



İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
UYGULAMALI MATEMATİK VE HESAPLAMALI BİLİMLER
ANABİLİM DALI

**Makine Öğrenmesi Algoritmaları ile Hazır Giyim
Perakendeciliğinde Satış Tahmini**

Yüksek Lisans Tezi

Abdül Meral

TEMMUZ 2023



İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
UYGULAMALI MATEMATİK VE HESAPLAMALI BİLİMLER
ANABİLİM DALI

**Makine Öğrenmesi Algoritmaları ile Hazır Giyim
Perakendeciliğinde Satış Tahmini**

Yüksek Lisans Tezi

Abdül Meral

Danışman

Prof. Dr. Rahmet Savaş

TEMMUZ 2023

TEZ JÜRİSİ ONAYI

Abdül MERAL tarafından hazırlanan "Makine Öğrenmesi Algoritmaları İle Hazır Giyim Perakendeciliğinde Satış Tahmini" başlıklı bu Yüksek Lisans Tezi, Uygulamalı Matematik Ve Hesaplamalı Bilimler Anabilim Dalı'nda hazırlanmış ve jürimiz tarafından kabul edilmiştir.

JÜRİ ÜYELERİ

İMZA

Tez Danışmanı:

Prof. Dr. Rahmet SAVAŞ

Kurumu: İstanbul Medeniyet Üniversitesi

Üyeler:

Dr. Öğr. Üyesi Nurullah Çalık

Kurumu: İstanbul Medeniyet Üniversitesi

Doç. Dr. Peyman MAHOUTİ

Kurumu: Yıldız Teknik Üniversitesi

Tez Savunma Tarihi: 07/ 07 / 2023

ÖNSÖZ

Akademik ve sektörel hayatımda başarılı olmamı destekleyen, öncelikle danışman hocam Prof. Dr. Rahmet Savaş'a ve bölümdeki tüm hocalarıma, ayrıca sevgili aileme teşekkürlerimi sunmak istiyorum.

Ayrıca, veri toplama sürecinde çalışmalarına destek olan saygı değer müdürüm Yavuz Selim DOLAŞ' a teşekkürü bir borç bilirim.

Abdül MERAL

BEYANLAR

Yazım ve Kaynak Gösterme Kılavuzu Beyanı

Danışmanlığımda yazılan bu tezin APA yazım ve kaynak gösterme kılavuzunda belirtilen kurallara uygun olarak yapılandırıldığı ve bu kılavuzun APA kaynak gösterme standartlarının bu tezde tutarlı olarak uygulandığı tarafımdan incelenerek teyit edilmiştir.

İmza

Prof. Dr. Rahmet SAVAŞ

Etik İlkeler Sadakat Beyanı

Hazırladığım bu tezin tamamen kendi çalışmam olduğunu, akademik ve etik kuralları gözeterek çalıştığımı ve her alıntıya kaynak gösterdiğimi beyan ederim.

İmza

Abdül MERAL

ÖZET

Makine Öğrenmesi Algoritmaları ile Hazır Giyim Perakendeciliğinde

Satış Tahmini

Meral, Abdül

Yüksek Lisans Tezi, Uygulamalı Matematik ve Hesaplamalı Bilimler Anabilim Dalı,
Uygulamalı Matematik ve Hesaplamalı Bilimler Programı

Danışman: Prof. Dr. Rahmet Savaş

Temmuz 2023

Günümüzde müşteriye doğru ürünü doğru zamanda ulaştırmak hem satış yapan firmalar için hem de müşteriler için önemli bir kazanımdır. Bununla birlikte yapılan çalışmalar veri biliminin önemini göstermektedir. Şirket içindeki verinin işlenip yararlı hale gelmesi de verinin ne kadar önemli olduğunu gösterecek ve rakipler karşısında daha verimli ve hızlı kararların alınmasını sağlayacaktır. Yapacağımız bu çalışma ile de doğru bir stok optimizasyonu sağlamış olacağımızdan dolayı, stoklama/depolama maliyetlerinin azaltılması amaçlanmaktadır. İlgili haftada satılma olasılığı yüksek olan ürünlerin yine ilgili mağazada yani ürünün hitap ettiği müşteri kesimi bakımından da en iyi satış oranını yakalayacağımız mağazaya gönderilmesi sağlanarak maksimum kar ve maksimum müşteri memnuniyeti elde edilmesi sağlanacaktır. Aynı zamanda şirket içerisinde şu anda mağaza ürün yöneticisi olarak çalışanlar tarafından yönetilen bu manuel ve tecrübeye dayalı yöntem, daha otomatize bir sistem ve veriye dayalı, insandan bağımsız bir iş süreci otomasyonunun oluşmasına büyük katkı sağlayacaktır. Oluşturacağı etki sayesinde diğer tecrübeye dayalı iş süreçleri de zamanla veri bilimi yaklaşımlarından etkilenecektir. Aynı zamanda yeni üretilmiş, henüz satış davranışı hakkında bilgisine sahip olmadığımız ürünler içinde, en iyi satışı yapacağı mağaza ve hafta önerisi yaparak en iyi verim elde edilmiş olacak. Regresyon ve sınıflandırma problemlerini barındıran bu proje, gelecek projeler içinde temel oluşturacak olup departmanın tecrübesini yükseltecek etkiye de sahiptir.

Anahtar Kelimeler: Makine Öğrenmesi, Derin Öğrenme, Görüntü İşleme, Regresyon, Sınıflandırma, Kümeleme

ABSTRACT

Sales Forecasting in Ready to Wear Retailing with Machine Learning Algorithms

Master's Thesis, Applied Mathematics and Computational Sciences Department,
Applied Mathematics and Computational Sciences Program

Advisor: Prof. Dr. Rahmet Savaş

July 2023

The delivery of the correct product to the customer at the appropriate time is of great significance for both the sales companies and the customers. Additionally, studies have demonstrated the importance of data intelligence. Processing the data within a company to make it useful highlights the significance of data and allows for quicker and more efficient decision making compared to competitors. Through this study, we will attain accurate stock optimization, leading to reduced storage and storage costs. By sending the products with the highest likelihood of selling to the store with the highest sales rate that caters to the target customer segment, we can achieve maximum profits and customer satisfaction.

Furthermore, replacing the current manual and experience-based process managed by store product managers with a more automated and data-driven system will greatly contribute to the company's efficiency. This automation will have a ripple effect on other processes within the company, leading to a gradual shift towards data science based approaches. Additionally, for new products with limited sales information, we will provide recommendations for the best store and week for maximum sales, thus ensuring the best outcome.

This project, which deals with regression and classification problems, will lay the foundation for future projects and enhance the department's expertise.

Keywords: Machine Learning, Deep Learning, Image Processing, Regression, Classification, Clustering

İÇİNDEKİLER

ÖNSÖZ	i
BEYANLAR.....	ii
ÖZET	iii
ABSTRACT.....	v
İÇİNDEKİLER	vi
TABLolar LİSTESİ.....	viii
ŞEKİLLER LİSTESİ	ix
KISALTMALAR	x
1 BÖLÜM I GİRİŞ.....	1
1.1 TEZİN AMACI	1
1.2 TEZİN HİPOTEZİ.....	1
1.3 LİTERATÜR TARAMASI.....	1
2 BÖLÜM II HAZIR GİYİM KAVRAMLARI.....	5
2.1 ALOKASYON	5
2.1 MAĞAZACILIK.....	5
2.2 MAĞAZACILIK TEMEL KAVRAMLAR.....	5
2.2.1 Doğru Ürün ve Doğru Zaman	5
2.2.2 Doğru Adet ve Doğru Zaman.....	6
3 BÖLÜM III VERİ TANIMI VE MAKİNE ÖĞRENMESİ	7
3.1 VERİ NEDİR?.....	7
3.2 VERİ MADENCİLİĞİ	7
3.3 MAKİNE ÖĞRENMESİ.....	8
3.3.1 Makine Öğrenmesi Algoritmaları	8
3.3.1.1 Denetimli öğrenme	8
3.3.1.2 Denetimsiz öğrenme	13

3.4	VERİ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ KAVRAMLARININ İLİŞKİSİ.....	15
3.5	GÖRÜNTÜ İŞLEME	17
3.5.1	Ürün Görsellerinin Gürültülerden Temizlenmesi	18
3.5.2	Ürün Görsellerinden Ürün Öznitelik Vektörlerinin Oluşturulması	19
3.5.2.1	Öğrenme aktarımı	19
3.6	ALGORİTMA BAŞARISI DEĞERLENDİRME METRİKLERİ.....	21
3.7	PYTHON PROGRAMLAMA DİLİ	24
4	BÖLÜM IV VERİ ANALİZİ.....	26
4.1	KULLANILACAK VERİ SETLERİNİN TANITILMASI	26
4.2	KEŞİFÇİ VERİ ANALİZİ.....	28
5	BÖLÜM V PROJE UYGULAMASI	33
6	BÖLÜM VI SONUÇ.....	39
	KAYNAKÇA.....	41
	ÖZGEÇMİŞ	45

TABLÖLAR LİSTESİ

Tablo 1. Karmaşıklık Matrisi	23
Tablo 2. Sınıflandırma Algoritmalarını Değerlendirme Metrikleri.....	23
Tablo 3. Ürün Satış Adet Veri Seti.....	26
Tablo 4. Ürün Görsel Öznitelikleri Veri Seti	27
Tablo 5. Mağaza Özellikleri Veri Seti.....	27
Tablo 6. Ürün Özellikleri Veri Seti	27
Tablo 7. Özel Günler	28
Tablo 8. Tabular Veri Seti Regresyon Modeli Sonuçları	35
Tablo 9. Ürün Görselleri Regresyon Modeli Sonuçları.....	35
Tablo 10. Tabular Veri Seti Regresyon Modeli Hiper Parametre Sonuçları.....	36
Tablo 11. Tabular Veri Seti Sınıflandırma Modeli Sonuçları	37
Tablo 12. Ürün Görselleri Sınıflandırma Modeli Sonuçları.....	38

ŞEKİLLER LİSTESİ

Şekil 1. Veri Türleri	7
Şekil 2. Veri bilimi, veri madenciliği, yapay zeka (AI), makine öğrenimi ve derin öğrenme (DL) arasındaki ilişki	15
Şekil 3. Yıllara göre Veri Bilimi ve Veri Madenciliği Google arama motorundaki aranma sayıları (2018-Nisan 2023).....	16
Şekil 4. CRISP Veri Madenciliği İşlem Şeması	16
Şekil 5. Başarılı örnekler	18
Şekil 6. Başarısız örnekler	19
Şekil 7. Makine Öğrenmesi ve Öğrenme Aktarımı Arasındaki Fark.....	20
Şekil 8. Gerçek Değer, Tahmin Değeri ve Artık Değer İlişkisi.....	21
Şekil 9. Programlama Dillerinin Yıllara Göre Karşılaştırması.....	25
Şekil 10. Mağaza Özelliklerinin Dağılımı	28
Şekil 11. Mağazanın Satış Adet Dağılımları ve Aykırı Değerler	29
Şekil 12. Mağaza Satış Adet Dağılımlarının Gözlemlenmesi	30
Şekil 13. Satış Stok İlişkisi	31
Şekil 14. Özel Günlere Göre Satış Adetlerinin Gözlemlenmesi.....	32
Şekil 15. K-Modes kümeleme modeli ile elde edilen örnek bir ürün kümesi.....	34
Şekil 16. K-Means kümeleme modeli ile elde edilen örnek bir ürün kümesi.....	34

KISALTMALAR

YSA	Yapay Sinir Ağları
CRISP-DM	Veri Madenciliği için Sektörler Arası Standart Süreç (Cross-Industry Standard Process for Data Mining)
GBM	Gradient Boosting Machine
LightGBM	Light Gradient Boosting
SVM	Destek Vektör Makinaları
DBSCAN	Density Based Spatial Clustering of Applications With Noise
MAE	Ortalama Mutlak Hata
MAPE	Ortalama Mutlak Yüzde Hata
RMSE	Hata Kareleri Ortalamasının Karekökü

BÖLÜM I GİRİŞ

1.1 TEZİN AMACI

Bu çalışma, perakende giyim sektöründe uzmanlaşmış bir şirket için makine öğrenmesi algoritmalarının kullanılması ile satış tahminlerinin yapılması ve stok yönetimi gibi kararların daha doğru bir şekilde verilmesine odaklanmaktadır. Amaç, hata oranlarını düşürmek ve daha etkili sonuçlar elde etmektir.

1.2 TEZİN HİPOTEZİ

Bu çalışmada, kişi bazlı tecrübeye dayalı tahminler ve veri setinden anlamlı bilgi elde etme tekniği olan makine öğrenmesi algoritmalarının kullanılması önerilmiştir. Böylece daha başarılı tahminler yapılabileceği iddia edilmiştir.

1.3 LİTERATÜR TARAMASI

Güncel koşullarda, şirketlerin ürün taleplerini önceden belirlemesi ve uygun zaman ve yerde bulunabilmesi, rekabet avantajı sağlar. Klasik zaman serisi algoritmaları satış tahmininde kullanılır, ancak doğrusal olmayan talepler için yetersiz kalır. Bu nedenle, doğru satış tahmini yapmak için yapay sinir ağlarının kullanılması tercih edilmiştir (Sevgi, 2012).

Beyaz eşya sektöründe satış tahmini yapılması üzerine odaklanılan çalışmada, yapay sinir ağı (YSA), doğrusal regresyon ve destek vektör makinaları kullanılmıştır (Polat, 2022). Müşteri beklentilerinin ve üretim maliyetlerinin şirketler için ne kadar önemli olduğu vurgulanmıştır. 2016-2021 yılları arasındaki aylık satış verileri veri seti olarak kullanılmış ve 2021 yılı tahmin yılı olarak belirlenmiştir. Tahminler Türkiye pazarı için yapılmıştır. Algoritmaların değerlendirilmesi, ortalama mutlak yüzde hata (MAPE), ortalama mutlak hata (MAE) ve kök ortalama kare hata (RMSE) metrikleri kullanılarak yapılmıştır. Weka programı kullanılmıştır.

