2.5.1. Veri Temizleme

İlk olarak veride yer alan tutarsız değerlerin belirlenmesi, gürültünün giderilmesi, kayıp verilerin tamamlanması için uygulanmaktadır. Modelin hatalı çıktı vermesini etkileyecek kayıtlar incelenerek kontrollü bir şekilde veri seti içerisinden temizlenir. Bu aşamada yapılan hatalar sonuç kısmını büyük oranda etkileyebilecek sorunlar doğurabilmektedir (OĞUZLAR, 2003).

Eksik verili kayıt sayısı, toplam sayıya bakarak değerlendirildiğinde sonucu değiştirmeyecek önemdeyse, eksik verilerin olduğu kayıtlar veri setinden silinebilir. Sonuçları değiştirecek kadar fazla ise bu yöntem kullanılmamalıdır. Bu durumda bu veriler temin edilebilir, ortalama bir değer girilebilir, benzer veriler taban alınarak değer verilebilir, eldeki veriler üzerinden tahmin edilerek doldurulabilir.

Uç veriler kümeleme yöntemi ile belirlenerek yeni değer verilebilir. Eldeki veriler üzerinden tahmin edilerek düzenlenebilir (ŞEKEROĞLU, 2010).

Veri düzeltme işlemlerinde aykırı değerlerin kümelenerek belirlenmesi için kümeleme tekniği, fonksiyon uygulanarak dışarıda kalan verilerin tespiti için regresyon

tekniği ve büyükten küçüğe veya tersi şekilde sıralanan verilerin düzeltilmesinde ise binning tekniği kullanılmaktadır (OĞUZLAR, 2003).

2.5.2. Veri Birleştirme

Bu adımda birçok veri kaynağından alınan verinin birleştirilerek veri setinin oluşturulması işlemleri gerçekleştirilir. Örneğin iki farklı veri deposunda bulunan aynı veriyi işaret eden değerlerin kolon isimleri farklı olabilir. Bu şekilde verilerin birleştirilmesi ile şema birleştirme hataları ortaya çıkmaktadır. Tutarsız verilerin olması konusunda korelasyon analizi yapılarak verilerin çıkarılması ile indirgeme işlemi yapılmış olur. Örnekteki gibi iki farklı isimde fakat aynı değerleri içeren kolonlardan sadece birinin veri seti içerisinde kalması sağlanacaktır (BUDAK, 2013).

2.5.3. Veri Dönüştürme

Kullanılacak modele göre verinin tiplerinin değiştirilmesidir. Algoritmada belirleyici olması için uygun hale getirilerek verinin dönüştürülmesi adımıdır. Bu adımın uygulanması için standartlaştırma ya da normalleştirme teknikleri kullanılmaktadır. En sık kullanılan veri dönüştürme tekniklerinden birisi veri normalleştirmesidir. Sayısal veriler üzerinde çalışan modeller için normalleştirme tekniklerinden sıklıkla min-max normalleştirmesi veya Z-Score normalleştirmesi uygulanmaktadır (BUDAK, 2013).

2.5.4. Veri İndirgeme

Sonucu değiştirmeyecek ise değişkenlerde veya veri sayısında azaltılma işlemi yapılabilir. Ana verinin özelliklerini barındıran fakat daha küçük ölçekli bir veri setinin elde edilmesi için indirgeme teknikleri kullanılmaktadır. İndirgenen veri üzerinde çalışılması zamandan tasarruf sağlayacak ve uygulanan modelin daha etkin çıktılar vermesini etkileyecektir. Kullanılan veri indirgeme teknikleri aşağıdaki gibidir.

- Veri Sıkıştırma
- Veri Küpü Birleştirme
- Boyut İndirgeme
- Kesikleştirme

2.6. Veri Madenciliği Modelleri

Veri madenciliği modelleri, tahmin edici ve tanımlayıcı modeller olarak ikiye ayrılmıştır. Modeller işlevlerine göre sınıflama, kümeleme ve birliktelik kuralları olmak üzere 3 ana başlıkta toplanmaktadır. Sınıflandırma işleminde eldeki örneklerin kullanılmasıyla modelin sınıflama yapması hedeflenir (AYGÜN & ADALI, 2005). Kümeleme ve birliktelik kuralı tanımlayıcı, sınıflama ve regresyon modelleri ise tahmin edici modeller olarak bilinmektedir (SAVAŞ, TOPALOĞLU, & YILMAZ, 2012).

2.6.1. Tanımlayıcı Modeller

Tanımlayıcı modellerde amaç karar vermeyi etkileyen verilerin arasındaki ilişkilerin tanımlanmasıdır. Eldeki verilerin kendi içindeki ilişkileri ile diğer veriler kıyaslanarak benzerlik olup olmadığının araştırılmasıdır. Kümeleme ve ilişki analizi tanımlayıcı modellerin ana bileşenleridir (YILDIRIM, 2019).

2.6.1.1. Kümeleme

Veri madenciliği kullanımında gerekli olan verilerin boyutu gün geçtikte artmaktadır. Büyük veri üzerinde çalışırken veri madenciliği teknikleri başarılı sonuçlar çıkaramayabilir. Bu durumda karmaşık olan bu verinin parçalara ayrılması gerekebilir. Bu parçalara ayırma işlemine kümeleme denir. Kümeleme analizinde, parçalama işlemi yapıldıktan sonra oluşan alt kümelerin elemanları kendi içlerinde benzer özellikler taşımaktadır. Kümeleme yöntemleri, farklı amaçlar ve farklı sektörlerde kullanılmaktadır. Fakat en fazla tercih edildiği kullanım alanı, müşterilerin ayrıştırılarak kümelere özel olarak satış stratejilerinin geliştirilmesi durumlarıdır (ÇELİK, 2009).

2.6.1.2. Birliktelik Kuralı

Veri kümesinde tahmin edilen birliktelikler haricinde kalan birlikteliklerin tespiti için kullanılan yöntemdir. Gerçekleşen bir olayın ardından diğerinin gerçekleşiyor olmasının tespit edilmesidir (ÇELİK, 2009).

Genellikle sepet analizlerinde kullanılan bu yöntem Pazar-sepet analizi de denmektedir. Müşterilerin satın alışkanlıklarının tespitinde kullanılan bu yöntem analizleri satış stratejilerinin oluşturulmasında büyük rol almaktadır. Amaçlanan birlikte

alınan ürünlerin tespit edilerek hangi ürünlere kampanya yapılması gerektiği bilgisinin elde edilmesidir (YILDIRIM, 2019).

2.6.2. Tahmin Edici Modeller

Tahmin edici modellerde, bilinen verilerden model geliştirilerek sonuçları bilinmeyen veri kümelerinden tahminlerin yapılması amaçlanmaktadır.

Tahmin edici modellerden bazıları;

- Karar ağaçları
- Regresyon
- Rastgele Orman
- Destek Vektör Makineleri
- K-en Yakın Komşu
- Zaman Serileri

2.6.2.1. Karar Ağaçları

Olayların sınıflandırılmasında kullanılan karar ağaçlarında, durumlar ve örnek kümeleri ile başlayarak ağaç yapıları oluşturulur. Ağaçlar, düğüm ve yapraklardan oluşur. Düğümlerden sonraki adım dallanmadır. Her bir durumun olasılığı dallanmada belirtilir. Sonraki dallanmaların nasıl olacağı test sonuçları ile belirlenir. Bu sebeple karar ağacının her bir düğümü test verisi içermektedir.

Karar ağaçlarında sınıflama yapmak için kullanılan kurallar ve karar ağaçlarının doğruluğu sınıflamada belirlenir. Karar ağacı doğruluğu uygun yükseklikte ise kurallar diğer verilerin sınıflandırılmasında kullanılabilir. Ağacın hangi sırada ve hangi veriler ile oluşturulacağının seçilmesinde en yaygın olarak entropi ölçümü tercih edilmektedir. Entropi ölçüsü düşük olanlar ağaç köklerinde kullanılır. Entropi formülü aşağıdaki gibidir (ÇALIŞ, KAYAPINAR, & ÇETİYOKUŞ, 2014).

$$H(p_1, p_2, \dots p_n) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{2.1}$$

Sınıflama algoritmalarının en önemlilerinden olan karar ağacında öğrenme algoritması diğerlerine göre daha basittir. Sonuçların gösterimi kolay ve anlaşılırdır. Bu ağaçlar bize kararlar ile birlikte ilgili kararların açıklamasını da verirler. Tümevarım yöntemi kullanılır. Eğitim kümesi kullanılarak karar ağacı oluşturulması, ağaç tümevarımı (tree induction) olarak adlandırılır. Bu yöntem en yaygın olanlardan biridir. Ağaç benzeri örüntüleri sınıflama veya tahmin etmek için kullanılabilir (EMEL & TAŞKIN, 2005).

Karar ağacı ile yapılan analizler

- Sınıfların üyesi olma ihtimalinin belirlenmesinde (Segmentation),
- Olayların düşük, orta ve yüksek risk grubunda olduğunun kategorik olarak ayrılmasında (Stratification),
- Gelecek tahminlemesi için kurallar oluşturulmasında,
- Veri kümesi içerisinde modellerde kullanılırken fayda sağlayacak verilerin seçilmesinde,
- Alt gruplardaki ilişkilerin tanımlanmasında,
- Kategori birleştirilmesi ve değişkenlerin dönüştürülmesinde yaygın olarak kullanılmaktadır (EMEL & TAŞKIN, 2005).

2.6.2.2. Regresyon

Regresyon analizi mevcut veri üzerinden, işlem adede olan etkisi üzerinden analiz edilmektedir. Regresyon analizi veri davranışlarının model ile tahmin edilmesi amacıyla kullanılmaktadır. Veri seti içerisindeki ilişkiler ve modele etkisi hesaplanarak veri analizinden çıkartılması konusunda fayda sağlamaktadır. Bir veya birden fazla değişken kullanılarak da regresyon analizi yapılabilmektedir.

Regresyon analizi ile hedeflenen, bir olayın gerçekleşmesinde hangi olayların etkisinin olduğunun belirlenebilmesidir. Bir veya daha fazla olay etkili olabilmektedir. Örneğin; müşterilerin bir krediyi çekme eğilimini kredi faiz oranlarıyla, kişinin geliriyle, vb faktörler ile ilişkilendirmek mümkündür. Bu değişken sayıları kurulacak matematiksel modelin belirlenmesinde etkilidir (YEGEN, 2020).

Doğrusal Regresyon:

Doğrusal regresyon, bir bağımlı değişken ile bağımsız değişken veya değişkenler arasındaki doğrusal ilişkiyi tahmin etmek üzere kullanılan bir modellerdir. En küçük kareler tahmin yöntemi en yaygın olarak kullanılandır. Doğrusal regresyon için formül aşağıdaki gibidir. Formülde yer alan X değişkeni, a0 ise değişkenin katsayısını ifade eder. Değişkendeki her artış veya azalış sonucu direkt olarak etkilemektedir.

$$Y_i = \beta_0 + \beta_l X_i + \varepsilon_i \qquad i = (l, n) \tag{2.2}$$

Y_i ; Bağımlı değişken

X ; Bağımsız değişken

β_0 ; Sabit değer

β_l ; X değişkeninin Y_i değişkenini açıklama derecesi

ε_i ; Hata terimi

Doğrusal regresyonun matematiksel ifadesi, birden fazla bağımsız değişken olması durumunda Denklem(2.3) ve Denklem(2.4) de gösterilmektedir.

$$Y_i = \beta_0 + \beta_{li} X_{li} + \cdots + \beta_{ki} X_{ki} + \varepsilon_i \quad i = (l, n) \tag{2.3}$$

$$Y_i = \sum \beta_k X_{ki} + \varepsilon_i \quad i = (l, n) \tag{2.4}$$

Lojistik Regresyon:

Doğrusal regresyondan farklı olarak, bağımlı değişkenin kategorik olduğu modelleridir. Bağımlı değişken ile bağımsız değişkenler arasındaki neden sonuç ilişkisinin belirlenmesinde kullanılmaktadır. Fonksiyonun amacı değişkenlerin sınıflara

ait olma olasılıklarının hesaplanmasıdır. Değişkenlerin üzerindeki etkilerinin olasılık olarak belirlenmesi sağlanır. Formülü aşağıdaki şekilde elde edilir.

Olasılık hesaplaması yapılacağı için doğrusal regresyonda kullanılan formülünün sol tarafı 0 ile 1 değeri arasında kalmalıdır. Bu değerleri alacak şekilde sağ tarafa lojistik fonksiyonu uygulanarak formülün aşağıdaki hali elde edilir.

$$y = \frac{e^{a_0+a_1x}}{1+e^{a_0+a_1x}} \tag{2.5}$$

Denklem üzerinde matematiksel eşleme yapılır.

$$\frac{y}{1-y} = e^{a_0+a_1x} \tag{2.6}$$

Her iki tarafın logaritması alınarak lojistik regresyon formülünün nihai hali elde edilmiş olur.

$$\log\left(\frac{y}{1-y}\right) = \log(e^{ao+a_1x}) \tag{2.7}$$

2.6.2.3. Rastgele Orman

Sınıflandırma konusunda iyi performans sergileyen Rastgele Orman algoritması diğer yöntemlere göre sıklıkla tercih edilmektedir. Rastgele orman algoritması, hızlı ve yüksek doğruluğa sahiptir. Topluluk öğrenme yöntemi olan rastgele orman algoritmasında, bireysel olarak oluşturulan karar ağaçları bir araya gelerek karar ormanını oluşturmaktadır (OK, AKAR, & GÜNGÖR, 2011).

Rastgele orman yönteminde gelecek tahmini ve tahminlerdeki değişkenlerin önem derecesi hesaplanmaktadır. Önem derecesinin hesaplanması çok değişkenli veri setleri için modelin indirgenmesinde fayda sağlamaktadır. Çok fazla değişkeni olan veri seti içerisinde önem dereceleri belirlenerek kurulacak model ile önem derecesi yüksek olan değişkenlerin kullanılması ile tahminlerin yapılması daha etkin olacaktır (KURUCAN, 2019).

Rastgele orman yönteminde, modeli test etmek için ayrı bir veri seti yoksa model sınıf dağılımına bağlı kalarak veri setinin 2/3 ‘ünü öğrenme veri seti, kalan 1/3’ü ise test veri seti olarak ayırarak kullanmaktadır. Eğer ayrı bir test veri seti varsa, modelin kurulması için ayrılan öğrenme veri seti, kendi içerisinde yine 2/3’ü öğrenme kalan 1/3’ü test verisi olacak şekilde ayrılır.

Rastgele orman algoritması aşağıdaki formülü kullanmaktadır. Bu değişken her bir dal için ayrı ayrı hesaplanarak üst düğümden alt düğüme azalacak şekilde dizilir. Sıfır olduğu durumda tamamen homojen olur ve dallanma sona erer (ERDEM, ve diğerleri, 2018).

$$Gini = 1 - \sum_{i=1}^{c} (p_i)^2 \tag{2.8}$$

2.6.2.4. Destek Vektör Makineleri

Veri seti içerisinde bulunan değişkenler arasındaki ilişkilerin bilinmediği veri setleri için sınıflandırmada kullanılan bir makine öğrenmesi yöntemidir. Eğitim veri setini kullanarak öğrenme yolu ile yeni veriler üzerinde tahmin yapmaya çalışan bir makine öğrenmesidir. Bu yöntem istatiksel öğrenme teorisine dayanarak regresyon, sınıflandırma ve aykırı değerlerin belirlenmesinde kullanılmaktadır.

İstatistiksel öğrenme teorisi alanında ortaya çıkmış bir öğrenme metodudur. Lineer olmayan örnek uzayının, lineer olarak ayrılabileceği bir boyuta aktararak maksimum sınırın bulunması destek vektör makinelerinde esastır (ATASEVER, 2011) .

Destek vektör makinelerinde problemler lineer olarak ayrılabilen ve ayrılamayan olmak üzeri ikiye ayrılır. Genellikle günlük yaşantımızdaki problemler lineer ayrılamamaktadır. Lineer olarak ayrılan problemlerde maksimum sınırın bulunması kolay iken lineer olmayan veriler için lineer olacak şekilde ayrıştırılabileceği bir uzaya aktarılması gerekmektedir (DEMİRCİ, 2007).

Destek vektör makineleri parametrik olmayan sınıflayıcılardır. Dağılım konusunda ön bilgi varsayımları yoktur. Girdi ve çıktılar eğitim setleri içerisinde eşlenir. Yapılan eşlemeler kullanılarak, sınıflama yapacak karar fonksiyonları elde edilir.

Avantajlar

• Yüksek doğruluk

• Karmaşık veri setlerini modelleyebilme

• Fazla bağımsız değişkenle çalışabilme

• Doğrusal olarak ayrılabilen veya ayrılamayan tüm veriler üzerinde uygulayabilme

• Diğer birçok yönteme göre yeni verilere uyum sağlama sorunun az olması

Dezavantajlar

• Noktasal ve olasılıksal tahminler yapılamaması

2.6.2.5. K-En Yakın Komşu

Aynı tip veriler veri uzayında birbirlerine yakın olarak bulunurlar. Bu durumda öğrenme veri kümesi ile var olan veri kümesinin benzerlikleri hesaplanarak birbirine yakın olduğu belirlenen n tane verinin ortalamasının alınarak elde edilen değere göre sınıflandırma yapıldığı modele k- en yakın komşu modeli denir.

Modelde tüm veriler bir örüntü uzayında saklanarak bilinmeyen bir örneklem geldiğinde hangi sınıfa ait olduğu belirlenerek en yakın olan k örneklemi bulunur. Algoritmada her bir sınıfın önceden net bir şekilde belirtilmiş olması gerekmektedir.

Bilinmeyen örneklemin en yakın komşuya yakınlığı öklid uzaklığı ile tanımlanır. Öklid uzaklığı en yaygın kullanılan uzaklık ölçüsüdür. Öklid uzaklık formülü aşağıdaki şekilde hesaplanır;

$$\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (2.9)$$

Öklid Formülü

Tüm örneklemler için sınıflandırma yapılana kadar uzaklık hesaplaması yapılır ve en yakın sınıfa yerleştirilir (ÇERKEZİ, 2013).

2.6.2.6. Zaman Serileri

Belirli aralıklarla hesaplanan, ölçülen değerlerin ardışık dizilimine zaman serisi denir. Belirli aralıklarla tekrar etmesi ve ardışık sıralı olması verinin zaman serisi olması için beklenen ölçütleridir. Veriler günlük, aylık, yıllık gibi tekrar eden veriler olabilir. Bir mağazanın günlük satışları, alışveriş merkezinin saatlik ziyaretçi sayısı, otobüs firmasının aylık sefer sayısı gibi veriler örnek olarak düşünülebilir (ETGİN, 2017).

Model içerisinde üzerinde çalışılan veriler birbiriyle ilişki göstermektedir ve bu veriler ile gelecek değerler için öngörü yapılmaktadır. Gelecekteki verinin tahmin edilmesi için, eldeki veriler için zaman serisi içerisindeki örüntüleri kullanarak en uygun zaman bağımlı model oluşturulur.

Elimizde var olan veri kümesindeki verilerin, geçmiş değerlerinin farklı yöntemlerle incelenmesi ile gelecek tahmini yapılması için kullanılan altı grupta toplanan analizler mevcuttur (IRMAK, KÖKSAL, & ASİLKAN, 2012)

I. Zaman serilerinin bileşenlerine ayrılması yöntemi (Decomposition methods)
II. Üstel düzgünleştirme yöntemleri (Exponential smoothing methods)
III. Otoregresif yöntemler (Autoregressive methods)
IV. Hareketli ortalama yöntemleri (Moving average methods)
V. Bileşik otoregresif hareketli ortalama yöntemleri (Autoregresive integrated moving average methods)
VI. Zaman serilerinde regresyon yöntemleri (Regression methods in time series)

Zaman serileri finansal sektörler için büyük bir öneme sahiptir. Veri madenciliğinde veri kümelerinin özellikleri zaman serisi analizleri ile çıkarılmaktadır. Zamana bağlı olarak hareketlerinin tespiti ile gelecek tahminlerinin daha doğru sonuçlar vermesi sağlanabilmektedir (CAN & GERŞİL, 2018).

