

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
VERİ YÖNETİMİ VE ANALİZİ PROGRAMI
YÜKSEK LİSANS TEZİ

KAMPANYA SEÇİMİNDE MÜŞTERİ
DAVRANIŞLARININ MAKİNE ÖĞRENMESİ İLE
ÖNGÖRÜLMESİ: CRM UYGULAMALARI

Semra ALTINCELEP

Danışman
Dr. Öğr. Üyesi EFE SARIBAY

2025

T.C.
DOKUZ EYLÜL ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
EKONOMETRİ ANABİLİM DALI
VERİ YÖNETİMİ VE ANALİZİ PROGRAMI
YÜKSEK LİSANS TEZİ

KAMPANYA SEÇİMİNDE MÜŞTERİ
DAVRANIŞLARININ MAKİNE ÖĞRENMESİ İLE
ÖNGÖRÜLMESİ: CRM UYGULAMALARI

Semra ALTINCELEP

Danışman
Dr. Öğr. Üyesi EFE SARIBAY

İZMİR- 2025

TEZ ONAY SAYFASI



YEMİN METNİ

Yüksek Lisans Tezi Olarak Sunduğum “Kampanya Seçiminde Müşteri Davranışlarının Makine Öğrenmesi ile Öngörülmesi: CRM Uygulamaları” adlı çalışmanın, tarafımdan, akademik kurallara ve etik değerlere uygun olarak yazıldığını ve yararlandığım eserlerin kaynakçada gösterilenlerden oluştuğunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve bunu onurumla doğrularım.

28.1.2025

Semra ALTINCELEP

İmza



ÖZET

Yüksek Lisans Tezi

Kampanya Seçiminde Müşteri Davranışlarının Makine Öğrenmesi ile

Öngörülmesi: CRM Uygulamaları

Semra ALTINCELEP

Dokuz Eylül Üniversitesi

Sosyal Bilimler Enstitüsü

Ekonometri Anabilim Dalı

Veri Yönetimi ve Analizi Programı

Son yıllarda artan veri hacmi ve gelişen teknoloji sebebiyle birçok ürün ve hizmet ulaşılabilir bir hale gelmiştir. Bu gelişmeler neticesinde verilerden anlamlar çıkarıp aksiyom alan çoğu şirket çağın gerisinde kalmayarak rekabetçi piyasa koşullarında ayakta kalmayı başarmışlardır. Gelecek bir zaman diliminde yaşanacak herhangi bir durumu önceden kestirebilmek şirketler açısından oldukça önemli ve öne çıkmayı sağlayan bir durumdur. Karı maksimize etmek en temel amaçlardan biridir. Bunun içinde ihtiyaç duyulan şey öncelikle bir müşteri ve müşteriye elde tutma sadakatidir. Markalar çoğu zaman yeni müşteri elde etmenin maliyetini göze almamak için var olan müşterilerine sürekli alışveriş yapmaları için kampanyalar yaparlar. Tam olarak bu noktada alım yapacak müşteri kitlesini belirlemek ve alım yapacakları fiyatları ön görebilmek için devreye makine öğrenmesi ve CRM kavramları girer. Müşteri kitlesini bilmek hem eldeki müşteriye sürekli alışverişe davet etmek için hem de daha sonra gelecek olan müşterinin hangi kampanyayı seçip alışveriş yapmaya meyilli olacağı makine öğrenmesi aracılığıyla tahmin edilerek doğru bir yönlendirme sağlar. Pazarlama kampanyalarında kişiselleştirilmiş kampanyalar ve bununla birlikte özelleştirilmiş fiyatlar sunabilmek amacıyla müşteri ilişkileri yönetiminde veri bilimi çalışmaları oldukça gündemdedir. Bu tez çalışmasında pazarlama stratejileri uygulayarak elde edilen müşterilerin şirket karını maksimize ederek müşterileri elde tutmayı hedeflerken müşterilerin satın alma davranışlarını makine öğrenmesi algoritmaları ile

analiz edip müşterilere uygun satın alabilecekleri kampanyaları tahmin eder. Böylelikle müşterilere alım yapacakları doğru fiyatlarla kampanyalar belirlenip var olan müşteriler elde tutularak kar maksimize edilecektir. Buna istinaden bu çalışmada beş farklı makine öğrenmesi algoritmasının performansı detaylı bir şekilde incelenmiştir. Random Forest, Decision Tree, Support Vector Machine, XGBoost ve CatBoost algoritmaları, kampanya seçiminde müşterilerin demografik özelliklerine göre seçilen kampanyayı belirlemek üzere kapsamlı bir karşılaştırmaya tabi tutulmuştur. Araştırma sonuçları, her algoritmanın performansını, doğruluk oranını ve genel etkinliğini değerlendirmiş ve özellikle Random Forest, XgBoost ve CatBoost algoritmalarının diğer yöntemlere kıyasla daha yüksek doğruluk oranı ve genel performans sunduğunu ortaya koymuştur. Bu bulgular, işletmelerin kampanya stratejilerini kişiselleştirilme açısından nasıl optimize edebileceği konusunda değerli bilgiler sunmaktadır.

Anahtar Kelimeler: Yapay Zeka, Makine Öğrenmesi, Kampanya Yönetimi, CRM, Topluluk yöntemleri.

ABSTRACT

Master's Thesis

Predicting Customer Behavior in Campaign Selection Using Machine Learning:

CRM Applications

Semra ALTINCELEP

Dokuz Eylül University

Graduate School of Social Sciences

Department of Econometrics

Data Management and Analysis Program

In recent years, the increasing volume of data and advancements in technology have made numerous products and services more accessible. As a result, many companies that extract meaningful insights from data and take action accordingly have managed to stay competitive in the evolving market landscape. The ability to predict future events in advance is crucial for businesses, as it provides a significant competitive advantage. Maximizing profit is one of the primary objectives, and achieving this requires acquiring and retaining loyal customers. Since acquiring new customers is often costly, brands frequently launch campaigns to encourage their existing customers to make repeated purchases. At this point, machine learning and Customer Relationship Management (CRM) play a vital role in identifying the target customer base and predicting the optimal pricing strategies for them. Understanding the customer base not only enables businesses to encourage existing customers to continue shopping but also allows them to predict which future customers are more likely to choose specific campaigns, thus ensuring effective targeting through machine learning. In marketing campaigns, the application of data science in customer relationship management has become increasingly relevant for offering personalized campaigns and customized pricing strategies. This thesis aims to maximize company profits by retaining customers acquired through marketing strategies while analyzing their purchasing behavior using machine learning algorithms to predict suitable campaigns for them. In this

way, campaigns will be designed with optimal pricing for customers, ensuring customer retention and maximizing profitability. Accordingly, this study conducts a detailed performance analysis of five different machine learning algorithms: Random Forest, Decision Tree, Support Vector Machine, XGBoost, and CatBoost. These algorithms are comprehensively compared to determine the most effective approach for selecting the appropriate campaign based on customers' demographic characteristics. The research findings evaluate the accuracy and overall effectiveness of each algorithm, revealing that Random Forest, XGBoost, and CatBoost outperform other methods in terms of accuracy and overall performance. These findings provide valuable insights into how businesses can optimize their campaign strategies through personalization and data-driven decision-making.

Keywords: Artificial Intelligence, Machine Learning, Campaign Management, CRM, Ensemble Learning.

KAMPANYA SEÇİMİNDE MÜŞTERİ DAVRANIŞLARININ MAKİNE ÖĞRENMESİ İLE ÖNGÖRÜLMESİ: CRM UYGULAMALARI

İÇİNDEKİLER

TEZ ONAY SAYFASI	ii
YEMİN METNİ	iii
ÖZET	iv
ABSTRACT	vi
İÇİNDEKİLER	viii
KISALTMALAR	xii
TABLOLAR LİSTESİ	xiii
ŞEKİLLER LİSTESİ	xiv
GİRİŞ	1

BİRİNCİ BÖLÜM

PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ

1.1. PAZARLAMA	6
1.1.1. Pazar Araştırması	7
1.1.2. Hedef Kitle Belirleme	7
1.1.3. Konumlandırma	7
1.1.4. Pazarlama Karması (4P)	8
1.1.4.1. Ürün (Product)	8
1.1.4.2. Fiyat (Price)	9
1.1.4.3. Dağıtım (Place)	9
1.1.4.4. Tutundurma (Promotion)	9
1.1.5. Müşteri İlişkileri Yönetimi (CRM)	10
1.1.6. Dijital Pazarlama	11
1.2. PAZARLAMA İLETİŞİMİ	12
1.3. E-TİCARET	13
1.3.1. Avantajları	14

1.3.2. Dezavantajları	14
1.4. MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM)	15
1.5. MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM) ÖZELLİKLERİ	17
1.6. MÜŞTERİ İLİŞKİLERİ YÖNETİMİNİN KAPSAMI VE YAŞAM DÖNGÜSÜ	19
1.6.1. Müşteri Değeri Yaratma	19
1.6.2. Müşteri Memnuniyeti	20
1.6.3. Müşteri Kazanma ve Tutma	21
1.6.4. Müşteri Sadakati	21
1.7. CRM’NİN TEMEL BİLEŞENLERİ	22
1.7.1. İnsan	23
1.7.2. Süreç	24
1.7.3. Teknoloji	24
1.7.3.1. Analitik CRM	25
1.7.3.2. Operasyonel CRM	26
1.7.3.3. İşbirlikçi CRM	26
1.8. MÜŞTERİ İLİŞKİLERİ YÖNETİMİNİN (CRM) FAYDALARI	27
1.9. PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM) İLİŞKİSİ	27
1.10. BÜYÜK VERİ (BIG DATA) VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ	28
1.11. KAMPANYA YÖNETİMİ	30

İKİNCİ BÖLÜM

YAPAY ZEKA VE MAKİNE ÖĞRENMESİ

2.1. YAPAY ZEKA	32
2.2. YAPAY ZEKANIN TARİHSEL GELİŞİMİ	33
2.3. YAPAY ZEKANIN KULLANIM ALANLARI VE ALT DALLARI	36
2.3.1. Kullanım Alanları	36
2.3.1.1. Sağlık	36
2.3.1.2. Finans	36
2.3.1.3. Eğitim	36
2.3.1.4. Otomotiv	37

2.3.1.5. E-ticaret ve Pazarlama	37
2.3.2. Alt Dalları	37
2.3.2.1. Makine Öğrenmesi (Machine Learning)	38
2.3.2.2. Derin Öğrenme (Deep Learning)	38
2.3.2.3. Yapay Sinir Ağları (Artificial Neural Networks)	39
2.3.2.4. Doğal Dil İşleme (Natural Language Processing)	39
2.3.2.5. Bilgisayarla Görü (Computer Vision)	40
2.3.2.6. Robotik	40
2.4. YAPAY ZEKA VE MAKİNE ÖĞRENMESİ İLİŞKİSİ	40
2.5. MAKİNE ÖĞRENMESİ	42
2.6. MAKİNE ÖĞRENMESİ SÜREÇLERİ	44
2.7. MAKİNE ÖĞRENMESİ TÜRLERİ	49
2.7.1. Denetimli Öğrenme (Supervised Learning)	49
2.7.2. Denetimsiz Öğrenme (Unsupervised Learning)	51
2.7.3. Yarı Denetimli Öğrenme (Semi Supervised Learning)	52
2.7.4. Pekiştirmeli Öğrenme (Reinforcement Learning)	53
2.7.5. Topluluk Öğrenmesi (Ensemble Learning)	53
2.8. MAKİNE ÖĞRENMESİ ALGORİTMALARI	55
2.8.1. Rastgele Ormanlar (Random Forest)	55
2.8.2. XgBoost Algoritması	56
2.8.3. Karar Ağaçları (Decision Tree)	57
2.8.4. Destek Vektör Makineleri (Support Vector Machine)	57
2.8.5. CatBoost Algoritması (Categorical Boosting)	58
2.9. PERFORMANS METRİKLERİ	59
2.9.1. Sınıflandırma	60
2.9.1.1. Doğruluk (Accuracy)	60
2.9.1.2. Hassasiyet (Precision)	61
2.9.1.3. Duyarlılık (Recall)	61
2.9.1.4. F1 Skoru	61
2.9.1.5. ROC Eğrisi ve AUC	62
2.9.1.6. Matthews Korelasyon Katsayısı (MCC)	62
2.9.1.7. Hamming Kaybı (Hamming Loss)	62

2.9.1.8. Jaccard Skoru	63
2.9.1.9. Hata Matrisi (Confusion Matrix)	63
2.9.2. Regresyon	64
2.9.2.1. R^2 Skoru (R-Squared Score)	64
2.9.2.2. Ortama Mutlak Hata (Mean Absolute Error - MAE)	64
2.9.2.3. Root Mean Squared Error (RMSE)	64
2.9.2.4. Kayıp Fonksiyonu (Loss Function)	65

ÜÇÜNCÜ BÖLÜM

YÖNTEM VE UYGULAMA

3.1. VERİ SETİ ÖZELLİKLERİ AÇIKLAMALARI VE İSTATİSTİKSEL ANALİZLERİ	66
3.1.1. Çalışmanın Amacı	66
3.1.2. Veri Setinin Detayları	67
3.1.3. Veri Ön İşleme Süreci (Data Pre-processing)	72
3.2. EKSİK GÖZLEMLER VE AYKIRI DEĞERLERİN GİDERİLMESİ	83
3.2.1. Eksik Gözlemlerin Giderilmesi	83
3.2.2. Aykırı Değerleri Temizlenmesi	85
3.4. KORELASYON ANALİZİ	89
3.5. VERİ AYIRMA, ÖLÇEKLENDİRME VE KATEGORİK VERİLERİN İŞLENMESİ	92
3.5.1. One-Hot Encoding İşlemleri	92
3.5.2. Veri Ölçeklendirme ve Eğitim- Test Verisine Ayırma	96
3.5.3. Hiperparametre Optimizasyonu	98
3.5.4. Makine Öğrenmesi Modeli Oluşturulması ve Performans Ölçütü	103
3.5.4.1. Random Forest Algoritması	103
3.5.4.2. Desicion Tree Algoritması	110
3.5.4.3. XgBoost Algoritması	115
3.5.4.4. CatBoost Algoritması	122
3.5.4.5. Support Vector Machine	128

SONUÇ

132

KAYNAKÇA

138



KISALTMALAR

AI	Artificial Intelligence (Yapay Zeka)
CRM	Customer Relationship Management (Müşteri İlişkileri Yönetimi)
SVM	Support Vector Machine (Destek Vektör Makinesi)
IQR	Interquartile Range (Çeyrekler Açıklığı)
EDA	Exploratory Data Analysis (Keşifsel Veri Analizi)
ROC	Receiver Operating Characteristic (Alıcı İşletim Karakteristiği)
AUC	Area Under The Curve (Eğri Altında Kalan Alan)
MCC	Matthews Correlation Coefficient
TP	True Positive (Doğru Pozitif)
FP	False Positive (Yanlış Pozitif)
TN	True Negative (Doğru Negatif)
FN	False Negative (Yanlış Negatif)
SVC	Support Vector Classifier (Destek Vektör Sınıflandırıcı)
CatBoost	Kategorik Örnekleme (Categorical Boosting)
XgBoost	Extreme Gradient Boosting
RF	Random Forest (Rastgele Ormanlar)
GBM	Gradyan Arttırma Makinesi (Gradient Boosting Machine)
Crips-DM	Cross-Industry Standard Process for Data Mining

TABLÖLAR LİSTESİ

Tablo 1: Müşteri İlişkileri Yönetimi (Crm) Tanımları	s. 10
Tablo 2: Yapay Zekanın Tarihsel Gelişimi	s. 34
Tablo 3: Performans Metriklerinin Formülleri	s. 60
Tablo 4: Veri Setindeki Değişkenlerin Tanıtılması	s. 68



ŞEKİLLER LİSTESİ

Şekil 1: Pazarlama Karması 4p	s. 8
Şekil 2: Müşterinin Algıladığı Değer	s. 20
Şekil 3: Müşteri İlişkileri Yönetimi Bileşenleri	s. 23
Şekil 4: Müşteri İlişkileri Yönetimi Teknoloji Uygulama Süreci	s. 25
Şekil 5: Yapay Zekanın Alt Dalları	s. 38
Şekil 6: Makine Öğrenmesi Ve Yapay Zeka İlişkisi	s. 41
Şekil 7: Klasik Programlama Ve Makine Öğrenmesi	s. 43
Şekil 8: Crips-Dm Metodolojisi	s. 45
Şekil 9: Makine Öğrenmesi Süreçleri	s. 46
Şekil 10: Makine Öğrenmesi Türleri	s. 49
Şekil 11: Denetimli Makine Öğrenmesi Çalışma Prensibi	s. 50
Şekil 12: Denetimsiz Makine Öğrenmesi Çalışma Prensibi	s. 52
Şekil 13: Karmaşıklık Matrisi (Confusion Matrix)	s. 63
Şekil 14: Gerekli Kütüphanelerin İçer Aktarılması	s. 73
Şekil 15: Veri Setinin İçer Aktarılması	s. 74
Şekil 16: Genel Veri Setinin Görüntülenmesi	s. 74
Şekil 17: Değişkenlere Ait Bilgilerin Görüntülenmesi	s. 75
Şekil 18: Veri Setinin Betimsel İstatistikleri	s. 76
Şekil 19: Kampanyaların Sayısı	s. 77
Şekil 20: Eğitim Durumu Değişkeni Histogram Grafiği	s. 78
Şekil 21: Yaş Değişkeni Dağılım Grafiği	s. 78
Şekil 22: Medeni Durum Değişkeni Histogram Grafiği	s. 79
Şekil 23: Çocuk Sayısı Değişkenleri Histogram Grafiği	s. 80
Şekil 24: Müşteri Grupları Dağılımı	s. 80
Şekil 25: Toplam Kampanya Sayısı	s. 81
Şekil 26: Aylar Bazında Kampanya Sayıları	s. 81
Şekil 27: Kampanya Katılım Durumunun Aylara Göre Dağılımı	s. 82
Şekil 28: Eksik Gözlemlerin Giderilmesi	s. 84
Şekil 29: Gelir Sınıfları Histogram Grafiği	s. 84
Şekil 30: Gelir Dağılımı	s. 85

Şekil 31: Box-Plot Grafiği İle Aykırı Değerlerin Gösterimi	s. 86
Şekil 32: Gelir Değişkeninin Dağılımı	s. 86
Şekil 33: Aykırı Değerlerin Temizlenmesi	s. 87
Şekil 34: Box-Plot Grafiği İle Aykırı Değerler Temizlendikten Sonra Gösterim	s. 88
Şekil 35: Aykırı Değerler Temizlendikten Sonra Gelir Değişkeni Dağılımı	s. 88
Şekil 36: Korelasyon Analizi	s. 90
Şekil 37: Veri Setindeki Değişken Tiplerinin Gösterimi	s. 93
Şekil 38: Kategorik Değişkenlerin Gösterimi	s. 94
Şekil 39: One-Hot Encoding İşlemi Uygulaması	s. 94
Şekil 40: One-Hot Encoding Uygulaması Sonrası Değişkenlerin Gösterimi	s. 95
Şekil 41: Hedef Değişkeni Liste Formatına Çevirmek	s. 95
Şekil 42: Yeni Oluşturulan Değişkenin Encode Edilmiş Data İle Birleştirilmesi	s. 96
Şekil 43: Dönüşüm Sonrası Hedef Değişken Gösterimi	s. 96
Şekil 44: Verilerin Ölçeklendirme Ve Eğitim - Test Veri Setlerinin Ayrılması	s. 97
Şekil 45: Random Forest Algoritması İçin Hiperparametre Optimizasyonu	s. 99
Şekil 46: Desicion Tree Algoritması İçin Hiperparametre Optimizasyonu	s. 100
Şekil 47: Support Vector Classifier Algoritması İçin Hiperparametre Optimizasyonu	s. 101
Şekil 48: Catboost Algoritması İçin Hiperparametre Optimizasyonu	s. 102
Şekil 49: Random Forest Algoritması İçin Performans Metrikleri Çıktısı	s. 105
Şekil 50: Random Forest Algoritması Classification Report Çıktısı	s. 106
Şekil 51: Kesinlik-Duyarlılık Grafiği	s. 107
Şekil 52: Karmaşıklık Matrisi (Confusion Matrix)	s. 108
Şekil 53: Özellik Önem Düzeyi Analizi	s. 109
Şekil 54: Desicion Tree Algoritması İçin Performans Metrikleri Çıktısı	s. 111
Şekil 55: Precision – Recall- F1 Score Çıktısı	s. 112
Şekil 56: Kesinlik - Duyarlılık Grafiği	s. 113
Şekil 57: Karmaşıklık Matrisi (Confusion Matrix)	s. 113
Şekil 58: Decision Tree Algoritması Öğrenme Eğrisi	s. 114
Şekil 59: Smote Yöntemi Uygulama Çıktısı	s. 117
Şekil 60: Xgboost Algoritması İçin Performans Metrikleri Çıktısı	s. 118
Şekil 61: Xgboost İçin Karmaşıklık Matrisi (Confusion Matrix)	s. 119

Şekil 62: Kesinlik – Duyarlılık Grafiği	s. 120
Şekil 63: Öznitelik Önem Düzeyi Analizi	s. 121
Şekil 64: Overfit Olmuş Performans Metrikleri	s. 123
Şekil 65: Catboost Algoritması İçin Performans Metrikleri Çıktısı	s. 124
Şekil 66: Catboost Algoritması İçin Sınıflandırma Çıktısı	s. 125
Şekil 67: Karmaşıklık Matrisi(Confusion Matrix)	s. 126
Şekil 68: Özellik Önem Düzeyi Analizi	s. 127
Şekil 69: Destek Vektör Makinesi Algoritması İçin Sınıflandırma Çıktısı (K=5)	s. 129
Şekil 70: Destek Vektör Makinesi Algoritması İçin Sınıflandırma Çıktısı (K=10)	s.130
Şekil 71: Destek Vektör Makinesi Algoritması İçin Performans Metrikleri	s. 131

GİRİŞ

Günümüz iş dünyasında, rekabet avantajı elde edebilmek için işletmelerin müşterilere kişiselleştirilmiş teklifler sunmaları büyük önem taşımaktadır. Bu kişiselleştirme, müşterilerin ihtiyaçlarını ve tercihlerine göre özel olarak hazırlanmış kampanyalar ve fiyatlandırma stratejileri anlamına gelir. Pazarlama kampanyalarının başarısı, doğru müşteriye doğru zamanda doğru fiyatı sunabilme yeteneği ile doğrudan ilişkilidir. Bu bağlamda, makine öğrenmesi algoritmalarının kullanımı, pazarlama stratejilerinin etkinliğini artırmak ve müşteri memnuniyetini yükseltmek için önemli bir araç haline gelmiştir. Makine öğrenmesi, büyük veri setlerinden anlamlı bilgiler çıkararak tahmin ve karar verme süreçlerini iyileştiren bir teknolojidir. İşletmeler, müşteri verilerini analiz ederek her bir müşteri için en uygun fiyatı belirlemeyi amaçlar. Bu süreçte, çeşitli makine öğrenmesi algoritmalarının kullanılması, pazarlama kampanyalarının daha etkili ve verimli olmasını sağlar (Johnson vd., 2010: 98). Özellikle, müşteri segmentasyonundan fiyat optimizasyonuna kadar birçok alanda makine öğrenmesi, işletmelere önemli avantajlar sunar. Bu araştırmanın temel amacı, çeşitli makine öğrenmesi algoritmalarını kullanarak pazarlama kampanyalarında kişiselleştirilmiş fiyat belirleme sürecini optimize etmektir. Bu tezde, dört farklı makine öğrenmesi algoritmasının (Random Forest, Decision Tree, XgBoost, CatBoost, Support Vector Classifier) kampanya seçiminde kişiselleştirilmiş fiyat belirleme konusundaki performansları incelenmiştir. Özellikle, müşteri davranışları ve geçmiş satın alma verileri gibi faktörleri dikkate alarak, her bir müşteri için en uygun kampanyayı belirleyebilecek en başarılı algoritmayı seçmek hedeflenmiştir. Müşteri ilişkileri yönetimi (CRM) sistemlerinde toplanan veriler, müşteri segmentasyonu ve kişiselleştirilmiş pazarlama stratejileri oluşturmak için değerli bir kaynak sağlamaktadır. Bu veriler, müşterilerin geçmişteki satın alma davranışları, harcama alışkanlıkları, demografik bilgileri ve diğer ilgili özellikleri içermektedir. Algoritmalar, CRM veri tabanından alınan geniş veri setleri üzerinde eğitilmiş ve test edilmiştir. (Rupnik vd., 2019: 260)

Her algoritma, müşterilerin geçmiş satın alma verileri ve davranışlarını analiz ederek, gelecekte satın alma olasılıklarını tahmin etmeye çalışmıştır. Performans değerlendirmesi, algoritmaların tahmin doğruluğu, hesaplama hızı ve verimliliği gibi

çeşitli kriterlere dayanarak yapılmıştır. Dolayısıyla şirketler müşterilerine kişiselleştirilmiş kampanyalar sunabilecek ve satış hacmini artırıp karını maksimize edecektir.

Literatür Taraması;

Güven ve Ayvaz Güven (2023) çalışmasında elektronik ve bilgi teknolojilerindeki hızlı gelişmelerin e-ticaret sektörüne etkisini ele alarak, yapay zekâ uygulamalarının işletmeler için nasıl stratejik bir avantaj sağladığını tartışmaktadır. Yapay zekâ, müşteri davranışlarını analiz ederek kişiselleştirilmiş pazarlama stratejileri geliştirme, dinamik fiyatlandırma yapma ve chatbotlar aracılığıyla müşteri hizmetlerini iyileştirme gibi birçok alanda kullanılmaktadır. Büyük veri ve makine öğrenimi sayesinde müşteri tercihleri daha iyi analiz edilerek işletmelere rekabet avantajı sağlanmakta, bu da küresel pazarda başarı için kritik bir faktör olarak değerlendirilmektedir. Yapay zekâ entegrasyonunun e-ticaret işletmelerinde verimliliği artırdığı ve müşteri deneyimini iyileştirdiği vurgulanarak, dijital dönüşüm süreçlerinde nasıl daha etkin hale gelinebileceğine dair öneriler sunulmaktadır. Çalışma, bu teknolojilerin entegrasyonu ile işletmelerin rekabet gücünü artırmalarına katkıda bulunmayı hedeflemektedir. Çalışmada yapay zekâ entegrasyonunun e-ticaretin verimliliğini ve etkinliğini artırdığı sonucuna varılmıştır.

Sumathi ve Sivanandam (2006) çalışmasında veri bilimi süreçlerinin iş dünyasındaki uygulamalarına değinerek çok önemli bir yere sahip olduğunu vurgulamaktadır. Finansal kurumlar ve diğer işletmeler, bu yöntemlerini kullanarak, verilerdeki desenleri ve ilişkileri keşfetmek suretiyle daha bilinçli ve etkili iş kararları alabilmektedir. Bu yöntemler, satış eğilimlerini analiz etmek, müşteri davranışlarını anlamak ve hedefe yönelik pazarlama kampanyaları geliştirmek için kritik bir rol oynamaktadır. Ayrıca, işletmelerin müşteri sadakatini artırmasına ve uzun vadeli müşteri ilişkilerini güçlendirmesine katkı sağladığını belirtmiştir. Bu bağlamda, veri madenciliği, yalnızca büyük veri kümelerinden anlamlı bilgiler çıkarma süreci olmakla kalmayıp, aynı zamanda stratejik karar alma süreçlerinde işletmeler için güçlü bir rehber niteliği taşımaktadır. Sumathi ve Sivanandam'ın çalışması, bu tekniklerin iş dünyasındaki uygulamalarına dair önemli bir referans sunmaktadır.

Sezer vd. (2022) çalışmasında kampanya başarısının tahmin edilmesinde önemli rol oynayan özniteliklerin seçimi ele alınmıştır. Çalışmanın temel amacı, pazarlama kampanyalarının etkinliğini artırmak ve başarı oranını yükseltmek için hangi özelliklerin (özniteliklerin) en etkili olduğunu belirlemektir. İşletmelerin yoğun rekabet ortamında müşteri odaklı stratejiler geliştirebilmesi için veri üretimi ve analitiği büyük önem taşımaktadır. Bu bağlamda, çalışmada CRM yazılımında kampanya başarısını tahmin etme sürecine odaklanılmıştır. Kampanya başarısını tahmin etmek için veri analizinin önemini vurgular ve pazarlama kampanyalarının etkinliğini artırmak için gerekli özniteliklerin belirlenmesine yardımcı olur.

Bach (2013) çalışmasında veri madenciliğinin bankacılık hizmetlerinin pazarlanmasında ve pazarlama stratejilerinin geliştirilmesinde doğru müşteri gruplarını belirlemenin önemini vurgulamıştır. Kendini Düzenleyen Haritalar (Self-Organizing Maps, SOM) gibi yöntemler, müşteri segmentasyonunda etkili bir şekilde kullanılarak geçmiş kampanyalara katılım gösteren müşterileri analiz etmiştir. Bu tür tekniklerle, hangi kampanyaların hangi müşteri grupları için daha uygun olabileceği öngörülebilir hale gelir. Ayrıca, çapraz satış stratejilerinde yüksek kârlılık potansiyeline sahip müşteri segmentleri belirlenebilir ve bu müşterilere yönelik özelleştirilmiş ürün ve hizmetler geliştirilebilir. SOM gibi yöntemlerin kullanımı, yalnızca müşteri segmentasyonunu kolaylaştırmakla kalmaz, aynı zamanda müşterilerin davranışlarında meydana gelen sapmaların ve eğilimlerin tespit edilmesini sağlayarak kurumsal yıpranma analizlerine de katkı sağlar. Böylelikle, müşteri kaybını önlemek ve daha kişiselleştirilmiş pazarlama stratejileri geliştirmenin daha mümkün olduğunu göstermiştir.

Acı ve Doğansoy (2022) çalışmasında, e-ticaret sektöründe talep tahmini için makine öğrenimi ve derin öğrenme yöntemlerinin performansları incelenmiştir. Araştırma, bu yöntemlerin e-ticarete satış tahmini süreçlerinde nasıl daha doğru ve etkili kullanılabileceğine odaklanmaktadır. Farklı makine öğrenimi algoritmaları ve derin öğrenme modelleri geçmiş satış verileri üzerinde test edilmiştir. Çalışmanın temel amacı, bu yöntemlerin talep tahmininde sağladığı doğruluk ve verimliliği değerlendirmek ve e-perakendeciler için daha etkili bir tahmin modeli sunmaktır. Elde edilen sonuçlar, derin öğrenme yöntemlerinin özellikle karmaşık zaman serisi verilerinde üstün bir doğruluk sağladığını göstermektedir. Makine öğrenimi

yöntemleri ise daha hızlı sonuçlar üretmekle birlikte, veri karmaşıklığının düşük olduğu senaryolarda başarılı bulunmuştur.

Han (2018) ve diğerleri çalışmasında veri madenciliğinin ilk olarak market alışverişlerinin analizinde kullanıldığını ve bu yöntemlerle müşterilerin alışveriş alışkanlıkları derinlemesine incelendiğini ele almıştır. Özellikle birlikte satın alınan ürünlerin belirlenmesi, market raflarının düzenlenmesinden promosyon stratejilerine kadar birçok alanda işletmelere değer katmıştır. Örneğin, birbirini tamamlayan ürünler, raflarda yan yana konumlandırılarak müşterilerin dikkatini çekmekte ve bu sayede satışlar artırılmaktadır. Bu analizler aynı zamanda doğru promosyon stratejileri oluşturulmasını sağlayarak, kampanyaların daha etkili ve hedefe yönelik olmasına katkıda bulunmuştur. Bu bağlamda, WalMart gibi öncü şirketler veri madenciliği tekniklerini pazarlama kampanyalarını optimize etmek için kapsamlı şekilde kullanmıştır. WalMart, günlük olarak milyonlarca müşteriden elde ettiği verileri analiz ederek hem müşterilerinin alışveriş alışkanlıklarını anlamış hem de tedarikçileriyle daha verimli iş birlikleri geliştirmiştir. Örneğin, "bebek bezi ve bira" gibi ilk bakışta mantıksız görünen ürün eşleşmeleri, veri madenciliği ile tespit edilmiş ve bu bilgi, kampanya stratejilerinde başarıyla kullanılmıştır. Bu tür içgörüler, müşterilere özel kampanyalar düzenlenmesine ve ürün yerleşimlerinin stratejik bir şekilde yapılmasına olanak tanımıştır. Pazarlama kampanyalarında bu bilgiler, hedef kitleye en uygun mesajların iletilmesini ve müşteri sadakatinin artırılmasını mümkün kılmaktadır. Günümüzde bu teknikler, sağlıktan perakendeye, bankacılıktan e-ticarete kadar geniş bir yelpazede kullanılmakta ve işletmelerin başarıya ulaşmasında kritik bir rol oynamaktadır.

Kestane (2024) çalışmasında CRM (Müşteri İlişkileri Yönetimi) kavramının evrimini ve bu alanda yapılan çalışmaları bibliyometrik analizle incelemektedir. 1996-2023 yılları arasında Web of Science veri tabanında yer alan 2869 makale incelenmiş ve CRM'in özellikle İşletme, Yönetim ve Bilgisayar Bilimleri alanlarında yoğunlaştığı görülmüştür. Son yıllarda CRM ile ilgili çalışmalarda yapay zeka, makine öğrenimi, müşteri katılımı gibi yeni teknolojilerin ve kavramların yer aldığı belirtilmiştir. Çalışma, araştırmacılara bu alandaki eğilimleri ve boşlukları göstermek için kapsamlı bir literatür haritası sunmaktadır.

Tang (2024), film giş e has ılatını tahmin etmek için optimize edilmiş XGBoost algoritmasını ele almıştır. Çalışmada XGBoost, Derin Sinir Ağı, Rastgele Orman, LightGBM ve CatBoost gibi modellerle karşılaştırılmıştır. Performans ölçütleri doğruluk, kesinlik, F1 skoru, stabilite ve genelleme hatasıdır. Optimize edilmiş XGBoost, özellikle büyük veri setlerinde üstün performans göstererek doğruluk ve F1 skorunda 0.9'a, stabilite skorunda ise 0.98'e ulaşmıştır. Bu model, film pazarlama ve dağıtım stratejilerinde ekonomik risklerin azaltılmasına ve daha isabetli kararlar alınmasına katkı sağlamaktadır.