Moda sektöründe bulunan bir perakende şirketi için yapılan satış tahmini çalışmasında, klasik istatistiksel yöntemler ile makine öğrenme algoritmaları arasındaki başarımlar karşılaştırması yapılmıştır. Talep tahmininin şirket için önemi açıklanmıştır. Tahminler, 30 farklı mağaza ve 2 ürün ailesi için yapılmıştır. İstatistiksel algoritmalar olarak lineer regresyon ve ARIMA ve makine öğrenme algoritmaları olarak ise Torbalama Regresyon Ağacı, Aşırı Rastgele Ağaçlar, Rastgele Ormanlar, Gradyan Artırma Algoritması, Uyarlamalı Artırma Algoritması ve Aşırı Gradyan Artırma Algoritması kullanılmıştır. Algoritma performansları MAE, RMSE ve R^2 metrikleri ile ölçülmüştür (Işık, 2020).

Türkiye’de gıda perakendeciliği yapan market zinciri üzerinde yapılan bir çalışmada, 2018-2019 tarihleri arasındaki verilerin kullanılmasıyla, bu market zincirinin iki mağazasının satış tahminlerinin yapılması için veriler gün ve fiş bazında toplanmış ve 2020'nin ilk haftası tahminleme haftası olarak kullanılmıştır. Kampanya ve özel günler de dikkate alınmıştır. IBM SPSS programı kullanılarak regresyon ve zaman serisi algoritmaları uygulanmış ve program tanıtımı yapılmıştır. Algoritma performansları MAPE (ortalama mutlak yüzde hata) kriterine göre değerlendirilmiştir.(Dinçoğlu, 2022).

Gıda perakendeciliği alanında yapılan bir diğer çalışmada, 2015 yılına ait fiş verileri baz alınarak oluşturulan veri seti, sadece 1 yıllık veriler içerdiğinden dolayı, 6, 9 ve 1 yıllık eğitim verileri kullanılarak gelecek 31 günlük tahmin yapılmıştır. IBM SPSS programı kullanılmış ve tanıtımı yapılmıştır. Zaman serisi algoritmaları kullanılmış ve değerlendirme metriği ortalama mutlak yüzde hata (MAPE) kullanılmıştır (Yegen, 2020).

Türker, çalışmasında, turizm sektörü açısından satış tahmininin önemine odaklanmıştır. Türkiye'deki bir turizm şirketinin verileri kullanılarak istatistiksel ve makine öğrenme yöntemleri kullanılmıştır. Lineer regresyon, Ridge regresyon, karar ağaçları, rastgele orman ve destek vektör makinaları gibi algoritmalar kullanılmış ve en iyi sonuçlar destek vektör makinaları yöntemi ile elde edilmiştir. Modelin değerlendirilmesinde ortalama mutlak yüzde hata (MAE), R-kare (R^2) ve ortalama mutlak hata (MSE) gibi metrikler kullanılmıştır (2021).

Wang, çalışmasında, satış tahminlerinin perakendeciler için daha doğru bir iş planı oluşturmasını sağlayabileceğini vurgulamıştır. Veriler, Amerika Birleşik Devletlerindeki Wal-Mart market zincirinden alınmıştır. Xgboost ve Lightgbm gibi makine öğrenme algoritmaları kullanılmıştır ve her birinin RMSE skoru sırasıyla 2.11 ve 2.09 olarak belirlenmiştir. Hibrid bir tahmin yöntemi kullanılarak, bir algoritmadan yarısı ve diğer algoritmadan diğer yarısı oluşan tahmin, RMSE skoru 2.07 olarak tahmin edilmiştir (2020).

Keung ve diğ. tarafından yapılan çalışmada, ürün sevkiyatlarında yaşanabilecek gecikmeler satış performanslarını etkileyebileceğinden dolayı, satış tahminlerinin daha iyi sonuçlar vermesi için verilerin anlaşılmasının önemi vurgulanmıştır. Ürün kategorileri ve lokasyon faktörlerinin dalgalı satışlar ve belirgin gecikmeleri belirleyebileceği her zaman göz önünde bulundurulması gerektiği belirtilmiş ve makine öğrenmesi algoritmalarının satışı etkileyen özelliklerin korelasyonu ele alınmıştır (2021).

Amerika'da bulunan bir alışveriş merkezinden 2013 yılına ait veri seti kullanılarak yapılan bir çalışmada satış tahminleri yapılmıştır. Bu çalışmada iki farklı yaklaşım denenmiştir: tek basamaklı ve çift basamaklı. Tek basamakta, temel makine öğrenmesi algoritmaları kullanılmıştır ve çift basamakta ise makine öğrenmesi modellerinin birlikte kullanımı ile oluşan algoritma uygulanmıştır. Bu yaklaşımın tahminlerin doğruluğunu artırdığı görülmüştür ve MAE kullanılarak modelin değerlendirilmesi yapılmıştır (Punam ve diğ., 2018).

Amerika'da bulunan market zinciri için yapılan bir çalışmada zaman serisi analizi yapılmış ve 5 yıllık veriler Facebook Prophet ve Lightgbm algoritmaları ile karşılaştırılmıştır. Her iki algoritmanın özellikleri tanıtılmış ve performansları RMSE skoru kullanılarak değerlendirilmiştir. Elde edilen sonuçlar, Facebook Prophet algoritmasının RMSE skoru 0.694 ve Lightgbm algoritmasının RMSE skoru 0.617 olarak belirlenmiştir (Jiang ve diğ., 2021).

Arif ve diğ. yaptığı çalışmada, tedarik zinciri yönetimi ve müşteri taleplerinin birbirine nasıl bağlı olduğu vurgulanmıştır. Tahminlerin doğruluğunun, tedarik zincirinin verimli bir şekilde çalışması için gerekli olduğu belirtilmiştir. Bu amaçla, makine öğrenmesi yöntemi oluşturulmuş ve uygulanmıştır. Bu yöntem ile verilerin anlamlı bir

şekilde yorumlanması ve sonuçların daha doğru tahmin edilmesi hedeflenmiştir. Değerlendirme metriği olarak ortalama mutlak yüzde hata (MAPE) kullanılmıştır (2019).

Zhaoweijie ve diğ.'nin yaptıkları bir çalışmada, bir süpermarket zincirinde 2013-2015 yılları arasında 1115 mağazadan oluşan veri seti analiz edilmiştir. Makine öğrenme yöntemi olarak Xgboost kullanılmış ve 48 günlük tahminler yapılmıştır. (2020)

Ramachandra ve diğ. yürüttükleri bu çalışmada, perakende sektöründeki satış performansının önemi vurgulanmıştır. Veri seti 8523 gözlemden ve 12 farklı özelliğinden oluşmaktadır. Veri seti kullanılarak karacumadaki satışların tahmini yapılmıştır. Değerlendirme için RMSE metriği kullanılmıştır (2021).

Amerika'da bulunan şirketin veri bilimi yarışmalarının düzenlendiği bir web sitesi üzerinden alınan veri seti kullanılarak yapılan bir çalışmada, perakende şirketlerinin satış tahmininde bulunması gereken öneme değinilmiştir. Yapılan önemli özellik mühendisliği sonucunda, xgboost algoritmasına girdi verilmiştir. 1913 gözlemin kullanıldığı ve gelecekteki 28 günlük satış tahmini yapılmıştır. Xgboost algoritması hakkında bilgi verilmiş ve performans karşılaştırması lineer regresyon, ridge regresyon ve xgboost algoritmaları için RMSSE (ve kök ortalama ölçeklenmiş kare hata) skoru kullanılarak yapılmıştır. Xgboost en düşük RMSSE skoru (0.655) ile en iyi performansı göstermiştir (Dairu & Shilong, 2021).

Hanzel ve diğ., çalışmasında, moda sektöründe talep tahmininin zorlukları ve güncel zorluklar incelenmiştir. Hızlı değişkenlikler ve satış adetlerinin azlığı talep tahmininin zorluklarını ortaya koymuştur. Çalışma gerçek veri seti üzerinde yapılmış ve veri seti, 2016 yılının Ocak ayından 2021 yılının Kasım ayına kadar haftalık olarak oluşmuştur. Test veri seti ise 2021 yılının Nisan ayından 2021 yılının Kasım ayına kadardır. Veri seti analiz edildiğinde güçlü sezonsallık ve ani artışlar gözlenmiştir (2022).

BÖLÜM II HAZIR GİYİM KAVRAMLARI

2.1 ALOKASYON

Depoda hazır olan stokların, müşterilerin ihtiyaçlarına en uygun şekilde dağıtılması için, bu stokların durumları ve mağazalardaki stok durumları göz önünde bulundurulmalıdır. Böylece, doğru ürün, doğru zamanda, doğru miktarda ve doğru mağazalarda müşterilerin hizmetine sunulabilir.

Depoda bulunan stokların müşterilerin ihtiyaçlarına en uygun şekilde dağıtılması, mağazaların stok durumlarının düzenli olarak izlenmesi ve doğru ürünlerin doğru zamanda doğru miktarda doğru mağazalarda bulundurulması süreci önemlidir. Bu sayede müşterilerin istekleri en hızlı ve efektif şekilde karşılanabilir.

2.1 MAĞAZACILIK

Ana hedef, müşteri taleplerinin belirlenmesi ve bu taleplerin en doğru şekilde karşılanmasıdır. Bunu yapmak için, sayısal analizler ve müşterilerin ihtiyaçlarının derinlemesine anlaşılması kullanılır. Bu anlamaya çalışma, bölgesel farklılıklara göre taleplerin belirlenmesine yardımcı olur. Bu belirlenen farklılıklar, bütçe, koleksiyon ve sipariş belirleme süreçlerine yansıtılır.

Ayrıca, ürün sevkiyatı, transfer, iade ve indirim gibi aksiyonların yönetimi, sezon içi süreçlerin yönetilmesi için önemlidir. Böylece, müşterilerin talepleri en hızlı ve etkin şekilde karşılanabilir ve stokların verimli bir şekilde yönetilmesi sağlanır.

2.2 MAĞAZACILIK TEMEL KAVRAMLAR

2.2.1 Doğru Ürün ve Doğru Zaman

Sevkiyat sürecinde, hangi ürünlerin ilk olarak sevk edileceği ve zaman aralıkları, satın alma ve planlama ekiplerinin direktifleri doğrultusunda belirlenir. Bu aşama, ürün stoklarının verimli bir şekilde yönetilmesi için çok önemlidir.

Satın alma ve planlama ekipleri, sevkiyat sürecinde kullanılacak ürünlerin belirlenmesi için direktiflerini ve zaman aralıklarını kullanırlar. Bu sayede, ürün stokları en verimli şekilde yönetilir ve müşterilerin ihtiyaçları en hızlı şekilde karşılanabilir.

Bu süreçte, sevkiyat sürecine konu edilecek ürünlerin belirlenmesi ve zaman aralıklarının doğru bir şekilde planlanması, ürün stoklarının verimli bir şekilde yönetilmesi için önemlidir. Bu sayede, müşterilerin ihtiyaçları en hızlı ve etkin şekilde karşılanabilir.

2.2.2 Doğru Adet ve Doğru Zaman

Ürünlerin belirlenen zaman aralığında sevk edilmesi için mağazalar-ürün eşlemesi yapılması gerekmektedir. Bu eşleme, analitik veriler ve trend analizleri gibi faktörlere dayanmaktadır. Alokasyon süreci, mağazalara hangi ürünlerin ne kadar gönderileceğinin belirlenmesi ile başlar ve bu eşleme en temel görevidir. Ürün-mağaza eşlemesi, müşteri talepleri, ürün stokları ve mağaza kapasitesi gibi birçok faktörü değerlendirerek yapılır. Bu sayede ürünlerin doğru zamanda doğru yere sevk edilmesi ve müşteri memnuniyeti sağlanması amaçlanır.

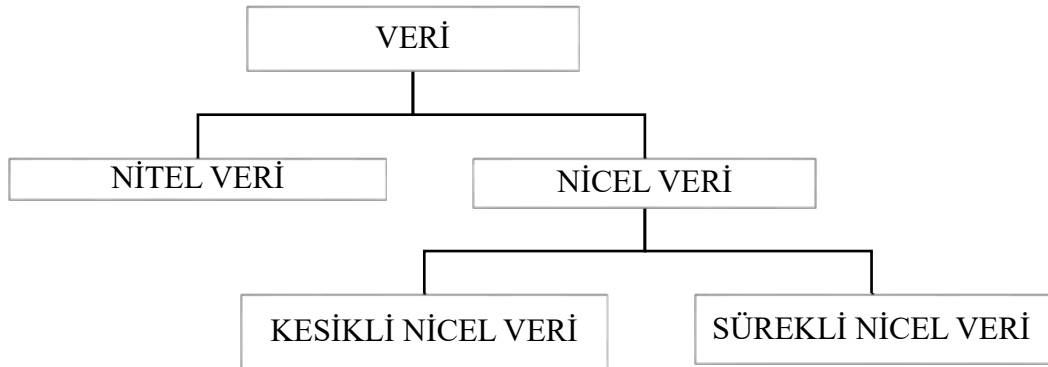
Bu analizler aşağıdaki gibi sıralanabilir.

- Mağaza ürün ailesi bazında stok satış analizi
- Sevke konu edilecek olan ürünlerin muadilleri var ise bu ürünlerin geçmiş dönem satış performansı analizi
- Gönderime konu edilecek olan ürün için genel bütçe değerlerinin kontrolü
- Ürün gruplarına ait karlılık analizi
- Gönderime konu edilecek ürünün iklime uygunluğunun analizi
- Gönderilecek olan ürünün trend ve geleneksel oluşuna göre doğru mağazanın seçiminin yapılması
- Mağazaların doluluk durumlarının analizi

BÖLÜM III VERİ TANIMI VE MAKİNE ÖĞRENMESİ

3.1 VERİ NEDİR?

Veri; sayılar, kelimeler, ölçümler, gözlemler veya sadece bir şeylerin açıklamaları gibi gerçeklerin bir koleksiyonudur. Veriler nitel veya nicel olabilir. Nitel veri, bir şeyin açıklaması, nicel veri ise sayısal bilgi içeren veridir. Nicel veri kendi içinde kesikli ya da devamlı veri olarak ikiye ayrılır (Şekil 1). Kesikli nicel veri sadece kesin sayısal değerler alırken, devamlı nicel veri ise belli bir aralıktaki herhangi bir değeri alabilir. Kısaca kesikli nicel veri sayılabilirken, devamlı nicel veri ölçülen bir değerdir.(DipCE BEng Rod Pierce, n.d.)