Zaman Serilerinin Bileşenlerine Ayrılması Yöntemi:

Bileşenlerine ayrılması yöntemi, zaman serilerindeki davranış biçimlerinin belirlenmesi amacıyla kullanılan en eski yöntemlerdendir.

Matematiksel olarak ifadesi: $z_t = f(T_t, M_t, C_t, \varepsilon_t)$ (2.10)

Serinin Tahmini: $\hat{z}_t = f(\hat{T}_t, \hat{M}_t, \hat{C}_t)$ (2.11)

Üstel Düzgünleştirme Yöntemi:

Üstel düzgünleştirme yöntemi, verilerin gelecek tahminlerinde, trend ve sezonsallık bileşenlerinin değişkenlik gösterdiği durumlarda etkin olarak kullanılmaktadır. Bu değişkenlikleri hesaba katarak işlem yapılmalıdır. Zaman serisi değerlerinin eşit olmayacak şekilde ağırlıklandıran yöntemdir (IRMAK, KÖKSAL, & ASİLKAN, 2012)

$$\hat{Z}_{T+1} = \hat{Z}_T + \alpha(e_t) \tag{2.12}$$

α : Düzleştirme katsayısıdır.

Formülde tahmin edilen dönem, bir önceki tahmin ve hata payının bir kısmı alınarak bulunmaktadır.

Denklem detaylandırılarak matematiksel olarak ifade edildiğinde aşağıdaki şekildedir.

$$S_T = S_{T-1} + \alpha(Z_T - S_{T-1}) \tag{2.13}$$

S_T: (T+1)'inci dönemin(Gelecek dönemin) tahmini

S_{T-1}: Son dönemin tahmini

Z_T: Son dönemin gerçek değeri

α: Düzleştirme katsayısı

(Z_T-S_{T-1}): Son dönemin tahmin hatası

Hareketli Ortalama Yöntemi:

Bu yöntem ile seri içerisinde yer alan beklenmedik verilerin, serinin hareketini etkilemeyecek şekilde azaltılarak düz bir doğruda olması sağlanmaktadır. Son dönemlerin ortalaması alınır ve değişim etkilerinin azaltılması için veri düzleştirilir. Kullanılan hareketli ortalama yöntemleri ve formülleri şu şekildedir (KARAKAŞ, 2019):

I. Basit hareketli ortalama

$$y_t = \frac{1}{n} \sum_1^n x_{t-i} \tag{2.14}$$

II. Ağırlıklı hareketli ortalama

$$y_t = \frac{nx_{t-1}+(n-1)x_{t-2}+\cdots+2x_{t-n-1}+x_{t-n}}{n+(n-1)+\cdots+2+1} \tag{2.15}$$

III. Üstel ağırlıklı hareketli ortalama

$$y_t = \alpha x_t + (1-\alpha)y_{t-1} \tag{2.16}$$

Bileşik Otoregesif Hareketli Ortalama Yöntemi:

Otoregresif modellerin hareketli ortalama modelleri ile birleştirilmesi 1970 yılında George E.P Box ve Gwilym M. Jenkins tarafından yapılarak ARIMA(Autoregressive Integrated Moving Average Model) olarak adlandırılan model

elde edilmiştir. ARIMA modeline, Box-Jenkins yöntemi de denilmektedir (YEGEN, 2020) .

Bu yöntem AR ve MA modellerinin birleşimi olarak ifade edilir. AR modeli t dönemindeki değerleri ve geçmiş dönemlerdeki değerleri ile hata teriminin doğrusal fonksiyonudur. MA modeli ise t dönemindeki değer ve hata terimleri ile geçmiş dönem hata terimlerinin doğrusal fonksiyonudur (KAYNAR & TAŞTAN, 2009).

ARIMA(p,d,q) modelinde , p,d,q değişkenlerin neleri ifade ettikleri aşağıdaki gibidir.

p: AR (otoregresyon) modelinin derecesi

d: Fark alma işlemi sayısı

q: MA (hareketli ortalama) modelinin derecesi

Genel ARIMA(p,d,q) modelinin gösterimi şu şekildedir;

$$w_t = \phi_1 w_{t-1} + \phi_2 w_{t-2} + \ldots + \phi_p w_{t-p} + a_t - \theta_1 a_{t-1} - \theta_1 a_{t-2} - \ldots - \theta_q a_{t-q} \quad (2.17)$$

∇ = Fark alma operatörü

d = Fark alma derecesi

$\{w_t\}$ = Farkı alınmış seri

Eğer birinci farklar sonrasında seri durağan hale geliyorsa (d = 1) fark operatörü aşağıdaki şekilde olur.

$$\nabla x_t = w_t = x_t - x_{t-1} \quad (2.18)$$

BÖLÜM 3. VERİ VE YÖNTEM

3.1. Veri Kümesinin Analizi

Migros Tic. A.Ş. firmasına ait 2 mağazanın 2018 ve 2019 yıllarına ait satış verileri satış tahminleme yapılması için kullanılmıştır. Bu çalışmada yer alacak mağazalar ve mağaza sayısı belirlenirken, mağazaların satış özellikleri dikkate alınmıştır. Seçilen mağazaların biri satışları az ve sezonsallık olmayan, kampanya, özel gün gibi durumlardan daha az etkilenen bir mağazadır. Diğeri ise satışları yüksek, kampanyaya duyarlı bir müşteri kitlesi olan ve özel günlerden etkilenen bir mağazadır.

Çalışmada kullanılmak üzere, belirlenen mağazaların ürün detayında günlük satış verileri alınmış olup, ham veri içerisindeki kayıt sayıları aşağıdaki tabloda gösterilmiştir. 1096 numaralı mağaza satış hacmi büyük olan, 5144 numaralı mağaza ise satış hacmi daha küçük olan bir mağazadır.

Tablo 3.1. Veri Sayıları

Mağaza Kodu	**2018**	**2019**
1096	10,1274,84	10,178,608
5144	916,229	1,089,154

Veri içerisinde ilgili mağazalara ait gün ve fiş bazında satış kayıtları yer almaktadır. Kampanyaların satış üzerinde yüksek oranda etkisi olacağı düşünülerek veriye dahil edilmesine karar verilmiştir. Modelde etkisi olabileceği düşünülerek özel gün, kampanya olup olmadığı, sanal/fiziksel satış bilgileri fiş bazında eklenmiştir. Yine modelde etkisinin olmayacağı düşünülerek kdv oran, kdv tutar, ödeme tipi gibi bilgileri çıkarılarak elde edilmiş veri deseni aşağıdaki gibidir.

Tablo 3.2. Veri Kümesi

Tarih	Yıl ay gün formatında satışın gerçekleştiği tarihi ifade etmektedir.
Mgz	Satış yapılan mağaza numarasıdır.
Mgz_adi	Satış yapılan mağaza adıdır.
Kasa	Satış yapılan yazarkasa numarasıdır.
Fisno	Satış işlem numarasıdır.
Saat	Satış işlem saatini ifade etmektedir.
Anagrupno	Satış yapılan ürünün ana grup numarasıdır.
Altgrupno	Satış yapılan ürünün alt grup numarasıdır.
Sınıfno	Satış yapılan ürünün sınıf numarasıdır.
Altsınıfno	Satış yapılan ürünün alt sınıf numarasıdır.
Urun	Satış yapılan ürün adını belirtmektedir.
İndirim	Satış işleminde herhangi bir indirim olduğu durumda bu alanda yer alıyor.
Miktar	Satış yapılan ürün miktarını ifade etmektedir.
Tutar	Satış tutarını belirtmektedir.

SanalFiziksel	Sanal mı fiziksel mi satış olduğunu ifade eder.
OzelGun	Özel gün bilgileri yer almaktadır.

3.2.Verilerin Hazırlanması

Bu çalışmada 2018-2019 yıllarına ait iki farklı mağazadan alınan satış bilgileri kullanılarak veri madenciliği algoritmalarından tahminleme algoritmaları ile satış tahminlemesi ve analiz çalışmaları gerçekleştirilmiştir. Gelecek döneme ait satış tutarının tahmin edilmesinde IBM SPSS Modeler üzerinde çalışılmıştır.

Satış yapılan günün özel bir gün olup olmadığı alınarak veri setine eklenmiştir.

Migros kampanyalarını büyük oranda temsil eden katalog kampanyaları olup olmadığı modelde kullanılmıştır. Kampanya bilgisi 0 yok 1 var şeklinde veri setine eklenmiştir.

Veri üzerinde tarih gün bazında olduğundan modelde daha anlamlı kullanılabilmesi için tarihten satışın haftanın hangi gününe ait olduğu bilgisi bulunarak veri setine eklenmiştir. Model içerisinde Pazartesi, Salı gibi kendi içlerinde örüntü yakalanması açısından gün bilgisi eklenmiştir.

Veriler IBM SPSS Modeler ile hazırlanmıştır. Veri seti Migros Tic. A.Ş. tarafından her mağaza için yıl bazında 3 parça halinde txt formatında verilmiştir. İki yıllık veri alındığı için toplamda 6 adet txt dosyası olarak alınmıştır.

Veri birleştirme, veri ön işleme ve modelleme için IBM SPSS Modeler uygulaması kullanılmıştır. SPSS birçok veri madenciliği süreçlerinde kullanılmaktadır ve veri analizleri yapılma imkanı sunmaktadır.

Alınan veri setine satışlar üzerinde büyük önem değeri taşıdığı düşünülen kampanya verisi eklenme kararı alınmıştır. Migros birçok kampanya düzenlediği için kampanyalar arasından en değerli olarak görülen ve 2 haftada bir yenilenen katalog kampanyası verilerinin modelde dikkate alınması için eklenmesi uygun görülmüştür.

Kampanya bilgilerinin yer aldığı kampanya tablosuna ait 16,298,143 adet olan kayıtları alınmıştır. Alınan kayıtlar içerisinde işlem tarihi, promosyon numarası, açıklaması, başlangıç tarihi, bitiş tarihi ve ilgili kampanyanın hangi üründe olduğu bilgisi yer almaktadır.

Elimizdeki her yıl için üç parçaya ayrılmış olan satış verisinin tarihlerine bakılarak kampanya verisi de üç parçaya ayrılarak ilgili ürünlerde kampanya olup olmadığı kontrol edildikten sonra veri setine katalog kampanyası var/yok bilgisi eklenmiştir.

Veri analizleri kapsamında en sık kullanılan fonksiyonlar:

Append: Veri kümesini birleştirmek için kullanılır.

Filter: Veri kümesinde filtreleme yapmak için kullanılır.

Derive: Değişkenler üzerinde hesaplamalar yapmak, koşullar koymak ve yeni değişkenler oluşturmak için kullanılır.

Type: Değişken tiplerini ve rollerini belirlemek ve okumak için kullanılır.

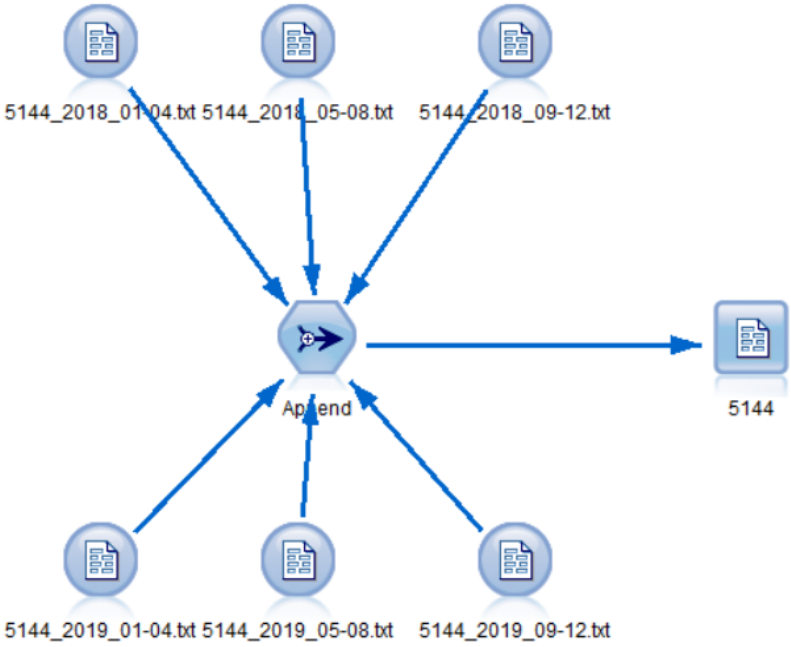
Filler: Veri setindeki değişkenlerin değerlerini yeni bir değer ile değiştirmek için kullanılır.

Select: Veri setindeki belirli değerlerin seçilmesi veya seçilmemesi için kullanılır.

Partition: Eğitim veya test amacıyla kullanılacak verilerin bölünmesi için kullanılır.

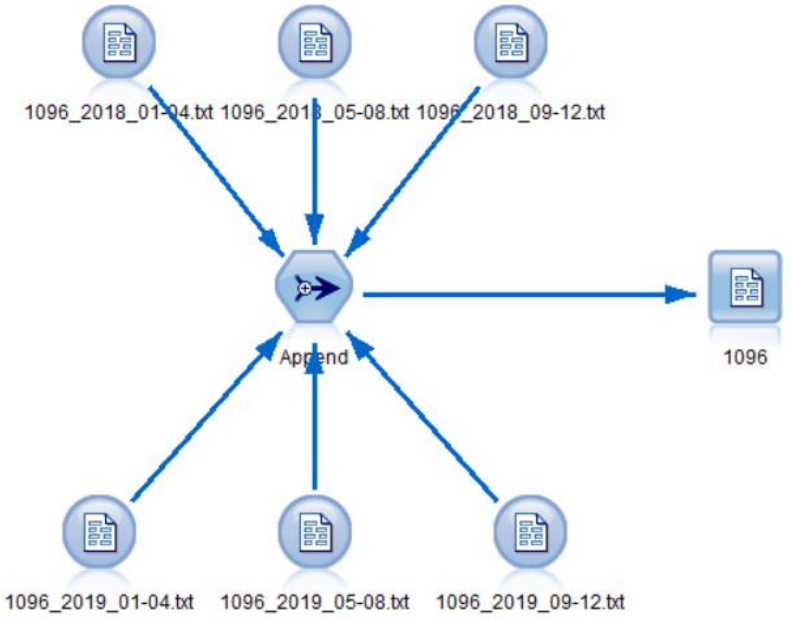
Çalışmada veri büyüklüğü sebebiyle iki mağaza için satış verileri ayrıştırılarak kullanıldı. 1096 numaralı mağaza için toplamda 20.306.092 kayıt, 5144 numaralı mağaza için ise 2.005.383 kayıt veri ile çalışılmış oldu.

5144 numaralı mağazanın iki yıllık verilen 6 adet txt dosyası SPSS kullanarak veri kümesi birleştirmede kullanılan “append” komutu ile birleştirilmiştir.



Şekil 3.4. 5144 Numaralı Mağazanın Satış Verilerinin Birleştirilmesi

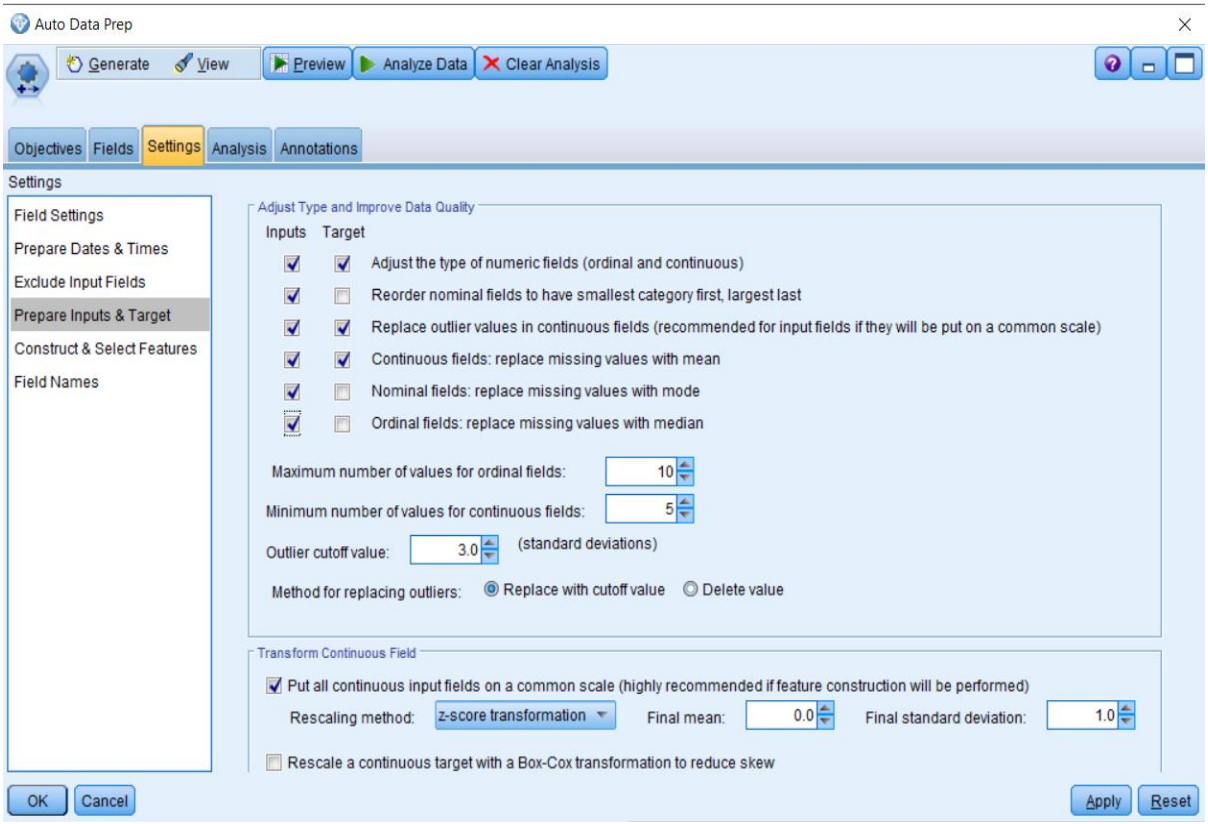
İki yılın verilerini içeren 1096 numaralı mağazanın satış verileri de veri kümesi birleştirmede kullanılan "append" komutu ile birleştirilmiştir.



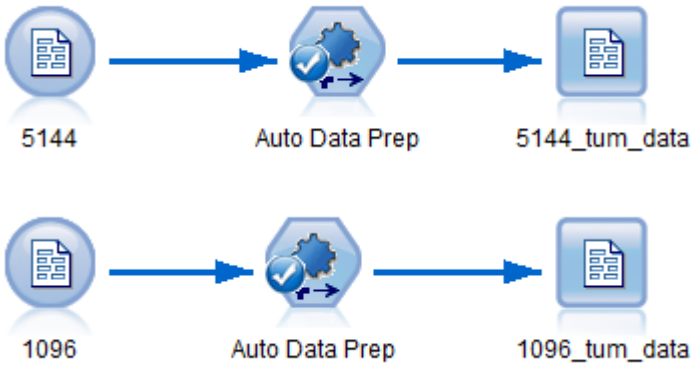
Şekil 3.5. 1096 Numaralı Mağazanın Satış Verilerinin Birleştirilmesi

Veri birleştirmeleri tamamlandıktan sonra model içerisinde kullanılacak miktar, tutar ve indirim alanlarında verilerin ön incelemesi yapıldığında ve şirket satış özellikleriyle ilgili deneyimler de dikkate alınarak, mağazalardan çeşitli dönemlerde ve kişiler tarafından toplu alımlar yapılabildiği belirlenmiş olup, bu toplu satışların tahminleme üzerinde yanıltıcı bir etki yaratmaması amacıyla IBM SPSS Modeler içerisindeki "Auto Data Prep" modülü kullanımı ile aykırı verilerin temizlenmesi sağlanmıştır.

Veri yapısı bozuk ya da eksik kayıtların, tahminlemeyi bozmaması amacıyla Auto Data Prep modülü kullanılarak modelin çalışma performansı ve doğruluğu arttırılmıştır. Temizlenen veriler dışarı alınmış ve modellerde nihai halleri kullanılarak daha performanslı ve hızlı şekilde işlem yapılmıştır. Kayıtlar temizlenerek analiz hızının optimize edilmesi bu modül ile sağlanmıştır.