Bu bilgiler ışığında çalışmanın geri kalan kısmında pazarlama ve müşteri ilişkileri kısaca açıklandıktan sonra makine öğrenmesi algoritmaları kullanılarak kampanya seçimleri için modeller tahmin edilecek ve bu tahmin edilen modellerin sonuçları değerlendirilecektir.

BİRİNCİ BÖLÜM

PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ

1.1. PAZARLAMA

Pazarlama, bir işletmenin ürün veya hizmetlerini potansiyel müşterilere tanıtmak için kullandığı stratejik bir yaklaşımdır. Bu süreç, tüketicilerin ihtiyaçlarını anlamakla başlar ve ürün veya hizmetlerin bu ihtiyaçları nasıl karşılayabileceğini belirlemekle devam eder. Sonrasında ise, bu mesajın etkili bir şekilde iletilmesi önemlidir. Her ne kadar pazarlama, yakın dönemde bir bilim dalı olarak kabul edilmiş olsa da (Erdoğan, 2014: 20), insanlık tarihinin en eski mesleklerinden ve faaliyetlerinden biri olarak karşımıza çıkar. İnsanların temel ihtiyaçlarını karşılama dürtüsüyle doğal bir şekilde ortaya çıkmış ve zamanla gelişmiştir. İnsan, yaşamı boyunca birçok mal ve hizmete ihtiyaç duyar; ancak bunların yalnızca çok azını kendisi üretebilir. Bu nedenle, her insan aslında özünde bir tüketicidir. Tüketici bireyler, ihtiyaç duydukları veya arzu ettikleri mal ve hizmetlere eski çağlarda pazar yerleri aracılığıyla ulaşmışlardır. Zannedilenin aksine, bu dönemde de oldukça hareketli bir ticaret hayatı yaşanmıştır. (Alankuş, 2021: 98).

İnsanlığın gelişim süreci, ihtiyaçları karşılama ve istekleri elde etme doğrultusunda kesintisiz bir şekilde devam etmektedir. Tarih boyunca bireylerin, toplulukların, yerel ve küresel işletmelerin ve siyasal otoritelerin ihtiyaçları ve talepleri doğrultusunda yaşanan değişimler, pazarlama faaliyetlerinin sürekli olarak yeniliklere ve gelişime açık olmasını sağlamıştır. Pazarlamanın tarihsel gelişimini etkileyen temel unsurlar arasında sanayi devrimleri, teknolojik ilerlemeler, iletişim olanaklarının artması, internetin günlük yaşamın bir parçası hâline gelmesi, eğitimin yaygınlaşması, küreselleşme ve kişi başına düşen gelirdeki artış yer almaktadır (Ünal ve Bozkurt, 2015: 9).

Pazarlama faaliyetleri, ürün tanıtımı, fiyatlandırma stratejilerinin belirlenmesi, dağıtım kanallarının seçimi ve tüketicilerle etkileşim kurma gibi aşamaları kapsar. Etkili bir pazarlama stratejisi, işletmelerin müşteri sadakatini artırmasına, satışlarını yükseltmesine ve rekabet üstünlüğü elde etmesine olanak tanır. Bu yüzden, pazarlama disiplini, işletmelerin büyümesi ve sürdürülebilir başarıya ulaşması açısından hayati bir öneme sahiptir (Cemalcılar, 2008: 25).

Pazarlama stratejileri, bir işletmenin ürün veya hizmetlerini hedef kitleye en etkili şekilde tanıtmak ve satmak için belirlediği plan ve yöntemler bütünüdür. Temel olarak, pazarlama stratejileri aşağıdaki unsurları içerir (Uyar,2024: 25).

1.1.1. Pazar Araştırması

Pazar araştırması, ürün tanıtımı, fiyatlandırma stratejilerinin belirlenmesi, dağıtım kanallarının seçimi ve tüketicilerle etkileşim kurma gibi aşamaları kapsar. Etkili bir pazarlama stratejisi, işletmelerin müşteri sadakatini artırmasına, satışlarını yükseltmesine ve rekabet üstünlüğü elde etmesine olanak tanır. Bu yüzden, pazarlama disiplini, işletmelerin büyümesi ve sürdürülebilir başarıya ulaşması açısından hayati bir öneme sahiptir (Cemalcılar, 2008: 5).

1.1.2. Hedef Kitle Belirleme

Bir ürün veya hizmetin en uygun ve potansiyel alıcılarını tanımlama sürecidir. Bu süreçte, tüketicilerin demografik özellikleri, coğrafi konumları, ilgi alanları, davranışları ve satın alma alışkanlıkları analiz edilerek belirli bir grup müşteriye odaklanılır.

Hedef kitle belirleme, şirketlerin pazarlama kaynaklarını daha verimli kullanmalarını ve doğru mesajlarla doğru insanlara ulaşmalarını sağlar. Ayrıca, müşteri ihtiyaçlarını anlamaya yardımcı olur ve pazarlama stratejilerinin daha etkili bir şekilde tasarlanmasına katkıda bulunur. (Kurnuç, 2020: 26)

Özetle, hedef kitle belirleme, bir ürün veya hizmet için en uygun müşteri grubunu bulmak ve bu gruba özel pazarlama çalışmaları geliştirmektir.

1.1.3. Konumlandırma

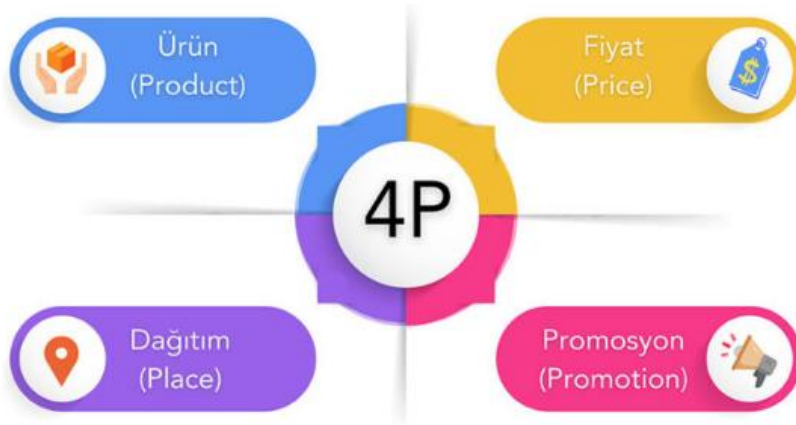
Ürün veya hizmetin piyasadaki yerini belirlemek ve rakiplerden farklılaştırmak için yapılan çalışmalardır. Bir ürünün veya markanın, hedef kitlenin zihninde benzersiz ve istenilen bir şekilde algılanmasını sağlama sürecidir. Pazarlama stratejisinin önemli bir parçası olan konumlandırma, müşterilerin ürünü rakiplerinden farklı ve değerli görmesini hedefler. Bu süreçte, markanın veya ürünün özellikleri, faydaları ve müşteri ihtiyaçlarına yönelik çözümleri vurgulanır. Konumlandırma,

doğru bir şekilde yapıldığında, markanın hedef kitlenin zihninde güçlü bir yer edinmesine ve rekabet avantajı kazanmasına yardımcı olur. (Tekel, 2006: 108-109)

1.1.4. Pazarlama Karması (4P)

Pazarlama, işletmenin hedeflerine ulaşmasını sağlamak amacıyla, ihtiyaçları karşılayacak ürünler, hizmetler ve fikirlerin geliştirilmesi, fiyatlandırılması, tanıtılması ve dağıtılmasına yönelik planlama ve uygulama sürecidir (Mucuk, 1999: 5). Pazarlama karması, İngilizce terimlerin (product, place, price, promotion) ilk harflerinden oluşan '4P' kısaltmasıyla sembolize edilmiştir (Cemalcılar, 1987: 23). Pazarlama faaliyetlerini yürütürken bir pazarlama yöneticisinin karar vermek zorunda olduğu temel unsurlar, pazarlama bileşenleri veya pazarlama karması olarak adlandırılır. Bu karma; ürün, fiyat, tutundurma ve dağıtım olmak üzere dört ana unsura sahiptir. Ancak, hizmetlerin pazarlanmasında bu dört bileşen yeterli olmayıp, buna ek olarak üç bileşen daha (süreçler, insanlar, fiziksel kanıt) eklenerek, genişletilmiş pazarlama karması oluşturulmuştur. (Tekel,2006:36-38)

Şekil 1: Pazarlama Karması 4P



Kaynak: Özay, M. (2024) tarafından alınmıştır.

1.1.4.1. Ürün (Product)

Pazarlamanın temel unsurlarından biri olan ürün, kalite, tasarım ve marka gibi önemli bileşenleri içeren ve işletmelerin hedef pazara sunduğu bir değer teklifidir. İşletmeler, ürünleriyle birlikte kiralama, dağıtım, onarım ve eğitim gibi çeşitli ek hizmetleri de sunarak müşterilerine kapsamlı bir deneyim sağlamalıdır.

Hedef pazara sunulacak ürün veya hizmetin tüm yönleri detaylı bir şekilde analiz edilmelidir. Ürün ve hizmetin temin edilmesinden tüketiciye ulaştırılmasına ve satış sonrası desteğe kadar her adımın dikkatlice planlanması, başarılı bir pazarlama stratejisi için kritik öneme sahiptir. (Kotler, 1997: 93; Yılmaz, 2018).

1.1.4.2. Fiyat (Price)

Fiyat, pazarlamanın en önemli unsurlarından biridir. Ancak, ürünün fiyatını belirlemek için yalnızca üretim maliyetleri yeterli değildir. Müşterilerin ürünü satın almayı kabul ettikleri fiyat aralığı ve pazardaki rekabetçi fiyatlandırma gibi unsurlar da dikkate alınmalıdır. Etkili bir fiyat stratejisi, satıcı ile tüketicinin beklentilerini ortak bir zeminde buluşturmayı amaçlar. Ayrıca, fiyat, müşterilerin bir ürün ya da hizmet karşılığında ödemeyi göze aldıkları tutardır. Belirlenen fiyat, ürünün sunduğu değerle uyumlu olmalıdır. Aksi halde, alıcılar rakip firmaların ürünlerini tercih edebilir (Kotler,1997; Kotler, 2015).

1.1.4.3. Dağıtım (Place)

Ürün ve hizmetin doğru zamanda ve doğru yerde müşteriye ulaşmasını sağlamak temel hedeftir. Bu, ürünün satışının gerçekleştirildiği noktayı ifade eder. Dağıtım noktası, müşteriler için büyük önem taşır; çünkü mesafe arttıkça, müşteriler rakip işletmelerin ürünlerini tercih edebilir. Dağıtım, ürünün müşteriler tarafından erişilebilir ve kullanılabilir hale gelmesi için gerekli olan tüm faaliyetleri kapsar (Kotler,1997; Yakar ve Baycık, 2001).

1.1.4.4. Tutundurma (Promotion)

Pazarlama karmasının son unsuru olan tutundurma, işletmelerin ürettikleri ürün veya hizmetleri hedef pazara tanıtmak ve bu pazarla etkili bir iletişim kurmak amacıyla gerçekleştirdiği çeşitli faaliyetleri kapsar. İşletmeler, kurumsal reklamcılık, halkla ilişkiler, doğrudan pazarlama ve çevrimiçi pazarlama gibi unsurları içeren kapsamlı bir iletişim ve tanıtım stratejisi oluşturmalıdır. Ürün ve hizmetlerin piyasada bilinirlik kazanması için tanıtım ve reklam kampanyaları düzenlenmeli ve bu faaliyetler sürdürülebilir olmalıdır. Aynı zamanda, tutundurma süreçleriyle paralel olarak markalaşma çalışmalarının da yürütülmesi önemlidir (Kotler, 1997: 93;

Yasak, 2017).

Pazarlama karmasının her bir bileşeninin kendine özgü bir önemi ve değeri bulunmakla birlikte bu unsurların tümünün 'insan' odaklı olması ve hedef kitlenin sosyolojik açıdan ele alınması başarının temel koşuludur.

1.1.5. Müşteri İlişkileri Yönetimi (CRM)

Pazarlamada dijital dönüşüm, dijital teknolojilerin pazarlama operasyonlarının tüm yönlerine entegre edilmesini ifade eder ve bu süreç, kuruluşların tüketicilerle etkileşim kurma ve iş büyümesini yönlendirme biçimlerini kökten değiştirmektedir. Bu dönüşüm, pazarlama stratejilerini geliştirmek, müşteri etkileşimlerini optimize etmek ve genel verimliliği artırmak için veri analitiği ve Müşteri İlişkileri Yönetimi (CRM) araçlarının kullanımını kapsamaktadır (Olaleye ve diğerleri, 2024).

Müşteri, bir ürün ya da hizmeti satın alan kişidir. Bugün itibarıyla müşteri ilişkileri yönetimi, işletmeleri birbirlerinden ayıran en önemli unsurlardan biri haline gelmiştir. Müşteriler potansiyel müşteri, mevcut müşteri, yeni müşteri ve hedef müşteri olarak genellikle dört kategoriye ayrılır (Yıldırım, 2012).

Tablo 1: Müşteri İlişkileri Yönetimi (CRM) Tanımları

Tanım	Yazar/Yazarlar
CRM, müşteri kazanımını, müşteriye elde tutmayı, müşteri sadakatini ve müşteri karlılığını artırmak amacıyla anlamlı iletişimler yoluyla müşteri davranışlarını anlamaya ve etkilemeye yönelik kurumsal bir yaklaşımdır.”	(Swift, 2001, s. 14)
“CRM, şirket ve müşteri için üstün değer yaratmak amacıyla seçilmiş müşterileri edinme, elde tutma ve onlarla ortaklık kurmaya yönelik kapsamlı bir strateji ve süreçtir.”	(Sheth vd., 2001, s. 12)
“CRM, her müşterinin firma için yaşam boyu değerini en üst düzeye çıkaracak kurumsal uygulamaları ve yöntemleri belirlemek için pazarlama veritabanlarını analiz etme ve kullanma ve iletişim teknolojilerinden yararlanma uygulamasıdır.”	(Kumar & Reinartz, 2012, s. 5)
“CRM, Müşterilerin ihtiyaçları ve davranışları hakkında daha fazla bilgi edinerek onlarla daha güçlü ilişkiler geliştirmek için kullanılan bir süreç veya metodolojidir.”	(Shanthi, 2013, s. 17)
“CRM, tüm müşterilerle ilişkileri en üst düzeye çıkarmak için insanları, süreçleri ve teknolojiyi entegre eden ve müşteriye dönük tüm işlevler arasında kusursuz iş birliği sağlayan bir iş yaklaşımıdır.”	(Goldenberg, 2015, s. 25)
CRM, bir firmanın en karlı şekilde hizmet verebileceği müşterileri seçme ve şirket ile bu müşteriler arasındaki etkileşimleri şekillendirmeye yönelik stratejik bir süreçtir. Nihai amaç, müşterilerin şirket için mevcut ve gelecekteki değerini optimize etmektir.	(Kumar & Reinartz, 2018, s. 5)
“CRM, süreç verimliliği, operasyonel maliyetlerin azaltılması ve müşteri etkileşimlerinin ve deneyiminin iyileştirilmesiyle ilgilidir.”	(Fatouretchi, 2019, s. 32)

Kaynak: Kestane (2024) çalışmasından alınmıştır.

Müşteri İlişkileri Yönetimi (CRM), ilişki pazarlaması teorisine dayansa da literatürde tam olarak ne olduğu, nasıl uygulanması gerektiği ve müşteri etkileşimini nasıl geliştirdiği konusunda farklı yaklaşımlar bulunmaktadır. CRM, bilgi teknolojileri (BT), stratejik yönetim ve farklı uygulayıcılar açısından farklı şekillerde ele alınmaktadır.

BT perspektifinden bakıldığında, CRM daha çok yazılım tabanlı bir sistem olarak değerlendirilir ve müşteri verilerinin toplanması, işlenmesi ve analiz edilmesine yönelik teknolojik çözümler sunar. Bu yazılımlar, işletmelerin müşterileriyle olan etkileşimini geliştirmek için çeşitli araçlar sağlayarak küresel iş ortamında kolayca benimsenebilir hale gelmiştir (Rîpa, 2022: 1067).

Stratejik ve yönetsel açıdan CRM, müşteri kazanımı, sadakat oluşturma, gelir artırma ve müşteri ilişkilerini geliştirme süreçlerini kapsayan bir yönetim yaklaşımıdır. Ancak, CRM kavramı pek çok danışmanlık firması, işletme ve analist tarafından farklı şekillerde yorumlanmakta ve çeşitli tanımlamalar yapılmaktadır. Bu nedenle, CRM'in tek bir ortak tanımını yapmak zor olmakla birlikte, müşteri odaklı stratejik bir yaklaşım olduğu konusunda genel bir görüş birliği bulunmaktadır. (Kestane, 2024: 315)

1.1.6. Dijital Pazarlama

Postmodern dönem pazarlamasının evriminde düşünce ve zeka faktörü, giderek daha merkezi bir rol oynamakta ve hızla yapay zekaya dönüşmektedir. Modern dönemde ekonomiyi doğrudan etkileyen, pazarlamayı ise dolaylı şekilde etkileyen teknolojik gelişmeler, dijital çağ ile birlikte pazarlamayı doğrudan etkileyen bir yapıya bürünmüştür. (Kozlu ve Karaca, 2019: 15; Şahin ve Kaya, 2019: 118; Şengül, 2019: 51). Dijital pazarlama, ürün ve hizmetlerin pazarlama faaliyetlerini teknolojik alanları kullanarak gerçekleştiren bir yöntemdir. İnternet, mobil cihazlar ve diğer interaktif platformlar aracılığıyla yapılan bu pazarlama, geleneksel yöntemlerden farklı olarak daha geniş kitlelere ulaşma ve hedefleme imkânı sunar. Bu sayede pazarlamacılar, dijital ortamda hızla etkileşimde bulunabilir, verileri anında analiz edebilir ve stratejilerini hızlıca güncelleyebilirler (Gökşin, 2017; Bulunmaz, 2016: 357). Geleneksel pazarlama, genellikle tek yönlü bir iletişim sunarken ve hedef kitlenin eş zamanlı olarak etkileşimde bulunmasına

izin vermezken, dijital pazarlama her aşamada hedef kitlenin görüş, düşünce ve geri bildirimlerine açık bir iletişim süreci sunar. Bu yaklaşım, pazarlama işlevinin alıcı tarafında yer alan müşterilere deneyim kazandırır. Dijital pazarlama, çeşitli kanallar aracılığıyla hedef kitleye ulaşarak, özellikle zaman, çaba ve maliyet açısından önemli avantajlar sağlar. (Bulunmaz, 2016: 357).

1.2. PAZARLAMA İLETİŞİMİ

Pazarlama iletişimi, ürün tasarımının yanı sıra fiyat, dağıtım, reklam ve mağaza içi faaliyetleri de kapsayan geniş bir süreçtir. Bu süreç, müşterinin ürünü satın almasıyla başlayıp, satın alma sonrası hizmetlerle devam eder. Pazarlama iletişimi, pazarlama disiplininin en hızlı gelişen ve en fazla tartışılan alanlarından biridir. Bu kavram, tutundurma ve satış çabalarından daha geniş bir anlam taşır; ürün kavramı ve şirket kişiliğinin tüketicilere sunuluşunu da içerir. Pazarlama iletişimi ile tüketicilerin ürünü fark etmeleri sağlanır ve satın alma kararlarını etkileyebilmek amacıyla pazarlamacıların gerçekleştirdiği tüm eylemler anlatılmaktadır. (Erciş, 2010: 111; Odabaşı ve Oyman, 2005: 36). Pazarlama iletişimi, iki yönlü bir iletişim sürecidir. Bu süreç, bilgi ve deneyim alışverişini sağlar. Özellikle yeni teknolojik gelişmeler, bu tür iletişim için yeni fırsatlar yaratmaktadır. Veri tabanlarının oluşturulması ve karar alma süreçlerinde kullanılması, pazarlama iletişimine büyük avantajlar sunmaktadır. Araştırmalarla iletişimin etkileri ve tüketici eğilimleri belirlenmeye çalışılmaktadır. Böylece tüketici, pazarlama iletişiminin kaynağı haline gelir. Elde edilen veriler, yeni pazarlama iletişim fırsatlarının ortaya çıkmasını sağlar. Pazarlama iletişimi, bir yandan işletmenin mesajlarını, hedef kitlelerine iletirken, diğer yandan bu kitlelerden aldığı geri bildirimler doğrultusunda stratejilerini yeniden şekillendirir. Pazarlama iletişimi faaliyetleriyle, bir markanın hedef kitlelere sunulması, tanıtımı, istek uyandırılması, alınması, satın alımın tekrarı, müşteri memnuniyeti ve sadakat oluşturulması hedeflenir. Bu bağlamda, pazarlama iletişimi, bir marka ile hedef kitle arasındaki bir iletişim biçimi olarak da tanımlanabilir (Erciş, 2010: 112; Yurdakul, 2006: 23). Bu sayede, hedef kitlenin talepleri, istekleri ve beklentilerindeki her türlü farklılık ve değişime yönelik bir hareket planı geliştirilmiş olur. Özellikle pazarlama alanında, geleneksel yöntemlerden faydalananarak pazarlama iletişimi unsurlarının kullanılmasının yanı

sıra, hedef kitleye yönelik teknolojik deęişimleri ve bütüncül yaklaşımları dikkate alarak yapılacak birleşik pazarlama iletişimi uygulamaları, günümüz bilgi, iletişim ve teknoloji çağında kurumların 'hayatta kalabilmesi' ve rekabet edebilmesi için temel stratejilerinden biri olmalıdır (Bulunmaz, 2016: 356).

1.3. E-TİCARET

Geçmişte internet pazarlaması olarak bilinen ve zamanla ticari işlemler ve satış süreçlerini de kapsayan e-ticaret gibi kavramlarla genişleyen dijital pazarlama, kurumların kendi markaları ve faaliyetleriyle ilgili olarak web sayfaları üzerinden içerik paylaşımı veya e-mail yoluyla tanıtım yapmalarını içeren bir süreç olarak tanımlanabilir. İnternet, dijital pazarlamanın temel kanalı olmanın yanı sıra, yeni medya ve iletişim teknolojilerinin sunduğu fırsatlar sayesinde diğer dijital pazarlama kanallarına da yeni uygulama alanları yaratmaktadır. Dijital çağda e-ticaretin hayatımıza girmesiyle, pazarlama biliminin metodolojisi deęişmiş, önemi artmış ve kapsamı genişlemiştir. E-ticaret, mal ve hizmet alım satımının elektronik ortamda gerçekleşmesini ifade eder. (Bulunmaz, 2016: 358; Çiçek, 2015: 285). Sadece Çin'de, Taobao'nun algoritmalarını yanıltmak amacıyla faaliyet gösteren ve Çin'in en büyük e-ticaret platformlarından birine ait olan, muhtemelen binden fazla sosyal medya şirketinin bulunduğu bilinmektedir. (Dixon, 2020: 175). Covid-19'un pazarlamaya etkisi, tüketicilerin e-ticarete daha fazla yönelmesi ve online pazarlamanın tercih edilmesiyle birlikte offline (fiziksel/geleneksel) pazarlamanın önemli ölçüde gerilemiştir. Pandemi sürecinde offline pazarlama, zorunlu olmadıkça tercih edilmeyen eski bir pazarlama yöntemi haline gelmiştir (Alankuş, 2021: 107). Amazon firması nominal deęerlerin dışında, yüz bin yeni çalışanı da işe alarak (Euronews, 2020) covid-19'un etkisiyle, dijital pazarlama ve yeni yaklaşımlar üzerine yapılan çalışmalar ile yaşanan olumsuzlukları aşarak kurumsal başarılar elde edilmesine katkı sağlamış ve e-ticaretin ekonomiye olan katkısının devlet hazinesi için ne kadar kritik olduğunu gözler önüne sermiştir. E-ticaret sektöründe, yapay zekâ teknolojileri, birçok kritik işlevi başarıyla yerine getirebilmektedir. Bu işlevler, müşteri tanıma ve yeniden hedefleme, öneri sistemleri, kişiselleştirilmiş deneyimler, akıllı arama motorları, sanal asistanlar, dinamik fiyatlandırma uygulamaları ve stok yönetimi gibi önemli süreçleri kapsamaktadır. Şirketler, potansiyel müşterilerin

verilerini doğru bir şekilde toplayıp analiz ederek, kullanıcıların e-ticaret platformlarını tercih etmelerini sağlayacak stratejik adımlar atmaktadır. Bu süreçte, yapay zeka teknolojilerinin çeşitli alanları, örneğin doğal dil işleme, görüntü ve ses tanıma, problem çözme teknikleri ve makine öğrenmesi, kullanıcı davranışlarını etkili bir şekilde analiz etmeye ve pek çok pazarlama faaliyetini otomatikleştirmeye olanak tanımaktadır. (Güven ve Ayvaz Güven, 2023: 73-74)

1.3.1. Avantajları

E-Ticaret, müşterilere 7/24 alışveriş yapabilme imkânı sunar. Bu ortamda alışveriş süreçleri hızlı bir şekilde tamamlandığı için zaman tasarrufu sağlanır. Müşteriler, farklı platformlardan ve uygun fiyatlarla alışveriş yapabilirler. Ürünler hakkında araştırma yapmak ve kullanıcı deneyimlerini incelemek oldukça kolaydır. E-Ticaretin sunduğu geniş ürün yelpazesi sayesinde mekân sınırlamaları ortadan kalkar. Ayrıca müşteriler kısa bir süre içinde pek çok ürünün stok durumunu kontrol etme olanağına sahip olurlar. Mağaza kalabalığından uzak daha kişisel bir alışveriş deneyimi yaşarlar. (Ganioğlu vd., 2024: 881).

E-Ticaret sayesinde satıcılar, 365 gün 24 saat açık bir sanal mağazaya sahip olur. Stok yapmadan ürün satışı yapabilme imkânına sahip olurlar. Sanal mağazaların giderleri, geleneksel mağazalara göre çok daha düşüktür. Ayrıca, sanal mağazaların müşteri potansiyeli çok daha geniştir. Satıcılar, farklı ülkelerden müşterilere ulaşma fırsatına sahip olur ve markalarını küresel çapta tanıtmaya imkânı elde ederler. Ürün çeşitliliği, promosyonlar ve indirimler aracılığıyla kolayca müşterilere ulaşabilirler (Kayhan, 2016).

1.3.2. Dezavantajları

Müşteriler, ilgilendikleri ürünleri doğrudan görme ve dokunma imkânına sahip olamazlar. Çevrimiçi platformların, sanal ortamda güvenlik açıklarını gidermeleri büyük önem taşır. Müşteri bilgileri üçüncü şahısların eline geçebilir ve bu da ciddi bir güvenlik riski oluşturur. Ayrıca, e-ticaret üzerinden yapılan alışverişlerde müşteriler, ürünlerin teslim edilmesi için kargo süresini beklemek zorunda kalırlar. Müşteriler, güven duydukları ve sorun yaşamadıkları sitelerden alışveriş yapmayı tercih ederler. Bu sebeple, sektöre yeni giren firmaların güven inşa

etmesi zaman alabilir (Akpunar, 2017).

E-ticaretin satıcılar açısından bazı dezavantajları da bulunmaktadır. Satıcı ile müşteri arasındaki iletişim, yüz yüze ve doğrudan gerçekleşmediği için, satış süreci daha az kişisel hale gelir. Ayrıca, e-ticaret platformlarında yapılan satışlarda, iade ve değişim oranları geleneksel mağazalara göre daha yüksek olabilir. Bu da kargo ve lojistik maliyetlerinde artışa neden olabilir. Müşteriler, alışveriş yapmadan önce kolayca farklı satıcıların fiyatlarını karşılaştırabilirler. Bu nedenle, satıcılar ürün fiyatlandırmalarını yaparken, rekabetçi piyasa koşullarını göz önünde bulundurmak zorundadırlar. (Erçetin, 2015).

1.4. MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM)

Customer Relationship Management (CRM), müşteri ilişkilerini yönetmek ve optimize etmek için geliştirilen bir yaklaşım olarak iş dünyasında önemli bir değişim ve gelişim süreci yaşamıştır. 1980'lerin sonlarına doğru ortaya çıkan CRM kavramı, şirketlerin müşteri verilerini daha sistemli bir şekilde yönetme ihtiyacından doğmuştur. Bu dönemde odak noktası, müşteri iletişimlerini ve satış süreçlerini daha verimli hale getirmektir. 1990'ların başlarında CRM sistemleri, müşteri verilerini bilgisayar tabanlı sistemlerde saklamak ve yönetmek için kullanılmaya başlanmıştır. 2000'lerin başında internetin yaygınlaşmasıyla birlikte CRM, bulut tabanlı çözümlerle daha sofistike hale gelmiş ve küçük- orta ölçekli işletmeler için de erişilebilir olmuştur. 2010'ların dijital çağıyla birlikte CRM, mobil teknolojiler ve sosyal medyanın etkisiyle daha da önem kazanmıştır. Günümüzde müşteri etkileşimleri sosyal medya platformları, mobil uygulamalar ve diğer dijital kanallar üzerinden yönetilebilmektedir. Yapay zeka ve veri analitiği gibi yenilikçi teknolojilerle desteklenmektedir. CRM, işletmelerin müşteri odaklı stratejilerini geliştirmelerine ve müşteri memnuniyetini artırmalarına yardımcı olan kritik bir işlevselliği ifade etmektedir. Sürekli olarak gelişen ve değişen teknolojiler yardımıyla fiyatlandırma stratejileri ve kampanya yönetimi müşterinin alım yapmaya meyilli oldukları fiyatları belirleme noktasında müşteri ilişkileri yönetimini oldukça geliştirmeye devam etmektedir (Şükrüoğlu, 2008: 49).

Müşteri ilişkileri yönetimi, İngilizce "Customer Relationship Management" kelimelerinin kısaltmasıdır ve "CRM" olarak bilinir. CRM, bir şirketin satış,

pazarlama ve satış sonrası destek faaliyetlerini günümüzün gelişmiş bilgi teknolojileriyle yürütmesidir. Bu süreç, tüketici verilerinin toplanması ve analizler sonucunda elde edilen bulgulara göre satış, pazarlama ve müşteri hizmetleri stratejilerinin belirlenmesidir. Kısacası, CRM, firmanın müşteri talepleri ve ihtiyaçları doğrultusunda kendini geliştirme ve buna göre cevap verme sürecini ifade eder. CRM, satış ve pazarlama kanallarının ayrılmaz bir parçası haline gelmiştir. Bugün, şirketlerin odak noktası, ürün ya da hizmeti daha ucuz hale getirmekten ziyade, rekabeti müşteri odaklı bir yapıya kavuşturmak olmuştur. Müşterilerle ilgili elde edilen değerli veriler hem günümüzde hem de gelecekte büyük önem taşımaya devam edecektir (Özdemir, 2018: 24; Üner, 2011).

Müşteri İlişkileri Yönetimi (CRM), modern iş dünyasında büyük önem kazanan bir stratejidir. CRM, bir işletmenin müşterileriyle olan ilişkilerini daha verimli bir şekilde yönetmek ve optimize etmek için kapsamlı bir yöntem sunar. Bu sistem, müşteri verilerini merkezi bir veri tabanında toplar, müşteri etkileşimlerini kaydeder ve analiz eder. Bu sayede, şirketler müşteri ihtiyaçlarını daha iyi anlayabilir, memnuniyeti artırabilir ve sadakati güçlendirebilir. CRM aynı zamanda satış süreçlerini iyileştirerek satış verimliliğini artırır ve pazarlama kampanyalarının etkinliğini artırmak için stratejik bilgiler sağlar. CRM, işletmelerin rekabet avantajı kazanmalarına ve sürdürülebilir büyümeyi sağlamalarına yardımcı olan önemli bir araçtır. Müşteri odaklı bir yaklaşım benimseyen işletmeler için kritik bir araç olan CRM'nin temel amacı, müşteri memnuniyetini ve sadakatini artırmak ve satış performansını yükseltmektir. Özel yazılımlar veya bulut tabanlı platformlar şeklinde sunulan CRM sistemleri, müşteri iletişimlerini kaydetme, talepleri yönetme, satış süreçlerini optimize etme, pazarlama kampanyalarını yönetme ve verileri analiz etme gibi işlevler içerir. Bu sistemler, müşteri ilişkilerini kişiselleştirir, taleplere hızlı yanıt verilmesini sağlar, satış süreçlerini iyileştirir ve pazarlama stratejilerini etkinleştirir. Sonuç olarak, CRM, işletmelerin rekabet avantajı elde etmelerine ve sürdürülebilir büyüme sağlamalarına yardımcı olur (Kestane, 2024: 4).

Veri analitiği, dijital dönüşümün temel bir bileşeni olarak, stratejik kararları yönlendirmek ve uygulanabilir içgörüler elde etmek amacıyla verilerin sistematik bir şekilde incelenmesini ifade eder (Odonkor ve diğerleri, 2024). Veri analitiğinden faydalanan kuruluşlar, tüketici davranışları, tercihler ve eğilimler hakkında daha

derin bir anlayış kazanarak daha hedefe yönelik ve etkili pazarlama kampanyaları oluşturabilir (Odonkor ve diğerleri, 2024). Bu yetenek, veri odaklı karar almanın rekabet avantajı için kritik bir faktör haline geldiği bir ortamda hayati bir öneme sahiptir (Adegbola ve diğerleri, 2024; Akinsulire ve diğerleri, 2024; Oriji & Joel, 2024; Ucha, Ajayi & Olawale, 2024).