Şekil 1. Veri Türleri

3.2 VERİ MADENCİLİĞİ

Veri madenciliği, hali hazırda var olan verileri analiz ederek sorunları çözmekle ilgilidir. Ayrıca, veri madenciliği, verilerdeki örüntüleri keşfetme süreci olarak tanımlanır. Bu örüntülerden keşfedilenler aynı zamanda bazı avantajlar sağlayabilecek anlamlara sahip olmalıdır. Veri madenciliğindeki mevcut yöntemlerde bu veride bulunan örüntüleri anlamamıza yardımcı olmaktadır.

Veri madenciliği, büyük veri kümelerindeki kalıpların hesaplamalı olarak keşfedilmesinin temel metodudur. Aynı zamanda makine öğrenimi, yapay zeka, istatistik ve veri tabanı sistemlerinin kesişme noktasındaki yöntemleri de içerir.

Günümüzde şirketler, büyük miktarda veriyi incelemek ve pazar araştırma raporlarını yıllarca analiz etmek için güçlü bilgisayarlar ya da bulut sistemlerini kullanırlar. Veri madenciliği, bu şirketlerin fiyat, personel becerileri ve rekabet, ekonomik durum ve müşteri demografisi gibi harici faktörler ile iç faktörler arasındaki ilişkiyi tanımlamalarına yardımcı olur. (Data Mining, n.d.)

3.3 MAKİNE ÖĞRENMESİ

Makine öğrenimi, insanların bilgisayarları programlamasını sağlayarak makinelerin kalıpları tanıyabilir ve neyin ne olduğunu öğrenerek gerekli öngörülerde bulunabilir hale gelmelerini sağlanmasıdır. Öğrenme süreci veriyi bilgiye dönüştürmektir. Bir makine öğrenmesi modelinde eğitim verisi tecrübeyi temsil ederken, bu modelin çıktısı da uzmanlığı temsil etmektedir. (Freeman ve diğ., 2019)

Makine öğrenmesi uygulamalarına örnek olarak, sahtecilik tespiti, öneri sistemleri, resim sınıflandırma hatta otonom araçlar verilebilir.

3.3.1 Makine Öğrenmesi Algoritmaları

Bütün problemleri çözebilen bir makine öğrenmesi algoritması bulunmamaktadır. Bu yüzden, problemin tipine göre de kullanılacak makine öğrenmesi modeli değişmektedir. (Rebecca Vickery, n.d.)

Makine öğrenmesinin üç ana kategorisi vardır: denetimli öğrenme, denetimsiz öğrenme ve takviyeli öğrenme. (Microsoft Corporation, n.d.)

3.3.1.1 Denetimli öğrenme

Bu öğrenmede kendi içerisinde ikiye ayrılır.

3.3.1.1.1 Sınıflandırma

Sınıflandırma problemlerinde, veri noktaları belirli bir etiket (kategori) ile ilişkilendirilir. Bu etiketler, veri noktalarının belirli bir kategoriye ait olduğunu belirtir. Örneğin, bir görsel sınıflandırma probleminde, görseller "kedi" veya "köpek" olarak

etiketlenir. Bu etiketler veri noktalarının kategoriye ait olduğunu belirtir. Sınıflandırma modelleri, veri noktalarının etiketlerini öğrenir ve sonra yeni veri noktalarının hangi etikete ait olduğunu tahmin eder. Örneğin, bir yeni görsel model tarafından "kedi" olarak sınıflandırılırsa, model kedi gibi görünen bir resmi tanımlamış demektir. Bu yolla, sınıflandırma modelleri veri noktalarını belirli kategorilere veya ilgilenilen değerlere göre sınıflandırır.

Lojistik Regresyon

Lojistik Regresyon, bir sınıflandırma modelidir ve özellikle kategorik değişkenleri tahmin etmek için kullanılır. Örneğin, bir müşterinin kredi kartı işlemlerinden elde edilen verileri kullanarak, müşterinin kredi notunun düşük veya yüksek olmasını tahmin etmek için kullanılabilir. Lojistik Regresyon, bağımlı değişkeninin kategori olmasına rağmen, bağımsız değişkenlerin sayısının çok olması durumunda kullanılır. Model, veriler arasındaki ilişkiyi öğrenir ve sonra veriye dayalı olarak bir sınıflandırma yapar. Lojistik Regresyon, sigmoid fonksiyonu kullanır. Bu fonksiyon, 0 ve 1 arasındaki değerleri geri döndürür ve sınıflandırma yaparken kullanılır. Örneğin, tahmin edilen değer 0,5'den büyük ise, müşterinin kredi notunun yüksek olduğu, aksi halde düşük olduğu söylenir. Lojistik regresyon modeli ayrıca, her bir bağımsız değişkenin katkısını ölçebilir ve hangi değişkenlerin sınıflandırma için önemli olduğunu belirleyebilir. (Robles-Velasco ve diğ., 2020a)

Destek Vektör Makinaları Sınıflandırma

Destek Vektör Makinaları (SVM) sınıflandırma, veriler arasındaki ayrımı en iyi şekilde yansıtan bir hiper düzlem üzerinde sınıflandırma yapar. SVM, veriler arasındaki ayrımı en iyi şekilde yansıtan hiper düzlemi bulmak için maksimum kenar payı (marjin) prensibini kullanır. Bu, veriler arasındaki en büyük boşluğu elde etmeye çalışır, böylece veriler arasındaki ayrım en belirgin hale gelir.

Örneğin, bir e-posta filtreleme uygulamasında, e-postaların "istenmeyen elektronik posta" olup olmadığını sınıflandırılması için kullanılabilir. Bir sağlık uygulamasında, hastaların "sağlıklı" veya "hasta" olarak sınıflandırılması için kullanılabilir. Bir finans uygulamasında, hisse senetlerinin "yükseliş" veya "düşüş" olarak sınıflandırılması için kullanılabilir. (Robles-Velasco ve diğ., 2020b)

LightGBM Sınıflandırma

LightGBM, Gradient Boosting Machine (GBM) algoritmasının hızlı ve efektif bir uygulamasıdır. GBM, birçok karar ağacı kullanarak verileri sınıflandırmak için kullanılan bir yöntemdir. LightGBM, GBM algoritmasının önceki sürümlerine göre daha hızlı çalışmasını sağlamak için bir dizi optimize edici yöntemi kullanır.

LightGBM, sınıflandırma ve regresyon problemleri için kullanılabilir. Özellikle, veri kümelerinde çok sayıda değişken veya özellik olduğunda veya veri kümesinde çok sayıda satır olduğunda, LightGBM daha hızlı çalışır. Ayrıca, LightGBM kategorik değişkenleri veya sınırlı sayıdaki eksik değerleri işleyebilir.

LightGBM, çok sayıda karar ağacı kullanarak verileri sınıflandırır. Her bir karar ağacı, veri kümesinden rastgele seçilen özellikleri kullanarak verileri sınıflandırır. LightGBM, her bir karar ağacının tahminlerini birleştirerek veya kümeleme yaparak, verileri daha doğru bir şekilde sınıflandırır. (Zeng ve diğ., 2019)

3.3.1.1.2 Regresyon

Regresyonda, bir bağımlı değişken (hedef değişken) bir veya daha fazla bağımsız değişken (açıklama değişkenleri) yardımıyla açıklanmaya çalışılır. Örneğin, ikinci el araba fiyatının arabanın modeli, yaşı, kilometre veya araba durumu gibi değişkenler ile ilişkisi incelenebilir. Bu değişkenler araba fiyatını açıklamak için kullanılır. Regresyon modelleri, bu değişkenler arasındaki ilişkiyi öğrenir ve sonra yeni verilerle birlikte, hedef değişkenin değerini tahmin etmeye çalışır. Örneğin, bir arabanın modeli, yaşı ve kilometresi bilindiğinde, arabanın fiyatını tahmin etmek için regresyon modeli kullanılabilir. Bu yolla, regresyon modelleri, belirli değişkenler arasındaki ilişkiyi öğrenir ve sonra yeni verilerle birlikte, hedef değişkenin değerini tahmin etmeye çalışır. Bu tahminler, genellikle bir doğru olmayan değer olarak ifade edilir, ancak regresyon modelleri, veriler arasındaki ilişkiyi anlamaya ve tahminleri daha doğru hale getirmeye çalışır. Regresyon modelleri, veri analitikte sıklıkla kullanılan ve çeşitli uygulamalar için kullanılabilen önemli bir araçtır. Örneğin, pazarlamada müşteri harcamalarını tahmin etmek, finansal piyasalarda hisse senedi fiyatlarını tahmin etmek veya sağlık sektöründe hastalıkların teşhis etmek gibi uygulamalar için kullanılabilir.

Lineer Regresyon

Lineer Regresyon, bir bağımlı değişkenin bir veya daha fazla bağımsız değişkenle açıklanmasını amaçlar.

Verilerinizdeki bağımlı değişken y eksenini, bağımsız değişkenler $x_1, x_2, x_3, \dots$ vb. x_n ekseninde gösterilir. Bu x değişkenlerine göre y değişkeninin değerleri tahmin edilir.

Lineer Regresyon modeli, veriler arasındaki ilişkiyi tanımlayan bir doğrusal fonksiyon olarak ifade edilir. Bu fonksiyon, veriler arasındaki ilişkiyi en iyi şekilde açıklayan katsayıları (a ve b) içerir.

Formül olarak $y = a + b \cdot x$ şeklinde gösterilir.

Eğer veri seti küçük ise katsayıların değerleri doğrusal denklem yardımı ile hesaplanabilir. Ancak veri seti büyüdükçe veya daha kompleks hale geldiğinde katsayıların değerleri çoklu doğrusal regresyon ile hesaplanır.

Lineer Regresyon yaparken dikkat edilmesi gereken en önemli nokta verilerin normal dağılıma sahip olmasıdır. Ayrıca, veriler arasındaki ilişki doğrusal olmalıdır. Eğer veriler arasında polinomial veya lineer olmayan bir ilişki varsa, lineer regresyon uygun bir model olmayabilir.

Lineer regresyon, çok sayıda veriye dayalı olarak oluşturulur ve bu veriler arasındaki ilişkiyi en iyi şekilde yansıtmak için kullanılır. Bu nedenle, veri kümesinin büyüklüğü çok önemlidir ve modelin doğruluğu ve güvenilirliği için yeterli veriye sahip olmak gerekir.

Lineer Regresyon, birçok farklı uygulama alanında kullanılabilir. Örneğin, ekonomik verileri analiz etmek, finansal piyasaları tahmin etmek, sağlık verileri analiz etmek veya müşteri davranışlarını anlamak gibi. (Gupta ve diğ., 2022)

Destek Vektör Makinaları Regresyon

Destek Vektör Makinaları (SVR) regresyon, lineer veya lineer olmayan veri kümelerinde kullanılan bir regresyon yöntemidir. SVR, veriler arasındaki ilişkiyi en iyi şekilde yansıtmak için kullanılan bir model olarak çalışır.

SVR, özellikle veri kümelerinde çok sayıda gürültü veya aykırı nokta içeren durumlarda etkilidir. Yani varyansın yüksek olduğu veri setlerinde tercih edilmesi doğru sonuçların elde edilmesini sağlayacaktır.

SVR'nin temel amacı, veri kümesinde belirli bir hata toleransı içinde en iyi uyum sağlamaktır. Bu, SVR'nin lineer regresyondan daha esnek ve uyumlu olduğu anlamına gelir.

SVR, çok sayıda veriye dayalı olarak oluşturulur ve bu veriler arasındaki ilişkiyi en iyi şekilde yansıtmak için kullanılır. Bu nedenle, veri kümesinin büyüklüğü çok önemlidir ve modelin doğruluğu ve güvenilirliği için yeterli veriye sahip olmak gerekir. (Li ve diğ., 2022).

LightGBM Regresyon

LightGBM, sınıflandırma ve regresyon problemleri için ayrı ayrı tasarlanmış algoritmalar sunar.

Sınıflandırma problemlerinde, hedef değişkeniniz bir kategori değişkenidir ve modelinizin tahmin etmesi gereken bir sınıf etiketidir. LightGBM sınıflandırma, birden fazla sınıf etiketi arasından en uygun olanını tahmin etmek için kullanılan bir ağaç bazlı algoritmadır.

Regresyon problemlerinde ise, hedef değişkeniniz bir sayısal değişken olduğu için modelinizin tahmin etmesi gereken bir sayısal değerdir. LightGBM regresyon, hedef değişkenin sayısal değerlerini tahmin etmek için kullanılan bir ağaç bazlı algoritmadır.

Her iki problem de de LightGBM, gradient boosting yöntemi kullanır ve ağaçların eğitimini hızlandırmak için birçok özelleştirmeye sahiptir. Ancak sınıflandırma ve regresyon problemleri için farklı tahmin yapma yöntemleri kullanır. Regresyon problemlerinde, ortalama değerleri tahmin etmek için kullanılan bir kayıp fonksiyonu kullanılırken, sınıflandırma problemlerinde ise sınıf etiketleri arasındaki olasılıkları tahmin etmek için kullanılan bir kayıp fonksiyonu kullanılır.

Ayrıca, LightGBM sınıflandırma ve regresyon modellerinde değişken seçimi ve önem derecesi gibi ek işlemler de farklılık gösterebilir.

Son olarak, LightGBM sınıflandırma ve regresyon modellerinin her ikisi de veri setleri üzerinde hızlı ve etkili bir şekilde çalışmasını sağlar ve genellikle diğer algoritmalarından daha iyi sonuçlar elde etme potansiyeline sahiptir (Yin ve diğ., 2023).

3.3.1.2 Denetimsiz öğrenme

Denetimsiz öğrenme, veri noktalarının etiketleri bilinmeyen bir veri kümesi üzerinde yapılan bir öğrenme türüdür. Bu öğrenme modeli, veri noktalarının etiketleri bilinmediğinden, veriyi organize etmek veya yapısını açıklayacak bir formata dönüştürmek için kullanılır. Örnek olarak, bir veri kümesinde bir grup veri noktasının benzerliklerini tespit etmek veya veri noktalarını kümeler halinde sınıflandırmak için kullanılabilir. Bu tür öğrenme için kullanılan yöntemler arasında kümeleme, küme analizi, boyut azaltma veya gizli Markov modelleri gibi yöntemler yer almaktadır.