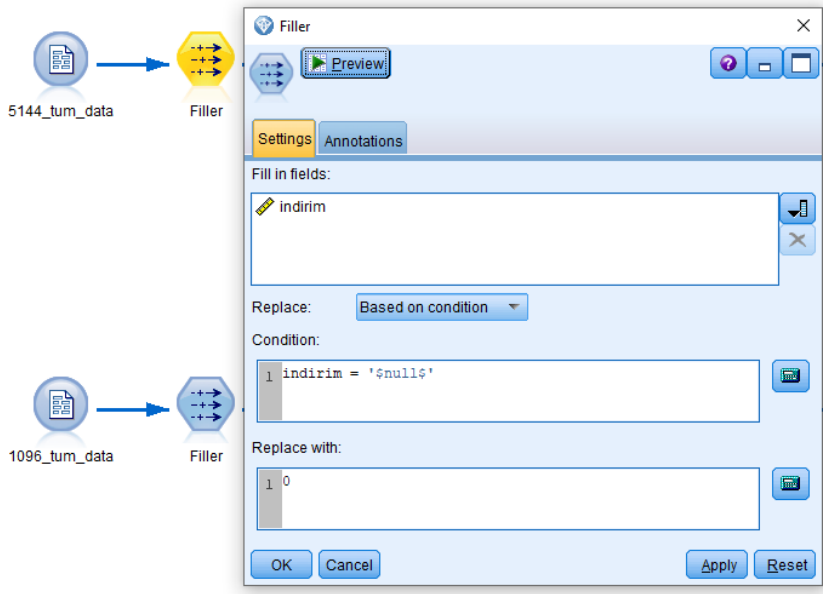
Şekil 3.6. IBM SPSS Modeler ile Auto Data Prep Kullanımı



Şekil 3.7. Aykırı Değerler Temizlenmiş Veriler

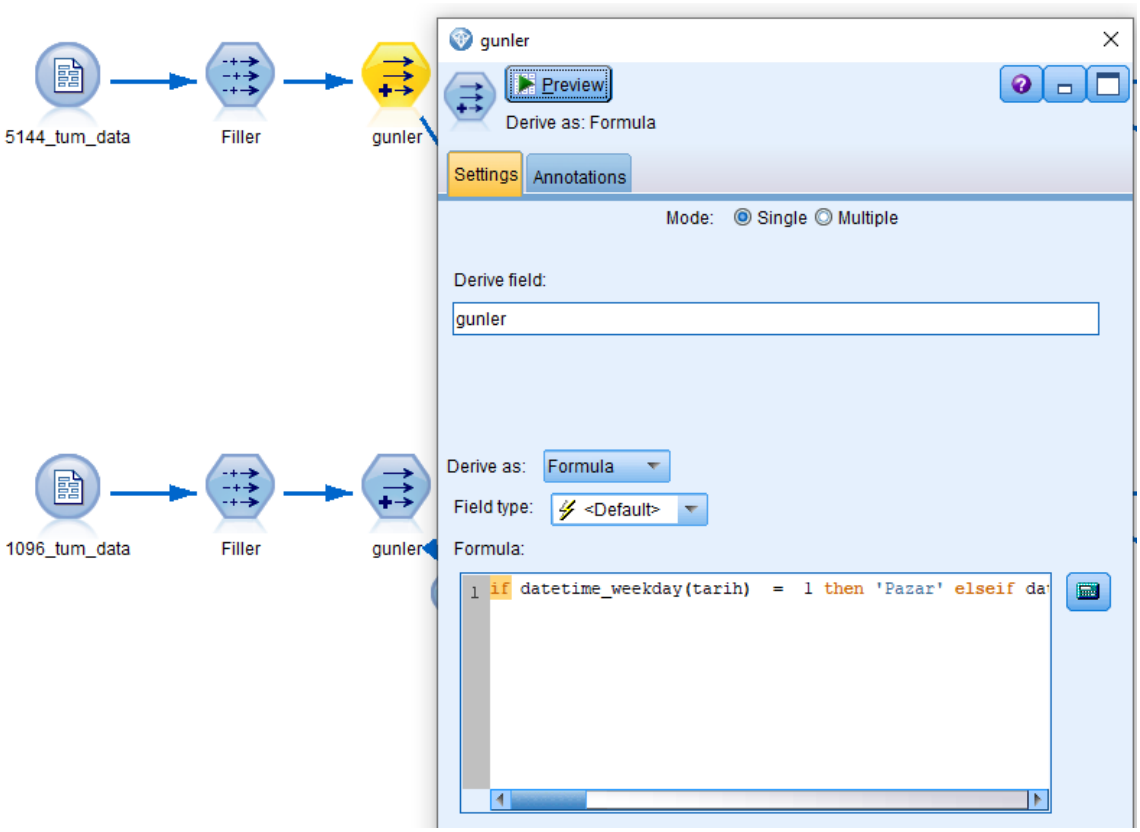
"Auto Data Prep" modülü ile verilerin düzenlenmesi, temizlenmesi ve performans arttırma işlemleri, her bir modelde tekrar tekrar çalıştırılarak sürenin uzamasının engellenmesi amacıyla, tek bir kez kullanılmış ve modellerde kullanılmaya hazır veri setinin txt formatında, mağazalar özelinde (1096 ve 5144 numaralı iki mağaza için de) çıktısı alınmıştır.

Birleştirilen veriler üzerinde, veri setindeki değişkenlerin değerlerini yeni bir değer ile değiştirmek için kullanılan "filler" komutu kullanılarak indirim olmayan (indirim değeri boş olan) kayıtların 0 olacak şekilde ataması yapıldı.



Şekil 3.8. IBM SPSS Modeler ile Filler Komutu Kullanımı

Verilerin temizlenmesi, seçilmesi ve birleştirilmesi işlemleri ayrı ayrı yapıldı. Veri kümesi içerisinde satışlar mağaza bazında günlük olarak yer almaktadır. Yeni örüntü oluşturulabileceği düşünülerek satışların haftanın hangi günü (Pazartesi, Salı vb.) olduğu bilgileri tespit edilerek veri kümesine eklenmiştir. Değişkenler üzerinde hesaplamalar yapmak, koşullar koymak ve yeni değişkenler oluşturmak için kullanılan "derive" komutu kullanarak gün bilgileri bulunup "günler" isimli kolon ile veri kümesine eklenmiştir. Gün bilgilerini eklemek amacıyla "Derive" komutu içerisindeki formül bölümünde "Datetime_weekday" fonksiyonu kullanılmış ve tarihten haftanın kaçıncı günü (1.2. gibi) olduğu bilgisi alınmıştır.



Şekil 3.9. IBM SPSS Modeler ile Derive Komutu Kullanımı

Veri kümesi içerisinde bulunan bayram, yılbaşı, kadınlar günü gibi özel gün bilgileri, özel günün var olup olmadığı kontrolü yapılarak veri kümesine ozelgunflag isimli bir kolon olarak eklenmiştir. Özel gün sezonsallığı oluşturabileceği için özel gün detayından var olup olmadığı bilgisi üretilmiştir.

Mağaza satış verileri üzerinde yapılan işlemler tamamlanarak tahminleme modellerinde kullanılmak üzere hazır hale getirilmiştir. Veri kümesinin eklenen kolonlar sonrasındaki nihai hali aşağıdaki tablodaki gibidir.

Tablo 3.3. Veri Kümesi Son Hali

Tarih	Yıl ay gün formatında satışın gerçekleştiği tarihi ifade etmektedir.
Mgz	Satış yapılan mağaza numarasıdır.
Mgz_adi	Satış yapılan mağaza adıdır.
Kasa	Satış yapılan yazarkasa numarasıdır.
Fisno	Satış işlem numarasıdır.
Saat	Satış işlem saatini ifade etmektedir.
Anagrupno	Satış yapılan ürünün ana grup numarasıdır.
Altgrupno	Satış yapılan ürünün alt grup numarasıdır.
Sınıfno	Satış yapılan ürünün sınıf numarasıdır.
Altsınıfno	Satış yapılan ürünün alt sınıf numarasıdır.
Urun	Satış yapılan ürün adını belirtmektedir.
İndirim	Satış işleminde herhangi bir indirim olduğu durumda bu alanda yer alıyor.

Miktar	Satış yapılan ürün miktarını ifade etmektedir.
Tutar	Satış tutarını belirtmektedir.
KatalogKampFlag	Katalog kampanyası olup olmadığı bilgisidir.
SanalFiziksel	Sanal mı fiziksel mi satış olduğunu ifade eder.
OzelGun	Özel gün bilgileri yer almaktadır.
Günler	Gün bilgileri yer almaktadır. (Pazartesi, Salı, vb.)
Ozelgunflag	Özel gün olup olmadığı bilgisi eklenmiştir.

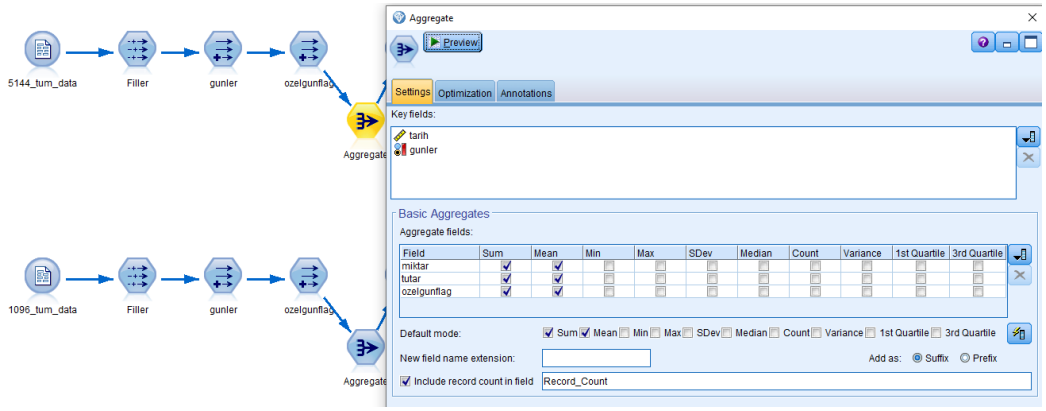
3.3. Yöntem

Ön işleme adımlarından geçirilerek hazırlanan veriler üzerinde IBM SPSS Modeler içerisindeki Regression (Regresyon) ve Time Series (Zaman Serisi) modellerinin satış tahminlemesinde kullanılmasına karar verilmiştir. Bu modeller belirlenirken, literatür taramalarında karşılaşılan araştırmalarda da bu modellerin öne çıkmış olması ve pek çok satış tahminleme modeli için hata payı düşük oranlar vermiş olması belirleyici olmuştur. İlgili modeller kullanılarak eldeki veriler üzerinden tutar hesaplaması yapılmıştır. Çıkan sonuçlar karşılaştırılmış ve doğruluk oranları hesaplanmıştır.

BÖLÜM 4. BULGULAR VE YORUMLAR

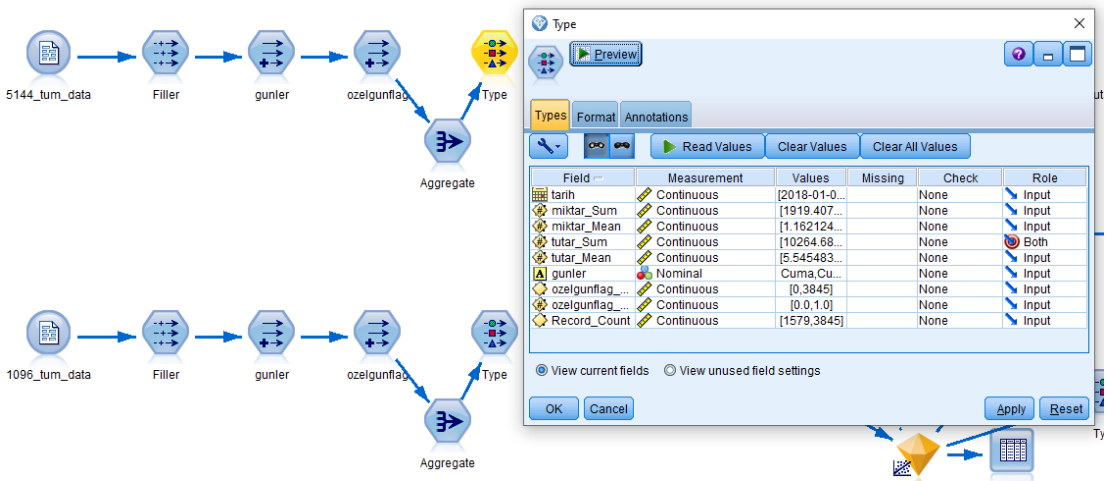
4.1.Regresyon İle Satış Tahmini Uygulaması

Çalışma içerisinde kullanılacak olan regresyon modeli için IBM SPSS Modeler üzerinde bulunan Aggregate komutu kullanılarak tarih ve günler bazında miktar, tutar ve ozelgunflag kolonlarının ortalama ve toplam değerleri alınmıştır.



Şekil 4.10. IBM SPSS Modeler ile Aggregate Komutu Kullanımı

Regresyonda kullanılacak girdi ve çıktı değerleri Type komutu ile belirlenerek veri tipleri düzenlenmiştir. Tahmin yapılacak alan olan Tutar verisi hem girdi hem de hedef alan olarak belirlenmiştir.



Şekil 4.11. IBM SPSS Modeler ile Type Komutu Kullanımı

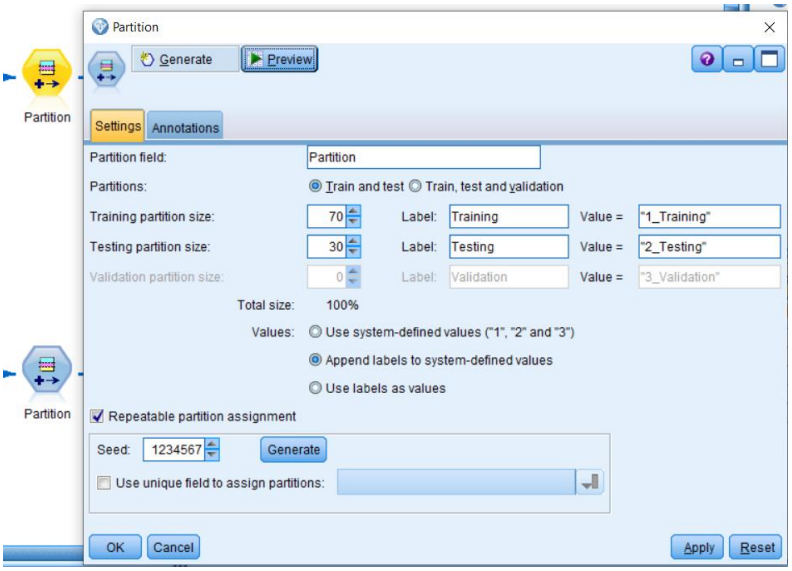
Modelin eldeki verileri kullanarak eğitilmesi için veri seti, test ve eğitim verisi olarak parçalanmıştır.

Veri seti: Veri elemanlarının bir araya gelerek oluşturduğu bir bütündür.

Test seti: Eğitim seti kullanılarak oluşturulan modelin değerlendirilmesinde kullanılan veri kümesidir.

Eğitim Seti: Modelde eğitmek için kullanılacak veri kümesidir.

Yukarıda belirtilen veri seti özellikleri dikkate alınarak, IBM SPSS Modeler üzerinde bulunan Partition komutu eklenerek %70 %30 oranında rastgele olarak test ve eğitim için bölünmesi yapılmıştır. Veri setine Partition değerinin eklenmesinden sonra Type komutu ile tekrar veri tipleri düzenlenerek modele uygun hale getirilmiştir.



Şekil 4.12. IBM SPSS Modeler ile Partition Komutu Kullanımı

Migros Tic. A.Ş firmasına ait satış hacimleri farklı iki ayrı mağazanın sezonsallık sağlaması açısından 2018 ve 2019 yıllarına ait iki senelik verileri kullanılarak regresyon modeli ile satış tahmini yapılması hedeflenmiştir

Satış hacmi yüksek olan 1096 numaralı mağaza için regresyon modeli kurularak eldeki iki yılın verileri ile satış tahmini yapılmıştır. Model beta katsayıları ve standart hata değerleri aşağıdaki şekildeki gibi bulunmuştur.

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	30569,371	12715,361		2,404	,017
	miktar_Sum	9,621	,056	,819	170,658	,000
	miktar_Mean	-301942,237	10937,110	-,116	-27,607	,000
	tutar_Mean	28973,031	353,025	,351	82,071	,000
	ozelgunflag_Sum	1,630	,122	,209	13,416	,000
	ozelgunflag_Mean	-42565,774	3519,480	-,184	-12,094	,000

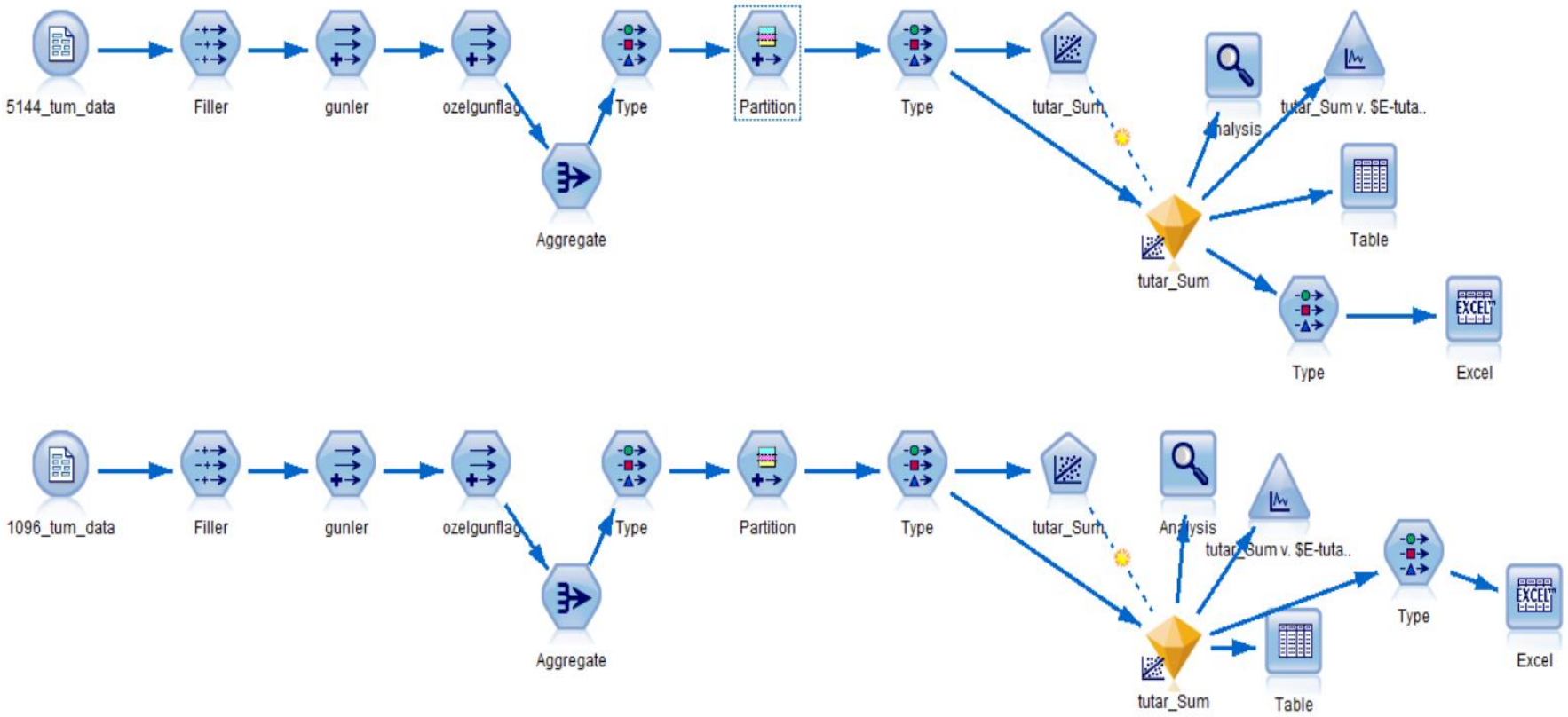
Şekil 4.13. 1096 Numaralı Mağaza Regresyon Model Katsayıları

Satış hacmi daha küçük olan 5144 numaralı mağaza için de regresyon modeli kurularak 2018 ve 2019 verileri ile satış tahmini yapılmıştır. Beta, standart hata gibi model değer sonuçları 5144 numaralı mağaza için aşağıdaki gibidir.