Benzer şekilde, Müşteri İlişkileri Yönetimi (CRM) araçları, müşteri ilişkilerinin yönetilmesi ve geliştirilmesinde önemli bir rol oynar. Müşteri odaklı yaklaşım, son yıllarda pazarlama alanında yükselen önemli bir trend olmuştur. Bu yaklaşıma göre, işletmelerin amacı, müşterilerin talepleri doğrultusunda üretilen ürün ve hizmetlerle beklentilerini karşılamaktır. Günümüzde şirketler, ürünleri üretmeden önce müşterilerin istek ve ihtiyaçlarını dikkate alarak üretim sürecini şekillendirirler. Bu anlayışa göre, pazarlama planlaması, tüketicilerin taleplerinin ve ihtiyaçlarının belirlenmesiyle başlar ve bu tespit edilen veriler üretim sürecine de dahil edilerek devam eder. İşletmelerin müşteri verilerini bir araya getirmesine, etkileşimleri takip etmesine ve iletişimi kişiselleştirmesine olanak tanıyarak daha güçlü ilişkiler kurmalarını ve müşteri sadakatini artırmalarına yardımcı olur (Urefe, Odonkor & Agu, 2024). CRM sistemleri, pazarlama stratejilerinin müşteri ihtiyaçlarıyla uyumlu hale getirilmesini kolaylaştırarak etkileşimi ve satış büyümesini destekler (Alabay, 2010; Ngai, 2005).

Günümüzün rekabetçi pazarında, müşterilerin çok sayıda seçeneğe sahip olmaları nedeniyle işletme değiştirmekte tereddüt etmemektedirler. Müşterilerin bir işletmeyi tercih etmeleri için, onlara istedikleri ürün ve hizmeti sunmak gerekmektedir. Müşteri davranışlarını analiz ederek, doğru stratejiler geliştirilmesi ve bu stratejilerle müşteri memnuniyetinin sağlanması mümkündür. Müşteri ilişkileri yönetimi (CRM), bir pazarlama stratejisi olarak, işletmenin rakiplerinden farklılaşmasını veya bir adım öne geçmesini hedeflemektedir (Ezerce, 2008).

1.5. MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM) ÖZELLİKLERİ

Müşteri ilişkileri yönetiminden maksimum verim elde edebilmek için bu kavramın doğru tanımlanması, anlaşılması ve işletme süreçlerine uygun bir şekilde uygulanması kritik öneme sahiptir. CRM, firmaların yoğun rekabet koşullarında başarısını belirleyen temel unsurlardan biri olarak görülmektedir (Elkordy, 2014:

131). İşletmelerin kârlarını artırabilmesi, doğru müşteri kitlesine ulaşmalarına ve bu müşterilerin memnuniyetini sürdürülebilir kılmalarına bağlıdır. Bu doğrultuda, sadık müşterilere odaklanmak, onların önceliklerine değer vermek ve değişen ihtiyaçlarına hızlı bir şekilde yanıt verebilmek başarının temel yapı taşlarını oluşturur. Ayrıca, teknolojinin sunduğu fırsatları etkin bir şekilde değerlendirirken, ortaya çıkabilecek tehditlere karşı hazırlıklı olunması ve işletmenin ihtiyaçlarına uygun bir teknik altyapının inşa edilmesi gerekmektedir. Bu unsurların bir araya gelmesi, işletmelerin müşteri ilişkilerini güçlendirerek uzun vadeli başarı elde etmelerini sağlar. (Demirel, 2006: 51)

Ayrıca Demirel'e göre müşteri ilişkileri yönetimi, doğru, anlaşılabilir ve tam zamanında bilgi girişine önem vererek müşterilerle olan ilişkilerden kazanılan deneyimlerle bir veri tabanı oluşturmayı, bu veri tabanından elde edilen bilgiler doğrultusunda müşterilerin istek ve ihtiyaçlarına uygun ürün veya hizmetler geliştirmeyi amaçlar. Bu yaklaşım, geçici bir süreçten ziyade sürekli yinelenen ve gelişen bir iş modeli olarak değerlendirilmelidir. CRM sistemleri de bu iş modelini destekleyen kritik araçlar arasında yer alır. Farklı kaynaklardan gelen müşteri verilerini merkezi bir veri tabanında birleştirerek satış, pazarlama ve müşteri hizmetleri gibi departmanların verilerini entegre eder ve tek bir müşteri profili sunar. Günümüz iş dünyasında mobil erişim bir gereklilik haline geldiğinden, CRM sistemleri mobil cihazlardan erişim imkanı sağlayarak saha ekiplerinin ve yöneticilerin müşteri verilerine her zaman ve her yerden kolayca ulaşabilmesine olanak tanır. Bu özellikler, müşteri ilişkileri yönetiminin sürekli gelişen bir yapı olduğunun altını çizer.

CRM sistemleri, bütünleşmiş analitik ve raporlama araçlarıyla müşteri verilerini derinlemesine analiz etme ve stratejik içgörüler elde etme olanağı sağlar. Bu sayede işletmeler, satış trendlerini, müşteri davranışlarını ve pazarlama kampanyalarının etkinliğini ölçerek daha bilinçli kararlar alabilirler. E-posta entegrasyonu ile müşteri iletişimini yönetmek kolaylaşır; otomatik e-posta gönderimi, kişiselleştirilmiş kampanyalar ve özel teklifler gibi özelliklerle pazarlama stratejileri güçlendirilir ve müşteri ilişkileri derinleştirilir. Tüm bu entegrasyon özellikleri, işletmelerin müşteri odaklı yaklaşımlarını destekleyerek müşteri memnuniyetini artırır. Operasyonel verimliliği yükseltir ve rekabet avantajı elde etmelerini sağlar

(BT vd., 2018: 4).

1.6. MÜŞTERİ İLİŞKİLERİ YÖNETİMİNİN KAPSAMI VE YAŞAM DÖNGÜSÜ

Literatürdeki tanımlara bakıldığında, müşteri ilişkileri yönetiminin temel amacı, müşteri sadakatini sağlamak, müşteri kaybını önlemek ve yeni müşteriler kazanarak işletmenin kârlılığını artırmaktır. Bu süreç, işletme genelinde etkili iletişim becerileri kullanılarak müşteri davranışlarının analiz edilmesi ve doğru şekilde yorumlanmasıyla şekillenmektedir. Böylece, müşteri odaklı stratejiler geliştirerek uzun vadeli başarılı ilişkiler kurmak mümkün hale gelmektedir.

Müşteri ilişkileri yönteminin yaşam döngüsü 4 ana başlık altında incelenir. Bunlar müşteri değeri yaratma, müşteri memnuniyeti oluşturma, müşteri kazanma-tutma ve müşteri sadakatidir.

1.6.1. Müşteri Değeri Yaratma

Yeni pazarlama yaklaşımında, işletmelerin rakipleriyle etkin bir şekilde rekabet edebilmesi için müşterilerine değer sunması gerekmektedir. Müşteri değer önerisi, bir işletmenin müşterilerine fiyat, kalite, hizmet gibi çeşitli değer unsurlarını bir arada sunarak oluşturduğu tekliften oluşur (Kurnuç, M.,2020: 19).

Müşteriye hitap etme, aslında müşteriye değer yaratma kavramının en temel şekilde ifade edilmesidir. Değer, müşterilerin ödemeye hazır oldukları fiyat karşılığında, işletmenizden beklentilerinin ötesinde bir ürün veya hizmet talep etmeleridir. Ürün ya da hizmetin düşük fiyatı, beklenen faydayı azaltmasa da yüksek fiyat, müşterilerin ürün veya hizmetten elde edecekleri faydaya ilişkin beklentilerini artırır. (Doyle, 2003: 151).

Yaratılan değer sayesinde, müşteri ile işletme arasında güçlü bir bağ kurulur. Bu bağ, müşterinin işletmeye olan sadakatini artırır, tekrar satın alma davranışını teşvik eder ve hatta işletmenin destekçisi olmasına yol açabilir. Ancak, müşteri açısından algılanan değer yalnızca fiyat gibi somut değerlerle açıklanamaz. Çünkü güven, güvence, sosyal statü ve ün gibi soyut unsurlar da duygusal değerlerin oluşumunda önemli rol oynar. Bu nedenle değer yaratma süreci, hizmet, kalite ve müşteri memnuniyetinin tüm unsurlarını kapsayan bir bütünsel yaklaşım gerektirir.

(Odabaşı,2005: 50)

Şekil 2: Müşterinin Algıladığı Değer



Kaynak: Odabaşı ve Oyman (2005) çalışmasından alınmıştır.

1.6.2.Müşteri Memnuniyeti

Müşteri memnuniyeti, işletmelerin müşterilerin ihtiyaç ve beklentilerini karşılama düzeyi olarak tanımlanabilir. Genel olarak müşteri memnuniyeti, müşterinin yaşadığı deneyimin beklentilerini karşılama veya aşmasıyla sağlanır. Bu nedenle, müşteri memnuniyetinin temel unsuru, müşterinin deneyim sürecinde tatmin olmasıdır.

İşletmeler açısından müşteri memnuniyetini sağlamak, satış öncesi, satış esnası ve satış sonrası süreçlerde müşterilerin istek ve ihtiyaçlarını karşılamaktan geçer. Müşteriler, memnuniyetlerini değerlendirme kararını genellikle bu süreçlerin tümünü deneyimledikten sonra verirler. Bu nedenle, işletmeler sadece ürün, hizmet veya servis sunmanın ötesine geçmeli, müşterilere olumlu bir deneyim sunmayı ve bu deneyimi geliştirmeyi hedeflemelidir. Müşterinin bireysel ya da kurumsal bir kimliğe sahip olmasından ziyade, işletmenin sunduğu değer ve hizmetlerin müşteri deneyimi üzerindeki etkisi esas alınmalıdır. Bu anlayış, müşteri memnuniyetini bir hedef olmanın ötesinde, işletmelerin sürdürülebilir başarıya ulaşmalarında stratejik bir araç haline getirmektedir (Kağnıcıoğlu, 2002: 80-81).

1.6.3. Müşteri Kazanma ve Tutma

Müşteri kazanımı ve elde tutulması, işletmelerin müşteri davranışlarını analiz ederken üzerinde durmaları gereken en önemli konular arasındadır. Müşteri kazanımı, yalnızca mevcut müşterileri rakiplere kaptırmamakla kalmaz, aynı zamanda yeni müşterileri de işletmeye çekmeyi hedefler. Bu süreç, uzun vadede müşteri sadakatini sağlamaya yönelik stratejik bir yaklaşımdır. Müşteri tutma, müşteriler için süreklilik yaratarak sadık müşteriler haline getirmektir (Odabaşı ve Oyman, 2005: 113).

Rekabet avantajı elde etmenin yolu, işletmelerin müşteri kazanımı ve mevcut müşterilerini elde tutabilmesinden geçmektedir. Bu başarı, büyük ölçüde müşterilerin güvenini kazanmakla sağlanır. Yeni müşteri kazanmak, mevcut bir müşteriyi elde tutmaktan daha pahalı olduğundan, işletmelerin sürdürülebilir kârlılık için müşteri bağlılığı oluşturmaları kritik öneme sahiptir. Ayrıca, işletmeler mevcut müşterileriyle güven temelli ilişkiler kurduğunda, yeni müşteri kazanma sürecinde de önemli bir ilerleme kaydetmiş olurlar (Hacıfendioğlu ve Çolular, 2008: 110).

1.6.4. Müşteri Sadakati

Müşteri sadakati, bir işletmeye, markaya, ürüne veya kişiye karşı süregelen bir bağlılık durumudur. Sadık müşteriler, işletme hakkında yaptıkları yorumlar ve verdikleri tavsiyeler ile aslında dolaylı yoldan reklam yapmış olurlar. (Gödekmerdan ve Deniz, 2010).

Müşteri sadakati ve müşteri memnuniyeti genellikle birbirine çok benzeyen kavramlar olarak düşünülse de aralarında önemli farklar bulunmaktadır. Bazı araştırmacılar, müşteri memnuniyetinin sağlanmasının müşteri sadakati için yeterli olduğunu öne sürse de bu görüş tam olarak doğru değildir. Kotler'a göre, müşteri memnuniyeti 1'den 5'e kadar bir ölçekle değerlendirildiğinde, müşterinin en yüksek seviyeye ulaşana kadar başka işletmeleri tercih etme olasılığı vardır. Sadece en yüksek seviyede, müşteri işletmeden tekrar ürün ve hizmet satın alma eğilimi gösterir ve çevresine işletme hakkında olumlu geri bildirimlerde bulunur. Bu bağlamda, müşteri sadakati ancak en üst düzeyde oluşmaktadır (Kotler, 2000: 197).

Müşteri sadakati davranışsal ve tutumsal olarak iki temel başlığa ayrılmıştır. Bu başlıklar ise sırasıyla:

a) Doğrusal yaklaşım: Tüketiciler, rakip işletmeler olsa dahi satıcıdan ve verdiği hizmetlerden tatmin olduğu sürece sadakatini değiştirmeyecektir. Ancak, tüketicilerin memnuniyetinin ve ürüne olan talebin azalması tüketici sadakatini olumsuz etkiler.

b) Tutumsal yaklaşım: Tüketicilerin, ürün veya hizmet ile arasında manevi bir köprü kurmasıdır. Bu sadık tüketici grubu, işletmeden birebir hizmet almasa dahi işletmenin ürün ve hizmetlerini tavsiye etmeye devam edecektir (Tassembledo, 2016).

1.7. CRM’NİN TEMEL BİLEŞENLERİ

Müşteri ilişkileri yönetimi (CRM), işletmelerin organizasyonel etkinliklerini ve verimliliklerini artırarak, müşterilerden elde edilen faydalı bilgiler ile müşteri ihtiyaçlarını anlamak, öngörmek ve yönetmekle ilgilidir (Tseng, 2016). Bu nedenle, CRM, bilgi teknolojilerindeki ilerlemeler ve müşteri odaklı süreçlerdeki organizasyonel değişikliklerle evrilmiştir. CRM’in büyük bir kısmı teknolojiyle ilişkilendirilsede, yalnızca teknolojik bir çözüm olarak ele alınması genellikle başarısız sonuçlar doğurabilir. Başarılı bir CRM uygulaması için; insan, teknoloji ve süreçlerin entegrasyonunu içeren dengeli bir yaklaşım gereklidir (Chen ve Popovich, 2003: 672). CRM teknolojisi, firmaların daha uygun pazarlama stratejileri geliştirmelerine ve pazarlama faaliyetlerini daha verimli ve hızlı bir şekilde yürütmelerine olanak tanıırken, özellikle üstün ön destek ve entegre müşteri verilerine erişim sunar (Chang vd., 2010: 850). Bir şirketin müşterilerini en iyi şekilde anlamayı ve ilişki yönetimini sağlamayı hedefleyen CRM, insan sermayesi ve teknolojinin entegrasyonu ile müşteri elde tutma odaklı, ilişki yönetimi için en yeni bütünleştirici yaklaşım olarak kabul edilmektedir (Gil-Gomez vd., 2020: 2734).

Müşterilere ait kişisel bilgiler, satın alma alışkanlıkları, alışveriş sıklığı ve tutarları gibi veriler, işletmelerin müşterileri daha iyi anlaması ve onların ihtiyaçlarına yönelik stratejiler geliştirebilmesi için bir veri tabanında toplanmalıdır. Bu veriler, müşterilerin alışkanlıklarını ve tercihlerini anlamayı kolaylaştırarak hem işletmelere gelir tahmini yapma hem de müşterilere özel seçenekler sunma imkânı sağlar. İşletmeler, bu bilgilerle müşterilerini değerlilik derecelerine göre sınıflandırır ve en kârlı müşteri gruplarına odaklanarak kaynaklarını daha etkili kullanabilir.

Sınıflandırmada yalnızca demografik veriler değil, aynı zamanda alışveriş alışkanlıkları da dikkate alınır. Bu süreç, müşterilerin gelecekteki davranışlarını öngörmeyi ve pazarlama faaliyetlerini daha etkili bir şekilde düzenlemeyi mümkün kılar. Ayrıca, müşteriyle kurulan her türlü iletişim, onların değerli olduğunu hissettirmek ve bu etkileşimlerden elde edilen verileri değerlendirmek için bir fırsat olarak görülmelidir. Bu tür analizler, müşteri ilişkilerinin uzun vadeli ve verimli bir şekilde geliştirilmesine katkı sağlar (Kolay, 2016; Yasak, 2017; Yiğit,2017).

Şekil 3: Müşteri İlişkileri Yönetimi Bileşenleri



Kaynak: Araştırmacı tarafından oluşturulmuştur.

1.7.1. İnsan

CRM'in merkezinde insan faktörü yer almaktadır. Müşteri ve çalışanlarla etkili bir iletişim kurmak hem pazarlama hem de işletmenin başarısı açısından kritik bir öneme sahiptir. Tengilimoğlu'na (2015: 199) göre, hizmetin kalitesini yalnızca hizmeti alan müşteriler değil, aynı zamanda hizmetle doğrudan ilişkili olmayan bireyler de etkileyebilmektedir. Benzer şekilde, Öztürk (2008: 99), tüketicilerin hizmetin değerini belirlerken çalışanları temel bir ölçüt olarak gördüğünü ve hatta çalışanların, sunulan hizmetin kendisi olarak algılandığını ifade etmektedir. Bu bağlamda, hizmet sektöründe çalışanların tutum ve davranışları, müşterilerin algısını doğrudan şekillendirerek hizmetin kalitesine dair önemli bir belirleyici olmaktadır. Müşteri ilişkileri yönetiminin başarılı olabilmesi için işletmeler, kaliteli, eğitimli, dürüst ve yetkin çalışanlara ihtiyaç duyarlar. İşletmeler, insan kaynaklarını etkili bir şekilde kullanabilmek için, müşteri odaklı ödüllendirme programlarını iç müşterilerine, yani çalışanlarına da uygulamalıdır. Bu yaklaşım, iç müşterileri

teşvik etmek ve onları ödüllendirmek için etkili bir yöntem olup, uzun vadeli operasyonlarda sadakatlerini sürdürebilmelerine olanak tanır. Ayrıca, iç müşteriye, yaptığı işlerde sorumluluk alabilmesi ve inisiyatif kullanabilmesi için fırsatlar sunulmalıdır. (Ezerce, 2008).

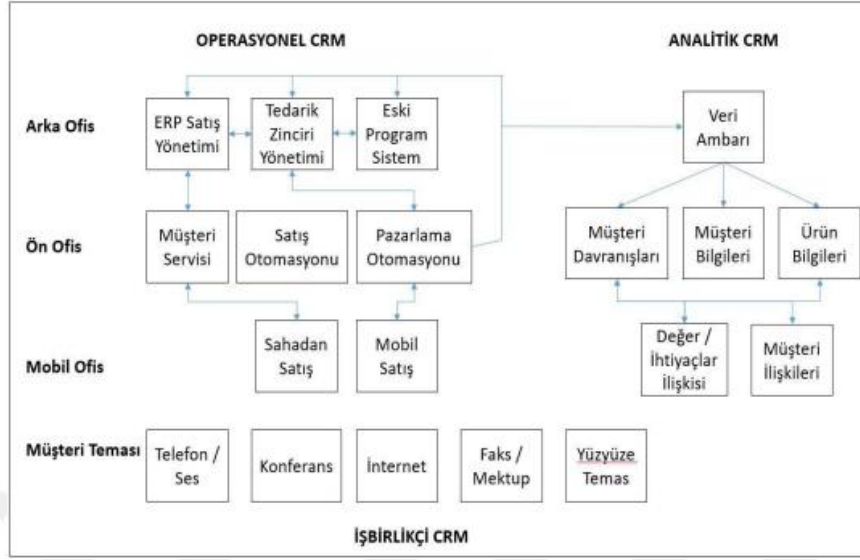
1.7.2. Süreç

Müşteri memnuniyetini artırmak ve rekabet avantajı sağlamak için hizmet süreçlerinin sürekli olarak gözden geçirilmesi ve iyileştirilmesi büyük önem taşır. Süreçler, işletmenin verimliliğini ve müşteri deneyimini doğrudan etkileyen unsurlar arasında yer alır. Bu nedenle, değişime ihtiyaç duyulan noktaların belirlenmesi ve süreçlerin yeniden tasarlanması, işletmelerin başarısı açısından kritik bir adımdır. Müşteri odaklı süreç iyileştirme çalışmalarında, öncelikle müşterilerin mevcut süreçleri nasıl deneyimlediği analiz edilir. Bu süreçte, müşteri geri bildirimleri toplanır ve çalışanlarla birlikte değerlendirilerek iyileştirme alanları belirlenir. Süreçlerin yeniden yapılandırılması genellikle bilgi teknolojileriyle desteklenerek gerçekleştirilir. Dijitalleşme sayesinde işletmeler, süreçlerini daha esnek, hızlı ve müşteri odaklı hale getirebilir. Bu süreçlerin yeniden yapılandırılması ve uygulanması, bilgi teknolojileriyle uyumlu bir şekilde gerçekleştirilir (Beliner, 2015).

1.7.3. Teknoloji

Müşteri ilişkileri yönetiminde, öncelikle müşterileri tanımak, sınıflandırmak ve onlarla iletişim kurmak için teknolojilere ihtiyaç duyulmaktadır. Müşterilerin duyarlılıkları ve satın alma alışkanlıklarıyla ilgili veriler, genellikle ham halde olup, öncelikle işlenmesi gerekir. Anlamlı bilgiye dönüştürülmesi ise teknoloji sayesinde mümkündür. Müşterilerden elde edilen veriler doğrultusunda, gelecekteki davranışları hakkında tahminler yapılabilir. Müşterilere yönelik pazarlama, satış, hizmet ve ürün geliştirme stratejileri, bu tahminler ışığında şekillendirilecektir. (Yıldırım, 2012).

Şekil 4: Müşteri İlişkileri Yönetimi Teknoloji Uygulama Süreci



Kaynak: Ezerce (2008) çalışmasından alınmıştır.

1.7.3.1. Analitik CRM

Müşteri ilişkileri yönetimi, analiz ve eylem olarak iki aşamaya ayrılır. Müşterileri ayırma ve sınıflandırma işlemi, analitik CRM aşamasına dahildir. Müşteri bilgilerine erişim sağlama ve müşteri ile etkileşimde bulunma ise CRM'in eylem aşamasını oluşturur. Ölçülemeyen bir şeyin yönetilmesi mümkün değildir. Analitik CRM, işletmelere rekabet avantajı sağlayarak müşteri analizlerini daha etkili hale getirir. Bu sayede müşteriler detaylı bir şekilde analiz edilerek segmentlere ayrılır ve her segment için özel aksiyon planları oluşturularak farklı kampanyalar yapılabilir. Analiz sonuçları, yürütülen faaliyetlerin hedef kitleye ulaşp ulaşmadığını belirlemeye yardımcı olurken, uygulanan satış ve pazarlama stratejilerinin etkinliğini de değerlendirme fırsatı sunar. Analitik CRM, operasyonel sistemlerden ve arka ofis uygulamalarından elde edilen verileri analiz ederek müşteri davranışlarını tahmin etme yöntemi kullanır. Bu süreç sonucunda müşteri segmentasyonuna göre her bir grup için özel stratejiler geliştirilerek daha odaklı ve etkili aksiyonlar alınır (Ezerce, 2008; Yiğit, 2017).

İşletmeler, Analitik CRM uygulamaları sayesinde çeşitli önemli avantajlar elde ederler. Bunlar arasında şunlar öne çıkmaktadır:

- Müşteri davranışlarını analiz ederek, onlara uygun aksiyonlar geliştirme,
- Müşterileri bireysel ve grup bazında detaylı bir şekilde analiz etme,
- Müşteri profilleri oluşturma,
- Yeni pazar fırsatları ve öneriler oluşturma,
- CRM çalışanlarının performans değerlendirmesini gerçekleştirme,
- Market sepeti analizleri yapma. (Özdemir, 2018).

1.7.3.2. Operasyonel CRM

Operasyonel CRM, müşterilere ait verilerin toplanmasıyla başlayan bir süreçtir ve temel amacı, potansiyel müşterileri belirlemek, müşteri bilgilerini elde etmek, farklı müşteri gruplarına yönelik etkileşim kurmak ve müşteri yaşam döngüsü boyunca destek sağlamaktır. Etkili bir operasyonel CRM için müşterilerin değerine odaklanmak oldukça önemlidir, çünkü tüketiciler kendilerini özel hissetmek ister. İşletmeler, müşterilerini kategorize ederek daha etkili ve odaklı operasyonel stratejiler uygulayabilir. Operasyonel CRM, satış, pazarlama ve hizmet süreçlerini bir araya getirir ve bu süreçlerin her biri tek başına değil, entegre bir şekilde çalışan bir otomasyon sistemi olarak işler. Bu bütünlük yaklaşım, işletmelerin operasyonlarını daha verimli hale getirmesine ve müşterilerle daha güçlü ilişkiler kurmasına olanak tanır (Gökmen, 2014).

1.7.3.3. İşbirlikçi CRM

İşbirlikçi CRM, tüketicilere ait bilgilerin işletmenin çalıştığı paydaşları ve diğer ilgili firmalar ile paylaşmasıdır. Tüketicilere özel promosyon ve kampanyaların sunulmasının sağlayan süreçlerdir. Tüketici ile her noktada temasa geçilebilir. İletişim araçları olarak, klasik yöntemler ve teknolojik uygulamalar seçilebilecektir. İşbirlikçi CRM; satış, teknik destek ve pazarlamayı geliştirmektedir. İşbirlikçi CRM'in asıl amacı müşteri hizmetlerinin kalitesini artırarak müşteri tatminini en üst düzeye taşımaktır (Tassembledo, 2016).

1.8. MÜŞTERİ İLİŞKİLERİ YÖNETİMİNİN (CRM) FAYDALARI

CRM'in süreç temelli yapısı, işletmelerin müşteri etkileşimlerini ve hizmetlerini daha düzenli yönetmelerine olanak tanır. Her müşteri etkileşimi, satış adımı veya pazarlama kampanyası belirli süreçlerle tanımlanır ve bu süreçler haritalandırılarak işletmelerin analiz, iyileştirme ve verimlilik artırma çalışmalarına destek olur. Süreç temelli CRM, müşteri deneyimini sürekli iyileştirme imkanı sunar. Müşteri etkileşimleri kaydedilir ve analiz edilir, bu veriler doğrultusunda kişiselleştirilmiş hizmetler sunulur ve müşteri memnuniyeti ve sadakati artırılır. Böylece işletmeler, rekabet avantajı elde ederler. Yapılandırılmış CRM süreçleri, stratejik yönetimde büyük fayda sağlar; süreçlerin her aşaması net tanımlandığı için stratejik hedeflere yönelik planlamalar yapılabilir. Süreç performansı sürekli izlenir ve analiz edilir, bu da işletmelerin hızla değişen pazar koşullarına uyum sağlamalarını ve rekabet güçlerini artırmalarını sağlar.

Ayrıca, süreç temelli CRM, işletmelerin eğitim ve personel entegrasyonunu kolaylaştırır. Belirli süreç adımları ve roller sayesinde yeni çalışanlar iş süreçlerine hızla uyum sağlayabilir ve görevlerini daha verimli şekilde yerine getirebilirler. Bu durum, operasyonel verimliliği artırır ve personel performansını iyileştirir (Şirzad, 2019: 100).

1.9. PAZARLAMA VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ (CRM) İLİŞKİSİ

Pazarlama ve CRM (Müşteri İlişkileri Yönetimi), modern iş dünyasında birbirini tamamlayan iki kritik kavramdır. Pazarlama, işletmelerin ürün veya hizmetlerini hedef kitleye tanıtmak, satışları artırmak ve müşteri taleplerini karşılamak amacıyla uyguladığı stratejik yöntemler bütünüdür. Bu süreç, pazar araştırmalarından başlayarak hedef kitlenin belirlenmesi, ürünün konumlandırılması, uygun fiyatlandırma politikalarının oluşturulması ve tanıtım faaliyetlerinin gerçekleştirilmesini içerir.

CRM ise, işletmelerin müşteri ilişkilerini yönetmelerine yardımcı olan bir sistemdir. CRM sistemleri, müşteriyle olan tüm etkileşimleri (satışlar, pazarlama kampanyaları, müşteri hizmetleri vb.) kaydederek merkezi bir veri tabanında toplar. Bu veriler, işletmelerin müşteri davranışlarını analiz etmelerine, kişiselleştirilmiş

kampanyalar düzenlemelerine ve müşteri memnuniyetini artırmalarına olanak tanır.

Pazarlama ve CRM' in entegrasyonu, işletmelerin müşteri odaklı stratejiler geliştirmelerine ve müşteri deneyimini sürekli olarak iyileştirmelerine imkan sağlar. Pazarlama faaliyetlerinin etkili bir şekilde yönetilmesi ve CRM sistemlerinin etkin kullanımı, işletmelere rekabet avantajı kazandırır ve uzun vadeli başarılarını destekler. Bu entegrasyon, işletmelerin müşteri ihtiyaçlarını daha iyi anlamalarını ve bu ihtiyaçlara uygun stratejiler geliştirmelerini mümkün kılar. (Demir & Kırdar, 2024: 10).

Pazarlama, bir ürün veya hizmeti müşterilere tanıtmak, pazarın ihtiyaçlarını anlama ve bu ihtiyaçlara cevap verecek stratejiler geliştirme sürecini kapsar. Pazarlama stratejileri genellikle geniş kitlelere yöneliktir ve müşteri edinimi, marka bilinirliği gibi temel hedefleri vardır.

CRM, bu süreçlerin daha ileri bir aşamasında, özellikle mevcut müşteri ilişkilerini sürdürme ve güçlendirme odaklıdır. CRM'in amacı, pazarlamanın başlattığı müşteri ilişkilerini derinleştirerek, müşteri sadakati yaratmak, uzun vadeli kârlılığı sağlamak ve müşterilere daha kişiselleştirilmiş hizmetler sunmaktır. CRM, pazarlama stratejisinin bir evrimi olarak düşünülebilir çünkü bir müşteri bir kez elde edildikten sonra, CRM devreye girerek bu ilişkiyi sürdürme, kişiselleştirme ve iyileştirme süreçlerini yönetir. Yani pazarlama, müşteri edinimi sürecini yönetirken CRM, bu müşteriyle olan uzun vadeli ilişkiyi inşa etmeye ve optimize etmeye çalışır. (Tekel, 2006: 10-17)

Özetle, CRM pazarlamadan bağımsız olarak değil, onunla güçlü bir sinerji içinde çalışır. Pazarlama kampanyaları yeni müşteriler çekmek için stratejiler geliştirir ve CRM bu müşterileri elde tutarak, onları tekrar eden müşterilere dönüştürmek için stratejiler uygular. İkisi birlikte çalışarak, müşteri deneyimini ve memnuniyetini maksimuma çıkarır.

1.10. BÜYÜK VERİ (BIG DATA) VE MÜŞTERİ İLİŞKİLERİ YÖNETİMİ

Büyük veri ve Müşteri İlişkileri Yönetimi (CRM), modern işletmelerin pazarlama iletişimi stratejilerini belirlemede kritik rol oynar. Büyük veri, işletmelere geniş veri kümelerini toplama ve bu verileri analiz etme imkanı tanır. Bu sayede işletmeler, müşteri davranışlarını daha derinlemesine anlama, piyasa trendlerini

belirleme ve stratejik kararlar alırken veriye dayalı içgörüler elde etme fırsatına sahip olur. Örneğin, müşteri satın alma alışkanlıklarını analiz ederek gelecekteki talepleri tahmin etmek veya müşteri geri bildirimlerini değerlendirerek ürün ve hizmetlerde iyileştirmeler yapmak mümkündür (Duran, 2021: 35)

CRM sistemleri ise, müşteri ilişkilerini yönetmek ve müşteri verilerini merkezi bir platformda organize etmek için tasarlanmıştır. Bu sistemler, işletmelerin müşterileriyle olan etkileşimlerini kaydetmelerine ve bu verileri analiz ederek kişiselleştirilmiş deneyimler sunmalarına olanak tanır. Müşteri profilleri, geçmiş satın alma verileri, iletişim geçmişi ve geri bildirimler gibi bilgiler, CRM sistemlerinde bir araya getirilir ve bu veriler, pazarlama kampanyalarını daha hedefli ve etkili hale getirmek için kullanılır. Böylelikle belirli bir müşteri segmentine özel teklifler sunarak satışları artırmak veya müşteri hizmetlerini kişiselleştirerek müşteri memnuniyetini yükseltmek mümkün hale gelir.

Büyük veri ve CRM'nin entegrasyonu, işletmelere müşteri odaklı stratejiler geliştirmede büyük avantaj sağlar. İşletmeler, büyük veri analizleri ile müşteri eğilimlerini ve tercihlerini daha iyi anlayarak pazarlama stratejilerini daha etkili bir şekilde planlayabilirler. CRM sistemleri, bu stratejilerin uygulanmasında kritik bir rol oynar. Çünkü müşteri etkileşimlerinin ve geri bildirimlerinin sürekli olarak izlenmesi ve değerlendirilmesi, stratejilerin etkinliğini artırır (Öztürk, 2021: 56). Ayrıca, müşteri ilişkilerini güçlendirerek müşteri sadakatini artırmak ve uzun vadeli başarıyı sağlamak için kişiselleştirilmiş iletişim ve hizmet sunmak mümkündür. Büyük veri ve CRM, işletmelerin rekabet avantajı elde etmelerini ve sürdürülebilir büyüme sağlamalarını destekleyen iki önemli bileşendir. Bu iki aracın etkin kullanımı, işletmelerin müşterilerine daha iyi hizmet sunmalarına, pazarlama iletişimini optimize etmelerine ve müşteri memnuniyetini artırmalarına yardımcı olur. Büyük veri analitiği, işletmelere müşteri tercihlerini ve alışveriş alışkanlıklarını derinlemesine inceleme fırsatı tanır. Bu değerli veriler, CRM sistemlerine entegre edildiğinde, işletmeler müşteriye özel pazarlama kampanyaları oluşturabilir ve iletişimlerini en uygun zaman ve kanallarla gerçekleştirebilir. Bu sayede işletmeler, müşteri memnuniyetini artırarak müşteri ilişkilerini güçlendirebilir ve rekabet avantajı elde edebilirler. Günümüzde büyük veri ve CRM entegrasyonu, işletmelerin stratejik hedeflerine ulaşmalarını destekleyen kilit bir araç olarak görülmektedir

(Kılıç vd., 2019: 15).