Kümeleme yöntemleri, verileri kümelere ayırmak için kullanılan algoritmalar arasında K-means, hiyerarşik kümeleme, dbscan gibi yöntemler bulunur. K-means, verileri belirli sayıda kümeye bölerken, hiyerarşik kümeleme, verileri ağaç yapısına benzer bir şekilde kümelere ayırır. DBSCAN ise, verileri benzer aralıklarda olan verileri aynı kümeye atar. Kümeleme yöntemleri genellikle, veri madenciliği, marketing, sağlık veya finans gibi farklı alanlarda kullanılır.

3.3.1.2.1 Kümeleme

Kümeleme, veri noktalarını benzer özellikleri olan gruplara ayırmak için kullanılan bir denetimsiz öğrenme tekniğidir. Bu teknik, veri noktalarının etiketlenmemiş olduğu durumlarda kullanılır ve amaç, veriyi benzer özelliklere sahip olan gruplara ayırmaktır. Örneğin, bir perakende mağazasının müşteri verilerini kullanarak, farklı harcama kategorilerine sahip müşteri gruplarını oluşturmak mümkündür. (Microsoft Corporation, n.d.)

K-Means

K-Means, veri noktalarını k kümeye bölmeye yarayan popüler bir denetimsiz kümeleme algoritmasıdır. Algoritma, veri noktalarını küme merkezlerine en yakın olanları seçerek veri noktalarını kümelere atama yapar. Bu atama işlemi, veri noktalarının küme merkezlerine olan uzaklıklarını minimize etmeye yönelik gerçekleştirilir.

Algoritmanın çalışma şekli şöyle özetlenebilir:

1. Başlangıçta, küme merkezleri rastgele seçilir.
2. Her veri noktası, küme merkezlerinden en yakın olanına atanır.
3. Her kümenin yeni merkezi, o kümedeki veri noktalarının ortalaması olarak hesaplanır.
4. Adım 2 ve 3, küme merkezlerinde herhangi bir değişiklik olana kadar tekrar edilir.

K-Means, küme sayısının önceden bilinmesi gerektiği için, veri noktalarının hangi kümelerle ait olduğunu belirlemede kullanılır. Ayrıca, bu algoritma sadece sayısal verilerle çalışabilmektedir. (Sinaga & Yang, 2020)

K-modes

K-modes, k-means algoritmasına benzer şekilde çalışan bir kümeleme algoritmasıdır. Ancak k-means gibi sayısal verileri kullanmak yerine kategorik verileri kullanır. K-modes, her bir veri noktasının bir veya daha fazla kategorik değişkeni olduğu veri kümelerinde kullanılabilir. Algoritma, veri kümelerindeki kategorik değişkenler arasındaki benzerlikleri kullanarak veri noktalarını kümelerle ayırır. Bu, veri noktalarının belirli kategori özelliklerine sahip olduğu kümelerde toplanmasını sağlar.

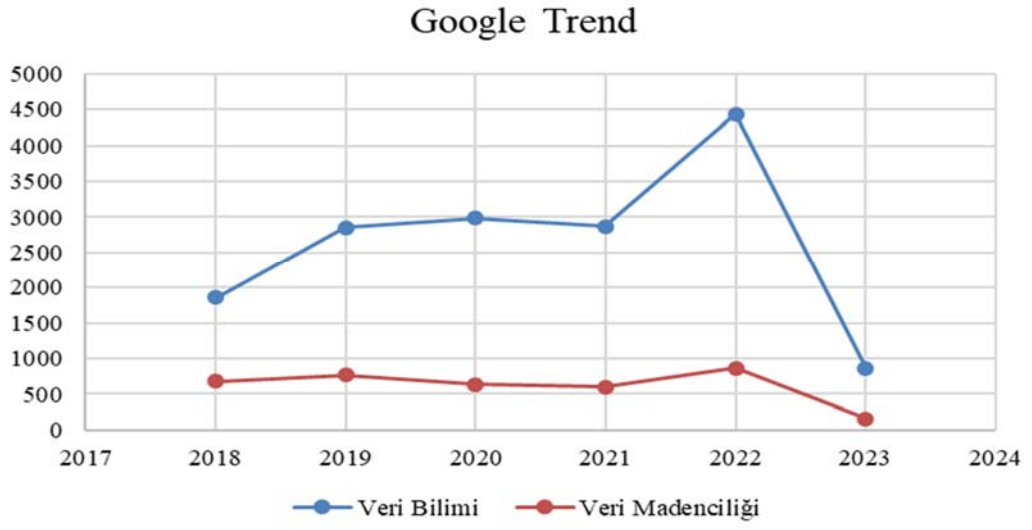
1. K-means algoritması gibi başlangıçta, küme merkezleri rastgele seçilir.
2. Merkezler ile diğer veri noktaları arasındaki benzersizlikler hesaplanır.
3. Oluşan kümelerden yeni modlar oluşturularak küme merkezleri güncellenir.
4. 2. ve 3. adımlar değişiklik olmayana kadar devam eder.

3.4 VERİ MADENCİLİĞİ VE MAKİNE ÖĞRENMESİ KAVRAMLARININ İLİŞKİSİ

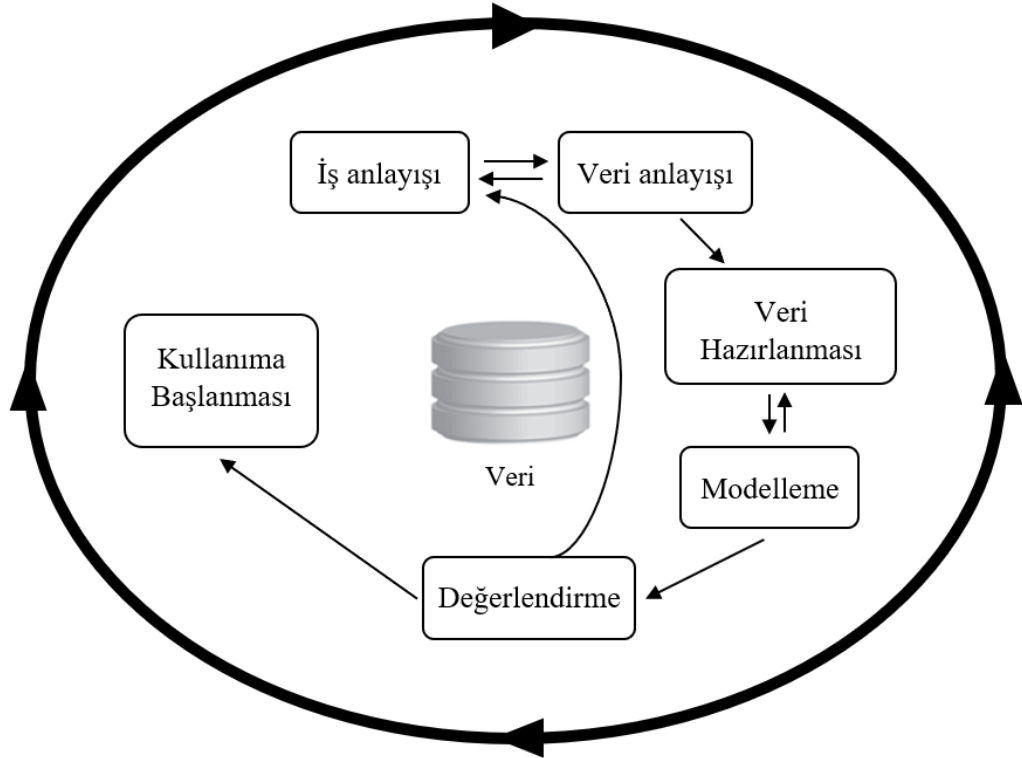


Şekil 2. Veri bilimi, veri madenciliği, yapay zeka (AI), makine öğrenimi ve derin öğrenme (DL) arasındaki ilişki

Bu kavramları kesin çizgilerle birbirinden ayırmak imkansızdır (Şekil 2). Bu kavramlar ortak olarak olasılık, istatistik ve sinir bilimi disiplinlerinden yararlanmaktadır. Veri madenciliğinde kullanılan bazı yaklaşımlar makine öğrenmesinde yapılan çalışmalar ile ortaya çıkmıştır. Örnek vermek gerekirse karar ağaçları ve en yakın komşu algoritmaları verilebilir. Ayrıca, veri madenciliğinde problem çözme adımları olarak benimsenmiş CRISP-DM yaklaşımı, makine öğrenmesi problemlerini çözmede de kullanılmaktadır (Saket Mittal, 2022). Makine öğrenmesi, veri madenciliğine göre yıllar içerisinde daha popüler hale gelmiştir (Şekil 3).



Şekil 3. Yıllara göre Veri Bilimi ve Veri Madenciliği Google arama motorundaki aranma sayıları (2018-Nisan 2023)



Şekil 4. CRISP Veri Madenciliği İşlem Şeması

Cross-Industry Standard Process for Data Mining (CRISP-DM) olarak kısaltılmıştır.

Veri madenciliği işlem sırası 6 adımdan oluşmaktadır (Şekil 4).

1. İş Anlayışı

- Hedeflerin ve ihtiyaçların belirlenmesi esastır.
- Mevcut çözümlerin analiz edilmesi, eksik ve iyileştirilmesi gereken kısımların belirlenmesi önemlidir.

2. Veri Anlayışı

- Projede kullanılacak verinin toplanması ve veri kalitesinin belirlenmesi.

3. Veri Hazırlama

- Veri özelliklerinin anlaşılması
- Problem için kullanılacak özelliklerin belirlenmesi
- Eksik verilerin tespit edilmesi
- Özellikler arası korelasyonlar
- Veri temizleme
- Yeni değişkenlerin oluşturulması

4. Makine Öğrenmesi Modellerinin Uygulanması

- Problem çözümüne, kullanılacak kaynak durumuna göre makine öğrenmesi algoritmasının uygulanması

5. Makine Öğrenmesi Modellerinin Performansının Değerlendirilmesi

- Problem türüne göre başarı kriterinin seçilmesi
- Sonuçların tutarlılığının incelenmesi
- Sonuçların tahmin ediciliğinin değerlendirilmesi

6. Kullanıma Başlanması

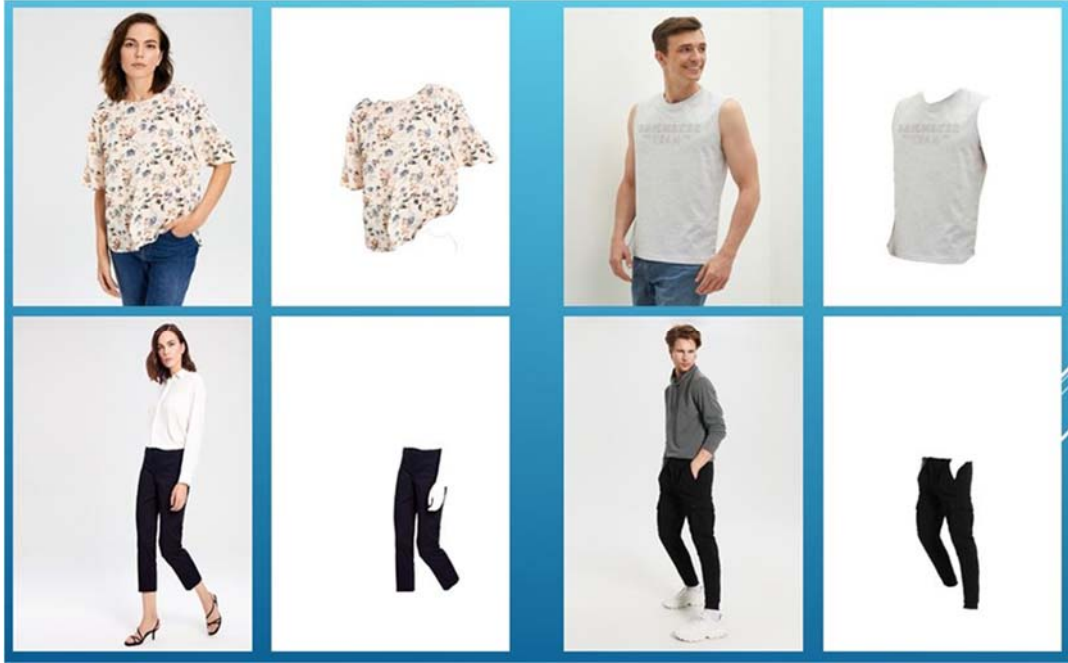
- Eldeki kaynaklar doğrultusunda modelin canlı sistemlerde çalıştırılması

3.5 GÖRÜNTÜ İŞLEME

Bilgisayar teknolojilerindeki ilerlemeler, görüntüler gibi çok boyutlu yapısal olmayan verilerin kullanılabilir hale gelmesini sağlamıştır. Derin öğrenme yöntemi sayesinde, görsellerin özellikleri daha ayırt edici hale getirilerek yüksek performanslı uygulamalar geliştirilebilmektedir.

3.5.1 Ürün Görsellerinin Gürültülerden Temizlenmesi

Ürün görsellerinde, ürünün haricinde insan veya nesnelerin yer alması durumunda, ürün özniteliklerinin gürültüden arındırılması için çeşitli teknikler kullanılabilir. Bu tekniklerden biri olan U-Net, evrişimli sinir ağı yaklaşımı kullanarak öncelikle biyomedikal görüntüler üzerinde daha iyi bir bölütleme (segmentasyon) yapmak için geliştirilmiştir. U-Net, evrişimli sinir ağı katmanlarından oluşan farklı bir mimariye sahip ve piksel tabanlı görüntü segmentasyonu/bölütleme konusunda çalışan bir derin öğrenme modelidir. U-Net ile yapılan başarılı ve başarısız örnekler sırasıyla Şekil 5 ve Şekil 6’da görülmektedir. (Jeremy Zhang, n.d.) (Levin Dabhi, n.d.)