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2490,306	885,044		2,814	,005
	miktar_Sum	6,698	,046	,811	144,476	,000
	miktar_Mean	-20188,851	745,086	-,129	-27,096	,000
	tutar_Mean	2733,241	30,171	,412	90,592	,000
	ozelgunflag_Sum	,646	,105	,107	6,144	,000
	ozelgunflag_Mean	-1697,631	306,101	-,094	-5,546	,000

Şekil 4.14. 5144 Numaralı Mağaza Regresyon Model Katsayıları



Şekil 4.15. IBM SPSS Modeler ile Regresyon Modelinin Kullanımı

Yapılan tahmin sonuçları ile analiz edilerek tahmin edilen tutarın da yer aldığı veri seti detaylı analiz amacıyla excel ile dışarı alınmıştır. Modelin başarı oranları ve gelecek tahmini model çıktıları kullanılarak hesaplanmıştır.

Excel üzerinde gün bazında modelde kullanılan girdilerin ortalama değerleri hesaplanmıştır.

Tablo 4.4. 1096 Numaralı Mağaza İçin Modelin Girdilerinin Gün Bazında Ortalamaları (2018-2019)

	miktar_Sum	miktar_Mean	tutar_Mean	ozelgunflag_Sum	ozelgunflag_Mean
Pazartesi	34.929,13	1,27	11,66	6.164,13	0,21
Salı	30.522,57	1,26	11,68	5.024,28	0,20
Çarşamba	29.535,94	1,26	11,67	3.581,68	0,16
Perşembe	30.246,39	1,27	11,88	4.254,36	0,17
Cuma	33.070,37	1,27	12,28	4.442,13	0,17
Cumartesi	43.081,78	1,27	12,49	5.982,92	0,17
Pazar	44.153,65	1,27	12,13	7.390,45	0,20

Modelin çıktılarından alınan ve Şekil 14 te gösterilen beta değerleri ile girdi değerlerinin ortalamaları katsayıları ile çarpılarak regresyon formülü ile 2018 ve 2019 yıllarının değerleri kullanılarak 2020 yılına ait ilk yedi günlük satışın tahmini hesaplanmıştır.

Çok değişkenli doğrusal regresyon formülü:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_K X_K + C$$

Bu formüle göre gün bazında hesaplamalar yapılmıştır.

Pazartesi = 1,079 + (34929,13 x 9,621) + (1,27 x -301942,24) + (11,66 x 28973,031) + (6164,13 x 1,63) + (0,21 x -42565,774) + 30569,371 = 322.245,57

Salı = 1,079 + (30522,57 x 9,621) + (1,26 x -301942, 24) + (11,68 x 28973,031) + (5024,28 x 1,63) + (0,20 x -42565,774) + 30569,371 =280.640,85

Çarşamba = 1,079 + (29535,94 x 9,621) + (1,26 x -301942,24) + (11,67 x 28973,031) + (3581,68 x 1,63) + (0,16 x -42565,774) + 30569,371 = 271.537,23

Perşembe = 1,079 + (30246,39 x 9,621) + (1,27 x -301942,24) + (11,88 x 28973,031) + (4254,36 x 1,63) + (0,17 x -42565,774) + 30569,371 = 283.091,68

Cuma = 1,079 + (33070,37 x 9,621) + (1,27 x -301942,24) + (12,28 x 28973,031) + (4442,13 x 1,63) + (0,17 x -42565,774) + 30569,371 = 321.931,63

Cumartesi = 1,079 + (43081,78 x 9,621) + (1,27 x -301942,24) + (12,49 x 28973,031) + (5982,92 x 1,63) + (0,17 x -42565,774) + 30569,371 = 427.277,75

Pazar = 1,079 + (44153,65 x 9,621) + (1,27 x -301942,24) + (12,13 x 28973,031) + (7390,45 x 1,63) + (0,20 x -42565,774) + 30569,371 = 426.591,17

2020 yılının ilk 7 günü için regresyon formülü ile hesaplanan tahmini satış bilgileri tabloda gösterilmiştir.

Tablo 4.5. 2018-2019 Verileri Kullanılarak Hesaplanan 1096 Numaralı Mağazanın 2020 Satış Verileri

1.Oca.20	Çarşamba	271.537,23
2.Oca.20	Perşembe	283.091,68
3.Oca.20	Cuma	321.931,63
4.Oca.20	Cumartesi	427.277,75
5.Oca.20	Pazar	426.591,17
6.Oca.20	Pazartesi	322.245,57
7.Oca.20	Salı	280.640,85

5144 numaralı mağaza için yine excel üzerinde gün bazında girdi değer ortalamaları alınarak regresyon formülünde yerine koyarak 2020 yılının ilk hafta satışları tahmin edilmiştir.

Tablo 4.6. 5144 Numaralı Mağaza İçin Modelin Girdilerinin Gün Bazında Ortalamaları (2018-2019)

	miktar_Sum	miktar_Mean	tutar_Mean	ozelgunflag_Sum	ozelgunflag_Mean
Pazartesi	3.277,52	1,25	8,21	613,47	0,21
Salı	3.053,30	1,25	8,16	546,15	0,20
Çarşamba	2.971,97	1,24	8,13	382,17	0,15
Perşembe	2.990,31	1,25	8,26	430,63	0,17
Cuma	3.137,67	1,25	8,17	448,89	0,17
Cumartesi	4.208,95	1,26	8,40	569,20	0,17
Pazar	4.276,91	1,27	8,49	733,25	0,20

Formüle göre yapılan hesaplama aşağıdaki gibidir.

Pazartesi = 1,107 + (3277,52 x 6,698) + (1,25 x -20188,851) + (8,21 x 2733,241) + (613,47 x 0,646) + (0,21 x -1697,631) + 2490,306 = 21.686,43

Salı = 1,107 + (3053,30 x 6,698) + (1,25 x -20188,851) + (8,16 x 2733,241) + (546,15 x 0,646) + (0,20 x -1697,631) + 2490,306 =20.076,15

Çarşamba = 1,107 + (2971,97 x 6,698) + (1,24 x -20188,851) + (8,13 x 2733,241) + (382,17 x 0,646) + (0,15 x -1697,631) + 2490,306 = 19.570,56

Perşembe = 1,107 + (2990,31 x 6,698) + (1,25 x -20188,851) + (8,26 x 2733,241) + (430,63 x 0,646) + (0,17 x -1697,631) + 2490,306 = 19.767,27

Cuma = 1,107 + (3137,67 x 6,698) + (1,25 x -20188,851) + (8,17 x 2733,241) + (448,89 x 0,646) + (0,17 x -1697,631) + 2490,306 = 20.632,62

Cumartesi = 1,107 + (4208,95 x 6,698) + (1,26 x -20188,851) + (8,40 x 2733,241) + (569,20 x 0,646) + (0,17 x -1697,631) + 2490,306 = 28.239,83

Pazar = 1,107 + (4276,91 x 6,698) + (1,27 x -20188,851) + (8,49 x 2733,241) + (733,25 x 0,646) + (0,20 x -1697,631) + 2490,306 = 28.871,44

2020 yılının ilk 7 günü için regresyon formülü ile 5144 numaralı mağaza için hesaplanan tahmini satış bilgileri tabloda gösterilmiştir.

Tablo 4.7. 2018-2019 Verileri Kullanılarak Hesaplanan 5144 Numaralı Mağazanın 2020 Satış Verileri

1.Oca.20	Çarşamba	19.570,56
2.Oca.20	Perşembe	19.767,27
3.Oca.20	Cuma	20.632,62
4.Oca.20	Cumartesi	28.239,83
5.Oca.20	Pazar	28.871,44
6.Oca.20	Pazartesi	21.686,43
7.Oca.20	Salı	20.076,15

Çalışmada hataların ölçümü için sıklıkla kullanılan ortalama mutlak yüzde hata (MAPE) hesaplamasının kullanılmasına karar verilmiştir. Tahmin hata oranlarını bulmak için ortalama mutlak yüzde hata MAPE ölçütü kullanılmıştır. Ortalama mutlak yüzde hata ile model başarısının ölçümü excel ile yapılmıştır. Tahmin modellerinin doğruluğunun ölçümlerinde sıkça kullanılan MAPE değerinin %10'un altında kalınması modelin yüksek doğruluğa sahip olduğunu göstermektedir. %10 ile %20 arasında kalan hata oranlarına sahip modeller ise doğru tahmin modeli olarak sınıflandırılmaktadır. (KARAHAN, 2015)

Her iki mağaza için çalıştırılan model sonuç grafikleri ve MAPE değerleri aşağıdaki gibidir. MAPE Değerleri Gün bazında da hesaplanmış ve aşağıdaki tabloda gösterilmiştir.

Tablo 4.8. Genel MAPE Oranları (2018-2019)

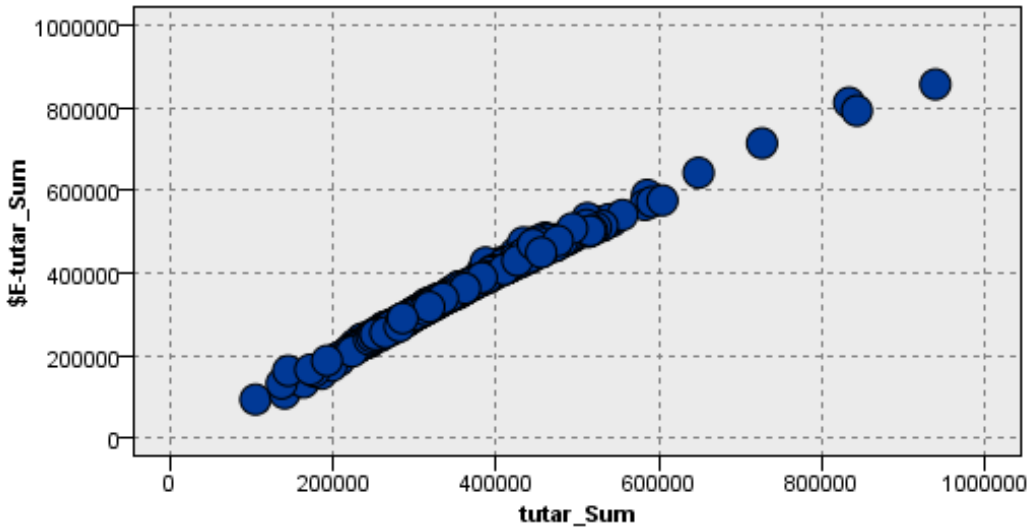
REGRESYON MODELİ (2018-2019)	
MAPE GENEL 1096	**MAPE GENEL 5144**
1,5%	2%

Tablo 4 bize modelin hata paylarının ne kadar düşük olduğunu göstermektedir. Bu durumda regresyon modeli tahnminlemede oldukça yüksek başarı yakalamıştır. Her iki mağaza için de regresyon modelinin hata oranları oldukça düşük çıkmış olup regresyon modeli satış tahminlemesinde tercih edilebilir olarak görülmektedir. Genel hata payları biraz daha detaylandırılarak gün bazında hata oranları tespit edilmiş olup her iki mağaza için ayrı ayrı Tablo 5 üzerinde gösterilmiştir.

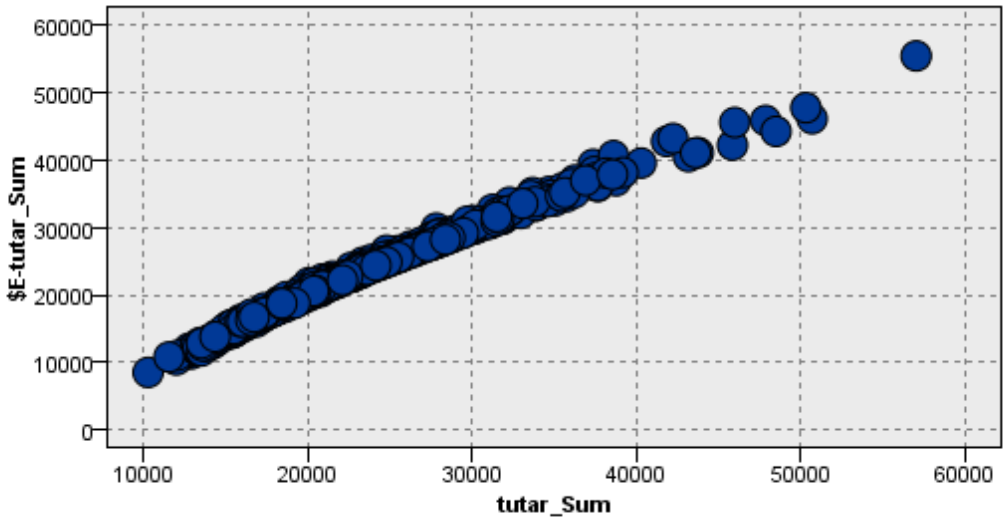
Tablo 4.9. Mağaza ve Gün Bazında Regresyon MAPE Değerleri

	MAPE Regresyon (2018-2019)	
	1096 numaralı mağaza	**5144 numaralı mağaza**
Pazartesi	1,00%	1,80%
Salı	1,70%	2,30%
Çarşamba	1,80%	2,20%
Perşembe	1,70%	2,50%
Cuma	1,00%	1,70%
Cumartesi	1,30%	1,90%
Pazar	1,80%	1,80%

Regresyon modelindeki genel ve gün bazında hata hesaplamaları tabloları incelendiği zaman satış hacmi yüksek olan 1096 numaralı mağazamız için modelin daha başarılı olduğu tespit edilmiştir.



Şekil 4.16. 1096 Numaralı Mağaza Regresyon Grafiği (2018-2019)



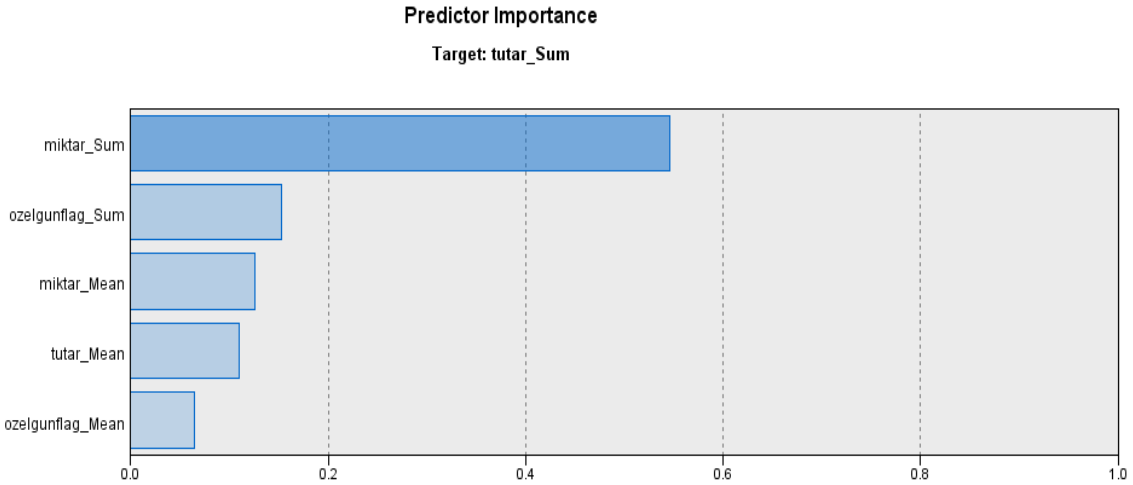
Şekil 4.17. 5144 Numaralı Mağaza Regresyon Grafiği (2018-2019)

İlgili grafikler ve MAPE değerleri incelendiğinde Regresyon modelinin başarı oranının oldukça yüksek olduğu görülmektedir. Hata oranları dikkate alındığında gerçeğe yakın tahminde bulunduğu kabul edilebilmektedir. Grafikler incelendiğinde doğrusal olarak ilerlediği ve gerçekleşen tutar ile tahmin edilen tutarın birbiri üzerine örtüştüğü görülmektedir. 1096 numaralı mağazada %99, 5144 numaralı mağazada ise %98 oranında başarı elde edilmesi sonucunda grafiklerde herhangi bir sapma tespit edilmemiştir.

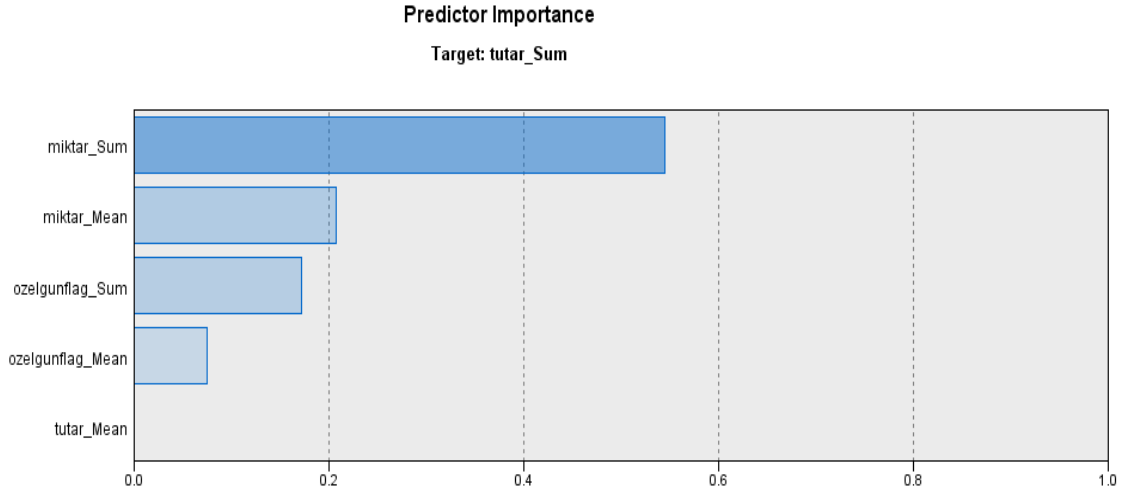
Tablo 4.10. Mağaza Bazında Regresyon Model Doğrulukları

	Regresyon Model Sonuçları			
	1096 Numaralı Mağaza		**5144 Numaralı Mağaza**	
	Eğitim Veri Seti	**Test Veri Seti**	**Eğitim Veri Seti**	**Test Veri Seti**
Minimum Hata	-39.298.659,00	-41.248.045,00	-1.847.838,00	-2112.0
Maksimum Hata	81.483.712,00	29.749.065,00	4.567.127,00	2.645.768,00
Ortalama Hata	0.0	-690.67	-0.0	-39.355,00
Ortalama Mutlak Hata	4.646.537,00	5.002.052,00	438.36	399.861,00
Standart Sapma	7.767.137,00	7.839.521,00	668.991,00	564.644,00
Doğrusal Korelasyon	0.996	0.996	0.995	0.996
Veri Sayısı	516,00	214,00	516,00	214,00

Oluşturulan modelin doğrulukları Tablo 6 içerisinde verilmiştir. Regresyon modelinin 1096 numaralı mağaza için eğitim ve test veri setinde korelasyon katsayısı 0.996, 5144 numaralı mağaza için eğitim veri setinde 0.995 test veri setinde 0.996 olarak hesaplanmıştır. Bağımlı değişkenler ile bağımsız değişkenler arasında çok güçlü sayılabilecek bir ilişki mevcuttur.