Müşteri İlişkileri Yönetimi (CRM) platformları, işletmelerin müşteri verilerini yönetmelerine ve ilişkilerini güçlendirmelerine yardımcı olur. Ancak, bu süreçlerin daha etkili hale gelmesi için yapay zeka (AI) entegrasyonu büyük önem taşır. AI, büyük veri setlerini hızlıca analiz ederek müşteri davranışlarını daha derinlemesine anlamayı sağlar, kişiselleştirilmiş pazarlama stratejileri geliştirilmesine olanak tanır ve müşteri memnuniyetini artırır. Ayrıca, yapay zeka destekli CRM sistemleri, operasyonel verimliliği artırarak tekrarlayan görevleri otomatikleştirir ve iş gücü verimliliğini yükseltir. Bu entegrasyon, işletmelere rekabet avantajı sağlar ve uzun vadeli başarılarını sürdürmelerine yardımcı olur (Çakır, 2024: 23).

1.11. KAMPANYA YÖNETİMİ

Pazarlama yönetimi, markanın oluşturulması ve yönetilmesini kapsar. Etkili bir marka yönetimi, işletmenin ürün veya hizmetlerini rakiplerinden ayıran benzersiz değer teklifleri geliştirmeyi, tüketicilerin zihninde olumlu bir imaj oluşturmaya ve marka sadakatini artırmayı amaçlar. Bu süreçte, işletmeler çeşitli pazarlama iletişimi kanalları aracılığıyla hedef kitlelerine doğru mesajları iletmeye çalışırlar. Bu kanallar arasında reklamcılık, halkla ilişkiler, dijital pazarlama, sosyal medya ve etkinlikler gibi yöntemler yer alır. Pazarlama yönetimi, müşteri ilişkilerini güçlendirmeyi ve müşteri memnuniyetini artırmayı da kapsar. Müşteri geri bildirimlerini toplamak, müşteri ile etkileşimleri yönetmek ve müşteri sadakatini sağlamak için CRM sistemleri kullanılır. Ayrıca, pazarlama yönetimi, pazarlama stratejilerini sürekli olarak izler, performansı değerlendirir ve gerektiğinde stratejik revizyonlar yaparak işletmenin hedeflerine ulaşmasına yardımcı olur. Bu yaklaşım sayesinde, pazarlama yönetimi, işletmelerin pazarlama faaliyetlerini sistemli ve etkili bir şekilde yönetmelerine olanak tanır, böylece rekabet avantajı kazanarak uzun vadeli başarıya ulaşmalarına destek olur (Tenekecioğlu, 2007).

CRM kampanya yönetimi, işletmelerin müşteri etkileşimlerini planlama, uygulama, izleme ve değerlendirme sürecini kapsayan bir stratejidir. CRM sistemleri, pazarlama departmanları için kritik bir araçtır. Çünkü bu sistemler doğru stratejilerle kişiselleştirilmiş kampanyalar oluşturmayı mümkün kılar. CRM kampanya

yönetiminde hedef belirleme, hedef kitle tanımlama, kampanya planlama, kampanya uygulaması, izleme ve analiz adımlarını içerir. Hedef belirleme adımı kampanyanın amacı net bir şekilde tanımlanır, bu hedefler genellikle satışları artırma, yeni müşteri kazanma veya mevcut müşteri sadakatini güçlendirme gibi unsurları içerir. İkinci olarak hedef kitle tanımlama kampanyanın ulaşmak istediği müşteri gruplarını belirler. CRM verilerini kullanarak demografik bilgiler, geçmiş satın alma alışkanlıkları ve müşteri tercihlerine dayalı olarak hedef kitle segmentasyonu yapılır. Üçüncü adımda ise kampanya planlama, kampanya içeriği, iletişim stratejisi ve kanal seçimleri belirlenir. Müşterilere hangi mesajların, hangi kanallar aracılığıyla ve hangi zaman diliminde iletileceği planlanır. Dördüncü adım da kampanya uygulaması doğrultusunda kampanya başlatılır. CRM sistemleri, kampanya mesajlarının otomatik olarak gönderilmesini ve müşteri etkileşimlerinin etkili bir şekilde yönetilmesini sağlar. Beşinci adım da ise kampanya süresince müşteri etkileşimleri, geri dönüşler ve performans düzenli olarak izlenir. CRM verileri kullanılarak kampanyanın başarısı değerlendirilir ve gerekli revizyonlar yapılır. CRM kampanya yönetimi, işletmelerin pazarlama stratejilerini müşteri odaklı ve veri destekli bir şekilde yönlendirmelerine yardımcı olur. Kişiselleştirilmiş kampanyaların doğru zamanlamayla iletilmesi, müşteri memnuniyetini artırır, satışları teşvik eder ve müşteri sadakatini güçlendirir. Bu nedenle, CRM kampanya yönetimi, işletmelere önemli bir rekabet avantajı sağlayan stratejik bir araçtır (Sezer vd, 2022).

Yapay zekâ, doğru stratejiler ve gerekli becerilerle tüketici etkileşimini dönüştürmekte ve tüketici alışkanlıklarının, tutumlarının ve davranışlarının tespitiyle doğru kampanyaların oluşturulmasına imkân tanımaktadır (Verma vd., 2021: 1-2). Ayrıca, yapay zekâ pazarlamayı daha veriye dayalı, gerçek zamanlı, kişiselleştirilmiş ve öngörülü hale getirirken, performansını da artırmaktadır (Nair ve Grupta, 2021). Kampanya yönetiminde yapay zekâ uygulamaları; kişiselleştirme, gerçekçi tahminlerde bulunma, yeniden hedefleme, yeniden pazarlama, programatik reklamcılık, bilgisayarlı görme, sosyal medya analizleri, pazar bölümlendirme, yerelleştirme ve görüntü tanıma gibi birçok alanda kullanılmaktadır (Dilmegani, 2022).

İKİNCİ BÖLÜM

YAPAY ZEKA VE MAKİNE ÖĞRENMESİ

2.1. YAPAY ZEKA

Günümüzün dijital çağında, veri, bilgi ekonomisinin temel taşı haline gelmiştir. Şirketler ve veriye ihtiyaç duyan diğer kurumlar dahil olmak üzere pek çok alanda doğru kararlar almak için verilerden yararlanılır. Her geçen gün artan veri hacmi, karmaşıklığı ve çeşitliliği, geleneksel veri işleme yöntemlerinin yetersiz kalmasına neden olmuş ve bu nedenle büyük veri kavramının ortaya çıkmasını zorunlu kılmıştır. Büyük veri, sadece veri miktarının büyüklüğünü ifade etmekle kalmaz, aynı zamanda bu verilerin hızlı bir şekilde işlenmesi ve anlamlı bilgilere dönüştürülmesi gerekliliğini de beraberinde getirir. Bu noktada, yapay zeka ve makine öğrenmesi gibi ileri düzey teknolojiler devreye girer. Bu teknolojiler, büyük veri analitiği ile entegre edilerek, daha önce mümkün olmayan öngörüler ve karar destek sistemleri oluşturulmasını sağlar. Oluşturulan bu karar destek sistemleri günümüz dünyasında hemen hemen her sektörde kullanılabilir hale gelmeye başlamıştır.

Veri, dijital dünyada toplanan ve işlenen her türlü bilgi parçasını ifade eder. Bu veriler, sosyal medya etkileşimlerinden, e-ticaret işlemlerine, sağlık kayıtlarından, bilimsel araştırma sonuçlarına kadar geniş bir yelpazeye yayılmaktadır. Ses verileri, görüntü verileri ve tüm meta veriler büyük veri dediğimiz kavramın içerisinde yer alır. Veri çeşitliliği, yapılandırılmış ve yapılandırılmamış veri formatları dahil olmak üzere çeşitli formlarda olabilir. Veri tabanlarında saklanan bu veriler bir araya gelerek analiz edildiklerinde amaca yönelik değerli bilgiler taşımaktadır. Veri analizinde kullanılan teknikler ve yöntemler, elde edilen bilgilerin doğruluğu ve güvenilirliği açısından büyük önem taşır. Doğru ve etkili veri analizinin yapılabilmesi, ham verinin anlamlı bilgiye dönüştürülmesi sürecinde kritik bir rol oynar.

Büyük veri ise, veri hacminin, hızının ve çeşitliliğinin artışıyla birlikte ortaya çıkan ve geleneksel veri işleme yöntemleriyle yönetilemeyen veri kümelerini ifade eder. Büyük veri, genellikle "3V" olarak adlandırılan hacim (volume), hız (velocity) ve çeşitlilik (variety) özelliklerine sahiptir. Hacim, büyük miktarda verinin depolanması ve işlenmesi gerekliliğini ifade ederken; hız, verinin gerçek zamanlı

olarak işlenmesi ve analiz edilmesi gereksinimini vurgular. Çeşitlilik ise, farklı kaynaklardan ve formatlardan gelen verilerin entegrasyonunu ve analizini ifade eder. Büyük veri teknolojileri, bu zorlukların üstesinden gelmek için geliştirilmiş olup, verinin anlamlandırılması ve işlenmesi sürecinde büyük kolaylıklar sağlar.

Esasen “11V” olarak nitelendirilen büyük veri kavramı en genel tabiriyle “3V” olarak ifade edilir. “3V” kavramının dışında kalan özellikler ise, Doğrulama (Verification), Değer (Value), Gerçeklik (Veracity), Oynaklık (Volatility), Geçerlik (Validity), Hassaslık (Vulnerability), Değişkenlik (Variability) ve Görselleştirme (Visualization) ‘nin birleşimidir (Karakuş, 2023: 16).

Yapay zeka, makinelerin insan benzeri zeka sergilemesini amaçlayan bir bilim dalıdır. Yapay zeka, geniş bir yelpazede uygulamalara sahiptir ve büyük veri ile birlikte çalışarak, karmaşık veri kümelerinden anlamlı öngörüler elde edilmesine yardımcı olur. Yapay zeka sistemleri, büyük veri analitiği ile beslenerek daha doğru ve etkili sonuçlar üretebilir. Makine öğrenmesi, yapay zekanın bir alt dalı olup, algoritmaların veriden öğrenerek performansını sürekli olarak iyileştirmesini sağlar. Bu süreçte, büyük veri setleri kullanılarak, modellerin doğruluğu ve genel performansı artırılabilir.

Yani veri, büyük veri ve yapay zeka kavramları, modern bilgi işlem dünyasında birbirleriyle yakından ilişkilidir ve birlikte kullanıldıklarında büyük potansiyele sahiptir.

2.2. YAPAY ZEKANIN TARİHSEL GELİŞİMİ

Son zamanlarda popüler hale gelen yapay zeka kavramı yeni bir kavram gibi görünse de aslında 1642 yılında ilk tohumu atılmış olup günümüz dünyasında hızlı bir ivmeyle yol kat etmiştir. Alan Turing’in makinelerin düşünme yeteneklerini test etmek amacıyla önerdiği Turing Testinden başlayarak Cahit Arf’ın “Makineler Düşünebilir Mi?” makalesinin yayınlanması, ilk endüstriyel robotun yapımı ve dünya şampiyonlarını satranç turnuvasında yenen akıllı yapay zeka sistemlerinin insan zekasını örnek alarak başarısını ispat edip gelişimine devam etmiştir. Bugün tüm bu gelişmeleri yakından takip ederken uzun yıllar önce başlayan yapay zeka geliştirmeleri günümüze kadar uzanan bir yolculuk içerisinde kendini kanıtlamıştır. (Pirim,2006:88)

Tablo 2: Yapay Zekanın Tarihsel Gelişimi

Yıl	Gelişme	Açıklama
1642	Pascaline	Blaise Pascal, mekanik hesap makinesi olan Pascaline'i icat etti. Bu, otomatik hesaplama cihazlarının öncüsü olarak kabul edilir.
1846	Analitik Makine	Charles Babbage, modern bilgisayarların temelini oluşturan Analitik Makine'nin tasarımını yaptı. Ancak, bu makine hiçbir zaman tamamlanmadı.
1950	Turing Testi	Isaac Asimov'un "Ben Robot" isimli bilim kurgu romanını yayınlaması ve Alan Turing'in makinelerin düşünme yeteneklerini ölçmek için Turing Testi'ni önermesi. Bu test, bir insan hakemin, makine ve insan arasındaki farkı ayırt edememesi durumunda, makinenin 'düşünebildiği' varsayımına dayanır.
1951	İlk Yapay Sinir Ağı	Manchester Üniversitesi'nin Ferranti Mark 1 makinesini kullanan Christopher Strachey bir dama programı yazdı. Aynı yıl Dietrich Prinz de bir satranç programı yazdı. Marvin Minsky ve Dean Edmonds tarafından bilgileri öğrenebilen ilk yapay sinir ağı olan SNARC makinesi yapıldı. Bu ağ beynin öğrenme sürecini simüle eden bir ağıdır.
1956	Dartmouth Konferansı	John McCarthy, Marvin Minsky, Nathaniel Rochester ve Claude Shannon tarafından düzenlenen konferans, yapay zekanın resmi doğuşu olarak kabul edilir. "Yapay zekâ" terimi burada tanıtıldı ve yapay zekâ araştırmalarının temelleri atıldı.
1958	LISP Programlama Dili	John McCarthy tarafından geliştirilen LISP, yapay zekâ araştırmaları için en yaygın kullanılan programlama dillerinden biri haline geldi.
1959	Makineler Düşünebilir Mi?	Cahit Arf'ın "Makineler Düşünebilir mi ve Nasıl Düşünebilir?" adlı makalesi yayınlandı.
1961	İlk Endüstriyel Robot	Unimate adlı robot, General Motors fabrikasında çalışmaya başladı. Bu, endüstriyel otomasyonun başlangıcı olarak kabul edilir.
1963	Proje MAC	Karar teorisi ve makine öğrenmesi gibi yapay zekâ teknolojilerini geliştirecek ve sonrasında MIT Yapay Zeka Laboratuvarı olacak Mac projesi başlatıldı.
1965	ELIZA Programı	Joseph Weizenbaum tarafından geliştirilen ELIZA, insan konuşmasını taklit eden ilk programlardan biridir. Temel metin işleme ve basit kalıplar kullanarak kullanıcılara psikoterapist rolü oynayan bir sistem sunmuştur.
1969	Shakey Robotu	SRI International tarafından geliştirilen Shakey, algı, hareket ve problem çözme yeteneklerine sahip ilk mobil robot olarak kabul edilir. Shakey, bir ortamda gezinme ve basit görevleri yerine getirme yeteneklerine sahipti.
1972	PROLOG Programlama Dili	Alain Colmerauer ve Robert Kowalski tarafından geliştirilen PROLOG, mantık programlama için kullanılır ve özellikle yapay zekâ ve doğal dil işleme alanlarında yaygın olarak kullanılır.
1980	Uzman Sistemler	Uzman sistemler, belirli bir alanda uzman bilgilerini kullanarak kararlar veren yapay zekâ sistemleridir. MYCIN ve DENDRAL gibi sistemler, tıp ve kimya gibi alanlarda başarılı uygulamalar sundu.
1986	Backpropagation Algoritması	Geoffrey Hinton ve David Rumelhart, yapay sinir ağlarının eğitiminde kullanılan backpropagation algoritmasını tanıttı. Bu algoritma, sinir ağlarının daha etkili öğrenmesini sağladı.
1987	Yapay Sinir Ağları	Yapay sinir ağlarının yeniden keşfi ve popülerleşmesi, yapay zekâ araştırmalarında önemli bir dönüm noktası oldu. Bu ağlar, insan beyninin çalışma şeklini taklit eden ve öğrenme kapasitesine sahip sistemlerdir.
1997	Deep Blue	IBM'in Deep Blue adlı bilgisayarı, dünya satranç şampiyonu Garry Kasparov'u yenerek yapay zekanın karmaşık oyunlarda insanları yenebileceğini gösterdi. Bu olay, yapay zekâ'nın stratejik düşünme yeteneklerinin bir kanıtı olarak görülür.
2002	Roomba Robotu	iRobot şirketi, ev içi temizlikte kullanan ilk ticari robot olan Roomba'yı tanıttı. Bu robot, çevresini algılayarak ve otonom hareket ederek temizlik yapabiliyordu.
2009	Google Sürücüsüz Araç	Google, sürücüsüz araç projesini duyurdu ve test etmeye başladı. Bu araçlar, yapay zekâ ve makine öğrenmesi kullanarak otonom sürüş yeteneklerine sahipti.
2011	IBM Watson	IBM'in Watson süper bilgisayarı, "Jeopardy!" adlı bilgi yarışmasında insan şampiyonları yenerek büyük bir başarı elde etti. Watson, doğal dil işleme ve büyük veri analizi yeteneklerini sergiledi.

2012	Derin Öğrenme Atılımı	Geoffrey Hinton ve ekibi, derin öğrenme algoritmalarının ImageNet yarışmasında büyük başarı elde etmesiyle yapay zekâ araştırmalarında yeni bir dönemi başlattı. Derin öğrenme, özellikle görüntü ve ses tanıma gibi alanlarda büyük ilerlemeler sağladı.
2014	Chat botların Yükselişi	Facebook ve diğer sosyal medya platformları, kullanıcılarla etkileşim kuran ve müşteri hizmetleri sağlayan Chat botları tanıttı. Bu Chat botlar, doğal dil işleme ve makine öğrenmesi kullanarak kullanıcı taleplerine cevap verir.
2015	Alpha Go	Google DeepMind'in Alpha Go programı, dünya Go şampiyonu Lee Sedol'u yenerek yapay zekanın karmaşık strateji oyunlarında da üstün olabileceğini kanıtladı. Go, hesaplaması zor stratejik derinliği nedeniyle satrançtan daha karmaşık olarak kabul edilir.
2016	OpenAI Kuruluşu ve İnsansı Robot	Elon Musk, Sam Altman ve diğerleri tarafından kurulan OpenAI, yapay zekanın güvenli ve faydalı bir şekilde geliştirilmesini ve kullanılmasını teşvik etmek amacıyla kuruldu. Aynı zamanda ilk insansı robot olan Sophia'ya vatandaşlık verildi.
2017	Otomom Araba Üretimi ve Chat Botların Konuşması	Tesla tarafından otonom arabalar üretildi. Aynı yıl içerisinde iki sohbet botu aralarında konuşmaya başladı.
2018	BERT Modeli	Google, doğal dil işleme alanında büyük bir atılım olan BERT (Bidirectional Encoder Representations from Transformers) modelini tanıttı. BERT, cümlelerin anlamını daha iyi kavrayarak arama motoru sonuçlarını iyileştirdi.
2020	GPT-3	OpenAI, dil işleme yetenekleriyle dikkat çeken ve geniş bir uygulama yelpazesi sunan GPT-3 modelini tanıttı. GPT-3, metin oluşturma, çeviri ve çeşitli dil görevlerinde yüksek performans gösterdi.
2021	DALL-E ve CLIP	OpenAI, metin girdilerinden görüntüler üretebilen DALL-E ve görsel-metin ilişkilerini anlamada güçlü olan CLIP modellerini tanıttı. Bu modeller, yapay zekâ'nın yaratıcı ve anlayış yeteneklerinde önemli ilerlemeler sağladı.
2022	ChatGPT Tanıtımı	OpenAI, GPT-3.5 tabanlı dil modeli olan ChatGPT'yi piyasaya sürdü. ChatGPT, metin tabanlı diyaloglarda insan benzeri yanıtlar verebilme yeteneği ile dikkat çekti ve geniş bir kullanıcı kitlesi tarafından benimsendi.
2023	GPT-4 ve Google Bard	OpenAI, GPT-4 modelini tanıttı. GPT-4, daha iyi dil anlama ve üretme yeteneklerine sahipti, daha fazla parametre içeriyordu ve daha geniş bir bilgi tabanı üzerinde eğitilmişti. Google, kendi dil modeli olan Bard'ı tanıttı. Bard, doğal dil işleme yetenekleriyle dikkat çekti ve Google'ın arama motoru sonuçlarını iyileştirme amacı taşıdı.
2024	Yapay Zekâ ve Sağlık	Yapay zekâ, sağlık sektöründe büyük ilerlemeler kaydetti. Özellikle kanser teşhisi ve tedavisinde, genomik analizlerde ve kişiselleştirilmiş tedavi planlarının oluşturulmasında önemli roller üstlendi.
2024	Etik ve Düzenlemeler	Yapay Zekanın kullanımı ve gelişimiyle ilgili etik ve düzenleyici çerçeveler oluşturulmaya başlandı. Birçok ülke, yapay zekanın güvenli ve etik kullanımı için yasalar ve yönetmelikler çıkardı.

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Yapay zeka bilim ve teknolojinin en hızlı gelişen alanlarından biri haline gelmiştir. Bu tarihsel sürecinde insan yeteneklerini aşan sistemler geliştirmeyi hedeflemiştir. Alt kategorilerinden biri olan makine öğrenmesi ile verilerden öğrenme ve model oluşturma becerisiyle oldukça parlamıştır. Makine öğrenmesi yapay zekanın gelişimin de kilit bir rol üstlenerek sistemlerin verilere uyum sağlamasını ve zaman içerisinde performanslarını iyileştirmesine yardımcı olur. Derin öğrenme gibi daha birçok alanıyla gelişimini sürdürmeye devam etmektedir. (Dhar,2016)

2.3. YAPAY ZEKANIN KULLANIM ALANLARI VE ALT DALLARI

2.3.1. Kullanım Alanları

Günümüzde bilgisayar teknolojilerindeki hızlı ilerlemeler, yapay zekanın kullanımını arttırmış ve uygulama alanlarını genişletip çeşitlendirmiştir. Yapay zeka problem çözme ve öğrenme gibi insanın bilişsel yeteneklerini taklit etme kapasitesiyle dikkat çeken birçok kavramı bünyesinde barındırır. İnsan gücüyle yapılması güç bazı çalışmaların birkaç dakika içerisinde bitirilmesini sağlar. Sağlık, finans, otomotiv, e-ticaret ve pazarlama gibi çeşitli alanlarda devrim niteliğinde uygulamalara sahiptir (LeCun vd., 2015: 16).

2.3.1.1. Sağlık

Sağlık sektöründe tedavi süreçlerinin iyileşmesinde çok önemli bir paya sahiptir. Tıbbi görüntü analizi, hastalık teşhisi, ilaç keşifi, ve hasta takibi gibi uygulamalarda kullanılmaktadır. Yapay zeka destekli görüntüleme sistemlerinde kanser gibi zührevi hastalıkların erken teşhisini yüksek doğruluk oranıyla sağlanmaktadır. (European Commission, 2018).

2.3.1.2. Finans

Finans sektöründe dolandırıcılık tespiti (fraud), kredi riski değerlendirmesi, piyasa trendlerini takip etme, portföy yönetimi optimizasyonu ve hatta müşteri hizmetleri alanlarında kullanılarak kurumların verimliliğini arttırabilir. Büyük veri setlerini analiz ederek anormallikleri tespit eder ve finansal kararların daha doğru alınmasına yardımcı olur. (European Commission, 2018).

2.3.1.3. Eğitim

Eğitim sektörün de kişiselleştirilmiş öğrenme deneyimi sunmak için kullanılmaktadır. Akıllı eğitim sistemleri, öğrencilerin öğrenme hızlarını analiz ederek

bireysel ders çalışma planları oluşturur. (European Commission, 2018).

2.3.1.4. Otomotiv

Otomotiv sektöründeki otonom araçlar yapay zekanın otomotiv sektöründeki en dikkat çekici uygulamasıdır. Otonom araçlar çevresel verilerin analizini yaparak güvenli bir şekilde sürücüsüz yol alır. İnsansız hareket edebilen bu araçlar trafik kazalarını azaltmayı hedefler. (European Commission, 2018).

2.3.1.5. E-ticaret ve Pazarlama

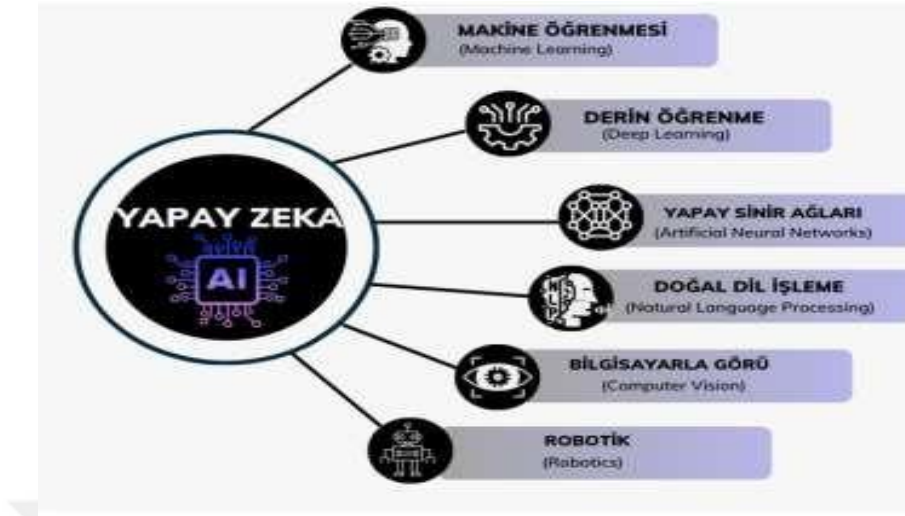
E-ticaret ve pazarlama sektöründe müşteri deneyimini geliştirmek için son derece gelişmiş uygulamalar bulunmaktadır. Kişiselleştirilmiş öneri sistemleri, müşterilerin ilgisini çekebilecek ürünleri önermek amacıyla kullanıcı verilerini analiz eder. Müşteri davranışını tahmin etmeye ve doğru fiyatlandırma politikası belirlemeye yardımcı olur. Ayrıca reklam ve kampanya yönetimi süreçlerinde müşterilerin talebine, stok durumuna ve müşterilerin satın alma davranışlarına göre fiyatı dinamik şekilde ayarlamak için yapay zeka kullanılır. (European Commission, 2018).

Sağlıktan finansa, eğitimden otomotive kadar geniş bir yelpazede kullanım alanı bulan yapay zeka aynı zamanda makine öğrenimi, derin öğrenme, doğal dil işleme, bilgisayarla görü ve robotik gibi alt dallarla çeşitli teknolojik çözümler sunmaktadır.

2.3.2. Alt Dalları

Yapay zeka makine öğrenmesi, derin öğrenme, yapay sinir ağları, doğal dil işleme, bilgisayarlarla görü ve robotik gibi alt dalları içerir.

Şekil 5: Yapay Zekanın Alt Dalları



Kaynak: Araştırmacı tarafından oluşturulmuştur.

2.3.2.1. Makine Öğrenmesi (Machine Learning)

Makine öğrenimi, bilgisayarların veriler aracılığıyla örüntüleri öğrenmesini ve bu örüntülerden yararlı bilgiler çıkarmasını sağlayan bir yapay zeka dalıdır. Diğer bir deyişle makine öğrenmesi, insan tarafından programlanmaya ihtiyaç duymadan, veriyi analiz ederek öğrenip, bunun üzerinden örüntüler tespit edip, çıkarım yapabilen algoritmaları çalışan yapay zekanın bir uygulama alanıdır (Güngör ve Yardımcı, 2019: 89). Makine öğrenmesi yapılandırılmış veri setleri üzerinde çalışır (Lindgren, 2020: 8). Yapılandırılmış veri seti, belirli bir mantıksal formata ve düzene göre organize edilmiş ilişkisel veri tabanlarında tablolar da sütun ve satır gibi yapılarla saklanan verilerdir.

2.3.2.2. Derin Öğrenme (Deep Learning)

Derin Öğrenme, makine öğrenmesinin bir alt dalı olup, çok katmanlı yapay sinir ağları kullanarak veri özelliklerini öğrenir. Diğer ifadeyle yapay zeka çatısı altındaki makine öğrenmesi algoritmaları sınıfıdır. Görüntü ve ses tanıma, doğal dil işleme, oyun oynama ve hatta sürücüsüz araç kullanma gibi karmaşık görevler de yüksek başarı gösterir. Derin öğrenme algoritmaları, büyük veri setleri üzerinde çalışarak insan benzeri sonuçlar elde edebilir. Esasen bu başarılı sonuçları elde etmek

için yine makine öğrenmesini kullanır. Makine öğrenmesinden derin öğrenmeyi ayıran en temel fark derin öğrenmenin yapılandırılmamış veri setleri üzerinde de çalışabilme yeteneğidir (Lindgren ,2020: 13). Yapılandırılmamış veri seti, belirli bir mantıksal formata veya yapıya göre organize edilmemiş dolayısıyla ilişkisel veri tabanlarında tablolar halinde tutulamayan ve kolayca işlenemeyen veri türleridir. Bu veriler genellikle metin belgeleri, videolar, ses kayıtları, sosyal medya verileri ve e-postalar gibi çeşitli kaynaklardan elde edilen verilerden oluşur.

2.3.2.3. Yapay Sinir Ağları (Artificial Neural Networks)

Yapay sinir ağları, insan beyninin öğrenme yeteneğini taklit eden ve örnekler aracılığıyla öğrenen derin öğrenme modellerinin temelini oluşturan yapay zeka sistemleridir. Aslında yapay zekanın kullandığı teknik ve yöntemler arasında olan bu yöntem çok başarılı sonuçlar verdiği için bazı kaynaklarda yapay zekanın bir alt dalı olarak geçmektedir. Geleneksel hesaplama yöntemlerinden farklı olarak eksik verilerle çalışabilen, belirsiz durumlar karşısında karar verebilen, hatalara karşı dayanıklı ve çevresine uyum sağlayabilen yapay sinir ağları çok çeşitli alanlarda uygulamalar sunar. İnsan beyninin çalışmasında etkili olan nöronların birbiri ile kurdukları bilgi aktarımından etkilenerek geliştirilen yapay sinir ağları girdi, ara ve çıktı katmanından oluşarak bilgiyi iletir. (Özen, 2022; Öztemel, 2003)

2.3.2.4. Doğal Dil İşleme (Natural Language Processing)

Doğal dil işleme, bilgisayarların insan dilini anlaması, yorumlaması ve üretmesi ile ilgilenen yapay zeka alt dalıdır. Elli yılı aşkın bir süre önce geliştirilmeye başlayan ve temelinde derin öğrenmeyi kullanan doğal dil işleme tekniği yapay zekanın en zorlu çalışmalarını içerir. Doğal dil işleme derin öğrenmeyi kullanarak insan ile makine arasındaki iletişimi kolaylaştırır. Sohbet botları, karmaşık bir yazının otomatik şekilde kısaltılması, duygu analizi, dil çevirisi, konuşma tanıma, bilgi ve belge yönetimi gibi çeşitli alanlarda kendine yer bulmuştur. Ancak dil sürekli gelişen ve değişen bir yapıya sahip olduğundan çalışmalar geçerliliğini yitirir. Konuşma içerisinde belirsiz ifadeler, ses tonu ve vurgunun her aksan için farklı olması, dil kurallarındaki değişimler doğal dil işleme alanındaki çalışmaları doğrudan etkilediği

için henüz mükemmel düzeyde çözülebilmüş değildir. Geliştirilmeye devam eden açık yönü olmasına karşın çok başarılı çalışmalar da mevcuttur (Lutkevich, 2023).

2.3.2.5. Bilgisayarla Görü (Computer Vision)

Bilgisayarla görü, görsel verileri analiz ederek anlam çıkarma yeteneği sağlayan bir yapay zeka alt dalıdır. Bilgisayar görüntülerinin insan gözünden algılanan görüntüyle birebir görülmesine, tanınmasına ve işlenmesine olanak tanır. Yapay görme sistemlerin de kullanılan teknikleri içerir. Nesne tanıma, yüz tanıma, hareket takibi ve görüntü segmentasyonu gibi uygulamalar bu alanda yer alır. (Aylak vd, 2020: 16). Bilgisayarlarla görü güvenlik, otomotiv ve sağlık alanlarında yaygın olarak kullanılmaktadır.

2.3.2.6. Robotik

Robotik, fiziksel görevleri gerçekleştirebilen akıllı makinelerin tasarımı ve geliştirilmesi ile ilgilenir. Endüstriyel otomasyon, cerrahi robotlar, insansız hava araçları ve hizmet robotları gibi çeşitli uygulamaları kapsar. Yapay zeka, robotların çevrelerini algılaması, karar vermesi ve otonom olarak hareket etmesini sağlar. (Gedik, 2020: 26)

2.4. YAPAY ZEKA VE MAKİNE ÖĞRENMESİ İLİŞKİSİ

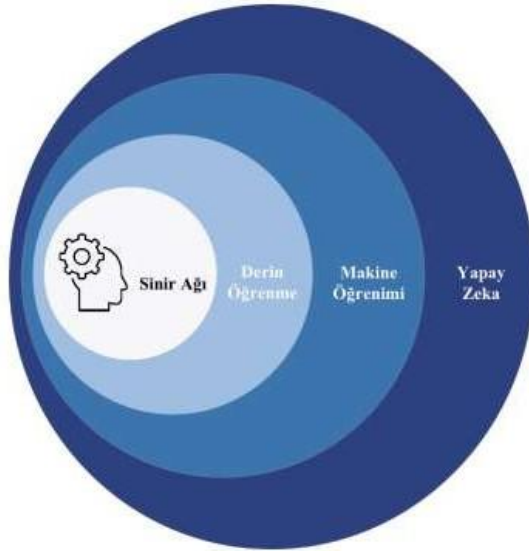
Profesör Andrew Moore tarafından yapay zeka makinelerin yakın bir gelecekte insan zekası gibi davranış geliştirmesini sağlayan mühendislik bilimi olarak tanımlanır. Profesör Tom Mitchell tarafından makine öğrenmesi bilgisayarların kendi deneyimleri aracılığıyla verilerden otomatik olarak öğrenmesini sağlayan algoritmaların çalışmasına olanak tanıyan yapay zekanın bilimsel bir dalı olarak tanımlanır.