Şekil 5. Başarılı örnekler



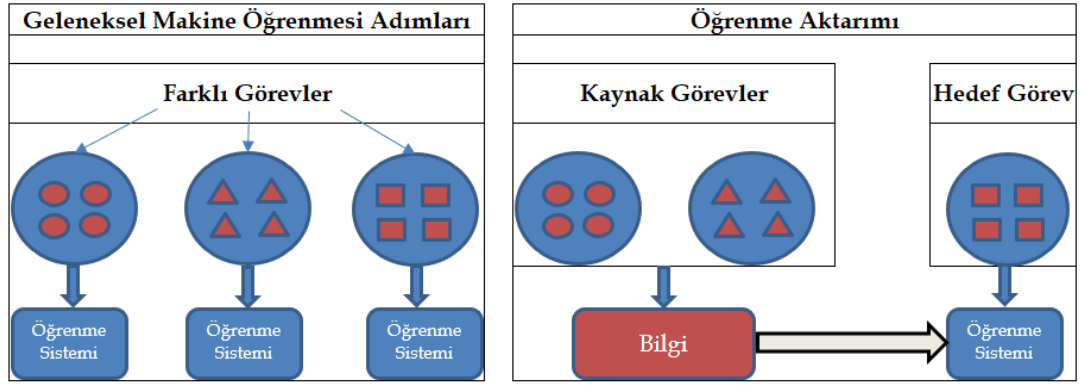
Şekil 6. Başarısız örnekler

3.5.2 Ürün Görsellerinden Ürün Öznelik Vektörlerinin Oluşturulması

3.5.2.1 Öğrenme aktarımı

Günlük yaşantımızda öğrendiğimiz bilgileri, farkında bile olmadan farklı durumlara uyarlayarak kullanabiliriz. Bu, karşılaştığımız durumların benzer olması durumunda daha kolay olabilir. Örneğin, yüzme bilen bir kişi, dalış yapmayı daha hızlı öğrenebilir çünkü sualtı aktiviteleri hakkında temel bilgilere sahiptir. Aynı şekilde, bir maraton koşucusu, bisiklete binmeyi öğrenirken kazandığı dayanıklılık ve konsantrasyon becerilerini kullanarak daha hızlı öğrenebilir. Özetle, yaşam boyu öğrenme sürecinde kazanılan beceriler, farklı durumlarda bize yardımcı olabilir ve bunları fark etmek ve kullanmak, daha hızlı öğrenme ve başarıya ulaşılmasına yardımcı olabilir (Paymode & Malode, 2022).

Transfer öğrenme, sinir ağı modeli için eğitim süresini azaltma faydasına sahiptir ve daha düşük genelleştirme hatası sonucuna yol açabilir. Tekrar kullanılan katmanlardaki ağırlıklar, yeni soruna yanıt olarak uyarlanmak üzere eğitim sürecinin başlangıç noktası olarak kullanılabilir. Bu kullanım, transfer öğrenmeyi bir tür ağırlık başlangıç şeması olarak ele alır (Şekil 7). Bu yaklaşım, ilgili ilk problemdeki etiketli veri miktarının ilgi duyulan problemde çok daha fazla olduğu durumlarda ve problem yapısındaki benzerlik her iki bağlamda da faydalı olabileceğinde kullanışlı olabilir.



Şekil 7. Makine Öğrenmesi ve Öğrenme Aktarımı Arasındaki Fark

Tensorflow kütüphanesinde bulunan imagenet veri seti ile eğitilmiş evrişimli sinir ağlarından VGG16, VGG19 ve ResNet152 derin öğrenme modelleri kullanılmıştır ve en iyi sonucu veren ResNet152 modeli ile devam edilmiştir.

VGG16, VGG19 ve ResNet152, görüntü sınıflandırma görevleri için kullanılan derin öğrenme modelleridir. Bu üç model arasındaki temel farklar ağırlıkları, mimarileri ve eğitim verileri hakkındadır.

1. VGG16:

VGG16, Oxford Üniversitesi'ndeki araştırmacılar tarafından geliştirilen bir modeldir. Bu model, 16 katmanlı bir derin sinir ağına sahiptir ve 138 milyon parametreye sahiptir. VGG16, 224x224 boyutundaki görüntülerle eğitilir ve 1000 sınıf için sınıflandırma yapabilir. Bu model, geniş ve düz bir mimariye sahiptir, yani her katmanda küçük boyutlu filtreler kullanır.

2. VGG19:

VGG19, VGG16'nın geliştirilmiş bir versiyonudur. VGG19, 19 katmanlı bir derin sinir ağına sahiptir ve 143 milyon parametreye sahiptir. VGG19, 224x224 boyutundaki görüntülerle eğitilir ve 1000 sınıf için sınıflandırma yapabilir. VGG19 da VGG16 gibi geniş ve düz bir mimariye sahiptir.

3. ResNet152:

ResNet152, Microsoft Research'teki araştırmacılar tarafından geliştirilen bir modeldir. Bu model, 152 katmanlı bir derin sinir ağına sahiptir ve 60 milyon parametreye sahiptir. ResNet152, 224x224 boyutundaki görüntülerle eğitilir ve 1000 sınıf için

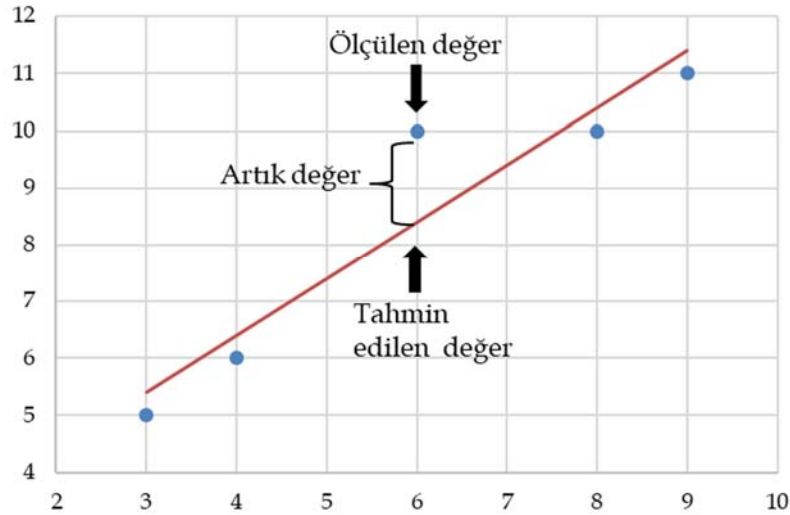
sınıflandırma yapabilir. ResNet152'nin mimarisi, birçok katmanda bağlantılar kullanır ve böylece daha derin ağların eğitilmesini sağlar. Bu bağlantılar, ağın çıktısı ile girdisi arasındaki farkları hesaplar ve bu farkları katmanlar boyunca aktararak, daha derin ağların eğitilmesine olanak tanır.

Genel olarak, VGG16 ve VGG19, basit ancak etkili bir mimariye sahipken, ResNet152 daha derin bir mimariye sahiptir ve bağlantıların kullanılması sayesinde daha iyi sonuçlar verir. Ancak, ResNet152 daha büyük bir model olduğu için eğitimi daha uzun sürer ve daha fazla kaynak gerektirir.

3.6 ALGORİTMA BAŞARISI DEĞERLENDİRME METRİKLERİ

Regresyon problemlerinde, tahminin hata oranı model performansı olarak tanımlanmıştır.

Aynı zamanda bu hata oranı gerçek değer ile tahmin değeri arasındaki farklılığı tanımlamaktadır. Bir regresyon modelinde bu hata oranını en düşük yapacak çizgiyi oluşturmaya çalışır (Şekil 8).



Şekil 8. Gerçek Değer, Tahmin Değeri ve Artık Değer İlişkisi

Artık değer = gerçek değer – tahmin değeri

- Bu değerler ile, değerlendirmelerimizi yapabiliriz, fakat elimizdeki veri setinin boyutları büyüdükçe yorumlamak zorlaşacağından dolayı, özet bir değer ile model başarıımızı ifade etmemiz daha kolay olacaktır.

Regresyon değerlendirme metrikleri (Tablo 2):

- Ortalama Mutlak Hata: Mutlak artık değerlerin toplamının gözlem sayısına bölünmesi ile elde edilir.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad 1$$

- Hata Kareleri Ortalaması: Artık değerlerin karesinin toplamının gözlem sayısına bölünmesi ile elde edilir.

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

- Hata Kareleri Ortalamasının Karekökü: Artık değerlerin karesinin toplamının gözlem sayısına bölünmesi ile elde edilen değerin kareköküdür.

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad 3$$

- Ortalama Mutlak Yüzde Hata: Gerçek değer ile tahmin değeri arasında oluşan yüzdelik hataların mutlak değerlerinin toplamının gözlem sayısına bölünmesi ile elde edilir. Dezavantaj olarak, gerçek değerler sıfır ya da sıfıra çok yakınsa çok yüksek hata elde edilir.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \quad 4$$

Ortalama Mutlak Yüzde Hata asimetriktir ve negatif hatalara (tahminler gerçeklerden yüksek olduğunda) pozitif hatalardan daha ağır cezalar verir. Bunun nedeni, çok düşük tahminlerde yüzde hatasının %100'ü geçememesidir. Çok yüksek tahminler için ise herhangi bir üst sınır yoktur. Sonuç olarak MAPE, fazla tahmin yerine eksik tahmin yapan modelleri tercih edecektir (Shweta Goyal, n.d.).

Sınıflandırma problemlerinde, gerçek değer ile tahmin değeri kıyaslanarak model başarımı belirlenir.

Sınıflandırma değerlendirme metrikleri:

- Karmaşıklık matrisi: Karışıklık matrisi, sınıflandırma problemlerini çözerken kullanılan çok popüler bir ölçüdür. İkili sınıflandırmaya uygulanabileceği gibi çok sınıflı sınıflandırma problemlerine de uygulanabilir (Tablo 1) (EMRE, 2021).

Tablo 1. Karmaşıklık Matrisi

Karmaşıklık Matrisi		Tahmin		Toplam
		Pozitif	Negatif	
Gerçek	Pozitif	TP doğru pozitif	FN yanlış negatif	P
	Negatif	FP yanlış pozitif	TN doğru negatif	N
Toplam		P'	N'	P+N
P: Gerçekte pozitif sınıfa ait olan örneklerin sayısı				
N: Gerçekte negatif sınıfa ait olan örneklerin sayısı				
P': Pozitif sınıfta olduğu tahmin edilen örneklerin sayısı				
N': Negatif sınıfta olduğu tahmin edilen örneklerin sayısı				

Tablo 2. Sınıflandırma Algoritmalarının Değerlendirme Metrikleri

Ölçüt	Açıklama	Formül
Doğruluk (accuracy, recognition rate)	Doğru tahmin edilen örneklerin, tüm örneklerin sayısına oranıdır.	$\frac{TP + TN}{TP + FP + TN + FN} = \frac{TP + TN}{P + N}$
Duyarlılık (sensitivity, recall, true positive rate-TPR)	Doğru tahmin edilen pozitif sınıftaki örneklerin, gerçekte pozitif sınıftaki örneklerin sayısına oranıdır.	$\frac{TP}{TP + FN} = \frac{TP}{P}$
Belirleyicilik (specificity, true negative rate-TNR)	Doğru tahmin edilen negatif sınıftaki örneklerin, gerçekte negatif sınıftaki örneklerin sayısına oranıdır.	$\frac{TN}{TN + FP} = \frac{TN}{N}$
Pozitif öngörü değeri/Keskinlik (positive predictive value, precision)	Doğru tahmin edilen pozitif sınıftaki örneklerin, pozitif sınıfta tahmin edilen örneklerin sayısına oranıdır.	$\frac{TP}{TP + FP} = \frac{TP}{P}$
F-ölçütü (F-score, F-measure)	Duyarlılık ve kesinliğin harmonik ortalamasıdır.	$\frac{(1 + \beta)^2 \times keskinlik \times duyarlılık}{(\beta^2 \times keskinlik) + duyarlılık}$

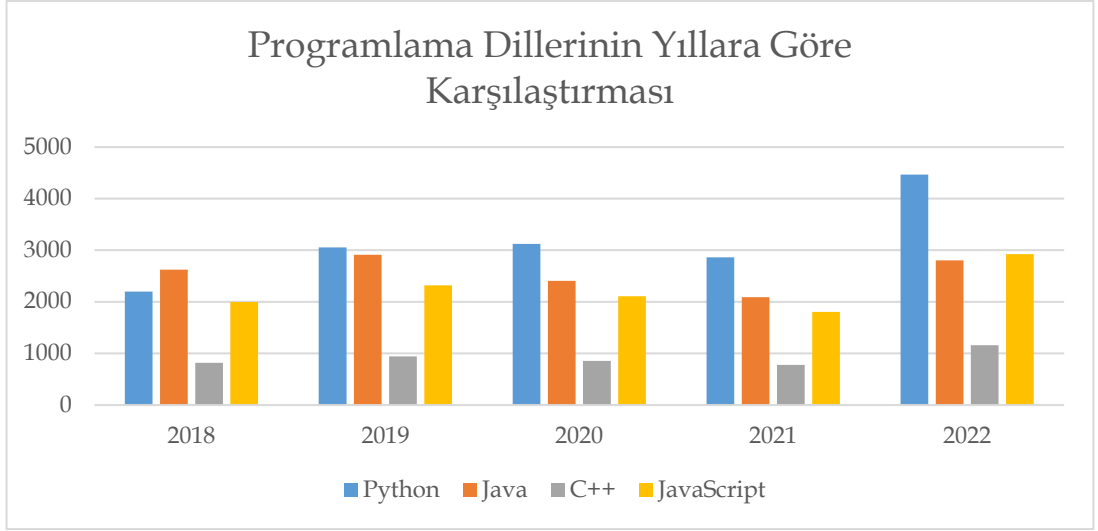
3.7 PYTHON PROGRAMLAMA DİLİ

Python, geniş kullanım alanlarına sahip, yüksek seviyeli bir programlama dilidir. Özellikle veri analitiği, makine öğrenimi ve uygulamalı matematik gibi alanlarda sıklıkla kullanılır. Python'un okunabilirliği ve kolay öğrenilebilirliği nedeniyle, yeni başlayanlar için idealdir. Ayrıca, çok sayıda kütüphane ve araç mevcut olduğu için, projelerinizi hızlı bir şekilde gerçekleştirebilirsiniz. Örneğin, NumPy ve pandas gibi kütüphaneler veri analitiği için, Scikit-learn ve TensorFlow gibi kütüphaneler ise makine öğrenimi için kullanılabilir. Python bağımsız bir platform olup, Windows, Mac ve Linux gibi farklı işletim sistemlerinde çalışabilir.