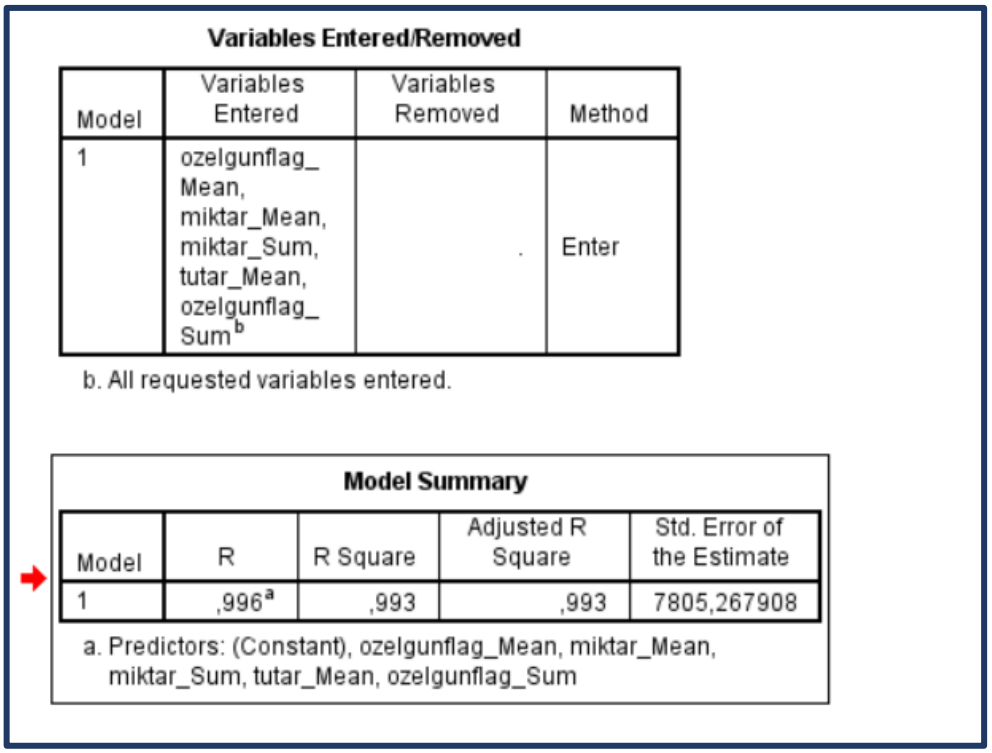


Şekil 4.18. 1096 Numaralı Mağaza Regresyon Modeli Girdi Önem Sıralaması



Şekil 4.190. 5144 Numaralı Mağaza Regresyon Modeli Girdi Önem Sıralaması

Regresyon modeli ile satış tahmininde girdi olarak kullanılan verilerin önem derecelerine bakıldığında miktar_sum (miktar toplamı) en önemli veri girdisi olarak kabul edilmektedir. 1096 numaralı mağaza için ikinci sırada özel gün bilgisi gelmektedir.



Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	ozelgunflag_ Mean, miktar_Mean, miktar_Sum, tutar_Mean, ozelgunflag_ Sum[b]	.	Enter

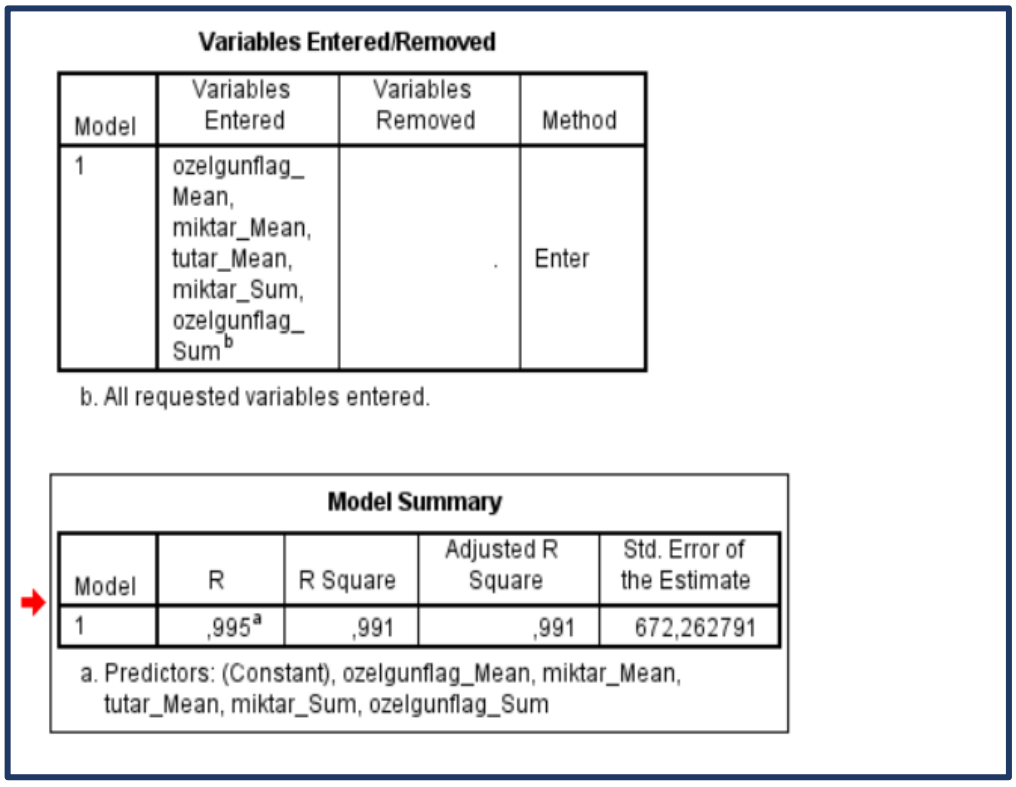
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,996[a]	,993	,993	7805,267908

a. Predictors: (Constant), ozelgunflag_Mean, miktar_Mean, miktar_Sum, tutar_Mean, ozelgunflag_Sum

Şekil 4.201. 1096 numaralı Mağazanın Model Çıktıları (2018-2019)



Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	ozelgunflag_ Mean, miktar_Mean, tutar_Mean, miktar_Sum, ozelgunflag_ Sum[b]	.	Enter

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,995[a]	,991	,991	672,262791

a. Predictors: (Constant), ozelgunflag_Mean, miktar_Mean, tutar_Mean, miktar_Sum, ozelgunflag_Sum

Şekil 4.212. 5144 numaralı Mağazanın Model Çıktıları (2018-2019)

Model çıktıları incelendiğinde her iki mağaza için de R kare değerlerinin çok yüksek olduğu gözlemlenmiştir. Satış hacmi yüksek olan 1096 numaralı mağazanın da daha düşük olan 5144 numaralı mağazanın da R-kare değerleri %99 olarak çıkmıştır. Bu oran, tahminleme konusunda modelin başarısını göstermektedir. Her iki mağaza için de modelin sildiği herhangi bir girdi verisi olmamıştır. Girilen tüm veriler model içerisinde kullanılmıştır. Yöntem olarak, tüm tahmin değişkenlerinin eş zamanlı olarak modele girdiğini varsayan enter metodu seçilmiştir.

4.1.1. 2019 Yılı Verisi İle Regresyon Modeli

Sezonsallık yakalaması amaçlanarak iki yıllık veri kullanılarak çalıştırılan regresyon modeli sonrasında sadece 2019 verisi ile modelin hata oranları hesaplanmıştır. Her iki mağaza için regresyon modeli kurularak eldeki 2019 yılının verileri ile satış tahmini yapılmıştır. Model beta katsayıları ve standart hata değerleri aşağıdaki şekillerdeki gibi bulunmuştur.

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	26626,933	14526,347		1,833	,068
	miktar_Sum	9,822	,047	,894	208,969	,000
	miktar_Mean	-295792,049	11044,821	-,101	-26,781	,000
	tutar_Mean	28094,708	325,200	,330	86,392	,000
	ozelgunflag_Sum	-,423	,134	-,057	-3,147	,002
	ozelgunflag_Mean	12011,058	3899,740	,055	3,080	,002

Şekil 4.223. 1096 Numaralı Mağaza Regresyon Model Katsayıları

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1596,383	1012,226		1,577	,116
	miktar_Sum	7,045	,052	,908	134,648	,000
	miktar_Mean	-23738,362	812,461	-,170	-29,218	,000
	tutar_Mean	3193,242	47,185	,363	67,676	,000
	ozelgunflag_Sum	,386	,122	,071	3,149	,002
	ozelgunflag_Mean	-1115,340	395,952	-,061	-2,817	,005

Şekil 4.234. 5144 Numaralı Mağaza Regresyon Model Katsayıları

Ver seti detaylı analiz ve hesaplamalar için dışarı alınmış ve excel üzerinde regresyon formülü kullanılarak 2020 yılının ilk gününe ait satış bilgileri tahmin edilmiştir.

Modelde kullanılan girdilerin gün bazında ortalamaları hesaplanmış ve tabloda gösterilmiştir.

Tablo 4.11. 1096 Numaralı Mağaza İçin Modelin Girdilerinin Gün Bazında Ortalamaları (2019)

	miktar_Sum	miktar_Mean	tutar_Mean	ozelgunflag_Sum	ozelgunflag_Mean
Pazartesi	35.990,53	1,29	12,07	5.272,46	0,17
Salı	31.338,48	1,29	12,12	5.409,23	0,21
Çarşamba	30.549,28	1,28	12,03	3.734,16	0,16
Perşembe	30.868,90	1,28	12,27	3.813,82	0,16
Cuma	34.559,85	1,29	12,84	5.243,44	0,19
Cumartesi	44.543,29	1,29	13,07	5.363,71	0,16
Pazar	45.017,50	1,30	12,59	7.767,56	0,21

Model çıktılarında yer alan beta ve katsayı değerleri ile hesaplanan girdi ortalamaları formülde yerine konularak hesaplamalar yapılmıştır.

Çok değişkenli doğrusal regresyon formülü:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_K X_K + C$$

1096 numaralı mağaza için gün bazında yapılan hesaplamalar;

Pazartesi = 1,121 + (35990,53 x 9,822) + (1,29 x -295792,05) + (12,07 x 28094,708) + (5272,46 x - 0,423) + (0,17 x 12011,058) + 26626,933 = 336.750,69

Salı = 1,121 + (31338,48 x 9,822) + (1,29 x -295792,05) + (12,12 x 28094,708) + (5409,23 x - 0,423) + (0,21 x 12011,058) + 26626,933 =294.909,02

Çarşamba = 1,121 + (30549,28 x 9,822) + (1,28 x -295792,05) + (12,03 x 28094,708) + (3734,16 x - 0,423) + (0,16 x 12011,058) + 26626,933 = 285.849,05

Perşembe = 1,121 + (30868,90 x 9,822) + (1,28 x -295792,05) + (12,27 x 28094,708) + (3813,82 x - 0,423) + (0,16 x 12011,058) + 26626,933 = 295.143,49

Cuma = 1,121 + (34559,85 x 9,822) + (1,29 x -295792,05) + (12,84 x 28094,708) + (5243,44 x - 0,423) + (0,19 x 12011,058) + 26626,933 = 344.894,53

Cumartesi = 1,121 + (44543,29 x 9,822) + (1,29 x -295792,05) + (13,07 x 28094,708) + (5363,71 x - 0,423) + (0,16 x 12011,058) + 26626,933 = 448.350,01

Pazar = 1,121 + (45017,50 x 9,822) + (1,30 x -295792,05) + (12,59 x 28094,708) + (7767,56 x - 0,423) + (0,21 x 12011,058) + 26626,933 = 437.632,12

2020 yılının ilk 7 günü için regresyon formülü ile hesaplanan 1096 numaralı mağazanın tahmini satış bilgileri tabloda gösterilmiştir.

Tablo 4.12. 2019 Verileri Kullanılarak Hesaplanan 1096 Numaralı Mağazanın 2020 Satış Verileri

1.Oca.20	Çarşamba	285.849,05
2.Oca.20	Perşembe	295.143,49
3.Oca.20	Cuma	344.894,53
4.Oca.20	Cumartesi	448.350,01
5.Oca.20	Pazar	437.632,12
6.Oca.20	Pazartesi	336.750,69
7.Oca.20	Salı	294.909,02

5144 numaralı mağaza için de gün bazında girdi değer ortalamaları excel üzerinde hesaplanarak regresyon formülü ile 2020 yılının ilk hafta satışları tahmin edilmiştir.

Tablo 4.13. 5144 Numaralı Mağaza İçin Modelin Girdilerinin Gün Bazında Ortalamaları (2019)

	miktar_Sum	miktar_Mean	tutar_Mean	ozelgunflag_Sum	ozelgunflag_Mean
Pazartesi	3.577,62	1,25	8,72	542,50	0,17
Salı	3.384,49	1,26	8,60	638,75	0,21
Çarşamba	3.219,77	1,24	8,62	414,77	0,15
Perşembe	3.252,18	1,25	8,79	416,46	0,15
Cuma	3.457,54	1,26	8,70	549,90	0,19
Cumartesi	4.676,59	1,27	9,00	555,81	0,15
Pazar	4.620,50	1,27	9,03	833,06	0,21

5144 numaralı mağaza için gün bazında yapılan hesaplamalar;

Pazartesi = 1,111 + (3577,62 x 7,045) + (1,25 x -23738,36) + (8,72 x 3193,242) + (542,50 x 0,386) + (0,17 x -1115,34) + 1596,383 = 24.924,25

Salı = 1,111 + (3384,49 x 7,045) + (1,26 x -23738,36) + (8,60 x 3193,242) + (638,75 x 0,386) + (0,21 x -1115,34) + 1596,383 = 23.111,78

Çarşamba = 1,111 + (3219,77 x 7,045) + (1,24 x -23738,36) + (8,62 x 3193,242) + (414,77 x 0,386) + (0,15 x -1115,34) + 1596,383 = 22.358,47

Perşembe = 1,111 + (3252,18 x 7,045) + (1,25 x -23738,36) + (8,79 x 3193,242) + (416,46 x 0,386) + (0,15 x -1115,34) + 1596,383 = 22.868,10

Cuma = 1,111 + (3457,54 x 7,045) + (1,26 x -23738,36) + (8,70 x 3193,242) + (549,90 x 0,386) + (0,19 x -1115,34) + 1596,383 = 23.935,00

Cumartesi = 1,111 + (4676,59 x 7,045) + (1,27 x -23738,36) + (9,00 x 3193,242) + (555,81 x 0,386) + (0,15 x -1115,34) + 1596,383 = 33.149,05

Pazar = 1,111 + (4620,50 x 7,045) + (1,27 x -23738,36) + (9,03 x 3193,242) + (833,06 x 0,386) + (0,21 x -1115,34) + 1596,383 = 33.010,89

2020 yılının ilk 7 günü için regresyon formülü ile 5144 numaralı mağaza için hesaplanan tahmini satış bilgileri tabloda gösterilmiştir.

Tablo 4.14. 2019 Verileri Kullanılarak Hesaplanan 5144 Numaralı Mağazanın 2020 Satış Verileri

1.Oca.20	Çarşamba	22.358,47
2.Oca.20	Perşembe	22.868,10
3.Oca.20	Cuma	23.935,00
4.Oca.20	Cumartesi	33.149,05
5.Oca.20	Pazar	33.010,89
6.Oca.20	Pazartesi	24.924,25
7.Oca.20	Salı	23.111,78

Her iki mağaza için çalıştırılan 2019 verilerine ait model sonuç grafikleri ve MAPE değerleri aşağıdaki gibidir. MAPE Değerleri Gün bazında da hesaplanmış ve tabloda gösterilmiştir.

Tablo 4.15. Genel MAPE Oranları (2019)

REGRESYON MODELİ (2019)	
MAPE GENEL 1096	**MAPE GENEL 5144**
1,2%	1,4%

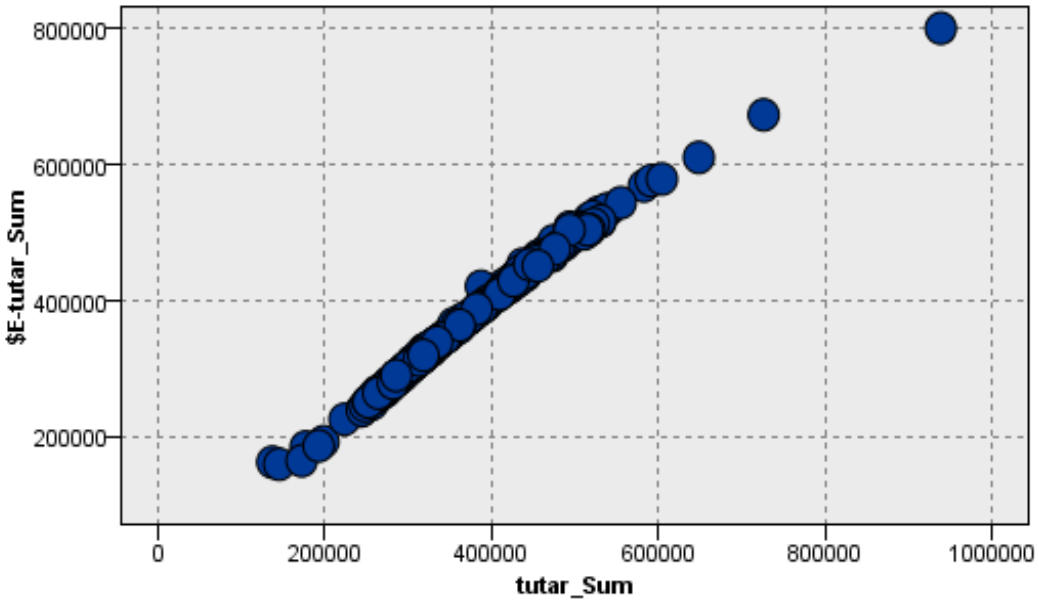
Tablo 4.16. Mağaza ve Gün Bazında Regresyon MAPE Değerleri (2019)

	MAPE Regresyon (2019)	
	1096 numaralı mağaza	**5144 numaralı mağaza**
Pazartesi	0,6%	1,0%
Salı	1,5%	1,1%
Çarşamba	1,3%	1,3%
Perşembe	1,5%	1,5%
Cuma	0,8%	1,0%
Cumartesi	1,1%	2,0%
Pazar	1,7%	1,9%

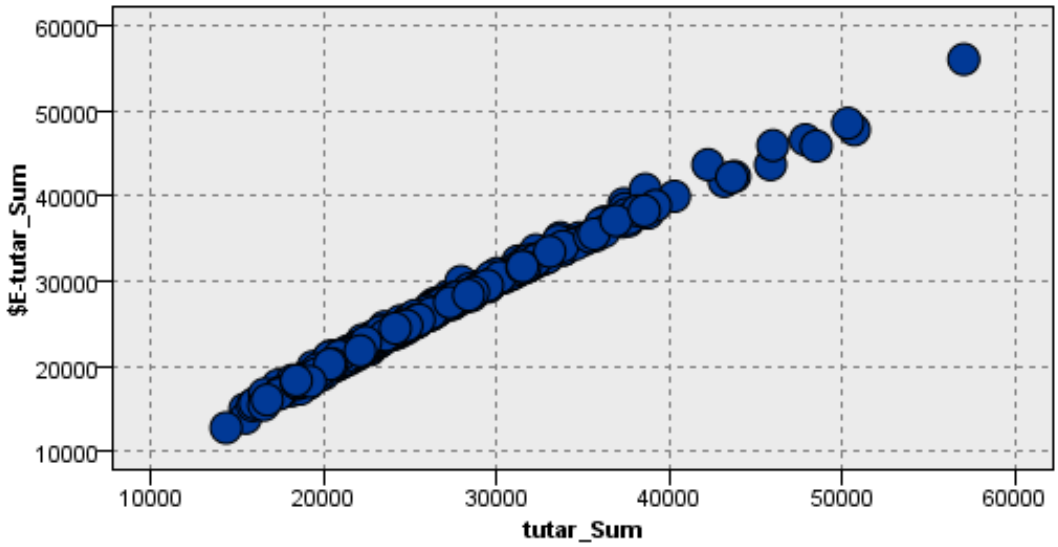
2019 MAPE tablosu incelendiği zaman tek yıllık veri ile hata oranlarının iki yıllık veriye kıyasla daha düşük olduğu tespit edilmiştir. Veride sezonluk yakalanması için iki yıllık veri kullanılmış. Fakat 2019 verisi kendi içerisinde değerlendirildiği zaman artan enflasyon ve yıl içerisinde yaşanan satış artışı trendi yakalanmış olup modelin başarısının arttığı görülmüştür.