Yapay zeka, makine öğrenmesini kullanmadan da karar verme ve öngörü yapabilme yeteneğine sahip olan teknolojilerin en genel adıdır. Makine öğrenmesi algoritmaları ortaya çıkana kadar ki süreçte bu çalışmalar yazılımcıların matematiksel ve mantıksal işlemleri kodlayarak yaptıkları bir yapıyla gerçekleşmekteydi. “Hard-coded” olarak nitelendirilen bu çalışmalar zaman içerisinde yerini verilerden öğrenen

makine öğrenmesi algoritmalarına bırakmaya başladı. Kendini kanıtlayan makine öğrenmesi algoritmaları yapay zekanın “hard-coded” olarak isimlendirilen sembolik yapay zeka türünü hızla geride bırakarak kendine büyük ölçüde literatürde yer bulmuştur.

Makine öğrenmesi algoritmalarını kullanarak örüntüleri yakalayabilen derin öğrenme modeli veriye dayalı öğrenme gerçekleştirebilmekle birlikte makine öğrenmesi algoritmalarındaki gibi tek bir matematiksel modelle sınırlı kalmayıp yapay sinir ağları aracılığıyla ağ diyagramlarına benzeyen karmaşık hesaplamalarla çalışmaktadır. Derin öğrenme algoritmalarının tümü makine öğrenmesi algoritmasıdır. Çünkü temelinde veriye dayalı öğrenme biçimini esas alır. Fakat tüm makine öğrenmesi algoritmaları derin öğrenme algoritması değildir. Büyük veri kümeleriyle yüksek bilgi işleme kapasitesine sahip yapay sinir ağlarının güçlü matematiksel modeller üretebildiği derin öğrenme, makine öğrenimi alt kümesidir. Derin öğrenme verilerin arasındaki örüntüyü ve parametrelerin ağırlığını kendisi bulabilirken makine öğrenmesi algoritmalarında “feature engineering” adı verilen bu aşama mevcuttur (Karakuş, 2023: 25-30).

Şekil 6: Makine Öğrenmesi ve Yapay Zeka İlişkisi



Kaynak: Gökalp (2022) çalışmasından alınmıştır.

Yapay zeka, makine öğrenmesi ve derin öğrenmenin birbirinin kapsayan alt kümeler olduğunu yukarıdaki şekilde açıkça görünmektedir. Bu süreçte karıştırılmaması gereken şey yapay zekanın kendi başına kararlar verebilen ve

verdikleri kararlar doğrultusunda insanlar gibi yeni şeyler öğrenebilen bir süreç olduğudur. Makine öğrenmesine ihtiyaç duymadan da çalışabilirler. Makine öğrenmesi ise veriler aracılığıyla çıkarımlarda bulunup yeni şeyler öğrenebilen bir sistemdir. Algoritmaların programlanması yerine kalıpların tanımlanarak etiket verileri ile beslenebilir (Gökalp, 2022: 3). Dolayısıyla veriler ne kadar çok olursa makine öğreniminin öğrenme süreci o kadar doğru olacaktır. Kısaca ifade etmek gerekirse yapay zekanın alanı makine öğrenmesiyle birlikte daha da kapsamlı hale gelmiştir. (SAP, 2021).

2.5. MAKİNE ÖĞRENMESİ

Arthur Samuel 1959 yılında makine öğrenmesini bilgisayarlara açıkça programlama yapmadan veriler aracılığıyla öğrenme yeteneği kazandıran bir çalışma alanı olarak tanımlamıştır.

Tom Mitchell 1997 yılında mühendislik odaklı bir tanım yaparak makine öğrenmesini, belli bir T görevinde P performans ölçütünün, E tecrübeleriyle artması durumunda bunun E tecrübelerinden öğrenen bir bilgisayar programı olduğunu söyleyen bir tanım yapmıştır. (Geron, 2022:3).

Öğrenme, bireyin gözlem ya da eğitim yoluyla bilgi ve beceri kazanması bu edinilen bilgileri gelecekteki bir durum ya da görevde kullanma yeteneğidir. Bilgi ve deneyimlerin kalıcı değişiklikler yaratmasına dayanır. Dolayısıyla önceki bilgilerin ve deneyimlerin yeni bağlamlarda uygulanmasına olanak tanır. Hem biyolojik hem de yapay sistemler için hayati öneme sahiptir. Makine öğrenmesi ise yapay öğrenme gerçekleştiren makinelerin verilerden öğrenerek değişme yeteneğini arttıran tekniklerin araştırılıp geliştirildiği mühendislik alanıdır. Amacı, verilerden yararlanarak geleceğe ışık tutup başarılı ve doğru kararlar vermesini sağlayan matematiksel modelleri geliştirmektir (Bonaccorso, 2017: 8).

Makine öğrenmesinde mevcut veriler algoritmaya öğretilir. Bu öğretilen veriler aracılığıyla algoritmanın genellemeler yapması beklenir. Bu süreçte, algoritmanın öğrenme yeteneğini geliştirmek ve henüz görülmemiş veriler karşısında doğru tahminler yapmasını sağlamak esastır. Ayrıca, tahminleme yapılırken modelin hata oranını minimum düzeyde tutarak yüksek doğruluk elde etmek amaçlanır. Bu

durum istatistiksel olarak modelin güvenilir ve verimli olmasını sağlar (Zhou, 2021).

Makine öğrenmesi multidisipliner bir alandır. Bu alanlar arasında bilgisayar bilimi, istatistik, matematik, bilgi teorisi ve mühendislik bulunmaktadır. Bu disiplinlerin her birinin makine öğrenmesinin geliştirilmesinde ve uygulanmasında önemli katkıları bulunur. Bilgisayar bilimi ve mühendislik, algoritmaların geliştirilmesine katkı sağlarken istatistik ve matematik, veri analizi ve modelleme de kullanılır. Bilgi teorisi ise insanın öğrenme süreçlerinin ve bilgilerin nasıl işlendiğinin anlaşılmasına katkıda bulunabilir (Ramu vd., 2022)

Şekil 7: Klasik Programlama ve Makine Öğrenmesi



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Klasik programlama ve makine öğrenmesinin nasıl işlediğini gösteren bu görselde yazılımcılar tarafından kurallar ve örüntüler makineye girdi olarak verilip bir çıktı oluştururken makine öğrenmesinde çıktılar girdi olarak verilip kuralların ve örüntülerin algoritma tarafından öğrenilip çıktı olarak verilmesi beklenir. Yapay zeka alanındaki yazılım geliştiriciler farklı çalışmalarda girdilere karşılık çıktıların tahmin edilmesinde, doğrudan programlamanın yerine sistemdeki girdi ve çıktı arasındaki ilişkilerin örnekler aracılığıyla sisteme tanıtılmasının daha etkili olduğunu ileri sürmüşlerdir (Jordan ve Mitchell, 2015: 255). Bu yöntem sistemin örneklerden öğrenerek otomatik olarak bu ilişkileri keşfetmesini ve doğru tahminler yapmasını sağlar. Böylece sistemin hem öğrenme süreci hızlanır hem de esneklik ve doğruluk artar.

Son yıllarda artan veri hacmi ve gelişen teknoloji sayesinde makine öğrenmesine duyulan ihtiyaç artmıştır. Karmaşık ve büyük hacimli olan bu verilerin

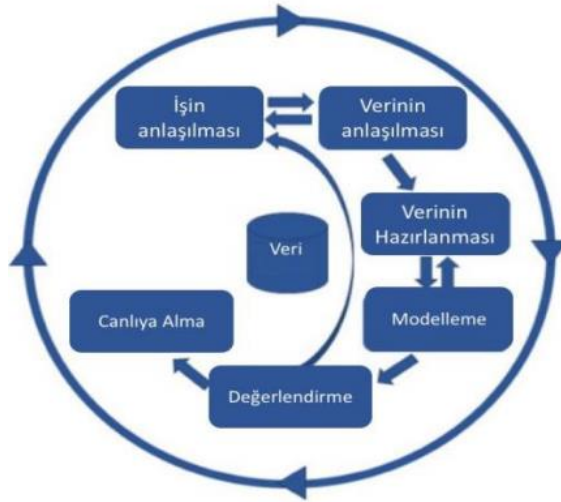
insanlar tarafından yorumlanması ve analiz edilmesi çok güç olduğu için zaman içerisinde popüler hale gelen makine öğrenmesi bu verilerin analizinde insan kapasitesinin çok üzerindedir. Günümüzde sıklıkla rastlayacağımız e-postanın spam olarak atanması, Trendyol'un baktığımız ürünlere benzer ürünleri bize sunması, Netflix ve YouTube'ın daha önce izlediğimiz kategorideki videoları ve filmleri bize önermesi, Amazon'un her müşterisi için ayrı bir fiyatlandırma politikası izlemesi, Google'un arama motoru optimizasyonu SEO kullanması, OpenAI'in ChatGPT dil modelini sunması, hemen hemen her şirketin müşteri hizmetleri için kullandığı chatbotlar, Apple'ın konuşan sohbet robotu Siri'yi kullanması, yaygınlaşan otonom araçlar ve yapılan insansız robotlar makine öğrenmesinin evrişimli sürecini açıkça kanıtlıyor. Hem kullanım açısından kolayca kullanılabilen hem de şirketler açısından zaman ve maliyet avantajı sağlayan bu teknolojik gelişmeler hayatı kolaylaştırmanın yanı sıra günümüz dünyasını çok farklı bir evreye taşımaya devam ediyor.

2.6. MAKİNE ÖĞRENMESİ SÜREÇLERİ

Makine öğrenmesi içerisinde çok fazla terimi barındıran bir kavramdır. Verilerin veri tabanlarından elde edilip modele uygun hale getirilene kadar ki süreç veri madenciliği olarak adlandırılır. Veri madenciliği aşamaları makine öğrenmesi aşamalarına çok benzerdir. Makine öğrenmesi verilerden öğrendiği öngörülere dayanırken veri madenciliği verilerden bilinmeyen özelliklerin keşfine odaklanır. Esasen veri tabanlarında bilgi keşfinin analiz basamağına daha çok odaklıdır. (Karakuş, 2023: 10-12)

Veri tabanlarından elde edilen veriler problem belirlendikten sonra tahminlemeye uygun hale gelmesi açısından veri madenciliği süreçleri ile temizlenir ve modele uygun hale getirilir. Makine öğrenmesini daha iyi kavrayabilmek, verimli ve başarılı bir sonuç elde etmek için makine öğrenmesi süreçlerinin iyi bilinmesi önemlidir.

Şekil 8: CRIPS-DM Metodolojisi



Kaynak: Yaman (2021) çalışmasından alınmıştır.

CRISP-DM (Cross-Industry Standard Process for Data Mining) metodolojisi, veri madenciliği projelerinin planlanması ve geliştirilmesi için kullanılan standart bir süreç modelidir. Bu metodoloji, veri madenciliği sürecindeki karmaşıklıkları düzenlemek ve projelerin başarıyla tamamlanmasını sağlamak amacıyla geliştirilmiştir. Veri madenciliği projelerinde karşılaşılan çeşitli disiplinler, veri çeşitliliği, büyük veri hacmi ve kirli veri kaynakları gibi zorluklar nedeniyle böyle bir standart metodolojiye ihtiyaç duyulmaktadır (Şahin, 2011: 16-18).

CRISP-DM, Daimler Chrysler AG, SPSS, NCR ve OHRA gibi öncü veri madenciliği kullanıcıları ve tedarikçileri tarafından oluşturulmuş ve dünya genelinde geniş bir kullanım alanına sahiptir. Bu metodolojide sırasıyla şu adımlar izlenmektedir: (Yaman, 2021: 31-32)

İşin Anlaşılması (Business Understanding): Projenin hedefleri belirlenir ve iş problemleri detaylı olarak anlaşılır.

Verinin Anlaşılması (Data Understanding): Veriler analiz edilir, kalitesi incelenir ve temel özellikleri anlaşılır.

Verinin Hazırlanması (Data Preprocessing): Veriler temizlenir, dönüştürülür ve analiz için uygun hale getirilir.

Modelleme (Modeling): Uygun makine öğrenmesi modelleri seçilir ve uygulanır.

Değerlendirme (Evaluation): Oluşturulan modellerin performansı değerlendirilir ve iş hedefleriyle uyumu kontrol edilir.

Canlıya Alma (Deployment): Model gerçek hayatta kullanılabilir hale getirilir ve süreç tamamlanır.

Bu adımlar, CRISP-DM'in sistematik yapısını oluşturur ve veri madenciliği projelerinde standart bir rehber görevi görür.

Şekil 9: Makine Öğrenmesi Süreçleri



Kaynak: Erboğa, M., 2023 çalışmasından alınmıştır

Geron (2022) kitabında tüm bu süreçleri ele almaktadır. Bir makine öğrenmesi süreci ilk olarak *problemin belirlenmesi*yle başlar. Belirlenen problemin türü çok önemlidir. Çünkü daha sonrasındaki adımlar belirlenen makine öğrenmesi probleminin türüne göre değişkenlik gösterir. Bu aşamada esasen problemin karmaşıklığı ve algoritmik zorlukları da belirlenir.

Ardından *veri toplama* süreci ile devam eder. Belirlenen probleme yönelik veri kaynakları tespit edilir ve veriler toplanır.

Daha sonra verileri tanımak, analize uygun hale getirmek ve ham veriden bilgi çıkarmak açısından *veri ön işleme* aşamasına geçilir. Verilerde eksik ya da aykırı değerlerin olup olmadığı tespit edilir. Veri setinde eksik değerler var ise ortalama ya da medyan değerleri ile bu eksikler giderilir. Önemsizmeyecek kadar az

eksik gözleme sahip bir veri seti ise eksik değerler silinir. Bu kısım problemi çözen kişinin bilgisine ve inisiyatifine bağlıdır. Veriyi modele uygun bir forma getirerek normalizasyon işlemleri yapılır. Özellik ölçekleme ile veri standart hale getirilir. Burada veri dönüşümleri ve feature engineering adı verilen özellik mühendisliği yapılır. Aslında var olan özelliklerden yeni özellikler türetilir. Veri ön işleme adımı en uzun ve en meşakkatli adımdır. Aynı zamanda modelin performansı açısından oldukça önemlidir. Çünkü model başarısının yüksek olmadığı durumlarda bu aşamalara geri dönerek farklı özellik mühendisliği adımları uygulanır. Sürekli devam eden bu süreç model en yüksek başarıya ulaşana kadar devam eder. Encoding yöntemleri de bu adıma dahildir. Değişkenlere one-hot-encoding ya da label-encoding yöntemlerinden uygun olanı kullanılarak özellikler modelin anlayacağı şekilde kodlanır. “Keşifsel veri analizi (EDA)” adımı daha iyi tanımak için betimsel istatistikleri çıkarılıp görselleştirmeler yapılır. Veri dağılımı incelenerek verinin genel yapısı hakkında fikir sahibi olunur. Korelasyon analizi yapılarak özelliklerin birbiri ile ilişkisi incelenir. Multicollinearity probleminin olup olmadığı incelenerek problemin varlığı durumunda bu adımda problem giderilir. EDA adımı model oluşturma aşamasından önce gerçekleştirilen kritik bir adımdır.

Daha sonrasında *model oluşturma* adımına geçilir. Bu aşama makine öğrenimi süreçlerinde önemli bir yere sahiptir. Çünkü bu aşamada seçilen algoritma veriye uygulanarak ilgili problemin çözüm süreci başlar. Model oluşturmak için belirlenen problemin türüne göre algoritma seçilir ve model eğitilir. Veri seti eğitim ve test veri seti olmak üzere ikiye ayrılır. Genellikle verilerin yaklaşık %70-80’i eğitim için %20-30’u test için ayrılır. Veri seti, eğitim ve test veri seti olmak üzere ikiye ayrılır. Ayrılan eğitim seti kullanılarak model eğitilir. Eğitim ve test veri setlerinin dışında bir doğrulama seti oluşturulabilir. Doğrulama seti, modelin hiperparametrelerini ayarlamak ve aşırı uyum (overfitting) sorunlarını tespit etmek için kullanılır. Doğrulama seti kullanmak, modelin daha sağlam ve genellenebilir olmasını sağlayan önemli bir adımdır, ancak doğrulama seti olmadan da başarılı sonuçlar elde edilebilir. Doğrulama setinin eklenmesi genellikle modelin genel doğruluğunu artırmak için faydalıdır. Test veri seti üzerinde ise modelin gerçek performansı değerlendirilir. Gerekli hiperparametre ayarlamaları yapılır. Grid Search ve Random Search gibi yöntemler yardımıyla en iyi hiperparametre kombinasyonları

bulunur. Hiperparametre ayarları yapıldığında modelin verileri ezberlemesi veya öğrenememesi problemleri minimize edilmiş olur. Bu aşamada dengesiz veri problemi varsa veriler smote yöntemi ile çoğaltılarak veri seti dengeli bir hale getirilmeye çalışılır. Veri sızıntısı problemleri ile karşılaşırsa yine bu adımda veri sızıntıları giderilir. Bu aşamada yapılan işlemleri çoğu modelin genelleme yeteneğini yükseltmek için yapılan ön çalışmalardır.

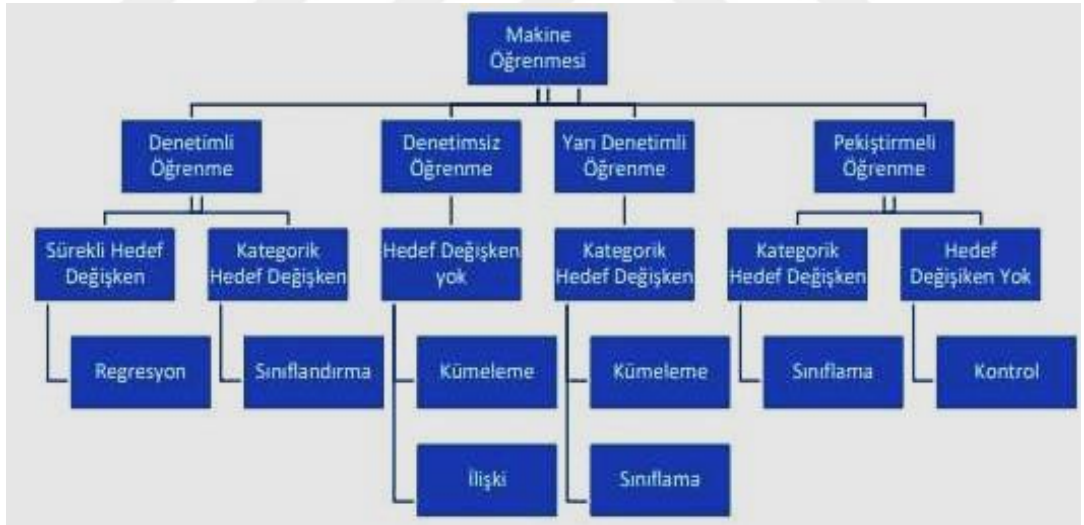
Model değerlendirme adımı olan bir sonraki aşamaya geçilir. Bu adımda modelin başarısı performans metrikleri aracılığıyla değerlendirilir. Makine öğrenmesi probleminin türüne göre performans metrikleri değişir. Sınıflandırma problemleri için accuracy, recall ve precision değerlerine bakılması gerekirken regresyon problemleri için R^2 , Mean Square Error ve Root Mean Square Error değerlendirme metriklerine bakılır. Ayrıca sınıflandırma problemleri için karmaşıklık matrisine de bakılır. ROC ve AUC eğrisi ile modelin kabiliyeti ölçülür. Modelin sınıflandırmaları ne kadar doğru ya da yanlış yaptığı analiz edilir. Ezberleme ya da verilerden hiç öğrenememe problemi oluşmuş mu tespit edilir. Modelin daha iyi sonuçlar vermesi açısından *model optimizasyonu* aşamasına geçilir.

Model optimizasyonu, modelin performansının artması için yapılan iyileştirme ve düzenleme sürecidir. Hiperparametre optimizasyonu ile bu aşamada en uygun hiperparametre değerleri çeşitli yöntemler kullanılarak belirlenir. En iyi hiperparametre kombinasyon değeri seçilir. Özellik seçimi ve boyut azaltma işlemleri de bu aşamada yapılır. Modelin performansını arttırmak ve karmaşıklığı azaltmak adına bağımlı değişkeni en iyi açıklayan özellikler belirlenir ve bu özellikler üzerinde model eğitilir. Principal Component Analysis (PCA) ile veri seti boyutu indirilerek kolay yönetilebilir bir hale getirilir. Aynı zamanda işlem maliyetini olabildiğince azaltarak büyük veri setini küçük boyutlu hale dönüştürür. Ensemble metodlar olan Random Forest, Gradient Boosting gibi yöntemleri uygulanarak modelin doğruluğunu arttırmak amaçlanır. Stacking yöntemiyle de farklı modellerin tahminleri birleştirilir Böylelikle farklı modellerin bir araya gelerek daha başarılı bir model oluşturması beklenir. Bu aşama, makine öğrenmesi süreçlerinde model performansını arttırmak için sıklıkla uygulanır. Eğer model doğru ve başarılı çıktılar üretiyor ise model canlı ortama aktarılabilir. Buna *model dağıtımı* ya da modelin deploy edilmesi denilir. Canlıya alınan model sürekli izlenir, güncellenir ve iyileştirilmeye devam eder. (Geron, 2022:

2.7. MAKİNE ÖĞRENMESİ TÜRLERİ

Makine öğrenmesi, makinelere verileri daha verimli bir şekilde işleme yeteneği kazandırmak için kullanılır. Veriler, ham haliyle çoğu zaman yorumlanamaz veya anlamlı sonuçlar çıkarılamaz. Bu gibi durumlarda, makine öğrenmesi teknikleri devreye girer ve verilerden değerli bilgiler elde edilmesini sağlar (Mahesh,2020: 20). Makine öğrenmesi denetimli öğrenme (supervised learning), denetimsiz öğrenme (unsupervised learning), yarı denetimli öğrenme (semi-supervised learning) ve pekiştirmeli öğrenme (reinforcement learning) olarak dörde ayrılır. Denetimli öğrenme, verilerin etiketlerinin bulunduğu veri setleri ve hedef değişkenler için kullanılırken denetimsiz öğrenme etiketsiz verilerin bulunduğu veri setleri ve hedef değişkenler için kullanılır. (Geron, 2022: 9-19)

Şekil 10: Makine Öğrenmesi Türleri



Kaynak: Bayram Arlı, Gürsakar, vd., 2022 çalışmasından alınmıştır.

2.7.1. Denetimli Öğrenme (Supervised Learning)

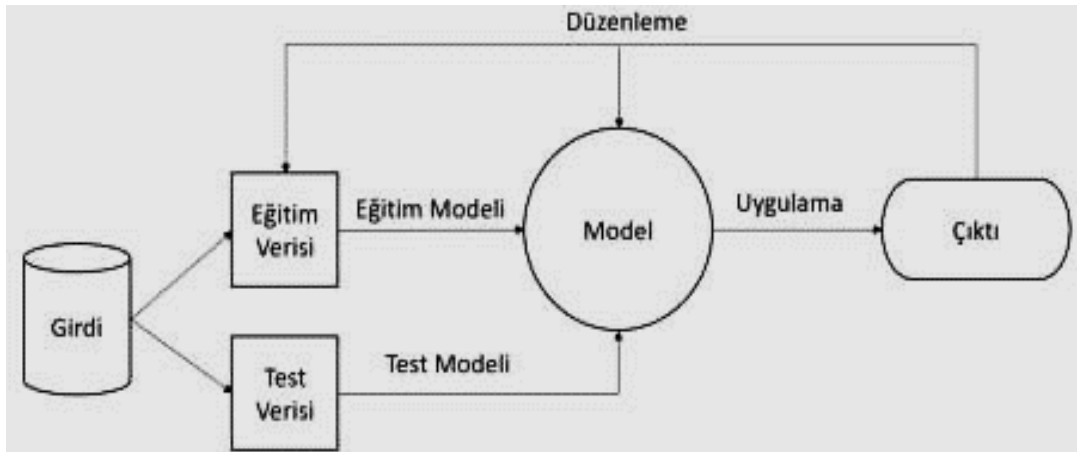
Denetimli öğrenme, makine öğrenmesi alanında en yaygın kullanılan yöntemlerden biridir. Bu öğrenme türü, etiketli veri setleri ile çalışır; yani, her bir giriş

(girdi verisi) ile ilişkili bir çıkış (etiket) vardır. Model, bu girdi-etiket çiftlerini kullanarak öğrenir ve bu öğrendiği bilgiyi yeni ve görülmemiş veriler üzerinde tahmin yapmak için kullanır. Denetimli öğrenmenin temel amacı, girdilerden öğrenilen örüntü ve ilişkileri kullanarak doğru tahminlerde bulunmaktır. Bu süreçte, veriler önce eğitim ve test seti olarak ikiye ayrılır. Eğitim seti algoritmayı eğitmek için kullanılırken, test seti algoritmanın performansını yeni verilere karşı değerlendirmek için kullanılır. Eğitim tamamlandıktan sonra, model test verileri üzerinde deneyerek etiketlenmemiş verilere karşı ne kadar doğru tahminlerde bulunduğu ölçülür. (Adalı, 2022: 70).

Denetimli makine öğrenmesinde, sisteme belirli kurallar vermek yerine, sisteme girdi-çıkı örnekleri sunulur ve bu örneklerden genelleme yapması beklenir. Bu yaklaşım, sistemin daha önce karşılaşmadığı durumları, verilen örneklerden öğrendiği bilgileri kullanarak doğru bir şekilde değerlendirebilmesini sağlar (European Commission, 2018).

Denetimli öğrenmede sıklıkla tahmin amacıyla sınıflandırma ve regresyon algoritmaları kullanılır. Bu iki algoritma arasındaki ana fark, hedef değişkenin türüdür. Regresyon algoritmalarında hedef değişken sürekli bir değer alırken, sınıflandırma algoritmalarında hedef değişken kategorik bir değerdir. (Bayram Arlı, 2021: 6).

Şekil 11: Denetimli Makine Öğrenmesi Çalışma Prensibi



Kaynak: Mahesh, 2020 çalışmasından alınmıştır.

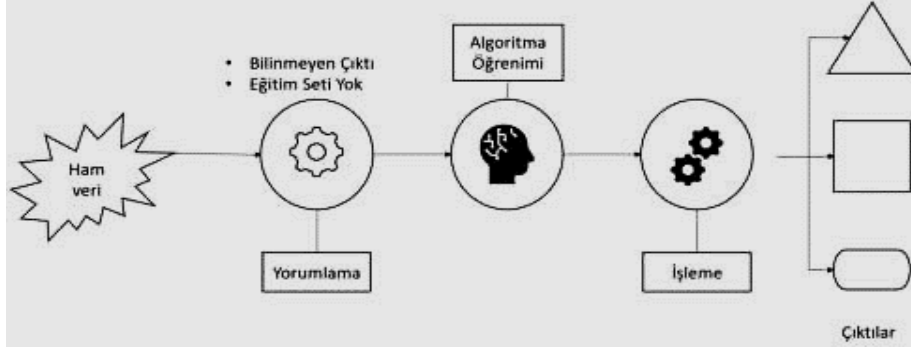
2.7.2. Denetimsiz Öğrenme (Unsupervised Learning)

Denetimsiz öğrenme, etiketsiz veriler arasındaki gizli kalmış yapıları ve örüntüleri ortaya çıkarmaya çalışan bir makine öğrenmesi yöntemidir. Bu yöntemde, verilerin etiketlerine dair herhangi bir bilgiye sahip olunmadığında denetimsiz öğrenme devreye girer. Yani, elimizde sadece ham veriler bulunur ve bu verilerden anlamlı sonuçlar çıkarılmaya çalışılır. Denetimsiz öğrenme algoritmaları, veriler arasındaki benzerlikleri ve farklılıkları analiz ederek gruplar veya örüntüler oluşturur. Bu nedenle, sistem veriler üzerinde kendi kendine öğrenme sürecini gerçekleştirir. Ancak, verilerin etiketsiz olması nedeniyle denetimsiz öğrenme algoritmalarının çıktıları kesin doğru olarak kabul edilmemelidir; bu çıktılar, daha fazla analiz ve doğrulama gerektiren başlangıç noktalarıdır. Denetimsiz öğrenme, özellikle büyük ve karmaşık veri setlerinde, gizli kalmış yapıların ve ilişkilerin keşfedilmesi açısından önemli bir araçtır. Bu yöntem, veri kümesi hakkında önceden bilgi sahibi olunmadan da yeni ve değerli bilgiler elde edilmesine olanak tanır. (Atalay, ve Çelik, 2017: 98)

Denetimsiz öğrenme, denetimli öğrenmeden farklı olarak, örnek çıktının olmadığı bir makine öğrenmesi yöntemidir. Bu yöntemde, algoritmalar verilerdeki ilginç yapıları keşfetmek ve sunmak üzere kendi haline bırakılır. Yeni veriler tanıtıldığında, algoritma daha önce öğrenilen özellikleri kullanarak verilerin sınıfını tanımlar. Denetimsiz öğrenme, esas olarak kümeleme ve ilişki tanımlama için kullanılır. Bu algoritmalar, insan müdahalesine ihtiyaç duymadan verilerdeki gizli kalıpları keşfeder, bu nedenle "denetimsiz" olarak adlandırılır. Yani, öğrenme süreci için herhangi bir "eğitim" veri setine ihtiyaç duymazlar. Algoritmalar, veriler arasındaki örüntüleri ve ilişkileri otomatik olarak bulur ve analiz eder. (Görentaş M.B, 2023: 98).

Denetimsiz öğrenme kümeleme ve ilişkileri belirleyen algoritmaları içermektedir.

Şekil 12: Denetimsiz Makine Öğrenmesi Çalışma Prensibi



Kaynak: Mahesh, 2020 çalışmasından alınmıştır.

2.7.3. Yarı Denetimli Öğrenme (Semi Supervised Learning)

Yarı denetimli makine öğrenmesi, denetimli ve denetimsiz öğrenme yöntemlerinin bir kombinasyonudur. Bu yaklaşım, etiketlenmemiş verilerin bol olduğu ve etiketli verilerin elde edilmesinin zor olduğu durumlarda özellikle faydalıdır. Yarı denetimli öğrenme hem etiketli hem de etiketlenmemiş verilerden öğrenme sürecini inceler. Geleneksel olarak, öğrenme ya tamamen etiketlenmiş verilerle (denetimli) ya da tamamen etiketlenmemiş verilerle (denetimsiz) yapılır. Yarı denetimli öğrenmenin amacı, etiketli ve etiketlenmemiş verilerin birlikte kullanıldığında öğrenme sürecini nasıl etkilediğini anlamak ve bu durumu avantaja çeviren algoritmalar geliştirmektir (Zhu & Goldberg, 2009). Bu yöntem, genellikle etiketsiz verilerin bol olduğu, ancak etiketli verilerin sınırlı ve zor elde edildiği durumlarda kullanılır. Model, öncelikle mevcut etiketli veriler üzerinde eğitilir ve ardından etiketsiz verileri kullanarak daha geniş bir genelleme yapar. Bu, modelin performansını artırır ve yeni, görülmemiş verilere karşı daha iyi tahminler yapmasını sağlar. Az miktarda etiketli veriyle bile yüksek performans elde edebilir. Etiketleme süreci genellikle zaman alıcı ve pahalıdır. Yarı denetimli öğrenme, bu maliyetleri azaltarak daha geniş veri setleriyle çalışmayı mümkün kılar. Yarı denetimli öğrenme, veri etiketleme zorluklarının üstesinden gelmek ve büyük veri setlerinden anlamlı bilgiler çıkarabilmek için güçlü bir yöntemdir. Bu yaklaşım, etiketsiz verilerin potansiyelini kullanarak daha verimli ve etkili makine öğrenme modelleri oluşturmayı sağlar (Learning, 2006).

2.7.4. Pekiştirmeli Öğrenme (Reinforcement Learning)

Pekiştirmeli öğrenme, bir ajanının çevresiyle doğrudan etkileşim kurarak öğrenme sürecini ve karar verme aşamalarını optimize etmeye yönelik bir hesaplamalı yaklaşımdır. Bu yöntem, ajanın bir dizi eylemi deneyimleyerek ve bu eylemlerden ödüller veya cezalar alarak hedeflere ulaşma stratejilerini geliştirmesine olanak tanır. Ajan, çevresindeki geri bildirimlere dayanarak zamanla daha iyi kararlar verir ve hedefe ulaşma sürecini otomatik olarak iyileştirir. (Sutton ve Barto, 1998:141).

Pekiştirmeli öğrenmede, temel kavramlar olarak "ajan" ve "çevre" ön plandadır. Ajan, öğrenme sürecinde kararlar veren ve eylemler gerçekleştiren bir varlıktır. Çevre, ajanın bu eylemlerine yanıt veren ve sürekli olarak değişen bir ortamı temsil eder. Ajan, çevredeki mevcut duruma göre eylemlerini seçerken, çevre bu eylemler sonucunda yeni durumlar sunar. Bu etkileşim döngüsü, ajanın çevre hakkında bilgi edinmesini ve bu bilgileri kullanarak çeşitli durumlarda hangi tepkilerin en uygun olduğunu belirleyen bir strateji geliştirmesini sağlar. Sürekli bir geri bildirim ve öğrenme süreci aracılığıyla, ajan çevresinin dinamiklerine uyum sağlar ve hedeflerine ulaşmak için en iyi eylem stratejilerini geliştirir (Geron, 2022: 16).

Pekiştirmeli öğrenme, özellikle oyunlar, robotik, otonom araçlar ve dinamik sistemlerin kontrolü gibi alanlarda kullanılır. Bu yöntem, ajanın deneyim yoluyla öğrenmesini ve sürekli olarak çevresine uyum sağlamasını mümkün kılar, böylece hedeflere ulaşmada etkili ve dinamik çözümler üretir. Ajan ceza ve ödül sistemiyle verdiği kararların doğru ya da yanlış olduğunu ayırt ederek oluşan benzer durumlarda vereceği kararı daha iyi öğrenir ve her bir denemede performansını artırarak başarılı kararlar vermeye başlar.