Popüler programlama dilleri arasında Python, genellikle en üst sıralarda yer alır. Örneğin, TIOBE indeksi, programlama dillerinin popülerliğini ölçen bir sistemdir ve Python, genellikle en üst sıralarda yer alır. Ayrıca, GitHub, Stack Overflow gibi platformlar tarafından yapılan yıllık anketlerde Python da sıklıkla en popüler diller arasında yer almaktadır. Örneğin, 2021 yılında Stack Overflow tarafından yapılan ankete göre Python, en popüler programlama dillerinden biridir.

Python, geniş kullanım alanlarına sahip, yüksek seviyeli bir programlama dilidir. Özellikle veri analitiği, makine öğrenimi ve uygulamalı matematik gibi alanlarda sıklıkla kullanılır. Python'un okunabilirliği ve kolay öğrenilebilirliği nedeniyle, yeni başlayanlar için idealdir. Ayrıca, çok sayıda kütüphane ve araç mevcut olduğu için, projelerinizi hızlı bir şekilde gerçekleştirebilirsiniz. Örneğin, NumPy ve pandas gibi kütüphaneler veri analitiği için, Scikit-learn ve TensorFlow gibi kütüphaneler ise makine öğrenimi için kullanılabilir. Python, ayrıca, platform bağımsızdır ve Windows, Mac ve Linux gibi farklı işletim sistemlerinde çalışabilir.

Popüler programlama dilleri arasında Python, genellikle en üst sıralarda yer alır. Örneğin, TIOBE indeksi, programlama dillerinin popülerliğini ölçen bir sistemdir ve Python, genellikle en üst sıralarda yer alır. Ayrıca, GitHub, Stack Overflow gibi platformlar tarafından yapılan yıllık anketlerde Python da sıklıkla en popüler diller arasında yer almaktadır. Örneğin, 2021 yılında Stack Overflow tarafından yapılan ankete göre Python, en popüler programlama dillerinden biridir.(Şekil 9)



Şekil 9. Programlama Dillerinin Yıllara Göre Karşılaştırması

BÖLÜM IV VERİ ANALİZİ

4.1 KULLANILACAK VERİ SETLERİNİN TANITILMASI

Satış veri seti olarak 2021 yılının birinci haftasından başlayarak 2022 yılının yirmi sekizinci haftasına kadar toplamda 80 haftadan oluşmaktadır. Ürün ailesi olarak kadın tişört grubu seçilmiştir. İstanbul ilinde bulunan 88 mağaza ve 2083 adet üründen oluşmaktadır. Veri setinde bir örneklem Tablo 3’de gösterilmiştir.

Tablo 3. Ürün Satış Adet Veri Seti

YılHafta	MagazaID	UrunID	Satış Adet	Stok	Birim Fiyat
202218	723	2386085	0	1	null
202136	673	1910203	0	13	null
202135	1162	2009397	3	8	35
202217	811	2302158	1	6	59
202208	639	2341987	1	11	69
202113	839	1908902	1	8	37
202118	1135	1944387	4	10	34
202128	704	2033236	2	7	44
202122	415	2068474	1	12	37
202116	267	1893323	0	1	null

Ürün görsellerinden Resnet152 derin öğrenme modeli ile elde edilen veri setinden bir örneklem Tablo 4’de gösterilmiştir. Bu modelin oluşturduğu bir vektör 4096 boyutundadır.

Mağazalara ait lokasyon bilgilerini ve özelliklerini belirten veri setinden bir örneklem aşağıda gösterilmiştir. Bütün mağazalar bu tez çalışmasında kullanılmayacak olup sadece İstanbul’da bulunan satış performansı en yüksek 10 mağaza seçilmiştir (Tablo 5).

Tablo 4. Ürün Görsel Öznitelikleri Veri Seti

Urun ID	0	1	2	4094	4095
1893231	-0.17591137	-0.08032271	0.94157684	-0.4846894	-0.7919965
1893234	-0.17591137	-0.08032271	0.36286342	0.6714894	1.9264202
1893238	-0.17591137	-0.08032271	2.5778666	-0.4846894	-0.7919965
1893263	-0.17591137	-0.08032271	1.1367652	3.8862936	-0.7919965
1893264	-0.17591137	-0.08032271	-0.48063087	-0.4846894	-0.22138041
1893268	-0.17591137	-0.08032271	0.40869504	4.707937	-0.7919965
1893269	-0.17591137	1.2613014	2.5102103	0.5119232	-0.7919965
1893274	-0.17591137	-0.08032271	1.2547071	0.41197652	1.7874188
1893276	-0.17591137	0.057821155	-0.48063087	-0.4846894	3.171392
1893277	-0.17591137	0.2206198	-0.48063087	-0.42019346	-0.5398096

Tablo 5. Mağaza Özellikleri Veri Seti

Mağaza ID	Mağaza Yerleşim Tipi	İlçe Adı	Operasyonel Bölge	Açılış Tarihi
1	Mall- Üstü Kapalı	ŞİŞLİ	İSTANBUL 3.BÖLGE	12.06.2020
2	Mall- Üstü Kapalı	BAYRAMPAŞA	İSTANBUL 2.BÖLGE	16.11.2009
3	Mall- Üstü Kapalı	ESENYURT	TRAKYA	18.10.2012
4	Cadde	BEYOĞLU	İSTANBUL 3.BÖLGE	17.01.2003
5	Cadde	FATİH	İSTANBUL 2.BÖLGE	06.12.2014
6	Mall-Üstü Açık	GAZİOSMANPAŞA	İSTANBUL 2.BÖLGE	31.07.2015
7	Cadde	BAKIRKÖY	İSTANBUL 2.BÖLGE	22.10.1999
8	Mall- Üstü Kapalı	SARIYER	İSTANBUL 3.BÖLGE	22.09.2007
9	Cadde	BAGCILAR	İSTANBUL 1.BÖLGE	24.03.2012
10	Cadde	FATİH	İSTANBUL 2. BÖLGE	03.07.20017

Ürünlere ait tanımlayıcı ve ayırt edici özellikler kullanılarak müşterilerin ürünleri hangi özelliklerine göre tercih ettiklerini belirlemek ve modelin bu özelliklere göre öğrenmesi için Tablo 6 örneklem veri seti kullanılmıştır.

Tablo 6. Ürün Özellikleri Veri Seti

UrunID	Moda Tanımı	Kumaş Tipi	Yaka tipi	Boy	Kalınlık	Fit
1406755	DC	SÜPREM	U YAKA	STANDART	İNCE	STANDART
1422105	DC	SUPREM	BİSİKLET YAKA	STANDART	İNCE	STANDART
1448372	DC	SÜPREM	KAYIK YAKA	STANDART	İNCE	STANDART
1453773	DC	SÜPREM	KAYIK YAKA	STANDART	İNCE	STANDART
1454157	DC	SUPREM	BİSİKLET YAKA	STANDART	İNCE	STANDART

Özel günlerin satış üzerindeki etkisini gözlemlemek için, her hafta özel gün olup olmadığı belirlenmiş, özel gün olan haftalar etiketlenmiştir. Bir haftada birden fazla özel gün var ise önem derecesine göre bir tanesi seçilmiştir. Tablo 7’de bu veri setinin bir örnekleme gösterilmiştir.

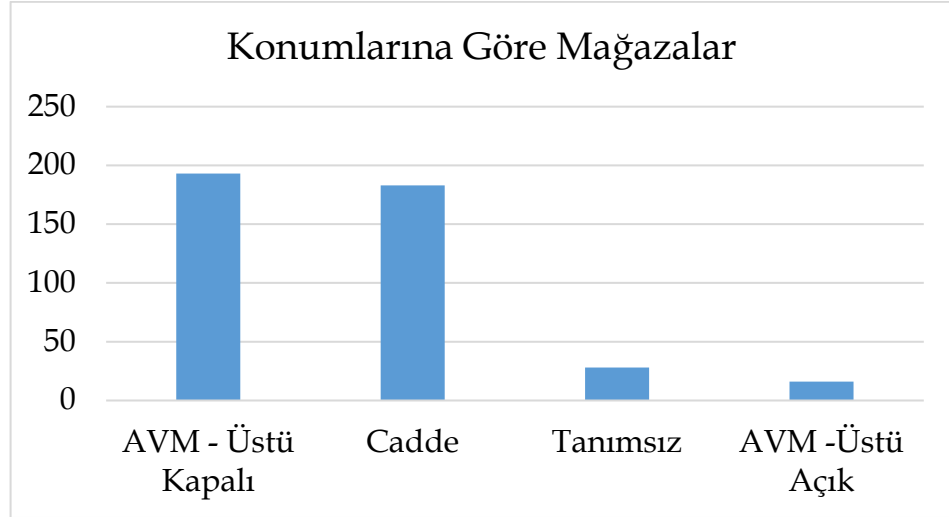
Tablo 7. Özel Günler

Yıl Hafta	Özel Gün
202101	YIL BASI
202102	NormalHafta
202103	NormalHafta
202104	YARI YIL TATILI
202105	NormalHafta
202106	NormalHafta
202107	SEVGILILER GUNU
202108	NormalHafta
202109	NormalHafta
202110	NormalHafta

4.2 KEŞİFÇİ VERİ ANALİZİ

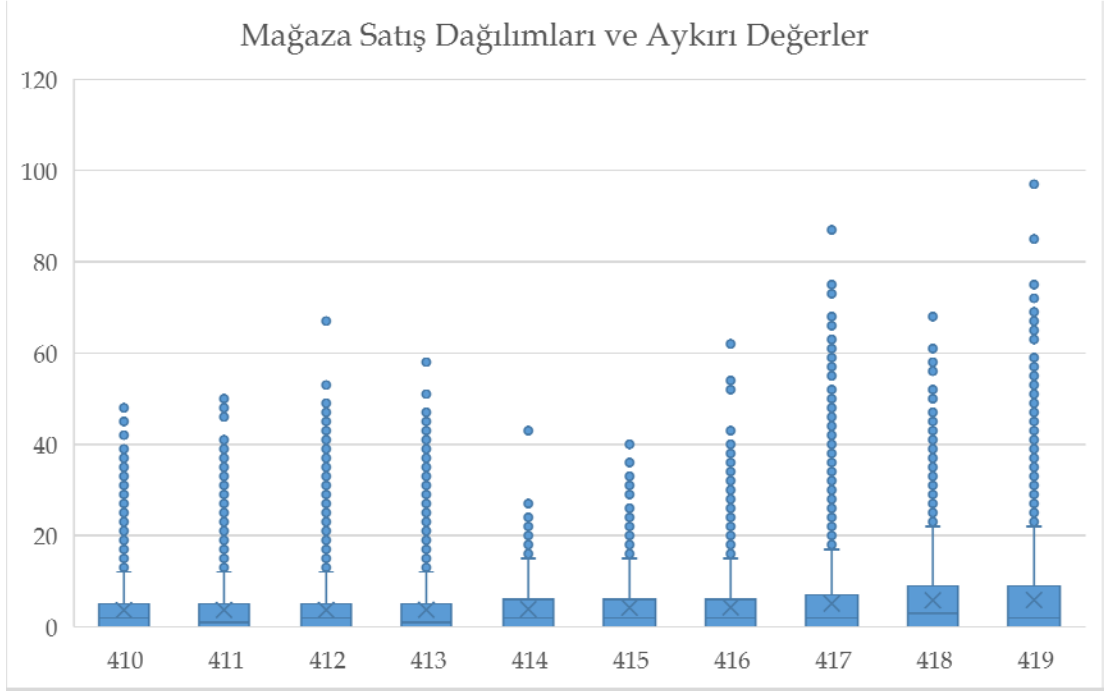
Veri setini tanımak ve veri setinde bulunan özelliklerden hangilerinin satış tahmini için kullanılacağını belirlemek için bazı analizlerin yapılması gerekir. Analiz yapılırken görselleştirmeler veri setini tanımada kolaylık sağlar.

Mağaza özelliklerinden satış tahmininde etkileyici bir etken olabilecek mağaza konumlarına göre tipleri Şekil 10’da gösterilmiştir.



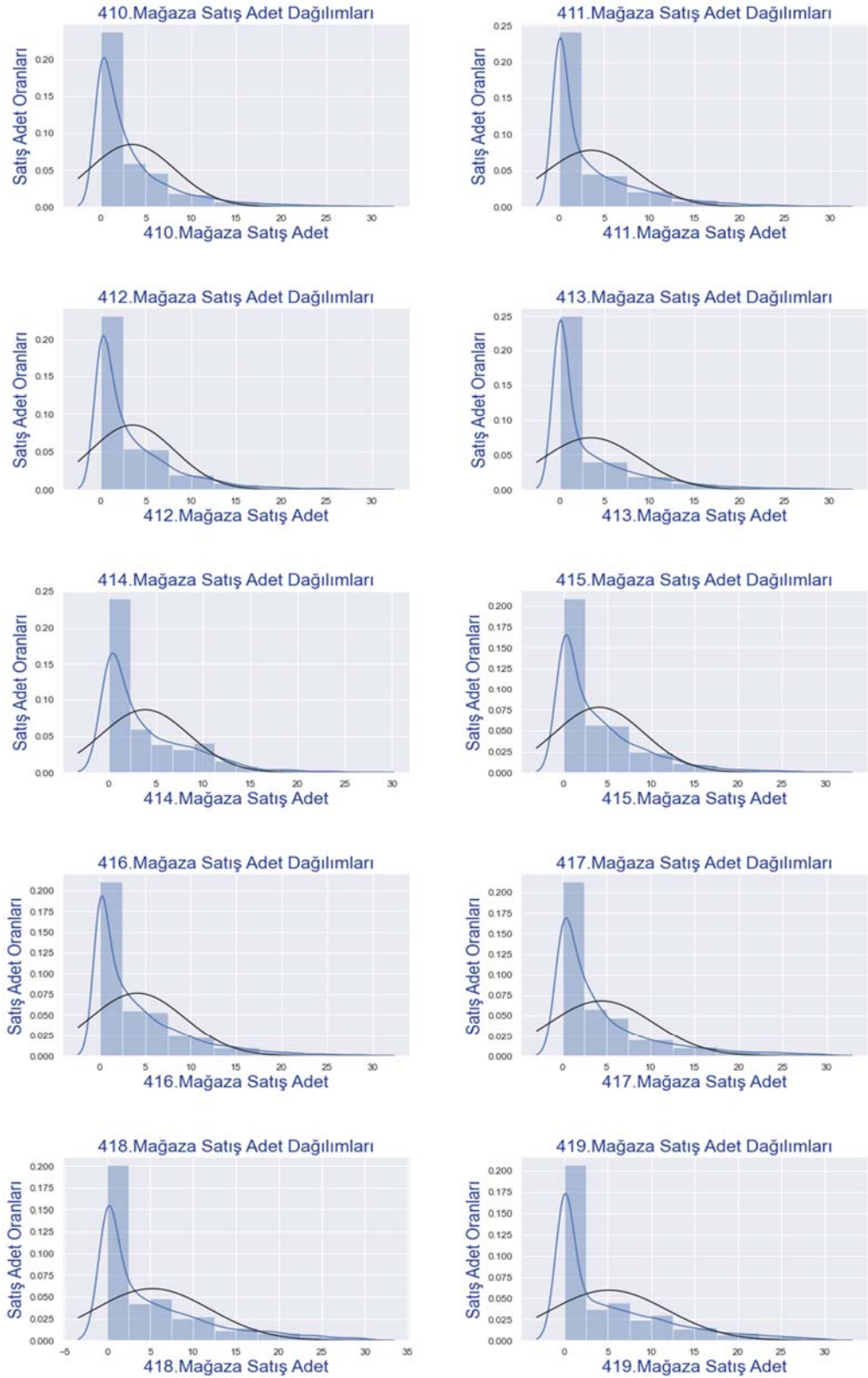
Şekil 10. Mağaza Özelliklerinin Dağılımı

Hedef değişkenimiz olan satış adetinin dağılımını gözlemlemek ve aykırı değer olup olmadığı kontrolü için kutu grafiğinde yararlanılmıştır (Şekil 11).