Hata oranlarının yüksek olduğu günler 1096 numaralı mağaza için Pazar günü olurken 5144 numaralı mağazada Cumartesi ve Pazar günleri olduğu tespit edilmiştir. İlgili günlerin verileri incelendiği zaman model, satış artışı sebebiyle hafta sonu yaşanan dalgalanmalardan dolayı hata oranlarında yükseliş yaşamıştır.

Her iki mağaza için 2019 yılı verileri kullanılarak oluşturulan modelin çıktıları analiz edilmek üzere IBM SPSS Modeler ile grafikleri alınmıştır.



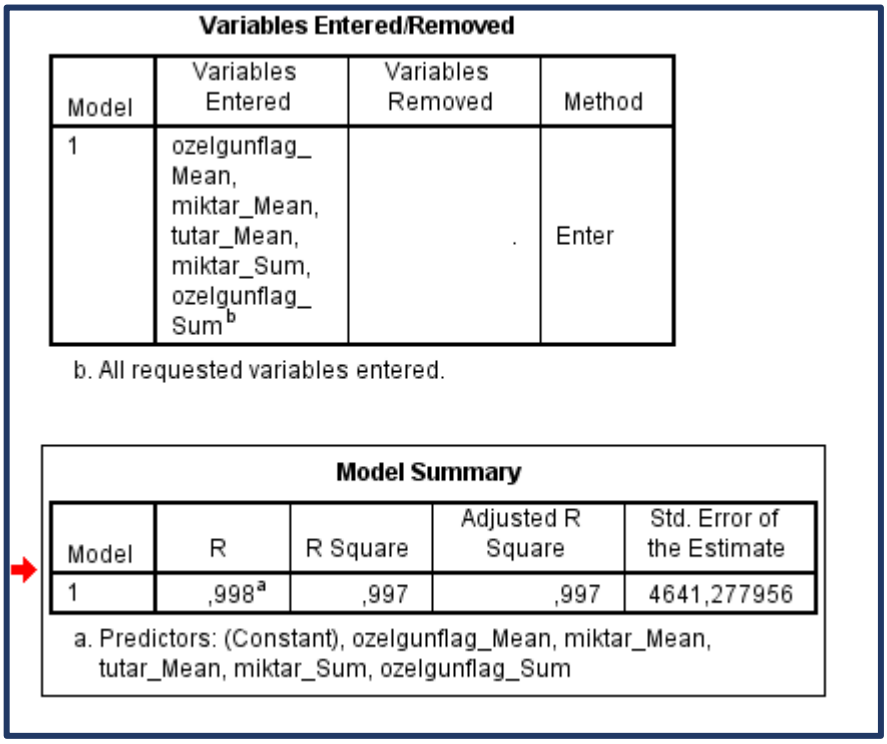
Şekil 4.245. 1096 Numaralı Mağaza Regresyon Grafiği (2019)



Şekil 4.256. 5144 Numaralı Mağaza Regresyon Grafiği (2019)

İlgili grafikler incelendiğinde 2018-2019 yılları verisi ile yapılan modele göre pek bir fark olmadığı tespit edilmiştir.

Model başarısı her iki uygulama içinde yüksek olmuştur. Sezonsallığın yakalanması için tek seneye ait olan verinin yeterli olduğu tespit edilmiş olup, özel gün işaretlerinin etkilerinin olduğu öngörülmüştür.



Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	ozelgunflag_Mean, miktar_Mean, tutar_Mean, miktar_Sum, ozelgunflag_Sum[b]	.	Enter

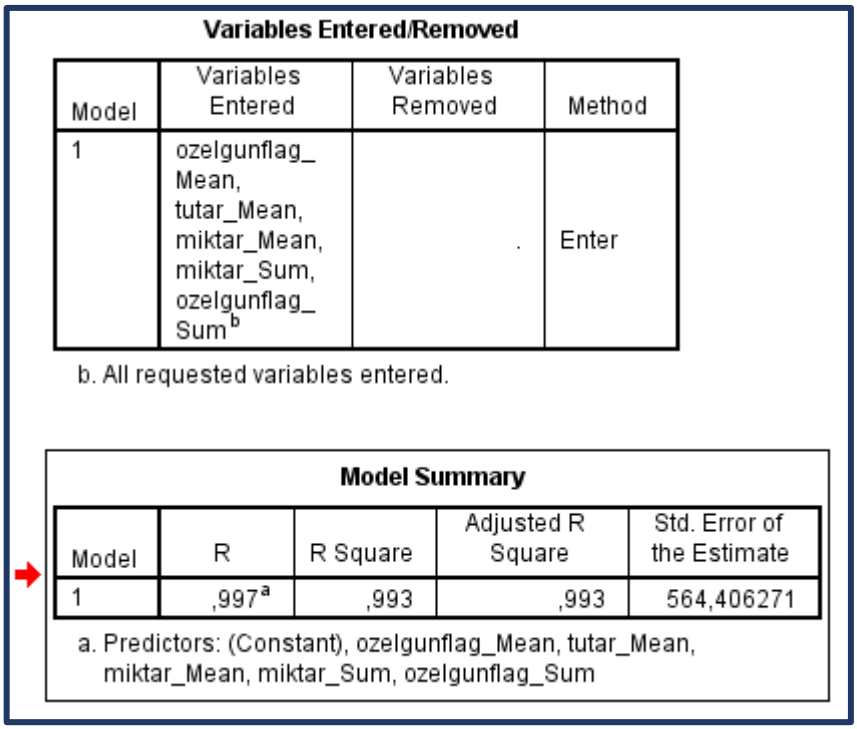
b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,998[a]	,997	,997	4641,277956

a. Predictors: (Constant), ozelgunflag_Mean, miktar_Mean, tutar_Mean, miktar_Sum, ozelgunflag_Sum

Şekil 4.267. 1096 numaralı Mağazanın Model Çıktıları (2019)



Variables Entered/Removed

Model	Variables Entered	Variables Removed	Method
1	ozelgunflag_ Mean, tutar_Mean, miktar_Mean, miktar_Sum, ozelgunflag_ Sum[b]	.	Enter

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,997[a]	,993	,993	564,406271

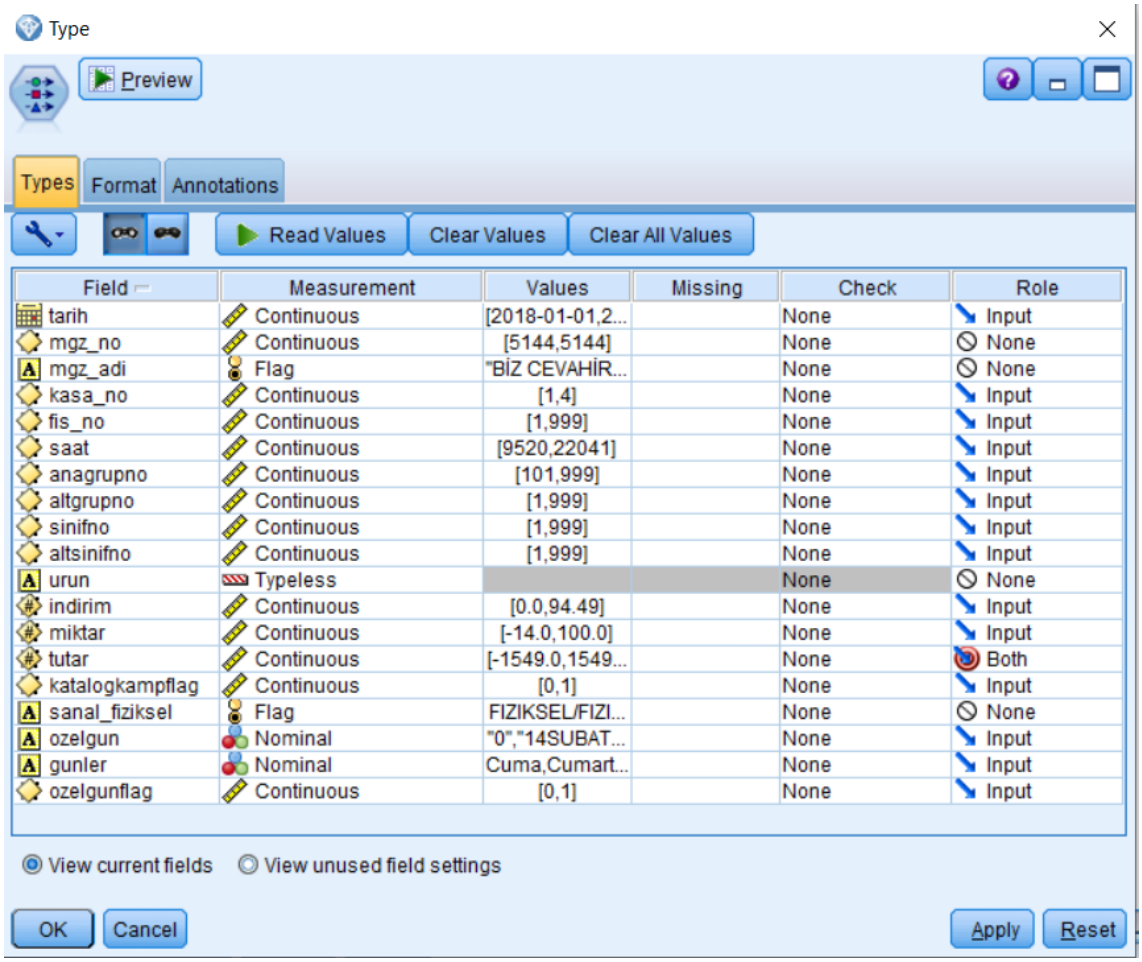
a. Predictors: (Constant), ozelgunflag_Mean, tutar_Mean, miktar_Mean, miktar_Sum, ozelgunflag_Sum

Şekil 4.278. 5144 numaralı Mağazanın Model Çıktıları (2019)

Model çıktıları R-kare değerlerinin her iki mağaza için de yüksek olduğunu göstermektedir. Her iki mağaza için R-kare değerleri %99 olarak çıkmıştır. Regresyon modeli elimizdeki veriler ile satış tahmininde yüksek başarı göstermiştir. Her iki mağaza için de girdi verileri içerisinden modelin sildiği bir veri olmamıştır ve tüm veriler kullanılmıştır.

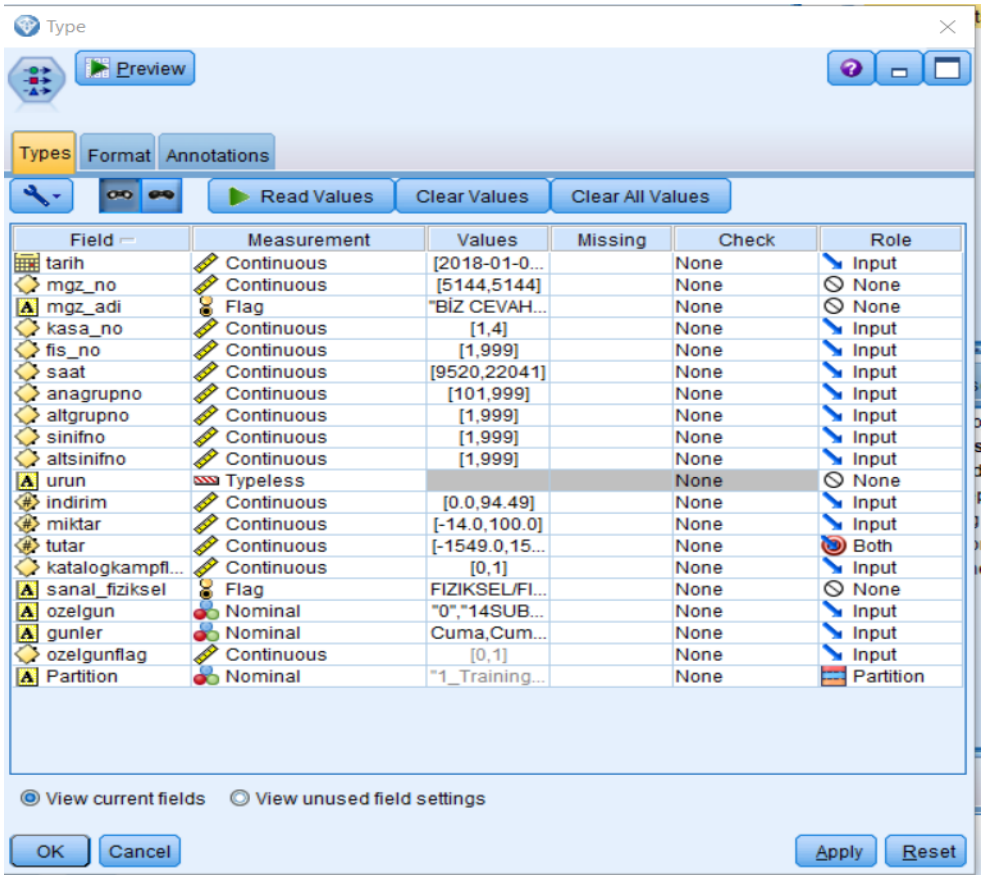
4.2.Zaman Serileri İle Satış Tahmini Uygulaması

Zaman serisinde(Time Series) kullanılacak girdi ve çıktı değerleri Type komutu ile belirlenerek veri tipleri düzenlenmiştir. Tahmin yapılacak alan olan Tutar verisi hem girdi hem de hedef alan olarak belirlenmiştir. Mağaza numarası, mağaza adi, ürün adı ve sanal fiziksel işareti veride örüntü kazandırmayacağı için çıkarılmıştır.



Şekil 4.289. Zaman Serisi Modeli Girdi Çıktı Değerleri

Model test ve eğitim veri seti olarak partition komutu kullanarak ayrılmıştır. Model eğitim veri setinin kullanarak eğitilecek ve test verisi ile doğruluğunu kontrol edecektir. IBM SPSS Modeler üzerinde bulunan Partition komutu eklenerek %70 %30 oranında rastgele olarak test ve eğitim seti olarak ayrılmıştır. Partition komutu ile ayırma işlemleri tamamlandıktan sonra Type komutu kullanılarak veri tipleri düzenlenmiş ve girdi çıktı verileri son halini alarak modele girmeye hazır hale getirilmiştir.



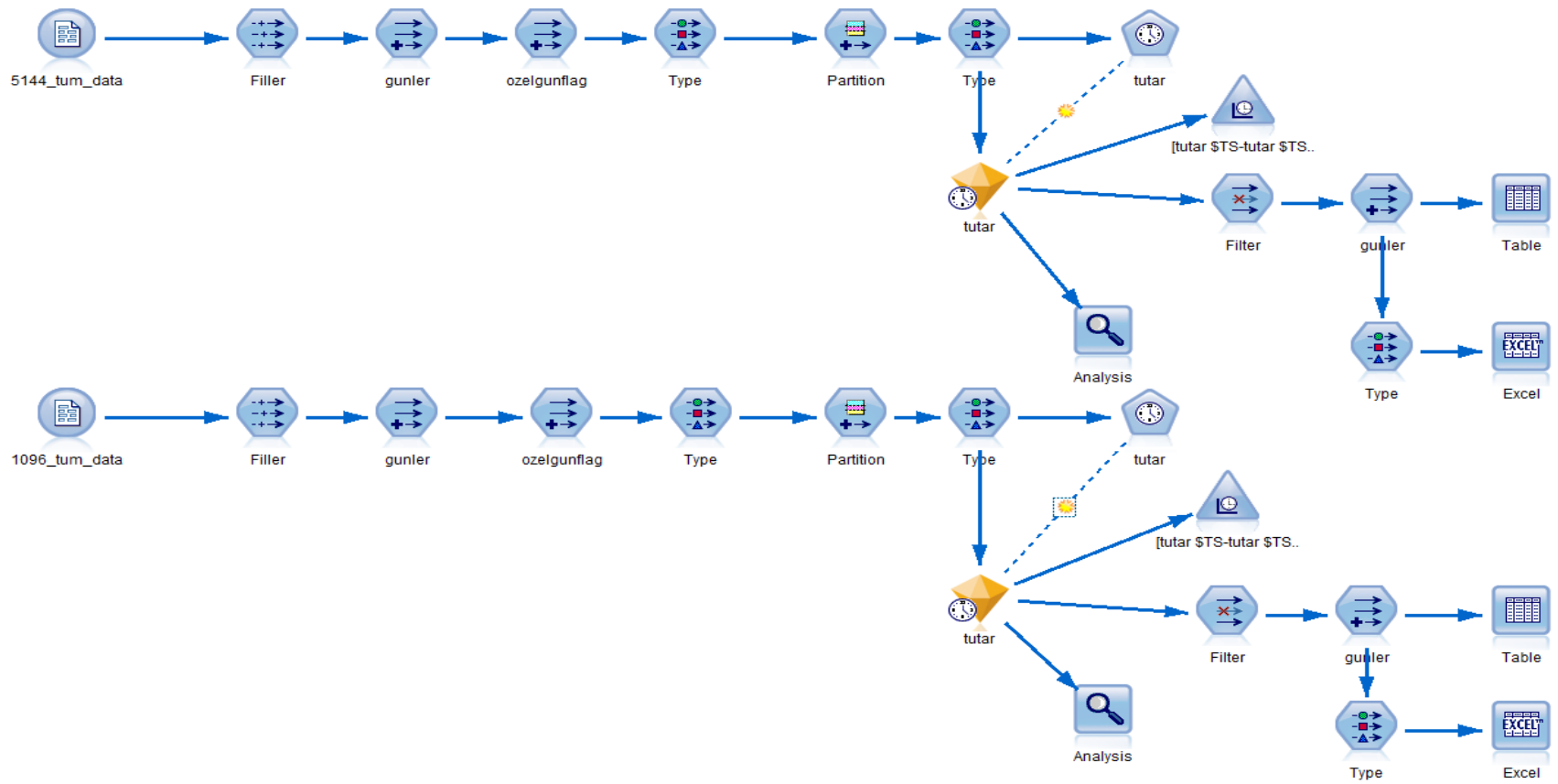
Şekil 4.290. Modele Girecek Veri Tiplerinin Nihai Hali

Migros Tic. A.Ş firmasına ait satış hacimleri farklı olan 1096 ve 5144 numaralı mağazaların 2018 ve 2019 yıllarına ait veriler kullanılarak zaman serisi modeli ile satış tahmini yapılması kararı verilmiştir. Veri setinde bulunan Tutar kolonu hem hedef hem de kaynak alan olarak değerlendirilmiştir.

Satış hacmi yüksek olan 1096 numaralı mağaza için zaman serisi modeli kurularak eldeki iki yılın verileri ile satış tahmini yapılmıştır.

Satış hacmi daha küçük olan 5144 numaralı mağaza için de zaman serisi modeli kurularak 2018 ve 2019 verileri ile satış tahmini yapılmıştır.

Model ile çıkan sonuçların analizi için sonuç veri seti excel ile dışarı alınarak belirlenen hata hesaplama yöntemi ile ilgili hesaplamalar yapılarak modelin başarısı analiz edilmiştir. IBM SPSS Modeler üzerindeki grafikler ile gerçek tutar ile tahmin edilen tutar karşılaştırmaları yapılmıştır.



Şekil 4.301. IBM SPSS Modeler ile Zaman Serisi Modelinin Kullanımı

Model kurallarında gelecek 7 gün için tahmin yapması seçilmiş olup 2018 ve 2019 yılları kullanılarak 2020 yılının mağaza bazında satış tahminleri tabloda gösterilmiştir.

Tablo 4.17. 1096 ve 5144 Numaralı Mağazaların 2020 Satış Tahminleri (2018-2019)

Tarih	Gün	1096 Numaralı Mağaza	5144 Numaralı Mağaza
1.Oca.20	Çarşamba	456.803,08	24.389,95
2.Oca.20	Perşembe	442.943,63	24.161,33
3.Oca.20	Cuma	507.347,83	24.642,14
4.Oca.20	Cumartesi	640.937,77	31.869,25
5.Oca.20	Pazar	637.009,02	31.567,70
6.Oca.20	Pazartesi	500.068,19	24.242,49
7.Oca.20	Salı	447.090,33	20.845,43

Çalışmada hataların ölçümü için karar verilen MAPE hesaplaması kullanılmıştır. Ortalama mutlak yüzde hata (MAPE) değerleri excel üzerinde hesaplanmıştır.