2.7.5. Topluluk Öğrenmesi (Ensemble Learning)

Toplu öğrenme, makine öğrenimi ve derin öğrenmede birden fazla modelin tahminlerini birleştirerek bireysel modellerin performansını aşmayı amaçlayan bir yaklaşımdır. Bu yöntem, farklı modellerin bir araya getirilmesiyle model hatalarının azaltılabileceği, varyansın düşeceği ve genelleme yeteneğinin artacağı fikrine dayanmaktadır. Toplu öğrenmenin en önemli avantajı, bireysel modellerin hatalarını diğer modellerin telafi etme kapasitesidir. Ayrıca, bu teknikler aşırı uyum ve yetersiz

uyum sorunlarını hafifleterek modelin daha kararlı ve doğru sonuçlar elde etmesini sağlar (Mienye ve Sun,2022).

Topluluk öğrenme yöntemleri, farklı modellerin tahminlerini birleştirerek daha güçlü ve doğru sonuçlar elde etmeyi hedefleyen yaklaşımlardır. Bu yöntemler, genellikle hata oranını düşürmek, modelin genelleme kapasitesini artırmak ve aşırı uyumu (overfitting) engellemek için kullanılır. Topluluk öğrenmesi, temel olarak heterojen ve homojen olmak üzere iki ana kategoriye ayrılır. Heterojen modeller, aynı veri seti üzerinde farklı algoritmaların veya aynı algoritmanın farklı parametre ayarlarının kullanılmasıyla çeşitlilik sağlar. Bu kategoride boosting, torbalama (bagging) ve rastgele alt uzaylar (random subspaces) gibi yöntemler öne çıkar. Homojen modeller ise bootstrap yöntemiyle alınan örneklemelerle aynı algoritmanın farklı veri alt kümeleri üzerinde uygulanmasıyla çeşitlilik yaratır. Bu yöntemler arasında en bilinen örneklerden biri istifleme (stacking) tekniğidir (Pulat ve Kocakoç, 2021).

Boosting, ardışık olarak oluşturulan modellerin yanlış tahminlere daha fazla ağırlık vermesi prensibine dayanır. İlk modelin yaptığı hatalar, bir sonraki modelin eğitimi sırasında daha önemli hale gelir. Bu süreç, hataların adım adım azaltılmasını sağlar. Son aşamada ise tüm modellerin tahminleri ağırlıklı bir oylama sistemiyle birleştirilir. Boosting, genellikle hata oranını düşürme konusunda oldukça etkili bir yöntemdir (Bartlett ve diğerleri, 1998; Liaw & Wiener, 2002).

Torbalama (Bagging) ise bağımsız olarak oluşturulan modellerin tahminlerini birleştirerek daha güçlü sonuçlar üretir. Bu yöntemde, veri kümesinden bootstrap yöntemiyle rastgele ve tekrarlı örneklemeler alınarak farklı eğitim setleri oluşturulur. Eğitim verisinin yaklaşık 2/3'ü (in-bag) ağaçları eğitmek için kullanılırken, geri kalan 1/3'lük kısım (out-of-bag) modelin performansını değerlendirmek ve değişken önemini ölçmek için kullanılır. Sınıflandırma problemlerinde çoğunluk oylaması, regresyon problemlerinde ise tahminlerin ortalaması ile nihai sonuç belirlenir. Torbalama, modelin varyansını düşürerek genelleme performansını artırır ve aşırı uyum sorununu azaltır (Breiman, 1996; Liaw & Wiener, 2002; Bartuğ ve Bulut, 2022).

Rastgele alt uzaylar (Random Subspaces) yöntemi ise veri setinden farklı özellik grupları seçerek modellerin eğitilmesini sağlar. Her bir alt uzayda oluşturulan

modellerin çıktıları birleştirilerek daha güçlü tahminler elde edilir. Bu yaklaşım, hem model çeşitliliğini artırır hem de daha dengeli sonuçlar üretir. Özellikle yüksek boyutlu veri setlerinde, rastgele alt uzaylar yöntemi modelin performansını artırmada etkili bir strateji olarak kullanılır (Liang ve diğerleri, 2018).

Bu yöntemlerin tümü, birden fazla modelin kombinasyonu yoluyla daha güçlü ve genel sonuçlar elde edilmesini sağlar. Boosting hataları düzeltme üzerine odaklanırken, torbalama ve rastgele alt uzaylar çeşitlilik ve genelleme performansı üzerine yoğunlaşır. Bu farklı yaklaşımlar, sınıflandırma ve regresyon gibi birçok problemde başarılı tahminler elde edilmesine olanak tanır.

2.8. MAKİNE ÖĞRENMESİ ALGORİTMALARI

Makine öğrenmesi algoritmaları, verilerden öğrenme ve bu öğrenme sürecini kullanarak tahminlerde bulunma veya kararlar verme yeteneğini geliştiren çeşitli teknikleri kapsar. Bu algoritmalar, verilerdeki gizli kalıpları ve ilişkileri keşfetmek, bilinmeyen durumlarla başa çıkmak ve karmaşık problemleri çözmek için kullanılır. Her bir algoritma, farklı veri türlerine ve problem gereksinimlerine göre özelleştirilmiş yöntemler sunar. Tez kapsamında kullanacağımız algoritmalar Random Forest, XgBoost, CatBoost, Decision Tree ve SVM algoritmalarıdır.

2.8.1. Rastgele Ormanlar (Random Forest)

Breiman tarafından 2001 'de geliştirilen rastgele orman (random forest) algoritması, karar ağaçları algoritmalarının bir topluluğundan oluşan güçlü bir makine öğrenmesi yöntemidir. Bu yöntem, birden fazla karar ağacını bir araya getirerek daha güvenilir ve genelleştirilebilir sonuçlar elde etmeyi amaçlar. Rastgele Orman algoritmasının temel prensibi, her bir karar ağacının eğitim sürecinde rastgele örnekler ve rastgele seçilmiş özellikler kullanmasıdır. Her karar ağacı, veri setindeki rastgele seçilen örnekler ve özellikler üzerinde eğitilir. Bu rastgele seçim süreci, her ağacın farklı bir özellik vektörü ile çalışmasını sağlar ve bu da modelin genel çeşitliliğini artırır. Özelliklerin rastgele seçilmesi, her bir ağacın veri setindeki farklı yönleri öğrenmesine olanak tanır ve dolayısıyla modelin genel performansını artırır. Rastgele orman algoritması, ağaçlar arasında bağımsızlığı teşvik ederek, her bir ağacın sadece

belirli özelliklerle eğitilmesini sağlar. Bu, modelin genel gürültüye karşı daha dayanıklı olmasına ve daha güvenilir sonuçlar vermesine yardımcı olur. Ayrıca, ağaçlar arasındaki korelasyonun düşürülmesi, modelin genelleştirme yeteneğini iyileştirir. Korelasyonun azalması, modelin daha farklı bakış açıları geliştirmesine ve daha dengeli tahminler yapmasına olanak tanır. (Geron, 2022)

Rastgele Orman algoritması, Karar Ağaçlarının bireysel güçlü yönlerini birleştirirken, aynı zamanda ağaçlar arasındaki korelasyonu minimize ederek genelleştirme hatasını azaltır. Bu özellikleri sayesinde, Rastgele Orman algoritması, çeşitli veri analizi problemlerinde yüksek performans ve dayanıklılık sunarak makine öğrenmesi uygulamalarında geniş bir kullanım alanına sahiptir (Breiman, 2001).

2.8.2. XgBoost Algoritması

Tianqi Chen tarafından 2014'te geliştirilen XGBoost algoritması, makine öğrenimi alanında son yıllarda geliştirilen ve birçok alanda başarılı sonuçlar veren güçlü bir yöntem olarak öne çıkmaktadır. Özellikle gereksiz e-postaların tespiti, reklam eşleştirme sistemleri, dolandırıcılık tespit mekanizmaları ve anomali olay tespit sistemleri gibi geniş bir uygulama yelpazesinde etkili bir şekilde kullanılmaktadır. XGBoost'un en önemli avantajlarından biri, tahminlerin tamamen veriye dayalı olarak gerçekleştirilmesidir. Regresyon ve sınıflandırma problemlerinde kullanılan boosting algoritmalarının gücünden yararlanarak yüksek performanslı sonuçlar elde edilmesini sağlamaktadır. Bununla birlikte, XGBoost algoritması, diğer makine öğrenimi yöntemlerine kıyasla daha hızlı ve daha az hesaplama maliyeti ile çalışması sayesinde, çok boyutlu ve büyük veri setlerinin analizinde son derece etkindir. (Chen ve Guestrin, 2016 : 15).

XGBoost algoritması, GBM algoritmasından türetilmiş ölçeklenebilir bir çerçeve olarak değerlendirilebilir. Bu algoritma, RF (Rastgele Ormanlar) ve GBM (Gradyan Artırma Makineleri) gibi ağaç tabanlı algoritmalar kullanan topluluk öğrenme tekniklerine dayanmaktadır. Aynı zamanda, regresyon ve sınıflandırma problemlerinde hızı artırmak ve tahmin performansını iyileştirmek amacıyla kullanılan GBM'nin optimize edilmiş bir versiyonudur. XGBoost algoritmasının temel amacı, aşırı öğrenme (overfitting) problemini azaltmak ve hesaplama süresini mümkün olduğunca en aza indirmektir. (Özüpek,2024: 14)

2.8.3. Karar Ağaçları (Decision Tree)

Karar Ağacı algoritması, verilerin sınıflandırılması veya tahmin edilmesi için kullanılan, hiyerarşik bir yapıya sahip bir modeldir. Bu yapı, bir dizi karar noktası ve bu kararların olası sonuçlarına dayalı olarak organize edilmiş düğümlerden oluşur. Karar Ağacı, verileri çeşitli özellikler ve bu özelliklere dayalı sorular kullanarak sistematik bir şekilde analiz eder. Karar ağacı, her bir düğümde bir soru veya test içerir. Bu sorular, veri kümesindeki özelliklere dayalıdır ve genellikle iki veya daha fazla olası cevaba sahiptir. Düğümler, verilerin özelliklerine göre çeşitli yolları temsil eder ve her bir yol, verilerin nasıl sınıflandırılacağını belirler. Her bir düğüm, olası yanıtların her biri için bir veya daha fazla alt düğüm içerir. Bu alt düğümler, bir sonraki soruyu veya testi temsil eder ve bu süreç, tüm olası yanıtlar ve yollar için devam eder (Geron, 2022:186). Her alt düğüm, bir önceki düğümdeki yanıtların sonuçlarına göre oluşturulur. Ağacın en alt seviyesinde yer alan düğümler, yaprak düğümleri olarak adlandırılır. Yaprak düğümler, önceden belirlenmiş sınıfları veya tahmin değerlerini temsil eder. Verilerin sınıflandırılması, sonunda bu yaprak düğümlerine ulaşarak gerçekleştirilir. Veriler, ağacın en üstündeki düğümden başlayarak işlem görür. Her adımda, verinin özellikleri doğrultusunda uygun bir yanıt seçilir ve bir sonraki alt düğüme geçilir. Bu süreç, yaprak düğüme ulaşana kadar devam eder. Yaprak düğümde ulaşılan sınıf veya tahmin değeri, verinin nihai sonuç olarak kabul edilir (Kingsford ve Salzberg, 2008).

2.8.4. Destek Vektör Makineleri (Support Vector Machine)

Destek Vektör Makineleri (SVM), veri dağılımı hakkında önceden varsayımlarda bulunmayan parametrik olmayan bir öğrenme yöntemi olarak dikkat çeker. İlk olarak ikili sınıflandırma problemlerini çözmek için geliştirilmiş olan SVM, verileri iki olası sınıftan birine ayırma amacı taşırken, zamanla çok sınıflı sınıflandırma problemlerinde de etkili bir şekilde kullanılabilir hale gelmiştir. Bu algoritma, başlangıçta verileri yüksek boyutlu bir özellik uzayına doğrusal olmayan bir şekilde eşler. Bu dönüşüm, verilerin daha karmaşık ilişkilerini ve sınırlarını temsil etmeyi mümkün kılar. Daha sonra, bu yüksek boyutlu uzayda, verileri en iyi şekilde ayıran bir hiper düzlem bulunarak sınıflar arasındaki marjin (ya da boşluk) maksimize

edilir. Hiper düzlem, verilerin destek vektörleri olarak bilinen en yakın noktalarına dayanarak oluşturulur ve bu, sınıflar arasındaki ayrımı en geniş şekilde sağlar. Ayrıca, çok sınıflı problemler için SVM, "bir-vs-bir" veya "bir-vs-diğer" gibi stratejiler kullanarak genişletilmiş bir uygulama alanı sunar. SVM'in parametrik olmayan yapısı ve yüksek boyutlu özellik uzayında çalışabilme yeteneği, onu karmaşık veri yapılarıyla başa çıkabilen güçlü bir sınıflandırma aracı yapar. Hiper düzleme, en yakın örneklerin "destek vektörleri" olarak adlandırıldığı ve maksimum marjini (kenar boşluğunu) sağlayan bir ayırım yüzeyidir. Destek vektörleri, hiper düzlemin konumunu ve yönünü belirleyen kritik veri noktalarıdır ve bu noktalar, sınıflandırma sınırının en yakınında yer alır. SVM algoritması, bu destek vektörlerini kullanarak doğrusal bir karar yüzeyi oluşturur ve bu yüzey, yanlış sınıflandırmaları en aza indirmeye çalışır. Bu süreçte elde edilen karar yüzeyi, algoritmanın veri setindeki farklı durumları doğru bir şekilde genelleyebilme yeteneğini artırır, bu da SVM'in geniş veri kümesi üzerinde etkili sonuçlar vermesine olanak tanır. (Mountrakis vd., 2011) (Cortes ve Vapnik, 1995).

2.8.5. CatBoost Algoritması (Categorical Boosting)

Yandex, 2017 yılında CatBoost algoritmasını geliştirerek XGBoost'un kategorik değişkenleri işleme konusundaki eksiklerini gidermeyi hedeflemiştir. Tianqi Chen tarafından 2014'te geliştirilen XGBoost'un ardından ortaya çıkan bu model, daha hızlı bir modelleme tekniği sunmasının yanı sıra kategorik değişkenleri etkili bir şekilde işleme yeteneğine sahiptir. CatBoost, kategorik değişkenleri doğrudan işleyerek bilgi kaybını önler ve bu sayede diğer boosting algoritmalarında olduğu gibi sayısal değerlere dönüştürme gereksinimini ortadan kaldırır. Kategorik değişkenlerin sıralanması ve ayrıştırılması gibi işlemler yapabilen bu yöntem, özellikle kategorik değişken sayısının fazla olduğu veri setlerinde daha hızlı çalışır ve daha doğru sonuçlar üretir. (Özüpek,2024: 16)

Catboost kategorik verileri doğrudan sayısal verilere dönüştürmeden işleyebilir. Bu, kategorik değişkenlerin özel bir ön işleme gerektirmediği anlamına gelir ve verilerin işlenmesini kolaylaştırır. CatBoost, veri sızıntısını (data leakage) engellemek için özel bir boosting t tekniği olan ordered boosting'i kullanır. Bu, modelin eğitim sürecinde veriyi rastgele sıralamak yerine, her bir veriyi sırasıyla ele

olarak doğru tahmin yapmasını sağlar.

CatBoost, özellikle büyük veri setlerinde hızlı eğitim süreçleri sağlayarak doğruluk ve genel performans açısından güçlü bir algoritma olarak öne çıkar. Genellikle diğer gradient boosting algoritmalarından daha iyi sonuçlar veren CatBoost, overfitting (aşırı uyum) sorununu minimize eden çeşitli teknikler kullanarak modelin yeni verilerle genellenabilirliğini artırır. Bunun yanı sıra, kullanıcıya model performansını optimize etmek için hiperparametreleri ayarlama imkânı tanır ve otomatik hiperparametre optimizasyonu yapabilen güçlü bir yapıya sahiptir. Karar ağaçları tabanlı bir model olması, CatBoost'un daha açıklanabilir (interpretable) sonuçlar üretmesine olanak tanır. Özellikle kategorik verilerin yoğun olduğu ve veri setindeki dengesizliklerin bulunduğu durumlarda etkili bir performans sergileyen CatBoost, kampanya yönetimi, müşteri segmentasyonu ve tahmin modelleri gibi uygulamalarda başarılı bir şekilde kullanılabilir. (Geron, 2022:200-209)

2.9. PERFORMANS METRİKLERİ

Makine öğrenmesi algoritmalarının performansını değerlendirmek için çeşitli metrikler kullanılır. Bu metrikler, modelin doğruluğunu, hassasiyetini, hatalarını ve genel etkinliğini ölçerek, hangi modelin en iyi sonucu verdiğini belirlemeye yardımcı olur. Modelin yeni gördüğü veriler üzerinde yapacağı tahminin ne kadar başarılı olduğunu açıklayan önemli metriklerdir. Makine öğrenmesinde sınıflandırma problemleri ve regresyon problemleri için farklı performans metrikleri kullanılır. (Özen vd.,2021).

Performans metriklerinin tümüne bakmak zorunlu değildir. Araştırmacı, hangi metriklerin önemli olduğunu belirlemek için derinlemesine analiz yapabilir. Bazı gözden kaçan hatalar, diğer metriklerin performanslarıyla gözlemlenebilir ve bu şekilde daha net anlaşılabilir. Bu nedenle, yalnızca ilgili metriklere odaklanmak ya da tüm metrikleri incelemek mümkündür.

Tablo 3: Performans Metriklerinin Formülleri

Kriter	Açıklama	Formül
Doğruluk	Doğru tahminlerin tüm örneklere oranı	$\frac{TP + TN}{TP + FP + TN + FN}$
Kesinlik	Doğru pozitif tahminlerin tüm pozitif tahminlere oranı	$\frac{TP}{TP + FP}$
Duyarlılık	Doğru pozitif tahminlerin toplam pozitif örneklere oranı	$\frac{TP}{TP + FN}$
F1-skor	Kesinlik ve duyarlılık değerlerinin harmonik ortalaması	$2 \times \frac{Kesinlik \times Duyarlılık}{Kesinlik + Duyarlılık}$

Kaynak: Bozkurt vd. Yalçın, 2024 çalışmasından alınmıştır.

Çok sınıflı sınıflandırma problemlerinde bu klasik metrikler performansı doğru ölçmemektedir. Bu sebeple incelenmesi gereken ek metriklerinde sonuçları değerlendirilmelidir.

2.9.1. Sınıflandırma

Sınıflandırma için performans metrikleri doğruluk, hassasiyet, duyarlılık, kesinlik ve F1 score'dur. Fakat dengesiz veri setleri söz konusu olduğunda ek performans metriklerine ihtiyaç duyulur. Modelin performansını belirlerken yanıltıcı metriklerden uzak durmak gerekir. Aksi halde yanlış ve gerçeği yansıtmayan bir tahminleme yapmış oluruz. Modelin veriyi ezberlediğini yada hiç anlayamadığını bu aşamada farkedebiliriz. Bu kısım bu sebeple kritik bir süreçtir.

2.9.1.1. Doğruluk (Accuracy)

Modelin doğru tahmin ettiği örneklerin toplam örneklere oranını ifade eder. Sınıflandırma problemlerinde genel performansı ölçmek için kullanılır.

Doğruluk, sınıflandırma problemlerinde genel model başarısını hızlı bir şekilde değerlendirmek için kullanılır. Ancak, bu metrik özellikle dengesiz sınıflandırma problemleri söz konusu olduğunda yanıltıcı olabilir. Eğer bir veri seti %90 negatif ve %10 pozitif sınıftan oluşuyorsa, model tüm örnekleri negatif olarak tahmin ettiğinde bile %90 doğruluk oranına ulaşabilir. Bu durumda, pozitif sınıf tamamen göz ardı edilmiş olur ve accuracy, modelin gerçek başarısını yansıtmaz.

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Samples}$$

Bu nedenle, accuracy metriği dengesiz veri setlerinde Precision, Recall ve F1-Skoru gibi daha dengeli ölçütlerle desteklenmelidir.

2.9.1.2. Hassasiyet (Precision)

Modelin pozitif olarak tahmin ettiği örneklerin gerçekten pozitif olma oranıdır. Pozitif sınıfların ne kadar doğru tahmin edildiğini ölçmek için kullanılır. Yanlış pozitiflerin maliyetinin yüksek olduğu durumlarda bu metrik kritik öneme sahiptir. Çok sınıflı sınıflandırma problemlerinde sınıfların arasında ağırlıklandırma yaparak sonuçları üretir.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

2.9.1.3. Duyarlılık (Recall)

Gerçek pozitiflerin model tarafından doğru tahmin edilme oranıdır. Modelin tüm gerçek pozitif örnekleri ne kadar iyi yakaladığını ölçer. Yüksek duyarlılık, pozitif sınıfları kaçırma oranını azaltır. Yanlış negatiflerin yüksek maliyete sahip olduğu durumlarda recall daha önemli hale gelir. Çok sınıflı sınıflandırma problemlerinde sınıfların arasında ağırlıklandırma yaparak sonuçları üretir

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

2.9.1.4. F1 Skoru

Hassasiyet ve duyarlılığın harmonik ortalamasıdır. Hem hassasiyeti hem de duyarlılığı dikkate alarak dengeli bir performans ölçüsü sağlar. Özellikle sınıfların dengesiz olduğu durumlarda kullanılabilir. F1-Skoru, her iki metriğe de eşit önem verir ve özellikle dengesiz veri setlerinde ağırlandırılmış şekilde öne çıkar. Pozitif sınıfın düşük bir orana sahip olduğu durumlarda, F1-Skoru hem doğru pozitif tahmin oranını hem de pozitif sınıfın model tarafından ne kadar iyi yakalandığını değerlendirmek için uygun bir ölçüttür.

2.9.1.5. ROC Eğrisi ve AUC

ROC Eğrisi, farklı sınıflandırma eşiklerinde duyarlılık ve 1 özgülük oranlarını gösterir. AUC, ROC eğrisinin altındaki alanı ifade eder. Modelin genel performansını ve sınıflandırma yeteneğini değerlendirmek için kullanılır. Özellikle ikili sınıflandırma problemlerinde önemlidir.

Bu performans metriklerine ek olarak dengesiz sınıflandırma problemlerinde Matthews Korelasyon Katsayısı, Jaccard score ve Hamming loss gibi ek performans metrikleri incelenebilir. Bu metrikler veri seti dengesiz olduğunda daha doğru sonuçlar verir.

2.9.1.6. Matthews Korelasyon Katsayısı (MCC)

Matthews Korelasyon Katsayısı (MCC), sınıflandırma problemlerinde kullanılan daha kapsamlı bir metriktir. Tüm sınıflandırma sonuçlarını (True Positive, True Negative, False Positive, False Negative) dikkate alarak pozitif ve negatif sınıflar arasında dengeli bir performans ölçümü sunar.

MCC, özellikle dengesiz veri setlerinde güvenilir bir ölçüt olarak öne çıkar. %90 negatif ve %10 pozitif sınıfların olduğu bir veri setinde, MCC hem baskın sınıfın hem de nadir sınıfın tahminlerini adil bir şekilde değerlendirir.

$$MCC = \frac{(TP \cdot TN) - (FP \cdot FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

2.9.1.7. Hamming Kaybı (Hamming Loss)

Hamming kaybı, yanlış sınıflandırılan etiketlerin oranını ifade eder. Bu metrik, özellikle çok etiketli sınıflandırma (multi-label classification) problemlerinde önemlidir. Modelin tahmin ettiği sınıflar ile gerçek sınıflar arasındaki farkı ölçerek performans hakkında bilgi verir. Hamming kaybı düşükse, modelin sınıflandırma başarısı yüksektir.

$$Hamming Loss = \frac{Yanlış Sınıflandırılan Etiket Sayısı}{Toplam Etiket Sayısı}$$

2.9.1.8. Jaccard Skoru

Jaccard skoru, pozitif tahminlerin doğruluğunu ölçmek için kullanılan bir metriktir. Gerçek pozitif tahminlerin hem doğru hem de yanlış pozitif tahminlere oranını ifade eder.

$$Jaccard\ Score = \frac{True\ Positives}{True\ Positives + False\ Positives + False\ Negatives}$$

Jaccard skoru, özellikle dengesiz veri setlerinde yanlış pozitiflere daha duyarlı bir ölçü olarak öne çıkar ve çok etiketli sınıflandırma problemlerinde sıklıkla kullanılır.

2.9.1.9. Hata Matrisi (Confusion Matrix)

Modelin tahmin sonuçlarını detaylı bir şekilde doğru pozitifler, yanlış pozitifler, doğru negatifler, yanlış negatif değerlerini gösteren bir tablodur. Sınıflandırma modelinin performansını daha ayrıntılı analiz etmek için kullanılır.

Modelin hangi sınıflarda iyi performans gösterdiğini ve hangi sınıflarda başarısız olduğunu görmeyi sağlar.

Özellikle dengesiz sınıflandırma problemlerinde, modelin azınlık sınıf üzerinde ne kadar iyi performans gösterdiğini analiz etmek için kullanılır. Örneğin, pozitif sınıfa ait TP ve FN değerleri, azınlık sınıfa yönelik tahmin başarısını anlamada kritik rol oynar.

Şekil 13: Karmaşıklık Matrisi (Confusion Matrix)

KARMAŞIKLIK MATRİSİ		TAHMİN SONUÇLAR	
		POZİTİF	NEGATİF
GERÇEK SONUÇLAR	POZİTİF	Doğru Pozitif (TP)	Yanlış Negatif (FN) (Tip II Hata)
	NEGATİF	Yanlış Pozitif (FP) (Tip I Hata)	Doğru Negatif (TN)

Kaynak: Feng ve diğerleri (2020: 733) çalışmasından alınmıştır.

Bu matris, modelin doğru ve yanlış tahminlerini dört farklı kategoriye ayırır.

True Positive (TP): Pozitif sınıfın doğru bir şekilde pozitif olarak tahmin edilmesidir.

True Negative (TN): Negatif sınıfın doğru bir şekilde negatif olarak tahmin edilmesidir.

False Positive (FP): Negatif sınıfın yanlış bir şekilde pozitif olarak tahmin edilmesidir. (tip 1 hata).

False Negative (FN): Pozitif sınıfın yanlış bir şekilde negatif olarak tahmin edilmesidir. (tip 2 hata).

2.9.2. Regresyon

Regresyon modellerinde modelin performansı sınıflandırmadan farklı metriklerle test edilir. Amaç modelin veriyi ne kadar anladığını test etmektir. Yeni göreceği bir veriyi ne kadar doğru tahmin edeceğini görmektir.

2.9.2.1. R² Skoru (R-Squared Score)

Regresyon modelinin açıklanan varyans oranını gösterir. Modelin bağımsız değişkenlerle bağımlı değişken arasındaki ilişkileri ne kadar iyi açıkladığını değerlendirmek için kullanılır.

2.9.2.2. Ortama Mutlak Hata (Mean Absolute Error - MAE)

Tahminler ile gerçek değerler arasındaki ortalama mutlak farkı verir. Modelin tahminlerinin ne kadar saptığını ölçer. Genellikle tahminlerin ortalama hata miktarını anlamak için kullanılır.

2.9.2.3. Root Mean Squared Error (RMSE)

Tahmin hatalarının karesinin ortalamasının kare köküdür. Tahmin hatalarının büyüklüğünü vurgular. Büyük sapmaların model performansına etkisini daha belirgin hale getirir.

2.9.2.4. Kayıp Fonksiyonu (Loss Function)

Modelin tahminlerinin ne kadar yanlış olduğunu ölçen fonksiyondur. Regresyon problemlerinde modelin tahmin hatalarını ölçmek için kullanılır. Hem sınıflandırma hem de regresyon problemlerinde kullanılabilir. Sınıflandırmada çapraz entropi gibi, regresyonda ise ortalama karesel hata gibi farklı çeşitleri vardır.



ÜÇÜNCÜ BÖLÜM

YÖNTEM VE UYGULAMA

3.1. VERİ SETİ ÖZELLİKLERİ AÇIKLAMALARI VE İSTATİSTİKSEL ANALİZLERİ

3.1.1. Çalışmanın Amacı

Şirketlerin pazarlama ve kampanya yönetiminde potansiyel müşterilerini tanıması çok önemlidir. Karı maksimize etmek ve müşterilerin alışveriş alışkanlıklarını yakından takip ederek müşterilere özel kampanya ve fiyatlar sunulması amaçlanır. Müşterilerin satın alma davranışlarını anlamak rekabetçi iş dünyasında oldukça önemlidir. Özellikle müşterilerin hangi kampanyalara yanıt vereceğini tahmin etmek işletmelerin pazarlama stratejilerini optimize etmelerine ve müşteri memnuniyetlerinin artmasına yardımcı olur. Geleneksel yöntemlerle müşteri davranışını tahmin etmek zor ve zaman alıcıdır. Ayrıca bu yöntemler genellikle büyük veri setleri ile çalışıldığında yetersiz kalmakta ve başa çıkılması zor bir süreç haline gelmektedir. Makine öğrenmesi yöntemleri kullanarak müşterilerin kampanyalara katılma olasılıkları tahmin edilebilir. Müşterinin hangi kampanyalara daha yatkın olduğunu bilmek, şirketlerin kişiye özel kampanyalar ve teklifler sunmalarını sağlar. Şirketler, hali hazırda alışveriş yapan mevcut müşteri kitlesini elde tutmanın, yeni müşteriler kazanmak için harcanan maliyetlerden daha düşük olduğunu bilirler. Bu nedenle şirketler, potansiyel müşterilerini sürekli olarak elde tutmayı amaçlar. Kampanya yönetimi ve müşteri ilişkileri yönetimi bu noktada oldukça kritik bir rol oynar. Kampanyalar, demografik özelliklerine göre benzer alışveriş davranışı gösteren müşterilere özel olarak uygulanabilir. Şirket stratejileri bu bağlamda pazarlama departmanları tarafından belirlenir. Ayrıca, bu veriler sayesinde önemli iç görüler elde edilebileceğinden, şirketler müşterilerini tanıırken rekabetçi ortamda müşteri verilerinin gizliliği konusunda da dikkatli olurlar. Bu tür veri bilimi çalışmaları şirketler tarafından genellikle gizli tutulur. Hatta şirketler, veri bilimi çalışmalarında kullandıkları algoritmalarla elde ettikleri başarıları dahi paylaşmak istemezler.

Geliştirilmeye açık ve sürekli değişen veri ışığında, kullanılan algoritmaların başarısının her problem özelinde aynı sonucu vermediği yapılan bazı çalışmalarda

açıkça görülmektedir. Fiyatlandırma stratejileri, müşteri satın alma davranışlarına göre değişkenlik gösterebilir ve aynı zamanda müşterilerin şirkete olan aidiyet duyguları ile de doğrudan ilişkilidir. Bu nedenle, müşteriler segmentlere ayrılarak en sadık müşteri kitlesi belirlenebilir. Müşterilerin harcadıkları toplam ücret ile kampanyaların fiyatları doğrudan ilişkilidir. Şirketler, müşterilerinin harcadığı fiyatları analiz ederek hangi müşterilerin şirkete daha fazla kar sağladığını tespit edebilirler. Bu çalışma, veri toplamanın zorluğu nedeniyle anonim bir veri seti kullanılarak yapılmıştır ve şirketlere, müşteri ilişkileri yönetiminde müşterilerini daha iyi tanıyabilme imkanı sağlar. Ayrıca, makine öğrenmesi algoritmaları yardımıyla, yeni gelen müşterilerin harcama alışkanlıklarına göre ne kadar fayda sağlayacağı belirlenebilir. Böylece şirketler, yeni müşterileri tanıyarak benzer alışveriş alışkanlıklarına sahip kişilerle aynı kategorideki kampanyaları sunabilirler. Bu süreçte kullanılacak makine öğrenmesi algoritmalarından en iyi sonucu verecek algoritma, şirketlerin kampanya yönetimi çalışmalarında model başarısı açısından fikir verecek ve uygulamalarına olanak sağlayacaktır. Çalışmada, Random Forest, Decision Tree, XGBoost, CatBoost ve SVM algoritmaları kullanılacak ve en yüksek performans ve başarıya ulaşan algoritma seçilecektir. Performans metrikleri ile en doğru sonuca ulaşacak algoritma, gelecekteki çalışmalara ve şirketlerin tümüne kampanya yönetimi süreçlerinde ışık tutacaktır.

3.1.2. Veri Setinin Detayları

Bu çalışmada kullanılan veri seti *Kaggle*'dan anonim olarak elde edilmiştir. *Kaggle* platformu gerçek zamanlı verileri içeren veri bilimciler ve makine öğrenimiyle yakından ilgili olan kişilerin kullandığı çevrimiçi bir internet sitesidir. Kampanya yönetimine dair müşterilerin demografik özellikleri ve hangi kampanyalardan alışveriş yaptıkları bilgisini içeren veri seti şirket adı verilmeden halka açık şekilde sunulmuştur. Kampanyaların türünün belli olmaması her şirket için farklı bir kampanya olabileceği anlamını taşımaktadır. Bu kampanyalar indirim, ücretsiz teslimat, ürün paketlenme, e-posta ve SMS, puan biriktirme, özel gün ve sezon sonu, arkadaş referanslı davet ve sınırlı süreli kampanyalar gibi şirketlerin kendi amacına yönelik kampanyalar olabilir. Her şirket kendi kampanyalarına göre veri seti içerisindeki kampanyaları uyarlayabilir.

Veri seti 2240 gözlemden oluşmaktadır. Müşterilere ait 26 değişken vardır. Bunlardan bazıları demografik, bazıları kampanya ve satın alma bilgilerini içeren verilerdir. Veri setinde belirli bir zaman aralığı mevcut değildir. Veri setinin içerdiği değişkenlerin açıklamaları aşağıda mevcuttur.