Şekil 11. Mağazanın Satış Adet Dağılımları ve Aykırı Değerler

Bu mağazaların satış adet dağılımına çizgi grafiği ile de baktığımızda sağa çarpıklığı daha net bir şekilde gözlemlenebilir (Şekil 12).



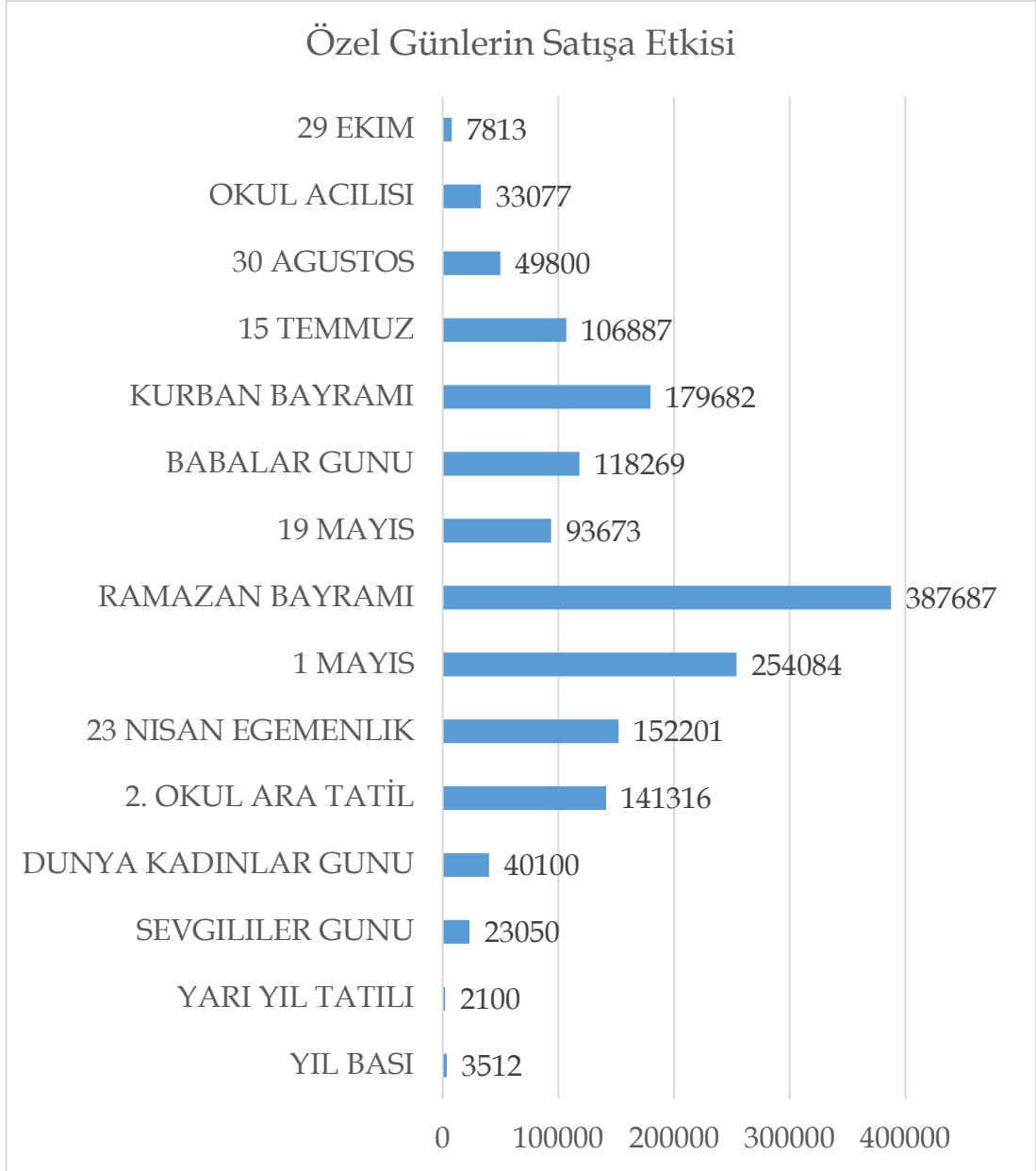
Şekil 12. Mağaza Satış Adet Dağılımlarının Gözlemlenmesi

Satış ile stok arasında güçlü bir korelasyonun olduğu Şekil 13'te görülmektedir. Fakat stok satışı tahminlerken bu korelasyon kullanılmamalıdır. Çünkü, gelecek tahmini yaparken eldeki stok bilgisi olmayacaktır. Aslında satış tahmini yapmaktaki amaç daha iyi bir stok optimizasyonu yapmaktır.



Şekil 13. Satış Stok İlişkisi

Özel günlerde ve yaklaşan haftalarında, insanların alış veriş yapma olasılıkları artmaktadır (Şekil 14).



Şekil 14. Özel Günlere Göre Satış Adetlerinin Gözlemlenmesi

BÖLÜM V PROJE UYGULAMASI

Moda sektöründe faaliyet gösteren bir perakende şirketi için, haftalık bazda kadın tişört ürünlerinin satış tahmini yapılmıştır. Bu amaçla, İstanbul ilinde bulunan mağazaların satış verileri kullanılmıştır. Veriler, belirli bir dönem içinde mağazalarda satılan kadın tişört ürünlerinin satış adetlerini, satış fiyatlarını, satış tarihlerini, mağaza özelliklerini, ürün özelliklerini ve özel günleri içermektedir.

Uygulama adımları:

- 1. Veri Setlerinin Toplanması ve Temizlenmesi:** Bu işlem, veri setinde yer alan boşlukları, eksik verileri veya hatalı girişleri temizlemek için yapılmıştır.
- 2. Veri Setinde Bulunan Ürünlerin Kümelenmesi:** Bu yöntem ile ürünlerin satış adetleri arasındaki benzerlikleri belirlememizi sağladı. Örneğin, elektronik cihazlar ve gıda ürünleri arasındaki satış adetleri arasındaki farklılıklar nedeniyle iki farklı kümeye ayrılabilir. Elimizde bulunan ürünler ile de benzer ürünler benzer satış performansı sergiler düşüncesiyle yapılmıştır. Aynı zaman da ürünlerimizin geçmiş haftalardaki satış performansları gözlemlenebilir hale geldi. İki farklı kümeleme yaklaşımı kullanılmıştır. Ürünlerin özelliklerini içeren tabular veri seti ile K-modes kümeleme yöntemi kullanılmıştır (Şekil 15). Ürün özelliklerinin derin öğrenme modelinden elde edilmesi ile oluşturulan veri setinin kümelenmesi de K-means kümeleme modeli ile yapılmıştır ve sonuçlar bu ayrım bazın da karşılaştırılmıştır (Şekil 16).



Şekil 15. K-Modes kümeleme modeli ile elde edilen örnek bir ürün kümesi



Şekil 16. K-Means kümeleme modeli ile elde edilen örnek bir ürün kümesi

3. **Eğitim Haftaları:** Eğitim veri seti olarak 2021 yılının 1.haftasından 2022 yılının 27.haftasına kadar alınmıştır.
4. **Hedef Hafta:** Tahmin edilecek hafta olarak 2022 yılının 28. haftası alınmıştır. Hem regresyon hem de sınıflandırma modelleri için bu hafta kullanılmıştır.
5. **Regresyon Modeli:** İkinci adımda yaptığımız kümeleme yöntemleri ile elde ettiğiniz veri setleri için regresyon modelleri uygulanmıştır. Bu modeller, ürünlerin satış adetlerini tahmin etmek için kullanıldı ve elde edilen sonuçlar ortalama mutlak hata, hata kareleri ortalaması ve ortalama mutlak yüzde hata metrikleri ile

ölçüldü. Ürün bazlı satış adetlerinin düşük olması sebebi ile ortalama mutlak yüzde hata metriğimiz çok yüksek çıkabilmektedir.

6. **Sınıflandırma Modeli:** Alternatif olarak, hedef değişkenimiz olan satış adeti, gruplayarak sınıflandırma modeline dönüştürme yaklaşımıdır. Bu yaklaşımın denenmesinin sebebi satış adetlerinin sayısal olarak küçük olmasından dolayı daha doğru tahminlerin elde edilememesidir. Sınıflandırma yaklaşımı, satış adeti hedef değişkenini belirli bir aralıkta kategorize ederek yapıldı. Sıfır satış "düşük" olarak, 1'den büyük ve 4'ten küçük satışlar "orta" olarak ve 4'ten büyük satışlarda "yüksek" satış olarak sınıflandırıldı.
7. **Hata Analizi:** Elde edilen sonuçlar değerlendirildi ve en iyi sonucu veren model seçildi.

Uygulama Sonuçları :

1. Regresyon Modeli:

Tabular veri seti ve ürün görselleri kullanarak farklı yöntemlerle elde edilen ürün özellikleriyle ilgili Tablo 8'de verilen regresyon sonuçlarına göre, LightGBM modelinin en başarılı olduğunu gözlemlenmektedir. LightGBM modeli, hem tabular veri seti hem de ürün görselleri kullanarak elde edilen ürün özellikleri yaklaşımlarında diğer modellere göre daha düşük hata değerleri elde etmiştir (Tablo 9). Bu sonuçlar, LightGBM'nin diğer regresyon modellerine göre daha iyi genelleştirme yeteneği olduğunu ve verilerdeki gürültüyü daha iyi ele aldığını göstermektedir.

Tablo 8. Tabular Veri Seti Regresyon Modeli Sonuçları

	MAE	RMSE	MAPE
Lineer Regresyon	6,75	7,96	134,77
Destek Vektör Makineleri	4,49	6,14	97,7
LightGBM	3,49	4,99	55,42

Tablo 9. Ürün Görselleri Regresyon Modeli Sonuçları

	MAE	RMSE	MAPE
Lineer Regresyon	8,1	9,88	140,8
Destek Vektör Makineleri	5,1	8,3	72,4
LightGBM	4	6,5	60

Özellikle, tabular veri seti ile elde edilen ürün özellikleri yaklaşımında LightGBM modelinin daha yüksek başarı oranı elde etmesi, tabular veri seti ile oluşturulan ürün özelliklerini tahmin etmede önemli bir role sahip olduğunu göstermektedir. Bu

durumun nedeni, tabular veri setinde daha belirgin bir veri yapısı olması ve bu yapının LightGBM modelinin işleyişine daha uygun olmasıdır. Diğer taraftan, ürün görsellerinden elde edilen veride, görsellerin daha fazla gürültüye sahip olması ve veri yapısının daha belirsiz olması nedeniyle, tahminlerin daha yüksek hata oranlarına sahip olması gözlemlenmiştir.

Diğer yandan, destek vektör makineleri (SVM) modeli de tabular veri seti kullanarak elde ettiği ürün özellikleri yaklaşımında iyi bir performans göstermiştir. SVM modeli, verilerdeki karmaşıklıkları ele almak ve daha az sayıda özellik ile doğru tahminler yapmak için iyi bir seçenek olarak görülmektedir.

Lineer regresyon modelinin sonuçları, diğer modellere kıyasla daha yüksek hata değerleri elde ettiğini göstermektedir. Bu sonuçlar, lineer regresyon modelinin daha az esnek olduğunu ve karmaşık veri yapılarındaki değişkenleri ele almakta sınırlı kaldığını göstermektedir.

Sonuç olarak, LightGBM modelinin tabular veri seti ve ürün görselleri kullanarak elde edilen ürün özelliklerinin tahmininde en başarılı yöntem olduğu görülmektedir. Bu sonuçlar, farklı yöntemlerin bir arada kullanılarak daha doğru tahminler yapabileceğini göstermektedir.

Başarılı model olarak Lightgbm'nin belirlemesinden sonra hiper parametre optimizasyonu için bu model ve tabular veri seti ile devam edilmiştir. Yeni sonuçlar Tablo 10'daki gibidir.

Tablo 10. Tabular Veri Seti Regresyon Modeli Hiper Parametre Sonuçları

	MAE	RMSE	MAPE
LightGBM	1,8	2,7	50

Hiper parametre optimizasyonu yaparak LightGBM modelinin performansını daha da artırmış olduk. Elde edilen sonuçlar, tabular veri seti ve LightGBM modeli kullanarak satış tahmini için oldukça başarılı bir yaklaşım olduğunu göstermektedir.

Yeni sonuçlar incelendiğinde, elde ettiğin MAE değerinin 1.8 olduğunu görmekteyiz. Bu değer, önceki sonuçlarla karşılaştırıldığında oldukça düşük bir hata oranıdır ve modelin satış tahminleri için oldukça doğru sonuçlar verdiğini göstermektedir. Ayrıca, elde edilen MAPE değeri 50 ve RMSE değeri 2.7'dir. Bu sonuçlar, modelin başarılı bir

şekilde hiper parametrelerin ayarlanması ile daha da doğru tahminler yapabildiğini göstermektedir.