İki mağaza için çalıştırılan model hata değeri ve grafikleri aşağıdaki gibidir.

Tablo 4.18. Mağaza Bazında Genel MAPE Oranları

	ZAMAN SERİSİ MODELİ (2018-2019)		
	MAPE TS	MAPE TSLCI	MAPE TSUCI
1096 NUMARALI MAĞAZA	3,3%	6,5%	6,5%
5144 NUMARALI MAĞAZA	4,8%	9,9%	9,6%

Time Series modelinde modelin başarısı(MAPE TS), alt limit(MAPE TSLCI) ve üst limit(MAPE TSUCI) değerlerinin başarısı ölçülmüştür. Alt ve üst sınırı hesaplanması için modelde seçim yapılmıştır. İyimser ve kötümser tahmin olan bu değerlerin arasında kalan modelin tahmini doğru bir tahminleme aralığında olduğunu göstermektedir.

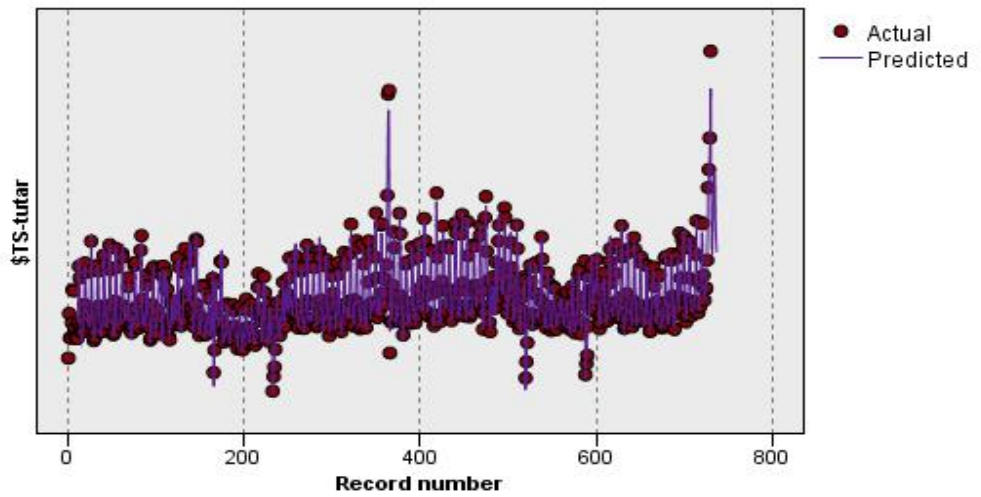
Modelin hata payı değerleri incelendiğinde yüksek başarı elde edildiğine karar verilmiştir. Modelin her iki mağazada da hata payı düşük çıkmış olup çalışılabilir bir model olduğu kabul edilmiştir. Genel olarak verilen hata değerleri gün ve mağaza bazında ayrıştırılarak Tablo 9 üzerinde gösterilmiştir. Tabloda gün bazında iki mağaza içinde hesaplanmış modelin hata payı, alt limit hata payı ve üst limit hata payı değerleri mevcuttur.

Tablo 4.19. Mağaza ve Gün Bazında Time Series MAPE Değerleri

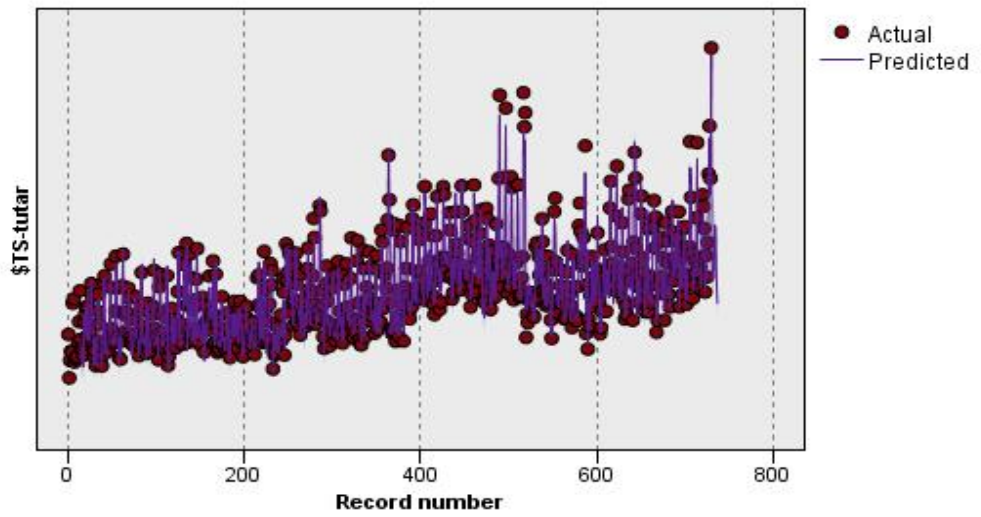
	Time Series Modeli (2018-2019)	
	1096 Numaralı Mağaza	5144 Numaralı Mağaza
	MAPE TS	MAPE TS
Pazartesi	2,9%	4,9%
Salı	3,9%	6,2%
Çarşamba	3,5%	4,4%
Perşembe	3,5%	5,0%
Cuma	3,1%	4,8%
Cumartesi	2,8%	4,5%
Pazar	3,1%	3,8%

Zaman Serisi modelindeki genel ve gün bazında hata hesaplamaları tabloları incelendiği zaman 1096 numaralı satışları daha fazla olan mağazamız için modelin hata payının daha düşük olduğu belirlenmiştir.

Her iki mağaza için analiz edilmek üzere IBM SPSS Modeler ile grafikler alınmıştır.



Şekil 4.312. 1096 Numaralı Mağaza Zaman Serisi Grafiği (2018-2019)

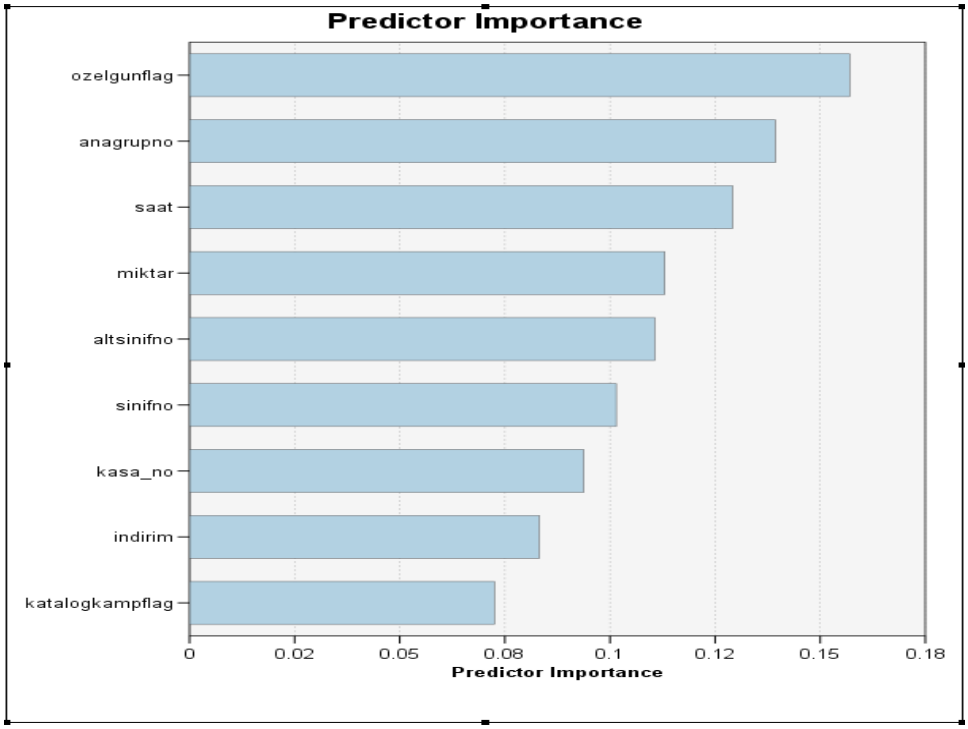


Şekil 4.323. 5144 Numaralı Mağaza Zaman Serisi Grafiği (2018-2019)

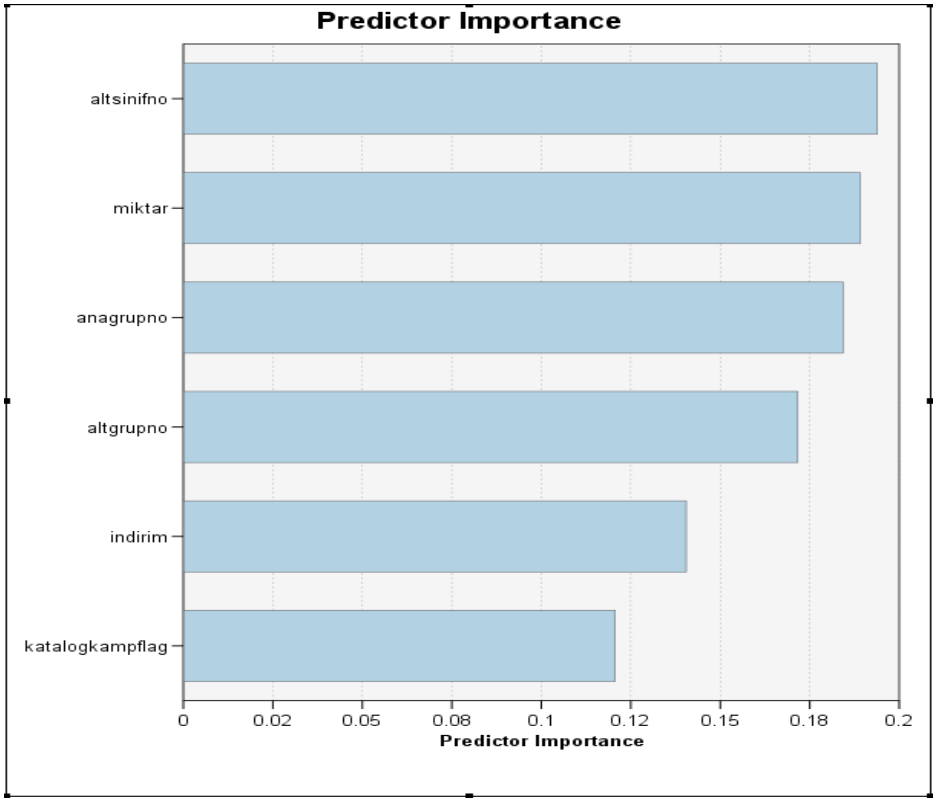
Zaman serisi grafikleri ve hesaplanan MAPE değerleri incelendiğinde zaman serisi modelinin başarı oranının yüksek olduğu ve tercih edilebilir olduğu tespit edilmiştir. Grafiklerde geçekleşen ve tahmin edilen arasında sapma farklıları görülmektedir.

5144 numaralı mağazanın sapma yaşadığı tarihler incelendiği zaman bayrama denk gelen tarihlerde normalin üzerinde ve bir önceki seneden de farklı bir artış olduğu için model bu tarihleri yakalayamamış ve gerçekleşen değerin altında kaldığı tespit edilmiştir. Bu durum yok satışına sebep olarak hata payını arttırmaktadır.

1096 numaralı mağazanın sapma tarihleri incelendiği zaman yılbaşı olduğu tespit edilmiş olup, satış hacmi büyük olan bu mağazada önceki yıldan çok daha fazla bir artış olduğu için sezonsallık yakalanamamış ve tahmin gerçekleşen değerin altında kalmıştır. Bu sapmalar modelin hata payını arttırmaktadır. Beklenenden farklı olarak gerçekleşen bu dalgalanmalar, ilgili tarihlerdeki hata payının yükselmesini açıklamaktadır.



Şekil 4.334. 1096 Numaralı Mağaza Zaman Serisi Modeli Girdi Önem Sıralaması



Şekil 4.345. 5144 Numaralı Mağaza Zaman Serisi Modeli Girdi Önem Sıralaması

Zaman Serisi modeli ile satış tahmininde girdi olarak kullanılan verilerin önem derecelerine bakıldığında 1096 numaralı mağaza içinozelgunflag (özel gün olup olmadığı) en önemli veri girdisi olarak kabul edilirken 5144 numaralı mağaza için altsinifno(alt sınıf numarası) veri girdisi kabul edilmektedir.

Seri içerisindeki veriler arasında doğrusal ilişkilerin olduğunu varsayan bir model olan ARIMA modeli zaman serisinde kullanılmıştır. ARIMA modeli en çok tercih edilen bir modeldir. Bu modelde tahmin işlemi, serideki değerler geçmiş verileri değerlendirilerek modellenir.

Model Information

Model Building Method		ARIMA
		Non-seasonal p=1,d=0,q=2; Seasonal p=0,d=1,q=1
Number of Predictors		9
Model Fit	MSE	95,422,505.577
	RMSE	9,768.444
	RMSPE	4.676
	MAE	6,683.808
	MAPE	3.373
	MAXAE	92,127.614
	MAXAPE	30.174
	AIC	13,232.437
	BIC	13,333.150
	R-Squared	0.966
	Stationary R-Squared	0.942
Ljung-Box Q(#)	Statistic	10.873
	df	14.0
	Significance	0.7

Şekil 4.356. 1096 Numaralı Mağazanın Zaman Serisi Model Bilgileri(2018-2019)

Model Information		
Model Building Method		ARIMA
		Non-seasonal p=1,d=0,q=1; Seasonal p=0,d=1,q=1
Number of Predictors		6
Model Fit	MSE	980,469.835
	RMSE	990.187
	RMSPE	7.138
	MAE	735.957
	MAPE	5.508
	MAXAE	4,102.671
	MAXAPE	30.053
	AIC	9,890.664
	BIC	9,950.122
	R-Squared	0.942
	Stationary R-Squared	0.910
Ljung-Box Q(#)	Statistic	27.462
	df	15.0
	Significance	0.0

Şekil 4.367. 5144 Numaralı Mağazanın Zaman Serisi Model Bilgileri(2018-2019)

IBM SPSS Modeler içeresinde yer alan model bilgileri incelendiği zaman Zaman Serisi modelinde R- kare değerinin 1096 numaralı mağaza için 0.966 olduğu, 5144 numaralı mağaza için 0.942 olduğu görülmektedir. R- kare değerinin iki mağaza için de yüksek değere sahip olması modelimizin yüksek başarı ile çalıştığını ifade etmektedir. 1096 numaralı mağaza için %96, 5144 numaralı mağaza için %94 oranında başarılı bir model olarak kabul edilebilir.

4.2.1. 2019 Yılı Verisi İle Zaman Serisi Modeli

İki yıllık veri kullanılarak sezonsallık yakalaması amacıyla çalıştırılan zaman serisi modelinin sadece 2019 verisi ile hata oranları hesaplanmıştır.

Model kurallarında gelecek 7 gün için tahmin yapması seçilerek 2019 yılı verileri kullanılarak 2020 yılının mağaza bazında 7 günlük satış tahminleri tabloda gösterilmiştir.

Tablo 4.20. 1096 ve 5144 Numaralı Mağazaların 2020 Satış Tahminleri (2019)

Tarih	Gün	1096 Numaralı Mağaza	5144 Numaralı Mağaza
1.Oca.20	Çarşamba	467.361,92	23.789,23
2.Oca.20	Perşembe	460.141,47	23.578,66
3.Oca.20	Cuma	463.761,03	24.313,46
4.Oca.20	Cumartesi	475.248,98	30.998,49
5.Oca.20	Pazar	478.798,78	30.685,03
6.Oca.20	Pazartesi	424.412,02	24.857,46
7.Oca.20	Salı	408.766,37	23.168,46

1096 ve 5144 numaralı mağazalar için 2019 verilerine ait zaman serisi modelinin MAPE hesaplaması ile hata oranları belirlenerek tablo oluşturulmuştur.

Tablo 4.21. Zaman Serisi Genel MAPE Değerleri (2019)

	ZAMAN SERİSİ MODELİ (2019)		
	MAPE TS	MAPE TSLCI	MAPE TSUCI
1096 NUMARALI MAĞAZA	3,4%	6,1%	6,5%
5144 NUMARALI MAĞAZA	4,6%	7,9%	8,8%

Gün ve mağaza bazında 2019 yılı için hata değerleri hesaplanmıştır. Tabloda gün bazında iki mağaza içinde hesaplanmış modelin hata payı(MAPE TS), alt limit hata payı(MAPE TSLCI) ve üst limit hata payı(MAPE TSUCI) değerleri mevcuttur. 2019 yılı kullanılarak çalıştırılan modelde sonuçlar iyimser ve kötümser tahmin aralığında kalacak şekilde çıktılar vermiştir.

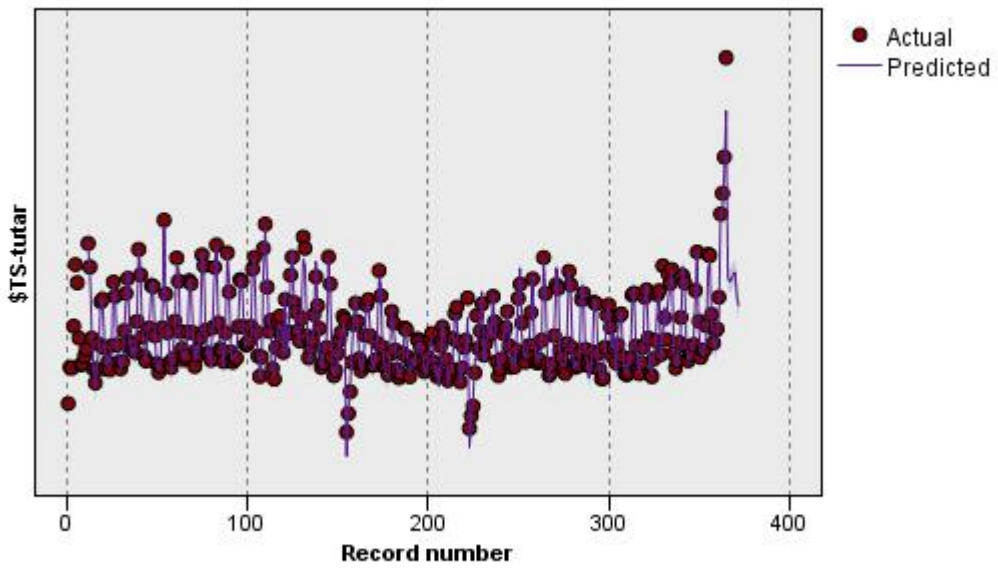
Tablo 4.22. Mağaza ve Gün Bazında Zaman Serisi MAPE Değerleri (2019)

	Time Series Modeli (2019)	
	1096 Numaralı Mağaza	5144 Numaralı Mağaza
	MAPE TS	MAPE TS
Pazartesi	3,1%	4,8%
Salı	4,3%	5,4%
Çarşamba	3,4%	4,1%
Perşembe	3,9%	4,6%
Cuma	3,5%	5,0%
Cumartesi	2,7%	4,0%
Pazar	3,0%	4,2%

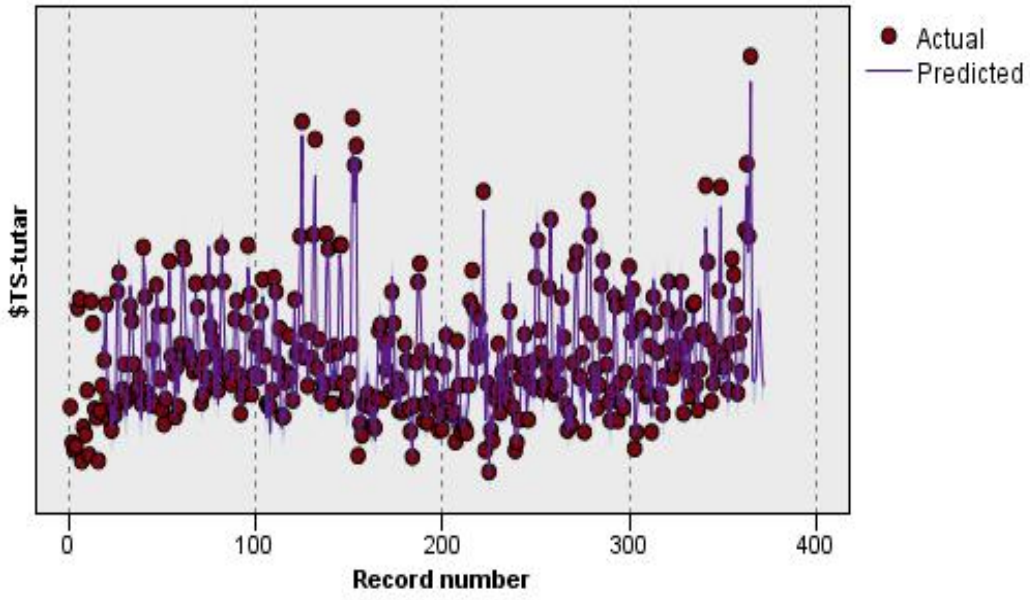
2019 MAPE tablosu incelendiği zaman daha önce iki yıllık veri kullanılarak hesaplanan hata oran tablosuna göre 5144 numaralı mağaza için hata oranlarında düşüş yaşanırken, 1096 numaralı mağazada hata oranlarında artış gözlemlenmiştir. Satış hacmi yüksek ve kampanya dönüşleri yüksek olan 1096 numaralı mağaza için zaman serisi kullanımında tek bir yıllık verinin dikkate alınması hata oranlarını yükseltirken satış hacmi küçük ve kampanya dönüşleri az olan mağazada tek yıllık veri kullanılması hata oranlarını azaltmıştır. Satış dalgalanmalarının yoğun olduğu mağaza olan 1096 numaralı mağaza için bir yıllık veri ile dalgalanmalar tespit edilememiş ve sapmalar yaşanmıştır.