Tablo 4: Veri Setindeki Değişkenlerin Tanıtılması

Değişken Adı	Açıklama
AcceptedCmp1	Müşteri 1. kampanyadaki teklifi kabul ettiyse 1, aksi takdirde 0.
AcceptedCmp2	Müşteri 2. kampanyadaki teklifi kabul ettiyse 1, aksi takdirde 0.
AcceptedCmp3	Müşteri 3. kampanyadaki teklifi kabul ettiyse 1, aksi takdirde 0.
AcceptedCmp4	Müşteri 4. kampanyadaki teklifi kabul ettiyse 1, aksi takdirde 0.
AcceptedCmp5	Müşteri 5. kampanyadaki teklifi kabul ettiyse 1, aksi takdirde 0.
Complain	Müşteri son 2 yılda şikayet ettiyse 1, aksi takdirde 0.
DtCustomer	Müşterinin şirkete kaydolduğu tarih.
Education	Müşterinin eğitim düzeyi.
Marital	Müşterinin medeni durumu.
Kidhome	Müşterinin evindeki küçük çocuk sayısı.
Teenhome	Müşterinin evindeki ergen çocuk sayısı.
Income	Müşterinin yıllık hane geliri.
MntFishProducts	Son 2 yılda balık ürünlerine harcanan miktar.
MntMeatProducts	Son 2 yılda et ürünlerine harcanan miktar.
MntFruits	Son 2 yılda meyve ürünlerine harcanan miktar.
MntSweetProducts	Son 2 yılda tatlı ürünlerine harcanan miktar.
MntWines	Son 2 yılda şarap ürünlerine harcanan miktar.
MntGoldProds	Son 2 yılda altın ürünlerine harcanan miktar.
NumDealsPurchases	İndirimle yapılan alışveriş sayısı.
NumCatalogPurchases	Katalog kullanarak yapılan alışveriş sayısı.
NumStorePurchases	Mağazalarda doğrudan yapılan alışveriş sayısı.
NumWebPurchases	Şirketin web sitesi üzerinden yapılan alışveriş sayısı.
NumWebVisitsMonth	Son ayda şirketin web sitesine yapılan ziyaret sayısı.
Recency	Son alışverişten bu yana geçen gün sayısı.
SelectedCampaign	Müşterinin hangi kampanyaları seçtiğini gösteren hedef değişken.

Kaynak: Araştırmacı tarafından oluşturulmuştur.

AcceptedCmp1, AcceptedCmp2, AcceptedCmp3, AcceptedCmp4, AcceptedCmp5: Bu değişkenler, müşterinin firmanın sunduğu kampanyalardaki teklifleri kabul edip etmediğini gösterir. Kampanya teklifini kabul eden müşteriler için değişken 1, reddedenler için ise 0 olarak işaretlenir. Bu veriler, işletmenin kampanya performansını değerlendirmesine, hangi tekliflerin daha cazip olduğunu analiz etmesine ve gelecekteki pazarlama stratejilerini optimize etmesine yardımcı olur. Her bir AcceptedCmp değişkeni (AcceptedCmp1, AcceptedCmp2, AcceptedCmp3, AcceptedCmp4, AcceptedCmp5), farklı kampanyaların başarısını karşılaştırmak, müşteri yanıtlarını takip etmek ve kampanya etkinliğini ölçmek açısından kritik öneme sahiptir. İşletme, bu veriler sayesinde müşteri segmentlerini analiz edebilir, pazarlama çabalarını daha hedefli hale getirebilir ve genel kampanya yönetimini geliştirebilirler.

Complain: Son 2 yılda müşteri şikayetini gösteren bu değişken, şikayet varsa 1, yoksa 0 değerini alır. Complain, müşteri memnuniyetini ölçmeye, şikayet yönetimini iyileştirmeye ve müşteri sadakatini artırmaya yardımcı olur. Bu veri, müşteri hizmetleri performansını ve ilişkilerini değerlendirmek için önemlidir.

DtCustomer: Müşterinin şirkete kaydolma tarihini gösteren bu değişken, üyelik başlangıcını belirtir. DtCustomer, müşteri yaşam döngüsünü analiz etmeye, yeni ve uzun süreli müşteriler arasındaki farkları belirlemeye ve bağlılık stratejilerini geliştirmeye yardımcı olur. Bu veri, müşteri ilişkisinin süresini ve davranış değişikliklerini izlemek için kullanılır.

Education: Müşterinin eğitim düzeyini belirten bu değişken, lise, üniversite, yüksek lisans gibi kategorilere sahiptir. Education, işletmenin müşteri demografiklerini analiz etmesine, pazarlama kampanyalarını özelleştirmesine ve hedef kitle segmentasyonunu geliştirmesine yardımcı olur. Eğitim düzeyi, müşteri davranışları ve tercihlerini etkiler.

Marital: Müşterinin medeni durumunu gösteren bu değişken, evli, bekar, boşanmış gibi durumları belirtir. Marital, müşteri segmentasyonunu iyileştirmeye, pazarlama stratejilerini özelleştirmeye ve müşteri ihtiyaçlarını anlamaya yardımcı olur. Bu veri, demografik analizler ve pazarlama hedeflemeleri için kritiktir.

Kidhome: Müşterinin evindeki küçük çocuk sayısını (genellikle 0-12 yaş) gösteren bu değişken, işletmenin müşteri segmentlerini analiz etmesine, çocuklu ailelere özel kampanyalar geliştirmesine ve ürün/hizmet sunmasına yardımcı olur. Bu veri, aile yapısına göre müşteri davranışlarını anlamak için kullanılır.

Teenhoe: Müşterinin evindeki ergen çocuk sayısını (genellikle 13-18 yaş) gösteren bu değişken, ergen çocuklu ailelere yönelik pazarlama stratejilerini geliştirmeye, müşteri segmentasyonunu iyileştirmeye ve ürün/hizmet tekliflerini optimize etmeye yardımcı olur. Bu veri, müşteri demografik analizleri için önemlidir.

Income: Müşterinin yıllık hane gelirini ifade eden bu değişken, müşterinin hane halkının toplam yıllık gelirini gösterir. Income, işletmenin müşteri segmentlerini gelir düzeyine göre analiz etmesine, fiyatlandırma stratejilerini belirlemesine ve gelir düzeyine göre özelleştirilmiş pazarlama kampanyaları geliştirmesine yardımcı olur. Bu veri, müşteri davranışları ve harcama alışkanlıkları üzerinde önemli bir etkiye sahip olabilir.

MntFishProducts: MntFishProducts, müşterinin son 2 yılda balık ürünlerine harcadığı miktarı gösteren bir değişkendir. Bu veri, işletmenin balık ürünleri talebini analiz etmesine, ürün stoklarını yönetmesine ve pazarlama stratejilerini balık ürünlerine yönelik optimize etmesine yardımcı olur. Müşteri harcama alışkanlıklarını ve tercihlerini anlamak için önemlidir.

MntMeatProducts: Müşterinin son 2 yılda et ürünlerine harcadığı miktarı gösteren bir değişkendir. Bu veri, işletmenin et ürünleri talebini analiz etmesine, stok yönetimini iyileştirmesine ve et ürünlerine yönelik pazarlama stratejilerini geliştirmesine yardımcı olur. Müşteri tercihleri ve harcama davranışlarını anlamak için önemlidir.

MntFruits: Müşterinin son 2 yılda meyve ürünlerine harcadığı miktarı gösteren bir değişkendir. Bu veri, işletmenin meyve ürünleri talebini analiz etmesine, stok yönetimini optimize etmesine ve meyve ürünlerine yönelik pazarlama stratejilerini geliştirmesine yardımcı olur. Müşteri harcama alışkanlıklarını ve tercihlerini anlamak için önemlidir.

MntSweetProducts: Müşterinin son 2 yılda tatlı ürünlere harcadığı miktarı gösterir. Bu veri, işletmenin tatlı ürünleri talebini analiz etmesine, stok yönetimini iyileştirmesine ve pazarlama stratejilerini bu ürünlere göre optimize etmesine yardımcı olur, müşteri tercihleri ve harcama alışkanlıklarını anlamak için önemlidir.

MntWines: Müşterinin son 2 yılda şarap ürünlerine harcadığı miktarı gösterir. Bu veri, işletmenin şarap ürünleri talebini analiz etmesine, stok yönetimini optimize etmesine ve şarap ürünlerine yönelik pazarlama stratejilerini geliştirmesine yardımcı olur. Müşteri harcama alışkanlıklarını ve tercihlerini anlamak için önemlidir.

MntGoldProds: Müşterinin son 2 yılda altın ürünlere harcadığı miktarı gösterir. Bu veri, işletmenin altın ürünleri talebini analiz etmesine, stok yönetimini iyileştirmesine ve pazarlama stratejilerini bu ürünlere göre optimize etmesine yardımcı olur. Müşteri tercihleri ve harcama davranışları üzerinde önemli bir etkiye sahiptir.

NumDealsPurchases: Kampanya süresince müşterilerin indirimli ürünler satın aldığı toplam alışveriş sayısını ifade eder. Bu veri, işletmenin indirim kampanyalarının etkinliğini değerlendirmesine, müşteri davranışlarını analiz etmesine ve gelecekteki indirim stratejilerini optimize etmesine yardımcı olur.

NumCatalogPurchases: Bu deęişken müşterilerin katalog aracılığıyla yaptığı toplam alışveriş sayısını ifade eder. Kataloglar genellikle yazılı ve dijital formatlarda sunulur. Müşterilere ürünlerin listesi, açıklamaları, fiyatları ve sipariş bilgileri hakkında bilgi sağlar. Bu deęişken katalogların etkinliğini deęerlendirmek, müşteri tercihlerini anlamak ve katalog kullanımını teşvik etmek için kullanılabilir.

NumStorePurchases: Bu deęişken belirli bir perakende mağazasında müşterilerin fiziksel mağazaya gelerek gerçekleştirdikleri toplam alışveriş sayısını ifade eder. Mağazanın ziyaretçi trafiğini deęerlendirmek, satış performansını izlemek, müşteri tercihlerini anlamak ve mağaza içi pazarlama stratejilerini geliştirmek için kullanılır. Ayrıca, online alışveriş veya dięer satış kanallarıyla karşılaştırmak amacıyla kullanılarak, farklı satış yöntemlerinin etkinliğini karşılaştırmak da mümkündür.

NumWebPurchases: Bu deęişken, bir işletmenin veya perakende mağazasının internet sitesi aracılığıyla gerçekleştirilen toplam alışveriş sayısını ifade eder. İşletmenin online satış performansını izlemesine, web sitesi trafiğini analiz etmesine, müşteri davranışlarını anlamasına ve online pazarlama stratejilerini geliştirmesine yardımcı olur.

NumWebVisitsMonth: Bu deęişken, belirli son bir ayda şirketin web sitesini ziyaret eden toplam ziyaretçi sayısını ifade eder. Şirketin web sitesi trafiğini ölçmesine, web sitesi etkinliğini deęerlendirmesine ve dijital pazarlama stratejilerini optimize etmesine yardımcı olur.

Recency: Müşterinin veya kullanıcının son alışverişini yapmasından itibaren geçen gün sayısını ifade eder. İşletmenin müşteri sadakatini deęerlendirmesine, alışveriş sıklığını analiz etmesine ve müşteri davranışlarını anlamasına yardımcı olur. Örneğin, bir müşterinin son alışverişiyile geçen süre ne kadar uzunsa, bu müşterinin potansiyel olarak kaybolma riski daha yüksek olabilir. Bu nedenle, bu deęişken işletmeler için müşteri ilişkileri yönetiminde ve pazarlama stratejilerinde önemli bir gösterge olarak kullanılır.

SelectedCampaign: Müşterinin hangi kampanyaya yanıt verdiğini gösteren hedef deęişkendir. Müşteri kampanyayı kabul ettiyse 1 aksi takdirde 0 deęerini döndüren binary bir deęişken türündedir.

3.1.3. Veri Ön İşleme Süreci (Data Pre-processing)

Veri ön işleme süreci makine öğrenmesi adımlarının en önemli bölümünü oluşturan adımlardan biridir. Karmaşık ve işlenmemiş veri temel (base) model olarak makine öğrenmesi algoritmasına doğrudan analiz yapılması için verildiğinde analizin tahmin performansı ve güvenilirliği düşük olacak ve bu da çalışmanın başarılı bir sonuç vermemesine neden olacaktır. Bu sebeple makine öğrenmesi yöntemleri kullanılmadan önce çalışmada kullanılacak veri setinin analiz için uygun hale getirilmesi esastır. Bu çalışmada makine öğrenmesi algoritmasına uygun hale getirilecek olan veri setinin ön işleme adımları detaylıca anlatılacaktır.

Bu çalışmada geniş kütüphane desteği bulunan, veri analizi ve bilimsel yetenekleri güçlü olan ve kod okunabilirliği kolay olan Python programlama dili kullanıldı. Çalışmada veri ön işleme adımları için kullanılan kütüphaneler import edildi. Bu kütüphaneler şunlardır:

NumPy: Sayısal hesaplamalar ve çok boyutlu diziler üzerinde yüksek performans işlemler için kullanılır. Bu kütüphane lineer cebir, istatistik ve matematiksel işlemlerde alt yapı sağlar.

Pandas: Veri manipülasyonu için kullanılan bu kütüphane veri kaynaklarıyla entegrasyon sağlar. Veri ön işleme sürecinde sıkça kullanılan bir kütüphanedir. Verileri filtreleme, gruptama ve dönüştürme gibi işlevleri sunar.

Matplotlib: Bu kütüphane verileri görselleştirmek için kullanılan bir kütüphanedir. Çizimler, grafikler ve çeşitli görseller oluşturarak özelleştirilmesine olanak tanır. Görsellerin kaydedilmesi ve paylaşılması için araçlar sunar.

Seaborn: Bu kütüphane matplotlib kütüphanesi üzerine inşa edilmiş kolay kullanılabilen grafikler oluşturan görselleştirme kütüphanesidir. Matplotlib kütüphanesine göre daha gelişmiş grafikler oluşturulmasına olanak tanır. İstatistiksel veri görselleştirmeleri için kullanılır. Veri keşfi ve verinin dağılımının gösterilmesi gibi işlemlerde kullanılır.

Scikit-learn (sklearn): Bu kütüphane makine öğrenimi algoritmalarını barındırır. Sınıflandırma, regresyon ve kümele gibi gelişmiş makine öğrenmesi yöntemlerini destekler. Model seçimi ve değerlendirilmesi adımlarında işlemleri kolaylaştırır.

Datetime: Bu kütüphane veri seti içerisindeki zamansal işlemlerin yapılmasını sağlayan Python kütüphanesidir. Python'da tarih ve saat bilgileriyle çalışmak için kullanılan yerleşik bir modüldür. Tarih ve saat verilerini oluşturma, formatlama, işlemler yapma ve dönüştürme gibi işlemler için kullanılır.

Smote: makine öğrenmesi için dengesiz veri setlerinde kullanılan bir over-sampling yöntemidir. Azınlık sınıfındaki verilerin sayısını artırarak sınıf dengesizliğini giderir. Bu işlem, mevcut azınlık sınıfı örneklerinden sentetik (yeni) örnekler oluşturarak yapılır.

Aşağıdaki tabloda görüldüğü gibi çalışma içerisinde kullanılacak olan tüm kütüphaneler, algoritmalar ve metrikler öncelikle içe aktarılmıştır.

Şekil 14: Gerekli Kütüphanelerin İçe Aktarılması

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.preprocessing import OneHotEncoder, MinMaxScaler
from sklearn.ensemble import RandomForestClassifier
from catboost import CatBoostClassifier
from xgboost import XGBClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.svm import SVC
from sklearn.metrics import f1_score, confusion_matrix, accuracy_score, precision_recall_curve, jaccard_score, matthews_corrcoef, auc, hamming_loss
from sklearn.multiclass import OneVsRestClassifier
from sklearn.metrics import multilabel_confusion_matrix
import matplotlib.pyplot as plt
from imblearn.over_sampling import SMOTE
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Ardından veri seti çağırılarak veri ön işleme süreçlerine geçilmiştir. Aşağıda veri setinin içe aktarılması için fonksiyon oluşturarak süreci dinamik hale getiren kod bloğu görülmektedir.

Şekil 15: Veri Setinin İçerik Aktarılması

```
path = "marketing_campaign.csv"

def load_data(path):

    df= pd.read_csv(path, sep=";")

    return df

# veri setini tanımla
df = load_data(path)
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Bu fonksiyon, belirtilen path parametresi ile verilen dosya yolundan csv dosyasını yükler ve bir pandas data frame olarak veriyi işler. sep=";"; "argümanı, csv dosyasındaki verilerin noktalı virgül ile ayrıldığını belirtir. İşlem süresini kısaltmak adına load_data adıyla oluşturulan bu fonksiyon daha sonrasında dosya yolunu başka bir veri seti adıyla tanımlayarak kolayca çağrılmasına olanak tanır.

Çağırdığımız bu kod ile data frame'in ilk 5 gözlemine göz atıp genel görüntü incelenir. Değişkenlerin nasıl olduğuna bakılır. Veri setinin ham hali başlangıçta 30 satırdan oluşuyor.

Şekil 16: Genel Veri Setinin Görüntülenmesi

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	NumCatalogPurchases	NumStorePurchases	NumWebVisitsMonth	
0	5524	1957	Graduation	Single	58138.0	0	0	2012-09-04	58	635	...	10	4	7	
1	2174	1954	Graduation	Single	46344.0	1	1	2014-03-08	38	11	...	1	2	5	
2	4141	1965	Graduation	Together	71613.0	0	0	2013-08-21	26	426	...	2	10	4	
3	6182	1984	Graduation	Together	26646.0	1	0	2014-02-10	26	11	...	0	4	6	
4	5324	1981	PhD	Married	58293.0	1	0	2014-01-19	94	173	...	3	6	5	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
2235	10870	1967	Graduation	Married	61223.0	0	1	2013-06-13	46	709	...	3	4	5	
2236	4001	1946	PhD	Together	64014.0	2	1	2014-06-10	56	406	...	2	5	7	
2237	7270	1981	Graduation	Divorced	56981.0	0	0	2014-01-25	91	908	...	3	13	6	
2238	8235	1956	Master	Together	69245.0	0	1	2014-01-24	8	428	...	5	10	3	
2239	9405	1954	PhD	Married	52869.0	1	1	2012-10-15	40	84	...	1	4	7	

2240 rows × 27 columns

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Aşağıda info kodu ile veri setinin genel yapısı hakkında bilgilendirmeler sunar. Veri setinde bulunan 27 değişkenden 4 tanesi kategorik türde iken 23 değişken sayısal türdeki verileri barındırmaktadır. Sayısal verileri barındıran 1 değişken ondalıklı sayı formatındadır. Kabul edilen 2 kampanya birleştirilerek selectedcampaign değişkeni oluşturulmuştur. Daha sonra bu kampanyalar veri setinde yer kaplamaması amacıyla dışarda bırakılmıştır. Aynı amaç ile analizde anlamlı bir etki yaratmayacak bazı değişkenler hacmi çoğalmaması için veri setinden atılmıştır. Veri setinde nihai kalan 22 değişken mevcuttur. Veri setinde 2240 gözlem mevcut olması beklenirken Income değişkeninde 24 adet eksik gözlem mevcuttur. Aynı zamanda bu rapor data frame'in bellek kullanımı hakkında da bilgi sunar.

Şekil 17: Değişkenlere Ait Bilgilerin Görüntülenmesi

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 22 columns):
#   Column                Non-Null Count  Dtype
---  -
0   ID                     2240 non-null   int64
1   Year_Birth             2240 non-null   int64
2   Education              2240 non-null   object
3   Marital_Status         2240 non-null   object
4   Income                 2216 non-null   float64
5   Kidhome                2240 non-null   int64
6   Teenhome               2240 non-null   int64
7   Dt_Customer           2240 non-null   object
8   Recency                2240 non-null   int64
9   MntWines               2240 non-null   int64
10  MntFruits              2240 non-null   int64
11  MntMeatProducts        2240 non-null   int64
12  MntFishProducts        2240 non-null   int64
13  MntSweetProducts       2240 non-null   int64
14  MntGoldProds           2240 non-null   int64
15  NumDealsPurchases      2240 non-null   int64
16  NumWebPurchases        2240 non-null   int64
17  NumCatalogPurchases    2240 non-null   int64
18  NumStorePurchases      2240 non-null   int64
19  NumWebVisitsMonth       2240 non-null   int64
20  Complain                2240 non-null   int64
21  selectedcampaign        2240 non-null   object
dtypes: float64(1), int64(17), object(4)
memory usage: 385.1+ KB
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Veri seti genel olarak tanıtılıp bazı genellemeler yapılmıştır. Fakat verilerin genel betimsel istatistiklerine bakarak farklı çıkarımlar da bulunarak verileri daha detaylı tanımak önemlidir. Aşağıdaki görselde veri setine ait değişkenlerin betimsel istatistikleri verilmiştir.

Veri setinde *Income* değişkeninde 24 adet eksik değer olduğu tespit edilmişti. Eksik değerlerin nedenini anlamak ve bu eksikliklerin analiz üzerindeki etkisini değerlendirmek önemlidir. *Income*, *MntWines*, *MntMeatProducts*, *MntFishProducts*, *MntSweetProducts* ve *MntGoldProds* değişkeninde maksimum değerler oldukça yüksektir. Bu, veri setinde aykırı değerlerin (outliers) varlığına işaret edebilir. Aykırı değerler üzerinde durulmalı ve aykırı değerler tespit edilip modelden uzaklaştırılmalı ya da başka bir teknik ile model başarısını etkilemeyecek duruma getirilmelidir. *Complain* değişkeninin ortalama değeri 0.009375'tir. Sıfıra yakın bir ortalama değere sahip complain değişkeni veri setinin büyük bir kısmında 0 değeri olduğunu söylemektedir. Bu, müşterilerin genellikle şikâyetle bulunmadığını gösterir. Ancak, bu düşük varyanslı bir değişken olduğu dikkat çekmektedir. *NumDealsPurchases* değişkeninin ortalama değeri 2.325 ve maksimum değeri 15'tir. Bu dağılım, bu değişkenin belirli müşteri segmentlerinde düşük alım yapma eğiliminde olduğunu ancak bazı müşterilerin daha fazla indirimli ürün alımı yaptığını gösterebilir. Bu değişkenin genel etkisi ve alışveriş davranışları üzerindeki yansımaları dikkate alınmalıdır.

Şekil 18: Veri Setinin Betimsel İstatistikleri

```
df.drop("ID",axis=1).describe().T
```

	count	mean	std	min	25%	50%	75%	max
Year_Birth	2240.0	1968.805804	11.984069	1893.0	1959.00	1970.0	1977.00	1996.0
Income	2216.0	52247.251354	25173.076661	1730.0	35303.00	51381.5	68522.00	666666.0
Kidhome	2240.0	0.444196	0.538398	0.0	0.00	0.0	1.00	2.0
Teenhome	2240.0	0.506250	0.544538	0.0	0.00	0.0	1.00	2.0
Recency	2240.0	49.109375	28.962453	0.0	24.00	49.0	74.00	99.0
MntWines	2240.0	303.935714	336.597393	0.0	23.75	173.5	504.25	1493.0
MntFruits	2240.0	26.302232	39.773434	0.0	1.00	8.0	33.00	199.0
MntMeatProducts	2240.0	166.950000	225.715373	0.0	16.00	67.0	232.00	1725.0
MntFishProducts	2240.0	37.525446	54.628979	0.0	3.00	12.0	50.00	259.0
MntSweetProducts	2240.0	27.062946	41.280498	0.0	1.00	8.0	33.00	263.0
MntGoldProds	2240.0	44.021875	52.167439	0.0	9.00	24.0	56.00	362.0
NumDealsPurchases	2240.0	2.325000	1.932238	0.0	1.00	2.0	3.00	15.0
NumWebPurchases	2240.0	4.084821	2.778714	0.0	2.00	4.0	6.00	27.0
NumCatalogPurchases	2240.0	2.662054	2.923101	0.0	0.00	2.0	4.00	28.0
NumStorePurchases	2240.0	5.790179	3.250958	0.0	3.00	5.0	8.00	13.0
NumWebVisitsMonth	2240.0	5.316518	2.426645	0.0	3.00	6.0	7.00	20.0
Complain	2240.0	0.009375	0.096391	0.0	0.00	0.0	0.00	1.0

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Veri bilimi ve makine öğrenmesi süreçlerinde veriyi anlamak ve analiz etmek oldukça önemlidir. Makine öğrenmesi algoritmalarının yüksek performans gösterebilmesi için veriyi iyi analiz etmek gereklidir. Çünkü makine öğrenmesi süreçlerinin en zaman alıcı kısmı veri manipülasyonu kısmıdır. Bu kısımda gözden kaçan detaylar modelin yetersiz kalmasına sebep olabilir. Bu nedenle veri seti titizlikle manipüle edilmelidir. Veri seti özelinde birleştirilerek oluşturulan selectedcampaign değişkeni 0 ve 1 değerlerinden oluşur ve kampanya katılımını temsil eder. 1 kampanyaya katılım durumunu temsil ederken 0 kampanyanın tercih edilmediğini gösterir.

Şekil 19: Kampanyaların Sayısı

```
#Seçilen kampanyaların sayısı
df[['selectedcampaign']].value_counts()

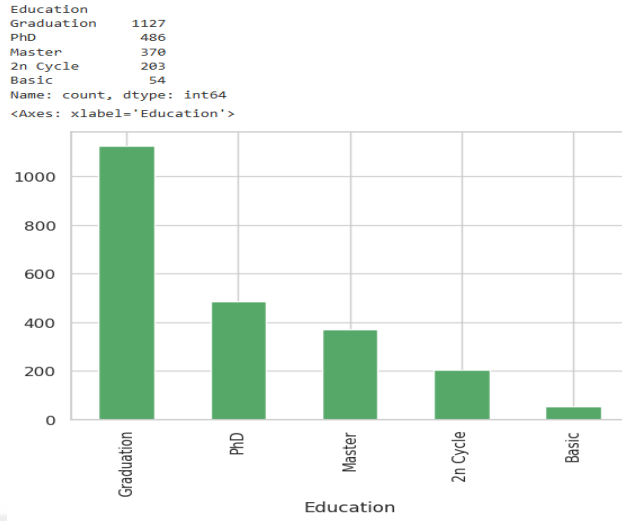
selectedcampaign
00      1969
10      108
01      104
11       59
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Selected campaign değişkeni içerisinde 4 farklı durumu barındırır. Bunlar birinci kampanyanın kabul edilip ikinci kampanyanın kabul edilmemesi, ikinci kampanyanın kabul edilip birinci kampanyanın kabul edilmemesi ve iki kampanyanın da kabul edilmesi ya da kabul edilmemesi durumlarını ifade eder.

Veri setinde hiçbir kampanyaya katılmayan 1969 kişi, her iki kampanyaya da katılan 59 kişi mevcuttur. Birinci kampanyayı tercih eden 108 kişi iken ikinci kampanyayı kabul eden 104 kişidir. Görüldüğü üzere bağımlı değişkenimiz çoklu sınıflandırma problemi içerir. Her sınıf arasında dengeli bir dağılım söz konusu değildir. Dengesiz veri dağılımı tez kapsamı içerisinde yer almadığından dikkate alınmayacaktır.

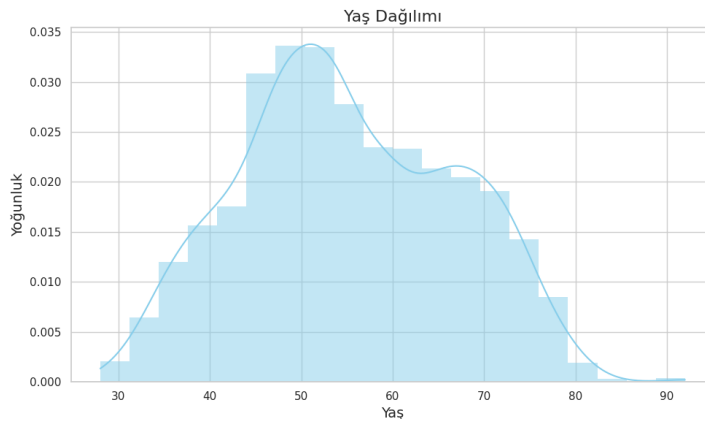
Şekil 20: Eğitim Durumu Değişkeni Histogram Grafiği



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Veri setimizde eğitim durumu değişkeni, 4 farklı kategoriye temsil etmektedir. Bu kategorilere göre, lise mezunu 54, üniversite mezunu 1127, yüksek lisans 370, doktora 486 ve çift ana dal yapan 203 müşteri bulunmaktadır. Veri dağılımına bakıldığında, üniversite mezunu müşterilerin online kanallar üzerinden daha fazla alışveriş yaptığı gözlemlenmektedir. Müşterilerimizin büyük bir kısmı üniversite mezunu olmasına rağmen, yüksek lisans ve doktora yapmış olanların sayısı da oldukça fazladır. Genel olarak, eğitim seviyesi yüksek bir müşteri kitlesine sahip olduğumuz söylenebilir. Bu durum, internet alışverişine olan eğilimin, eğitimli kitle arasında daha yoğunlaştığı şeklinde yorumlanabilir.

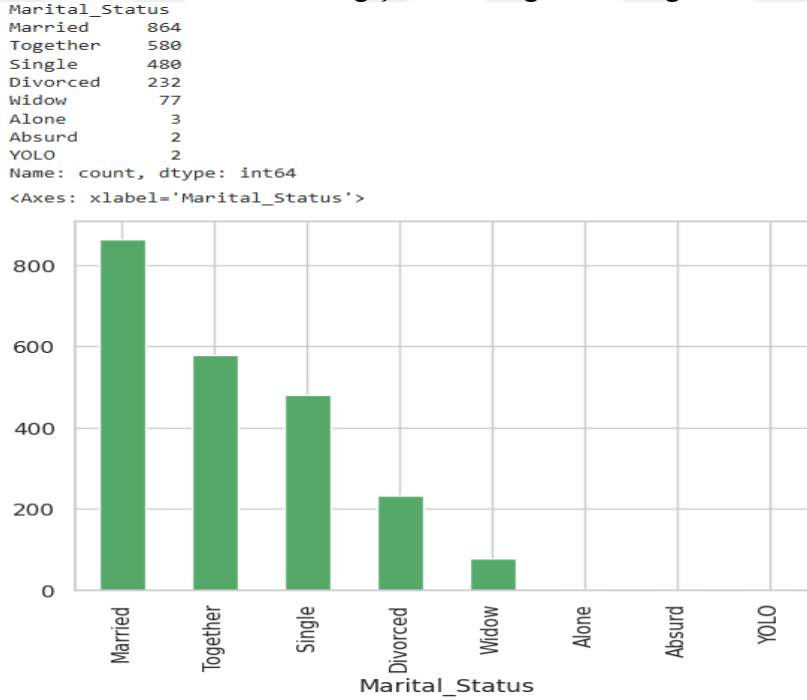
Şekil 21: Yaş Değişkeni Dağılım Grafiği



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Yaş değişkeni, Year_Birth değişkeninden türetilmiştir. Bu tarih formatındaki değişkenden, müşterilerin doğum yılı ve ayı kullanılarak yaşları hesaplanmış ve yeni bir yaş değişkeni oluşturulmuştur. Hesaplamalar, içinde bulunduğumuz yıl göz önünde bulundurularak yapılmıştır. Yaş değişkeni, genel olarak normale yakın bir dağılım sergilemekte ve çoğunlukla müşteriler 45-60 yaş aralığındadır. 80 yaş üzeri yalnızca 2 müşteri bulunmaktadır ve bu durum veri dağılımını kısmen etkilemektedir. Ancak, azınlıkta kaldıkları için bu değerler veri setinden dışarıda bırakılabilir. Araştırılan konu ve model açısından önemli olan değişkenlerin, model performansını artırmak amacıyla aykırı değerlerden arındırılması gerektiği unutulmamalıdır.

Şekil 22: Medeni Durum Değişkeni Histogram Grafiği

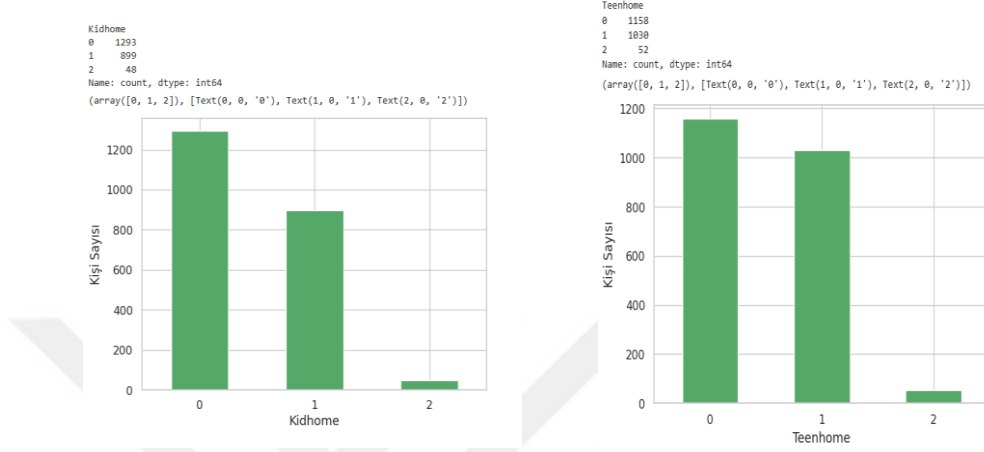


Kaynak: Araştırmacı tarafından oluşturulmuştur.

Veri setimizdeki medeni durum değişkenine göre yapılan incelemeye göre, en fazla müşterinin evli kategorisinde yer aldığı, en az müşterinin ise yolo kategorisinde bulunduğu görülmektedir. Birlikte yaşayanlar ve yalnız olanlar kategorileri ise, müşteri kitlesinin önemli bir kısmını oluşturmaktadır. Özellikle balık, et, tatlı gibi ürünlerin satışa sunulduğu düşünüldüğünde, alışveriş yapan kitlenin daha çok ev

hayatına sahip bireylerden oluştuğu söylenebilir. Evli ve birlikte yaşayanlar kategorilerinde yer alan kişiler, hane halkı açısından daha fazla tüketime sahip oldukları için, şirketin sürekli müşterileri arasında yer alabilirler.

Şekil 23: Çocuk Sayısı Değişkenleri Histogram Grafiği



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Yukarıdaki görselde veri setinde yer alan Kidhome değişkenine göre, 1293 kişinin evinde küçük çocuğu bulunmazken, 899 kişinin 1 ve 48 kişinin 2 küçük çocuğu bulunmaktadır. Teenhome değişkenine ilişkin analizde ise, 1158 kişinin evinde ergen çocuğu olmadığı, 1030 kişinin 1 ergen çocuğa, 52 kişinin ise 2 ergen çocuğa sahip olduğu görülmüştür.