Satış tahmini için LightGBM modeli, tabular veri seti ve modelin hiper parametre optimizasyonu yapılması, daha doğru sonuçlar elde etmek için önemli faktörler olduğu gözükmemektedir. Bu sonuçlar ayrıca, müşteri taleplerinin öngörülmesi gibi birçok iş uygulaması için de kullanılabilen başarılı bir regresyon modeli oluşturmanın mümkün olduğunu göstermektedir.

Sonuç olarak, elde ettiğin başarı metrikleri oldukça etkileyici ve LightGBM modelinin, satış tahmini gibi birçok iş uygulamasında yani lineer bir dünyanın olmadığı veri setleri için uygun bir model olduğunu göstermektedir.

2. Sınıflandırma Modeli:

Elde edilen sonuçlara göre, sınıflandırma yaklaşımında regresyon modelinde olduğu gibi ürün özelliklerini elde etmek için hem görsel veri hem de tabular veri kullanılmıştır. Tabular veri ile elde ettiğiniz ürün özellikleri için, lojistik regresyon, destek vektör makineleri ve LightGBM algoritmaları kullanarak sınıflandırma performansı değerlendirilmiştir. Bu değerlendirme sonucunda, LightGBM'nin diğer iki yöntemden daha iyi performans gösterdiği görülmüştür (Tablo 11).

Tablo 11. Tabular Veri Seti Sınıflandırma Modeli Sonuçları

Sınıflandırma Modeli	F1-Skor
Lojistik Regresyon	0,33
Destek Vektör Makineleri	0,32
LightGBM	0,39

Öte yandan, görsel verilerden elde edilen ürün özellikleri için de aynı yöntemleri kullanarak sınıflandırma performansını değerlendirilmiştir. Sonuçlar, tabular veriden elde edilen sonuçlara kıyasla daha düşük performans gösterdiği görülmüştür. Ancak yine de LightGBM, diğer iki yöntemden daha iyi bir performans göstermiştir (Tablo 12).

Elde edilen sonuçlardan biri, farklı veri türlerinin farklı sınıflandırma performansı sağladığıdır. Tabular veri setinin, görsel veri setinden daha iyi sonuçlar verdiği görülmektedir.

Tablo 12. Ürün Görselleri Sınıflandırma Modeli Sonuçları

Sınıflandırma Modeli	F1-Skor
Lojistik Regresyon	0,21
Destek Vektör Makineleri	0,3
LightGBM	0,37

Genel olarak, sınıflandırma yaklaşımı kullanarak ürünlerin satış adetlerini öngörmek için farklı veri kaynaklarının kullanımı, sınıflandırma performansını artırabilir. Sonuç olarak, bu yaklaşımın satış tahmini performansını artırmak için daha fazla araştırma yapmak gerektirir.

BÖLÜM VI SONUÇ

Tabular verilerle yapılan regresyon yaklaşımı sonuçlarından farklı olarak, sınıflandırma yaklaşımı sonuçları daha düşük performans göstermiştir. F1 skoru, 0 ile 1 arasında bir ölçümdür ve 1'e yaklaştıkça daha iyi bir performans gösterir. Lojistik regresyon ve destek vektör makineleri modellerinde, tabular verilerle elde edilen ürün özellikleri yaklaşımı, görsel verilerle elde edilen yaklaşımdan daha düşük bir F1 skoru ile sonuçlanmıştır. Ancak, LightGBM modelinde görsel verilerle elde edilen yaklaşım, tabular verilerle elde edilen yaklaşımdan daha düşük bir F1 skoru ile sonuçlanmıştır.

Sınıflandırma yaklaşımı, satış adetleri hakkında daha ayrıntılı bilgi sağlamak için kullanılabilir. Özellikle, sıfır satış adetlerinin ayrı bir sınıf olarak ele alınması, sıfır satışa neden olan sebeplerin belirlenmesine yardımcı olabilir ve satış stratejilerinin geliştirilmesine katkıda bulunabilir. Ancak, sınıflandırma yaklaşımının, regresyon yaklaşımından daha iyi sonuçlar vermediği görülmüştür. Bu nedenle, satış tahmini yapmak için regresyon yaklaşımı daha uygun olur.

Sınıflandırma yaklaşımı ile regresyon yaklaşımı arasında bir karşılaştırma yaparak, her bir yaklaşımın güçlü ve zayıf yönleri hakkında daha iyi bir fikir edinilebilmektedir. Bu da daha doğru ve verimli bir satış tahmini yapmak için farklı yaklaşımların birleştirilmesine yardımcı olabilmektedir.

Öneri olarak ise; tabular ve görsel verinin aynı anda kullanılması, ürünlerin özelliklerini daha kapsamlı bir şekilde temsil etmeyi amaçlamaktadır. Bu yaklaşım, hem tabular verideki sayısal ve kategorik özellikleri hem de görsel verideki görsel özellikleri bir araya getirerek daha zengin bir özellik vektörü oluşturmayı hedefler. Bu iki veri kaynağı bir araya getirilerek, ürünün hem sayısal ve kategorik özellikleri hem de görsel özellikleri aynı anda kullanılabilir. Böylece, özellik vektöründe hem tabular hem de görsel veriye ait bilgiler yer alır. Bu yaklaşım, özellikle ürünlerin satış tahmini gibi problemlerde daha kapsamlı ve genelleyici bir özellik vektörü elde etmeyi sağlar. Hem tabular verinin içerdiği yapısal bilgileri hem de görsel verinin sağladığı görsel

özellikleri kullanmak, daha fazla bilgi ve derinlik sağlar. Bu da satış tahmini performansını artırabilir ve daha doğru sonuçlar elde edilmesini sağlayabilir.

KAYNAKÇA

- Arif, Md. A. I., Sany, S. I., Nahin, F. I., & Rabby, A. S. A. (2019). Comparison Study: Product Demand Forecasting with Machine Learning for Shop. *IEEE Conference ID: 46866*, 171–176.
- Dairu, X., & Shilong, Z. (2021). Machine Learning Model for Sales Forecasting by Using XGBoost. *2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021*, 480–483. <https://doi.org/10.1109/ICCECE51280.2021.9342304>
- Data Mining. (n.d.). Retrieved February 8, 2023, from <https://tr.weblogographic.com/difference-between-data-mining-and-machine-learning-3290>
- Dinçoğlu, P. (2022). *PERAKENDE SEKTÖRÜNDE VERİ MADENCİLİĞİ TEKNİKLERİ İLE SATIŞ TAHMİNİ*. Maltepe Üniversitesi.
- DipCE BEng Rod Pierce, with contributions from many others. (n.d.). *What is Data?* Retrieved February 8, 2023, from <https://www.mathsisfun.com/data/data.html>
- EMRE, İ. E. (2021). PSİKİYATRİK HASTALIKLARIN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE AYRIŞTIRILMASI [İstanbul Üniversitesi]. In *Psikiyatride Guncel Yaklasimler - Current Approaches in Psychiatry* (Vol. 13, Issue 2). <https://doi.org/10.18863/pgy.779987>
- Freeman, I., Haigler, A., Schmeelk, S., Ellrodt, L., & Fields, T. (2019). What are they Researching? Examining Industry-Based Doctoral Dissertation Research through the Lens of Machine Learning. *Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018*, 1338–1340. <https://doi.org/10.1109/ICMLA.2018.00217>
- Gupta, I., Mittal, H., Rikhari, D., & Singh, A. K. (2022). *MLRM: A Multiple Linear Regression based Model for Average Temperature Prediction of A Day*. <http://arxiv.org/abs/2203.05835>
- Henzel, J., Wawrowski, L., Kubina, A., Sikora, M., & Wrobel, L. (2022). Demand forecasting in the fashion business - an example of customized nearest neighbour

- and linear mixed model approaches. *Proceedings of the 17th Conference on Computer Science and Intelligence Systems, FedCSIS 2022*, 61–65. <https://doi.org/10.15439/2022F256>
- Işık, H. (2020). *SALES FORECASTING IN FASHION RETAIL INDUSTRY WITH CLASSICAL AND MACHINE LEARNING METHODS*. ISTANBUL TECHNICAL UNIVERSITY.
- Jeremy Zhang. (n.d.). *UNet — Line by Line Explanation*. Retrieved February 8, 2023, from <https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5>
- Jiang, H., Ruan, J., & Sun, J. (2021). Application of Machine Learning Model and Hybrid Model in Retail Sales Forecast. *2021 IEEE 6th International Conference on Big Data Analytics, ICBDA 2021*, 69–75. <https://doi.org/10.1109/ICBDA51983.2021.9403224>
- Keung, K. L., Lee, C. K. M., & Yiu, Y. H. (2021). A Machine Learning Predictive Model for Shipment Delay and Demand Forecasting for Warehouses and Sales Data. *2021 IEEE International Conference on Industrial Engineering and Engineering Management, IEEM 2021*, 1010–1014. <https://doi.org/10.1109/IEEM50564.2021.9672946>
- Levin Dabhi. (n.d.). *Cloth Segmentation*. Retrieved February 8, 2023, from <https://github.com/levindabhi/cloth-segmentation>
- Li, Q., Li, D., Zhao, K., Wang, L., & Wang, K. (2022). State of health estimation of lithium-ion battery based on improved ant lion optimization and support vector regression. *Journal of Energy Storage*, 50. <https://doi.org/10.1016/j.est.2022.104215>
- Microsoft Corporation. (n.d.). Machine Learning Algorithm Cheat Sheet. In 2021. Retrieved February 5, 2023, from <https://learn.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet>
- Paymode, A. S., & Malode, V. B. (2022). Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG.

Artificial Intelligence in Agriculture, 6, 23–33.
<https://doi.org/10.1016/j.aiia.2021.12.002>

POLAT, E. D. (2022). *BEYAZ EŞYA SEKTÖRÜNDE SATIŞ TAHMİNİ: BİR VERİ MADENCİLİĞİ UYGULAMASI*. ULUDAĞ ÜNİVERSİTESİ.

Punam, K., Pamula, R., & Jain, P. K. (2018). A Two-Level Statistical Model for Big Mart Sales Prediction. *IEEE*, 617–620.

Ramachandra, H. v., Balaraju, G., Rajashekar, A., & Patil, H. (2021). Machine learning application for black friday sales prediction framework. *2021 International Conference on Emerging Smart Computing and Informatics, ESCI 2021*, 57–61. <https://doi.org/10.1109/ESCI50559.2021.9396994>

Rebecca Vickery. (n.d.). *Machine Learning Algorithms Cheat Sheet*. Retrieved February 8, 2023, from <https://towardsdatascience.com/machine-learning-algorithms-cheat-sheet-2f01d1d3aa37>

Robles-Velasco, A., Cortés, P., Muñuzuri, J., & Onieva, L. (2020a). Prediction of pipe failures in water supply networks using logistic regression and support vector classification. *Reliability Engineering and System Safety*, 196. <https://doi.org/10.1016/j.ress.2019.106754>

Robles-Velasco, A., Cortés, P., Muñuzuri, J., & Onieva, L. (2020b). Prediction of pipe failures in water supply networks using logistic regression and support vector classification. *Reliability Engineering and System Safety*, 196. <https://doi.org/10.1016/j.ress.2019.106754>

Saket Mittal. (2022, May 10). *Data Mining vs Machine Learning: 6 Critical Differences*. <https://hevodata.com/learn/data-mining-vs-machine-learning/>

Sevgi, P. (2012). *HAZIR GİYİM PERAKENDECİLİĞİ YAPAN BİR FİRMADA YAPAY SİNİR AĞLARI İLE SATIŞ TAHMİNİ*. YILDIZ TEKNİK ÜNİVERSİTESİ.

Shweta Goyal. (n.d.). *Evaluation Metrics for Regression Models*. Retrieved February 8, 2023, from <https://medium.com/analytics-vidhya/evaluation-metrics-for-regression-models-c91c65d73af>

- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Türker, N. D. (2021). *HOTEL ROOM SALES PREDICTION FOR A TRAVEL AGENCY*. Sabancı University.
- Wang, J. (2020). A hybrid machine learning model for sales prediction. *Proceedings - 2020 International Conference on Intelligent Computing and Human-Computer Interaction, ICHCI 2020*, 363–366. <https://doi.org/10.1109/ICHCI51889.2020.00083>
- Yegen, N. (2020). *PERAKENDE SEKTÖRÜNDE VERİ MADENCİLİĞİ İLE SATIŞ TAHMİNİ*. BAHÇEŞEHİR ÜNİVERSİTESİ.
- Yin, Z., Shi, L., Luo, J., Xu, S., Yuan, Y., Tan, X., & Zhu, J. (2023). Pump Feature Construction and Electrical Energy Consumption Prediction Based on Feature Engineering and LightGBM Algorithm. *Sustainability (Switzerland)*, 15(1). <https://doi.org/10.3390/su15010789>
- Zeng, H., Yang, C., Zhang, H., Wu, Z., Zhang, J., Dai, G., Babiloni, F., Kong, W., & Chuang, L. (2019). A LightGBM-Based EEG Analysis Method for Driver Mental States Classification. *Computational Intelligence and Neuroscience*, 2019. <https://doi.org/10.1155/2019/3761203>
- Zhaowei jie, Chenliang, & Hujiangmin. (2020). Forecast Rossmann Store Sales Base on Xgboost Model. *Proceedings - 2020 2nd International Conference on Economic Management and Model Engineering, ICEMME 2020*, 521–525. <https://doi.org/10.1109/ICEMME51517.2020.00110>

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı: Abdül MERAL

Uyruğu: T.C.

EĞİTİM

Derece	Kurum	Mezuniyet Yılı
Yüksek Lisans	İstanbul Medeniyet Üniversitesi Hesaplamalı Bilimler ve Uygulamalı Matematik	2023
Lisans	Sakarya Üniversitesi Elektrik Elektronik Mühendisliği	2013

DENEYİMLER

Tarih	Kurum	Görev
2020-2023	Lcwaikiki	Veri Bilimci
2019-2019	Rish Yazılım	Veri Analisti
2018-2019	Asaş Alüminyum	Bakım Mühendisi
2016-2018	Elektres Elektrik	Üretim Müdürü