Hata oranlarının yüksek olduğu günler 1096 numaralı mağaza için Salı günü olurken 5144 numaralı mağazada Pazartesi ve Salı günleri olduğu tespit edilmiştir. 5144 numaralı mağaza için ilgili günlerin verileri incelendiği zaman model, satış artışı sebebiyle hafta sonu yaşanan dalgalanmalardan dolayı hafta başında gerçekleşenden yüksek tahminlerde bulunarak hata oranlarında yükseliş yaşamıştır.

Her iki mağaza için 2019 yılı verileri kullanılarak oluşturulan modelin çıktıları analiz edilmek üzere IBM SPSS Modeler ile grafikleri alınmıştır.



Şekil 4.378. 1096 Numaralı Mağaza Zaman Serisi Grafiği (2019)



Şekil 4.389. 5144 Numaralı Mağaza Zaman Serisi Grafiği (2019)

İlgili grafikler incelendiğinde 1096 numaralı mağaza için sadece yılbaşında sapma olduğu belirlenmiştir. Beklenenden yüksek satışların gerçekleşmesi ile model sadece yılbaşında olan satışı yakalayamamış altına kalmıştır. İlgili mağaza satışları normale göre yüksek ve kampanyalara duyarlı bir müşteri kitlesine sahiptir. Yılbaşında gerçekleşen satış 938.963,390 TL iken modelin tahmini 824.320,616 TL olmuştur.

5144 numaralı mağaza için bayram ve yılbaşı haricinde herhangi bir sapma yaşanmamıştır.

Model Information

Model Building Method		ARIMA
		Non-seasonal p=0,d=0,q=4; Seasonal p=0,d=1,q=1
Number of Predictors		2
Model Fit	MSE	138,283,289.290
	RMSE	11,759.392
	RMSPE	5.574
	MAE	7,964.116
	MAPE	3.855
	MAXAE	73,484.996
	MAXAPE	37.604
	AIC	6,644.548
	BIC	6,679.372
	R-Squared	0.952
	Stationary R-Squared	0.916
Ljung-Box Q(#)	Statistic	24.522
	df	13.0
	Significance	0.0

Şekil 4.390. 1096 Numaralı Mağazanın Zaman Serisi Model Bilgileri (2019)

Model Information

Model Building Method		ARIMA
		Non-seasonal p=0,d=0,q=3; Seasonal p=0,d=1,q=1
Number of Predictors		5
Model Fit	MSE	1,317,686.885
	RMSE	1,147.905
	RMSPE	7.330
	MAE	885.521
	MAPE	5.721
	MAXAE	4,685.907
	MAXAPE	27.424
	AIC	4,915.591
	BIC	4,961.818
	R-Squared	0.917
	Stationary R-Squared	0.917
Ljung-Box Q(#)	Statistic	45.524
	df	14.0
	Significance	0.0

Şekil 4.401. 5144 Numaralı Mağazanın Zaman Serisi Model Bilgileri (2019)

2019 yılı için çalıştırılan modelin bilgileri incelendiği zaman Zaman Serisi modelinde R- kare değerinin 1096 numaralı mağaza için 0.952 olduğu, 5144 numaralı mağaza için 0.917 olduğu görülmektedir. R- kare değerinin iki mağaza için de yüksek değere sahip olması modelimizin yüksek başarı ile çalıştığını ifade etmektedir. 1096 numaralı mağaza için %95, 5144 numaralı mağaza için %91 oranında başarılı bir model olarak kabul edilebilir. İki yılın verileri kullanılarak oluşturulan modele göre başarı değerleri düşmüştür. 1096 numaralı mağaza için %1 oranında 5144 numaralı mağaza için ise %3 oranında modelin başarısında düşüş olmuştur.

4.3. Elde Edilen Bulgular

Uygulanan modellerin hata oranları hesaplanmış olup tüm senaryolar için tek bir tablo üzerinde toplanmıştır.

Tablo 4.23. Mağaza, Model ve Yıl Bazında MAPE Değerleri

	REGRESYON MODELİ		**ZAMAN SERİSİ MODELİ**	
	2018-2019 MAPE	**2019 MAPE**	**2018-2019 MAPE TS**	**2019 MAPE TS**
1096 NUMARALI MAĞAZA	1,5%	1,2%	3,3%	3,4%
5144 NUMARALI MAĞAZA	2%	1,4%	4,8%	4,6%

Tablo incelendiği zaman iki modelde de yüksek başarılar elde edilmiş olup 2019 verileri ile daha yüksek başarı yakalanmıştır. Satış trendi 1 yıllık veri üzerinden belirlenmiş ve modellerde etkisini göstermiştir.

Kullanılan modeller incelendiği zaman her iki modelde kullanılabilir oranda hata payı üretmekle birlikte regresyon modelinin hata payı daha düşüktür. Ama zaman serileri modelinden çıkan değerler değerlendirildiği zaman kullanılamaz gibi bir yorum yapılamaz. Her iki model, her iki mağaza için de kullanılabilir hata oranlarına sahiptir.

Kullanılan iki model içerisinden regresyon modelinin 1096 numaralı mağaza için %98,8 ve 5144 numaralı mağaza için ise %98,6 olduğu tespit edilmiştir. Bu sonuca göre

regresyon modelinin Migros verileri için kullanımı zaman serisi modelinden daha uygun olduğu görülmektedir.

Çalışmada ciro üzerinden tahmin yapıldığı için enflasyonun direkt ciroyu etkilediği düşünülerek sadece 2019 verileri kullanılarak da modeller çalıştırılmıştır. Her iki veri seti için de modellerin kullanımı ile 2020 yılının tahmini sonuçları mağaza bazında ayrı tablolarda gösterilmiştir.

Tablo 4.24. 1096 Numaralı Mağaza için 2020 Satış Tahmini

		1096 numaralı mağaza				
		2018-2019 Verileri ile Tahmin		2019 Verileri ile Tahmin		2020
		Regresyon	Zaman Serisi	Regresyon	Zaman Serisi	Gerçekleşen
1.Oca.20	Çarşamba	271.537,23	456.803,08	285.849,05	467.361,92	235.175,42
2.Oca.20	Perşembe	283.091,68	442.943,63	295.143,49	460.141,47	288.678,35
3.Oca.20	Cuma	321.931,63	507.347,83	344.894,53	463.761,03	321.796,40
4.Oca.20	Cumartesi	427.277,75	640.937,77	448.350,01	475.248,98	433.781.21
5.Oca.20	Pazar	426.591,17	637.009,02	437.632,12	478.798,78	456.715,50
6.Oca.20	Pazartesi	322.245,57	500.068,19	336.750,69	424.412,02	325.316,84
7.Oca.20	Salı	280.640,85	447.090,33	294.909,02	408.766,37	299.873,80

1096 numaralı mağaza için tablodaki değerler incelendiğinde zaman serisi modelinde tahmini satış değerlerinin daha yüksek olduğu görülmektedir. Zaman serisi bir önceki gün olan satışların yüksek olduğu 31 Aralıkta yakaladığı trendi geri alamadığı için daha yüksek tutarlar hesaplamıştır. İlgili veri seti içerisine bakıldığında da 2019 yılının başında yapılan tahminler için hata oranlarının yüksek olduğu tespit edilmiştir.

Yılbaşında yapılan satışların ve özel günlerin tekrarlanması ile tespit edilebilir duruma gelmesinden dolayı iki yıl verileri kullanılarak yapılan tahminin daha gerçeğe yakın olduğu tespit edilmiştir. Bu durumda 2018 ve 2019 verileri kullanılarak oluşturulan regresyon modelinde yüksek başarı yakalanmıştır.

Modelin tahmin ettiği tutar, gerçek satıştan çok yüksek değer aldığı durumda stok gün artacağı, tedarikçiden fazla mal alınması durumunda imha sürecine çıkılması ile zarar oluşabileceği için elimizdeki veriler incelendiğinde gerçeğe daha yakın sonuçlar verdiği düşünülerek 1096 numaralı satış hacmi ve kampanya duyarlılığı yüksek olan mağaza için regresyon modelinin kullanılmasının daha uygun olduğuna karar verilmiştir.

Tablo 4.25. 5144 Numaralı Mağaza için 2020 Satış Tahmini

		5144 Numaralı Mağaza				
		2018-2019 Verileri ile Tahmin		2019 Verileri ile Tahmin		2020
		Regresyon	Zaman Serisi	Regresyon	Zaman Serisi	Gerçekleşen
1.Oca.20	Çarşamba	19.570,56	24.389,95	22.358,47	23.789,23	28.019,17
2.Oca.20	Perşembe	19.767,27	24.161,33	22.868,10	23.578,66	18.347,55
3.Oca.20	Cuma	20.632,62	24.642,14	23.935,00	24.313,46	24.393,98
4.Oca.20	Cumartesi	28.239,83	31.869,25	33.149,05	30.998,49	33.748,07
5.Oca.20	Pazar	28.871,44	31.567,70	33.010,89	30.685,03	36.978,40
6.Oca.20	Pazartesi	21.686,43	24.242,49	24.924,25	24.857,46	21.705,11
7.Oca.20	Salı	20.076,15	20.845,43	23.111,78	23.168,46	22.687,74

Koşullar çok değişmediği durumlarda daha verimli olan zaman serisi, satışları hızlı artıp azalmayan ve kampanya duyarlılığı olmayan 5144 numaralı mağaza için daha gerçeğe yakın sonuçlar vermiştir. Yine iki yıl verisi kullanılarak yapılan tahminlerin daha doğru olduğu görülmüş olup bu durumun yılbaşı gibi özel gün satışlarının sezonsallık oluşturarak yakalanabilir olmasından kaynaklı olduğu tespit edilmiştir.

BÖLÜM 5. SONUÇ

Günümüzde depolanan ve bilgiye dönüştürülmek için tutulan birçok veri mevcuttur. Elde edilmesi planlanan bu bilgiler şirketlere amaçlarına yönelik fayda ve kazanç sağlamaktadır. Veri madenciliği bilgiyi elde etmede kullanılan popüler bir tekniktir.

Bu çalışmada Migros Ticaret A.Ş firmasının satış verileri kullanılarak satış tahminlemesi yapılmasında hangi modelin kullanılabileceği bilgisinin bulunması hedeflenmiştir. Çalışma içerisinde model olarak, daha önce yapılan çalışmalar incelenerek pek çok satış tahminlemesinde düşük hata oranı ile kullanılan modeller olan regresyon ve zaman serileri modelleri kullanılmıştır. Bu tekniklerin birlikte kullanıldığı çalışmalara bakıldığında kullanılan teknik ve algoritmaların daha doğru ve anlamlı sonuçlar vermeleri bakımından karşılaştırıldığı görülmüştür.

Çalışmada 1096 ve 5144 numaralı mağazaların satış verileri alınarak ciro bazında satış tahminlemesi yapılmıştır. Kullanılan mağazalar, satış hacmi ve kampanya duyarlılığı konusunda birbirinden farklıdır.

Çalışmada sezonsallık yakalanması amacıyla 2018 ve 2019 yıllarına ait veriler kullanılmıştır. Daha sonra satış trendinin yakalanması, enflasyon artışlarının olması ve artan fiyatların daha gerçeğe yakın sonuç verebilmesi amacıyla sadece 2019 verisi kullanılarak sonuçlar karşılaştırılmıştır. Çalışmada iki farklı model üzerinde çalışılmış olup iki yıllık ve tek yıllık veri kullanılarak uygulamalar yapılmıştır.

Satış tahmini perakende şirketlerinde stok gün sayısını, stok maliyetini ve nakit para akışını azaltmak, mağazalarda bulunurluğu arttırarak müşterinin geldiğinde almak istediği ürünü bulması ile satışın arttırılıp müşteri kaybının önüne geçilmesi konularında çok büyük önem taşımaktadır.

KAYNAKÇA

ARICI, M. S. (2017). Büyük Veri Analizi / Veri Madenciliği ve Makine Öğrenmesi.

ATASEVER, Ü. H. (2011, Ocak). Uydu Görüntülerinin Sınıflandırılmasında Hızlandırma (Boosting), Destek Vektör Makineleri, Rastgele Orman (Random Forest) Ve Regresyon Ağaçları Yöntemlerinin Kullanılması. *Yayınlanmamış Yüksek Lisans Tezi*. Kayseri.

AYDEMİR, E. (2018). *Weka ile Yapay Zeka.* Ankara, Türkiye.

AYGÜN, H., & ADALI, E. (2005). Dempster'in Birleştirme Kuralı İle Sınıflandırma Algoritmalarının Birleştirilmesi. *TBV Bilgisayar Bilimleri Ve Mühendisliği Dergisi*, 99-103.

BİLEKDEMİR, G. (2010). Veri Madenciliği Tekniklerini Kullanarak Üretim Süresi Tahmini Ve Bir Uygulama. *Yayınlanmamış Yüksek Lisans Tezi*. İzmir, Türkiye.

BUDAK, E. Ç. (2013). Üniversite Öğrencileri İçin Bilgisayar Okuryazarlığını Etkileyen Faktörlerin Etkisinin Veri Madenciliği İle Analizi. *AJIT-e: Online Academic Journal of Information Technology*.

BÜCHNER, A., ANAND, S., & HUGHES, J. (1997). Data Mining in Manufacturing Environments: Goals, Techniques and Applications. *Studies in Informatics and Control* .

CAN, Ş., & GERŞİL, M. (2018). Manisa Pamuk Fiyatlarının Zaman Serisi Analizi Ve Yapay Sinir Ağı Teknikleri İle Tahminlenmesi Ve Tahmin Performanslarının Karşılaştırılması. *YÖNETİM VE EKONOMİ*.

ÇALIŞ, A., KAYAPINAR, S., & ÇETİYOKUŞ, T. (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları İle Bilgisayar Ve İnternet Güvenliği Üzerine Bir Uygulama. *Endüstri Mühendisliði Dergisi*, 2-19.

ÇELİK, M. (2009). Veri Madenciliğinde Kullanılan Sınıflandırma Yöntemleri ve Bir Uygulama. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul.

ÇERKEZİ, M. (2013, Haziran). Veri Madenciliği Yöntemlerini Kullanarak Diyabetik Retinopati Hastalığının Teşhisi. *Yayınlanmamış Yüksek Lisans Tezi*.

DEMİRCİ, D. A. (2007). Destek Vektör Makineleri İle Karakter Tanıma. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul.

DURMAZ, C., & KOCAMIŞ, M. (2008, 05). İstanbul Teknik Üniversitesi Fen Edebiyat Fakültesi Oracle Data Miner İle Öğrenci Kayıtları Üzerine Bir Veri Madenciliği Uygulaması. *Bitirme Ödevi*. İstanbul, Türkiye.

EMEL, G. G., & TAŞKIN, Ç. (2005). Veri Madenciliğinde Karar Ağaçları Ve Bir Satış Analizi Uygulaması. *Eskişehir Osmangazi Üniversitesi*.

ERDEM, F., DERİNPINAR, M. A., NASIRZADEHDIZAJI, R., OY, S., ŞEKER, D. Z., & BAYRAM, B. (2018). Rastgele Orman Yöntemi Kullanılarak Kıyı Çizgisi Çıkarımı İstanbul Örneği. *Geomatik Dergisi,* 100-107.

ETGİN, E. E. (2017). Zaman Serilerinde Veri Madenciliği Öngörü Algoritmalarının Etkinlik Ve Verimliliğinin Bist100 Hisse Senetleri Üzerinde Gerçeklenmes. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul.

GEMİCİ, B. (2012). Veri Madenciliği Ve Bir Uygulaması. *Yayınlanmamış Yüksek Lisans Tezi*. İzmir, Türkiye.

IRMAK, S., KÖKSAL, C. D., & ASİLKAN, Ö. (2012). Hastanelerin Gelecekteki Hasta Yoğunluklarının Veri Madenciliği Yöntemleri İle Tahmin Edilmesi. *Uluslararası Alanya İşletme Fakültesi Dergisi*, 101-114.

KARAHAN, M. (2015). Yapay Sinir Ağları Metodu İle İhracat Miktarlarının Tahmini: ARIMA ve YSA Metodunun Karşılaştırmalı Analizi. *EGE AKADEMİK BAKIŞ*, 165-172.

KARAKAŞ, E. (2019). Çocuk Yoğun Bakım Ünitesine Olan Talebin Zaman Serisi Yöntemleri ile Tahmin Edilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, 454-462.

KAYA, Ç., & YILDIZ, O. (2014). Makine Öğrenmesi Teknikleriyle Saldırı Tespiti: Karşılaştırmalı Analiz. *Marmara Fen Bilimleri Dergisi*.

KAYNAR, O., & TAŞTAN, S. (2009). Zaman Serisi Analizinde Mlp Yapay Sinir Ağları Ve Arima Modelinin Karşılaştırılması. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 161-172.

KURUCAN, F. (2019). Büyük Veri Ve Finansal Piyasalarda İstatistiki Makine Öğrenmesi Metodlarının Yatırım Kararlarında Kullanılması. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul.

OĞUZLAR, A. (2003). Veri Ön İşleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 67-76.

OK, A. Ö., AKAR, Ö., & GÜNGÖR, O. (2011). Rastgele Orman Sınıflandırma Yöntemi Yardımıyla Tarım Alanlarındaki Ürün Çeşitliliğinin Sınıflandırılması. *TUFUAB 2011 VI. Teknik Sempozyumu.* Antalya.

ORAKÇI, M., CİYLAN, B., KÖK, İ., & SEVRİ, M. (2019). Suç Analizinde Veri Madenciliği Teknikleri ve Makine Öğrenmesi Algoritmalarının Kullanılması. *Yayınlanmamış Yüksek Lisans Tezi*. Ankara, Türkiye.

SAVAŞ, S., TOPALOĞLU, N., & YILMAZ, M. (2012). Veri Madenciliği Ve Türiye'deki Uygulama. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*.

ŞEKEROĞLU, S. (2010, Eylül). Hizmet Sektöründe Bir Veri Madenciliği Uygulaması. *Yayınlanmamış Yüksek Lisans Tezi*.

YEGEN, N. (2020). Perakende Sektöründe Veri Madenciliği İle Satış Tahmini. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul.

YILDIRIM, B. (2019). Modern Perakendecilik Sektöründe Veri Madenciliği Tekniklerinin Uygulanması. *Yayınlanmamış Yüksek Lisans Tezi*. İstanbul.