Şekil 24: Müşteri Grupları Dağılımı



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Yukarıdaki görselde, müşterilerin Recency (yenilik) değerlerine göre üç kategoriye ayrılmış dağılımı sunulmaktadır. Aktif Müşteriler, son 30 gün içinde alışveriş yapmış müşterileri ifade ederken, Potansiyel Müşteriler son 31-90 gün arasında alışveriş yapmış müşterileri temsil etmektedir. Kayıp Müşteriler ise 90 günden daha uzun süredir alışveriş yapmamış olan müşterilerden oluşmaktadır. Aktif müşteri sayısı 1318, potansiyel müşteri sayısı 724 kişi ve kayıp müşteri sayısı ise 198 kişiden oluşmaktadır. Grafikten görüleceği üzere, müşterilerin büyük bir kısmı 'Aktif Müşteriler' kategorisinde yer alırken, 'Kayıp Müşteriler' kategorisi daha az bir yoğunluk göstermektedir. Bu bulgular, müşteri sadakatini artırmaya yönelik stratejilerin belirlenmesi ve kayıp müşterilerle yeniden etkileşim kurmaya yönelik çalışmalar için yol gösterici olabilir.

Şekil 25: Toplam Kampanya Sayısı

```
# Her kampanyanın seçilme toplamını hesaplama
campaign_totals = {}
for col in campaign_columns:
    campaign_totals[col] = df[col].sum()
```

Campaign_1 Seçilme Sayısı: 167
Campaign_2 Seçilme Sayısı: 163

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Şekil 26: Aylar Bazında Kampanya Sayıları

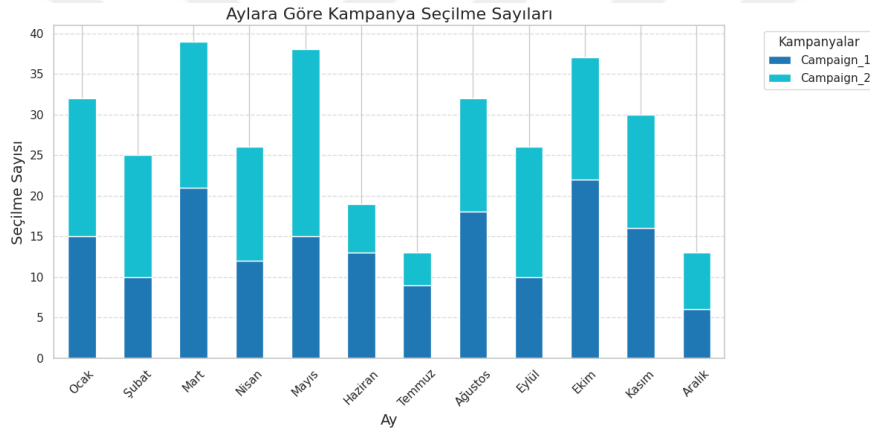
```
# Kampanya bazında seçilme sayısını yazdırma
for campaign, total in campaign_totals.items():
    print(f"{campaign} Seçilme Sayısı: {total}")
```

	Campaign_1	Campaign_2
Month		
1	15	17
2	10	15
3	21	18
4	12	14
5	15	23
6	13	6
7	9	4
8	18	14
9	10	16
10	22	15
11	16	14
12	6	7

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Veri setimizdeki kampanya tercihleri incelendiğinde, Kampanya 1'i 167 kişinin, Kampanya 2'yi ise 163 kişinin seçtiği gözlemlenmiştir. Her bir kampanyanın aylara göre seçilmesini içeren kişi sayıları tek tek incelenmiştir. Bu veriler, her iki kampanyanın da birbirine oldukça yakın bir katılım aldığını ve müşterilerin kampanya seçiminde belirgin bir eğilim göstermediğini ortaya koymaktadır. Burada belirtmek gerekir ki, kampanyalar anonimdir ve her şirketin kendine özgü kampanyalarını oluşturabileceği bir yapıya sahiptir. Bu durum, her şirketin kendi stratejilerine göre özelleştirilmiş kampanyalar tasarlayabileceği anlamına gelmektedir. Kampanya seçimlerinin yanı sıra, bu tercihler zamanla değişiklik gösterebileceğinden, aylara göre kampanya seçimlerinin dağılımı daha ayrıntılı bir şekilde incelenmiştir.

Şekil 27: Kampanya Katılım Durumunun Aylara Göre Dağılımı



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Grafik incelendiğinde, her iki kampanyanın (Kampanya 1 ve Kampanya 2) seçim sayıları arasında benzer bir trend izlediği görülmektedir. Kampanya 1 ve Kampanya 2 arasındaki farklar, her ayda belirli ölçüde değişkenlik göstermektedir, ancak her iki kampanya da yıl boyunca benzer seviyelerde ilgi görmektedir. Grafikte dikkat çeken bir diğer önemli nokta, özellikle Mart, Mayıs ve Eylül aylarında Kampanya 2'nin, diğer aylara kıyasla daha yüksek seçim sayısına sahip olduğudur. Haziran ve Temmuz gibi yaz aylarında ise her iki kampanyanın seçim sayıları daha düşük kalmaktadır. Bu durum, sezonluk etkenlerin veya kampanya içeriğinin müşterilerin tercihlerinde nasıl bir etki yarattığını gösterebilir. Kampanya 1 ise yılın diğer aylarında, özellikle Ocak, Nisan ve Kasım aylarında Kampanya 2 ile oldukça yakın seviyelerde seçim sayıları göstermektedir. Bununla birlikte, Kampanya 1'in bazı

aylarda daha fazla ilgi gördüğü, özellikle Mart ve Ekim gibi bahar aylarında belirgin bir şekilde görülmektedir. Bu grafikteki veriler, kampanya stratejilerini belirlerken zaman dilimlerine göre değişen müşteri ilgisini dikkate almanın önemini vurgulamaktadır. Şirketler, mevsimsel eğilimlere ve kampanyaların aylık performansına göre kampanya içeriklerini optimize edebilir, belirli dönemlerde daha fazla dikkat çekici kampanyalar sunabilirler.

3.2. EKSİK GÖZLEMLER VE AYKIRI DEĞERLERİN GİDERİLMESİ

Eksik gözlemler ve aykırı değerler, veri analizi ve modelleme süreçlerinde sonuçların doğruluğunu ve güvenilirliğini etkileyen önemli unsurlardır. Eksik gözlemler, veri setindeki boş veya eksik değerlerin varlığı nedeniyle analitik süreçleri olumsuz etkileyebilirken, aykırı değerler ise veri setinin genel dağılımından önemli ölçüde sapma gösteren uç noktalar olarak tanımlanır. Bu tür anormalilerin tespit edilmesi ve uygun yöntemlerle düzeltilmesi, modelin performansını artırmak ve daha doğru sonuçlar elde etmek için kritik bir adımdır.

Eksik gözlemlerin yönetiminde, ortalama veya medyan gibi istatistiksel yöntemlerle doldurma, regresyon temelli tahmin yöntemleri veya eksik veri içeren gözlemlerin çıkarılması gibi stratejiler uygulanabilir. Aykırı değerlerin temizlenmesinde ise Z-skoru, IQR (Interquartile Range) yöntemi veya görsel analiz gibi tekniklerle aykırı noktaların belirlenmesi ve gerektiğinde veri setinden çıkarılması ya da dönüştürülmesi sağlanır. Bu adımlar, veri kalitesini artırarak modelleme süreçlerine sağlam bir temel oluşturur.

3.2.1.Eksik Gözlemlerin Giderilmesi

Income değişkeninde tespit edilen 24 adet eksik değer değişken özelinde verilerin sayısı genel veri setine oranla az olduğu için değerlerimizi Income sütunundaki değerlerin genel ortalaması ile doldurduk. Income değişkeninin ortalaması ve medyanı birbirine çok yakın olduğu için değişken içerisinde uç değerler mevcut değildir. Dolayısıyla değişkenin genel yapısının etkilenmeyeceği göz önüne alınarak verilerin ortalama ya da medyan değerleriyle doldurulmasında bir

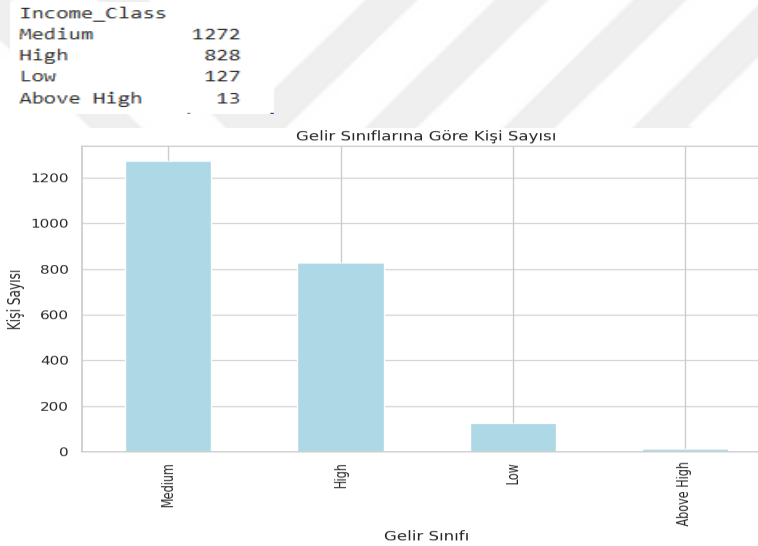
sakınca görülmemektedir. Aşağıdaki kod satırıyla eksik değerleri ortalama ile dolduruyoruz.

Şekil 28: Eksik Gözlemlerin Giderilmesi

```
df["Income"].fillna(df["Income"].mean(),inplace=True)
```

Eksik değerleri ortalama ile doldurduktan sonra 24 adet olarak gözlemlenen eksik değer sayısı sıfır olmuştur. Bu durumda gelir değişkeni sınıflara ayrılarak alışveriş yapan müşteri kitlesinin gelir durumları ve gelir sınıfları hakkında bilgi sahibi oluruz. Bu inceleme analiz için çok önemli iç görüler elde edilmesini sağlar.

Şekil 29: Gelir Sınıfları Histogram Grafiği

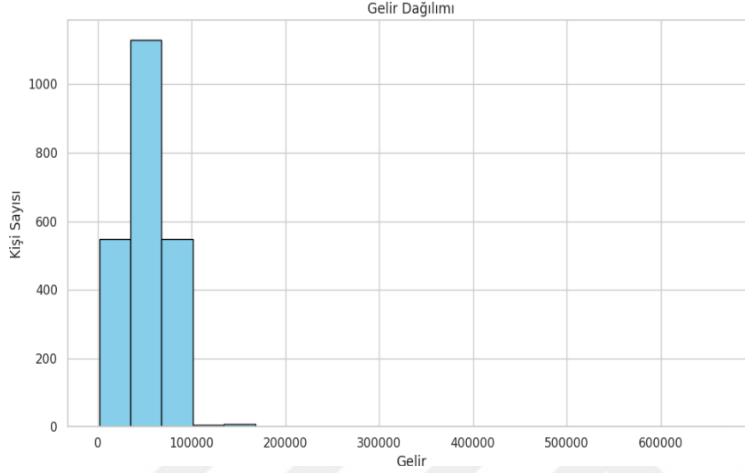


Kaynak: Araştırmacı tarafından oluşturulmuştur.

Veri setimizdeki gelir değişkeni incelendiğinde 4 ayrı kategoriden oluştuğu görülmektedir. Kategoriler veri seti içerisinde gelir aralıklarına göre elde edilmiştir. Bu sınıflandırma için gelir aralıkları 0-20,000 arası Low, 20,000-60,000 arası Medium, 60,000-100,000 High ve 100,000'in üzeri Above High olarak adlandırılmıştır. Income değişkeni, bu aralıklara göre Income_Class adlı yeni bir kategorik değişkene dönüştürülmüştür. Her müşteri, gelir seviyesine göre uygun sınıfa atanmıştır. Veri seti içerisinde en çok Medium kategorisi yer alırken en az Above High kategorisi yer almaktadır. Dolayısıyla 1272 müşterinin gelir durumu aylık 20,000 ile 60,000 arasında seyretmektedir. 13 müşteri ise 100,000 üzerinde

gelir durumuna sahiptir. High kategorisine ait 828 müşterinin aylık geliri 60,000 ile 100,000 arasındadır. Veri setimizde 0 ile 20,000 arasında olan Low kategorisinde 127 kişi bulunmakta ve azınlığı oluşturmaktadır. Buradan alışveriş yapan müşterilerimizin çoğunun iyi bir gelir düzeyine sahip olduğu çıkarımı yapılabilir. Değişkenin genel dağılım yapısına bakılarak daha net anlaşılabacaktır.

Şekil 30: Gelir Dağılımı



Kaynak: Araştırmacı tarafından oluşturulmuştur.

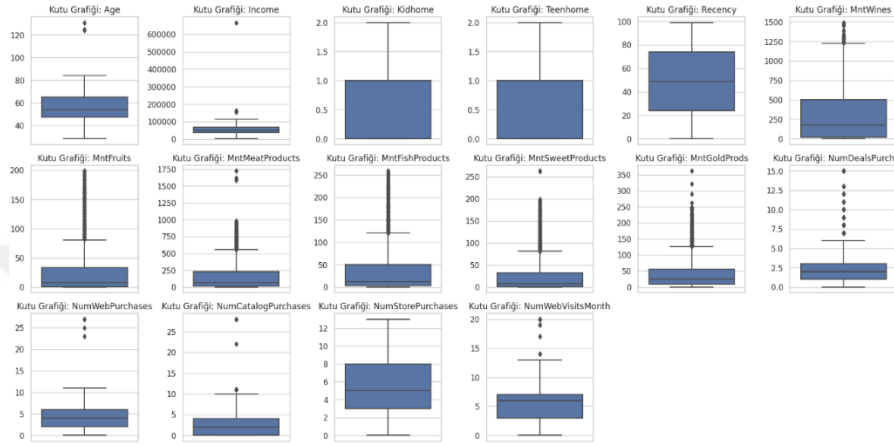
Yukarıdaki histogram grafiği, gelir dağılımını göstermektedir. Grafikten görüldüğü üzere, verilerin büyük bir kısmı 0 ile 100.000 aralığında yoğunlaşmış olup, daha yüksek gelir seviyelerinde gözlem sayısı oldukça azdır. Ayrıca, 100.000 ve üzerindeki değerler, veri setinde aykırı değerler olabileceğini düşündürmektedir. Bu durum, gelir dağılımında asimetric bir yapı ve potansiyel aykırı değerlerin varlığına işaret etmektedir. Aykırı değerlerin doğrulanması ve gerekli temizleme işlemlerinin yapılması, analizin doğruluğunu artırmak için önemlidir.

3.2.2 Aykırı Değerleri Temizlenmesi

Veri setimizde bazı değişkenlerde aykırı değerlerin bulunabileceği tespit edilmiştir. Bu durumu daha ayrıntılı bir şekilde incelemek için Box-plot yöntemi kullanılmıştır. Aykırı değerler, model performansı açısından büyük bir öneme sahiptir ve bu tür veriler modelin doğruluğunu olumsuz şekilde etkileyebilir. Bu nedenle, aykırı değerlerin tespiti ve giderilmesi kritik bir adımdır. Box-plot grafiği,

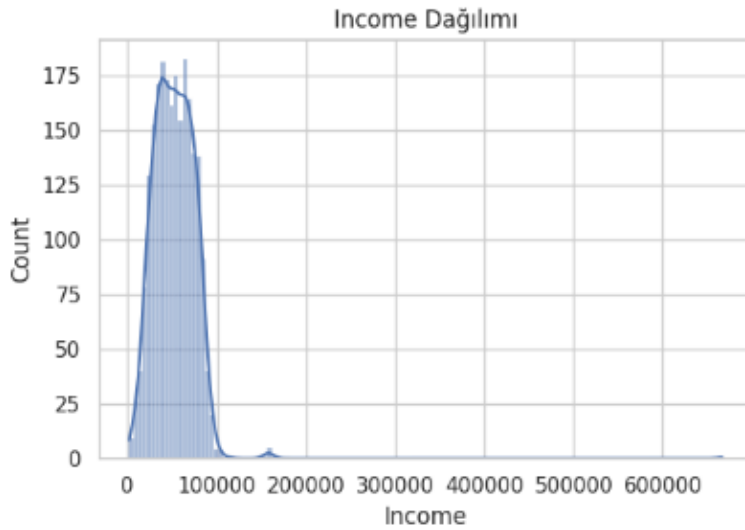
veri setindeki aykırı değerleri belirlemede etkili bir araçtır; çünkü box-plot üzerinde aykırı değerler genellikle nokta olarak gösterilir. Bu sayede, aykırı değerlerin nerelerde yoğunlaştığı görsel olarak analiz edilebilir ve modelin doğruluğunu artıracak önlemler alınabilir.

Şekil 31: Box-Plot Grafiği ile Aykırı Değerlerin Gösterimi



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Şekil 32: Gelir Değişkeninin Dağılımı



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Yukarıdaki box-plot grafiğinde görüldüğü üzere 9 değişkenimizde aykırı değerler mevcuttur. Aykırı değerler, bir veri setindeki diğer gözlemlerden belirgin şekilde farklı olan veri noktalarıdır. Bu tür değerler, veri analizinde ve model

oluşturma süreçlerinde çeşitli sorunlara neden olabilir. Özellikle ortalama, varyans gibi istatistiksel ölçümleri yanıltabilir ve veri modellemelerinde hatalı sonuçlara yol açabilirler. Bu nedenle aykırı değerlerin belirlenmesi ve uygun şekilde ele alınması büyük önem taşır. Bu aykırı değerlerden kurtulmak için aşağıdaki kod satırında yazılan fonksiyon ile IQR (Interquartile Range) yöntemi kullanılır.

Şekil 33: Aykırı Değerlerin Temizlenmesi

```
def outlier_handle(df, columns):  
    for column in columns:  
        # IQR hesapla  
        Q1 = df[column].quantile(0.25)  
        Q3 = df[column].quantile(0.75)  
        IQR = Q3 - Q1  
  
        # Aykırı değerlerin alt ve üst sınırlarını hesapla  
        lower_bound = Q1 - 1.5 * IQR  
        upper_bound = Q3 + 1.5 * IQR  
  
        # Aykırı değerleri alt ve üst sınırlara eşitle  
        df[column] = df[column].apply(lambda x: lower_bound if x < lower_bound else upper_bound if x > upper_bound else x)  
  
    return df
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

IQR, verinin ortadaki %50'lik bölümünü temsil eden çeyrekler arası aralığı ifade eder. Bu yöntem, aykırı değerleri tespit etmek için sıkça kullanılır. IQR yöntemiyle, verinin 1. ve 3. çeyrekleri arasındaki fark hesaplanır ve bu fark kullanılarak aykırı değerlerin tespit edilmesi için alt ve üst sınırlar belirlenir. Q3 ile Q1 arasındaki fark alınarak hesaplanır ve bu değer veri setinin ortadaki %50'sindeki değerlerin dağılımını ölçer. Kısaca IQR, veri setinin ilk çeyreği ile son çeyreğini kapsayan değerlerin medyanıdır.

$IQR = Q3 - Q1$ şeklinde ifade edilir.

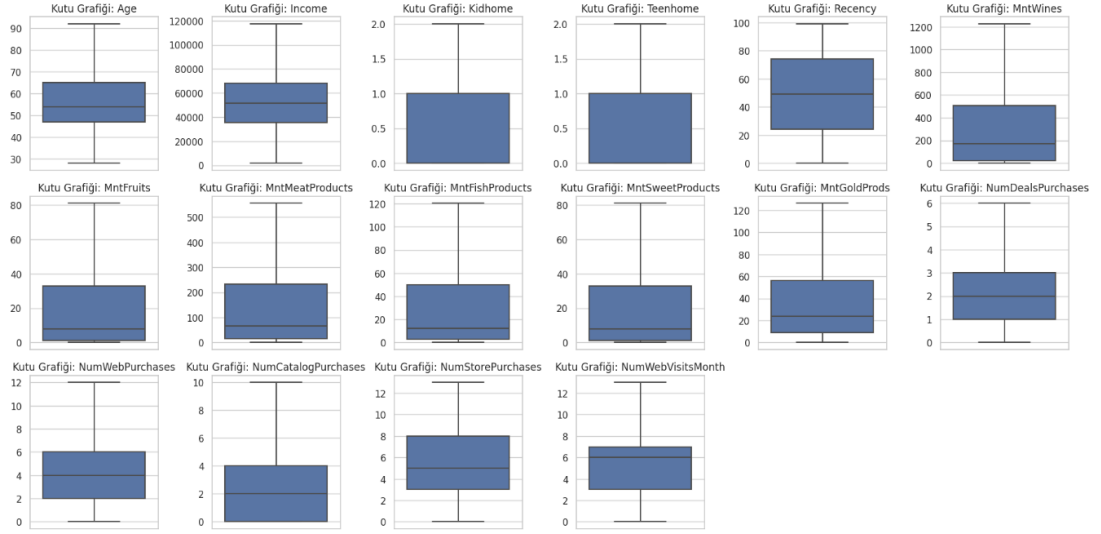
Aykırı değerlerin sınırlarını ise alt sınır ve üst sınır belirleyerek yapar. Eğer bir veri noktası alt veya üst sınırların dışında kalıyorsa o veri noktası aykırı değer olarak kabul edilir.

Alt Sınır: $Q1 - 1,5 * IQR$ Üst Sınır: $Q3 + 1,5 * IQR$

Böylelikle IQR yöntemi ile alt sınırın altında olan değerler alt sınırı üst sınırın üstünde olan değer üst sınırı eşitlendi. IQR yöntemi hesaplaması kolay ve anlaşılır bir yöntemdir. Ortalama ve standart sapma gibi ölçülere göre uç değerlere karşı daha az duyarlıdır. Çünkü medyan ve çeyrek değerler kullanılarak hesaplanır.

Aykırı değerlere sahip olduğu tespit edilen 9 değişken IQR yöntemi ile temizlenmiştir. Aykırı değerlerden temizlenen veri setindeki değişkenlerin box-plot grafikleri aşağıdaki gibidir.

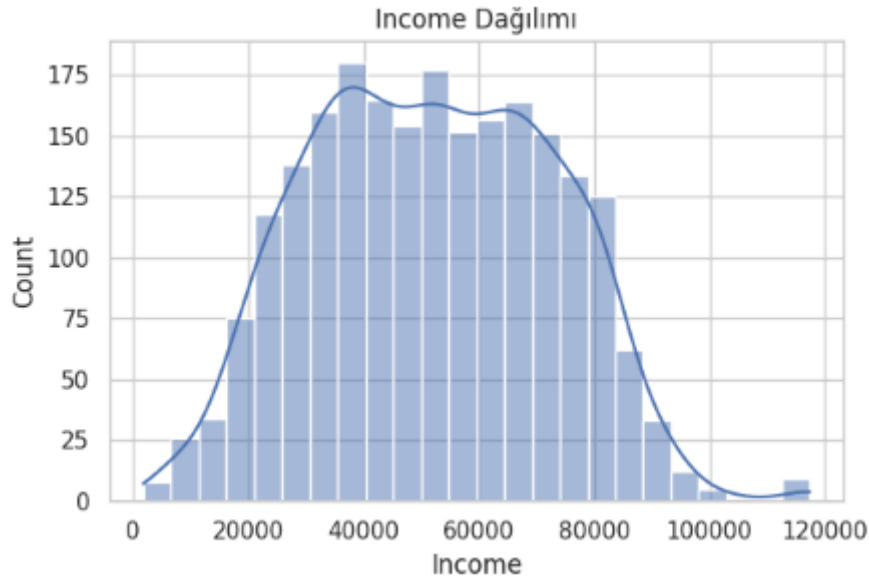
Şekil 34: Box-Plot Grafiği ile Aykırı Değerler Temizlendikten Sonra Gösterim



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Görüldüğü gibi veri setinde artık aykırı değerler söz konusu değildir. Bu durum modelin daha doğru çalışmasını sağlayacaktır. Aykırı değerlerden arındırılan Income değişkeni normal dağılıma yaklaşmıştır. Veri setinin aykırı değerlerden arındırılması işleminden sonra normal dağılıma yaklaşan Income değişkeni aşağıdaki gibidir.

Şekil 35: Aykırı Değerler Temizlendikten Sonra Gelir Değişkeni Dağılımı



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Gelir deęiřkeninin daęılımını gösteren histogram grafięinden de anlařılacaęı gibi müşterilerimizin aylık geliri 40,000 ile 60,000 arasında yoğunlařmaktadır. Normal daęılıma yakın olan bu grafikten müşterilerimizin çoęunun gelir durumunun ortalama olarak birbirine yakın olduęunu söyleyebiliriz.

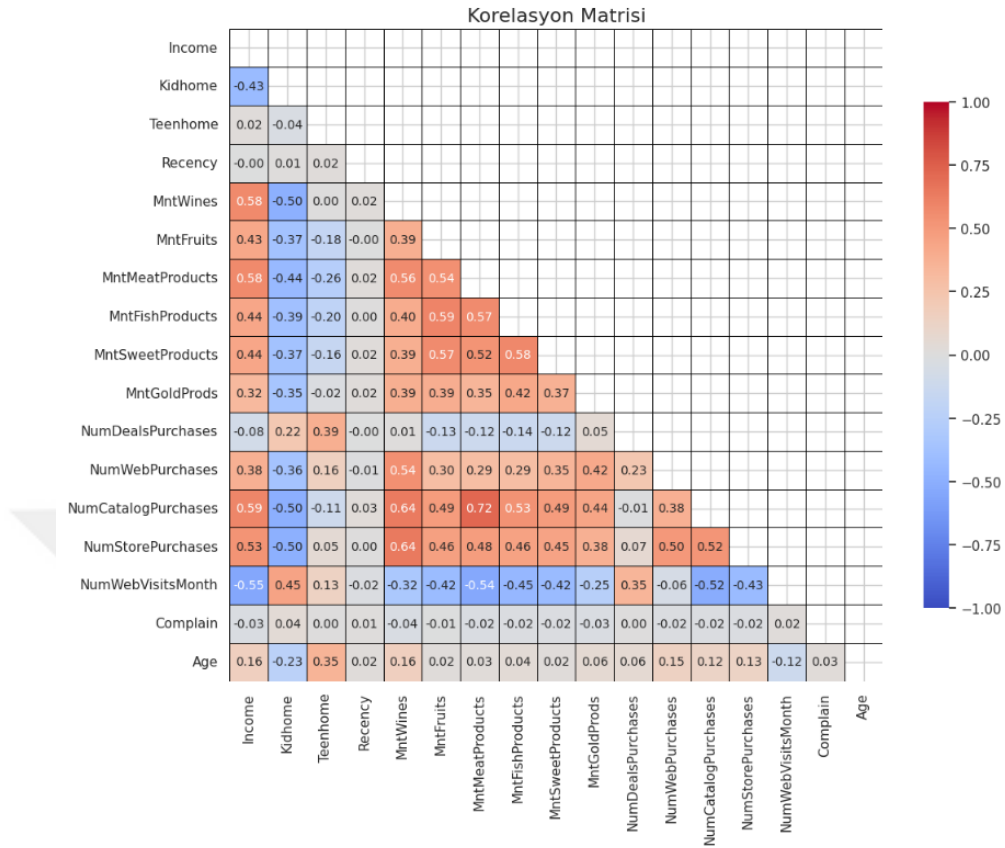
Gelir sınıflarına göre yapılan bu daęılım, müşteri segmentasyonu açısından önemli bir veri sunmakta olup, řirketlerin farklı gelir düzeylerine yönelik stratejiler geliřtirmelerine yardımcı olabilir. Bu tür bir analiz, müşteri gruplarının alışveriş alışkanlıkları ve tüketim davranıřları üzerinde etkili olabilir, böylece daha hedeflenmiř kampanyalar oluřturulabilir.

3.4. KORELASYON ANALİZİ

Korelasyon, iki deęiřken arasındaki doęrusal iliřkinin yönünü ve gücünü gösteren bir istatistiksel ölçüdür. Korelasyon katsayısı genellikle -1 ile $+1$ arasında bir deęer alır:

- Pozitif korelasyon ($0 < r \leq 1$): Bir deęiřken artarken dięerinin de arttıęını ifade eder.
- Negatif korelasyon ($-1 \leq r < 0$): Bir deęiřken artarken dięerinin azaldıęını belirtir.
- Korelasyon yok ($r \approx 0$): İki deęiřken arasında anlamlı bir iliřki olmadıęını gösterir.

Şekil 36: Korelasyon Analizi



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Bu grafikte değişkenler arasındaki doğrusal ilişkilerin korelasyon katsayıları gösterilmektedir. Pozitif değerlere sahip katsayılar değişkenler arasında doğru yönlü ilişki olduğunu, negatif katsayılar ise ters yönlü ilişkiyi ifade eder. Grafikte renk skalası kullanılarak korelasyon değerleri görsel olarak desteklenmiştir. Kırmızı tonları pozitif korelasyonları, mavi tonları ise negatif korelasyonları göstermektedir.

Gelir (Income) değişkeni ile MntWines değişkeni arasında 0.58 gibi yüksek pozitif bir korelasyon bulunmaktadır. Bu durum, geliri yüksek bireylerin şarap harcamalarının daha fazla olduğunu göstermektedir. Benzer şekilde gelir değişkeni ile MntMeatProducts değişkeni arasında 0.58 , NumCatalogPurchases değişkeni ile 0.59 güçlü pozitif korelasyonlar gözlenmiştir. Gelir seviyesi arttıkça et harcamalarının ve satın almaların arttığı görülmektedir.

Kidhome (evdeki çocuk sayısı) ile Income arasında -0.43 gibi anlamlı bir negatif korelasyon görülmektedir. Bu, çocuk sayısının artmasının gelir seviyesini sınırlandırabileceğini göstermektedir.

Kidhome ve harcama değişkenleri arasında negatif korelasyonlar gözlenmiştir. Özellikle MntWines (-0.50), NumCatalogPurchase (-0.50) ve NumStorePurchase (-0.50) gibi değişkenlerle ters yönlü ilişkiler, çocuk sayısı arttıkça lüks ve gıda harcamalarının azaldığını göstermektedir.

Buradan dikkat edilecek olursa çocuk sayısı ile en yüksek pozitif korelasyona sahip olan değişken web site ziyareti sıklığı değişkeni olduğu görülebilir. Aynı şekilde çocuk sayısı değişkeninin indirimli alışverişlerle de diğer değişkenlere nazaran güçlü olmayan pozitif bir ilişkisi vardır. Dolayısıyla evde çocuğu olan müşterilerin web siteyi sürekli ziyaret ederek indirim olduğu zamanlarda alışverişlerini gerçekleştiren müşteriler olduğu görülür.

NumWebVisitsMonth değişkeni ile harcama değişkenleri arasında güçlü negatif korelasyonlar dikkat çekmektedir. Özellikle MntFruits (-0.42), MntMeatProducts (-0.54) ve MntFishProducts (-0.45) gibi harcama kalemleri ile belirgin ters yönde ilişkiler gözlemlenmiştir. Bu durum, web sitesi ziyaret sayısının artmasının harcamalarla ters yönde bir ilişki gösterdiğini ortaya koymaktadır. Yani, sık web ziyareti yapan bireylerin daha düşük satın alma eğiliminde oldukları söylenebilir.

NumCatalogPurchases değişkeni ile harcama değişkenleri arasında ise yüksek pozitif korelasyonlar bulunmuştur. Örneğin, MntMeatProducts (0.72), MntFishProducts (0.53) ve MntWines (0.64) gibi harcama kategorileriyle güçlü ilişkiler mevcut olup, bu da katalog üzerinden yapılan satın alımların harcama tutarlarını doğrudan etkilediğini göstermektedir.

Yaş değişkeni ile harcama değişkenleri arasında zayıf negatif korelasyonlar gözlemlenmiştir. Özellikle MntWines (-0.16) ve Teenhome (-0.35) gibi harcama kalemlerinde yaşın ilerlemesiyle ve çocuk sahibi olmakla birlikte harcama miktarlarının azaldığı görülmektedir. Diğer yandan, Income ile Age arasındaki 0.16 değeri, yaş ile gelir arasında hafif pozitif bir ilişki olduğunu ortaya koymaktadır.

Özetle, bu korelasyon matrisi, gelir seviyesinin bireylerin harcama alışkanlıkları üzerindeki etkisini net bir şekilde ortaya koymaktadır. Gelir arttıkça, özellikle lüks tüketim ve katalog alışverişlerinde bir artış gözlenirken, çocuk sayısı ve web ziyaret sıklığının harcamaları sınırlayıcı bir etkiye sahip olduğu anlaşılmaktadır. Yaşın ilerlemesi ise harcamaları azaltırken, gelir ile zayıf bir pozitif

ilişki gösterdiği tespit edilmiştir. Bu bulgular, bireylerin tüketim alışkanlıkları ve gelir seviyelerine göre satın alma eğilimlerini anlamak açısından önemli ipuçları sunmaktadır.

Complain ve Recency değişkenlerinin tüm değişkenlerle nötr bir ilişkisi olması sebebiyle model aşamasında modele dahil edilmeyebilir.

3.5. VERİ AYIRMA, ÖLÇEKLENDİRME VE KATEGORİK VERİLERİN İŞLENMESİ

Veri madenciliği ve makine öğrenimi süreçlerinde, veri hazırlığı kritik bir aşamadır ve bu süreçte veri ayırma, ölçeklendirme ve kategorik verilerin işlenmesi önemli adımlar arasında yer alır. Veri ayırma, modeli eğitmek ve değerlendirmek için veri setini eğitim, doğrulama ve test alt kümelerine bölmeyi içerir. Genellikle %70-80'lik bir kısım eğitim, geri kalanı ise test ve doğrulama için ayrılır. Bu yöntem, modelin genelleme kabiliyetini ölçmede büyük önem taşır. Ölçeklendirme ise verilerdeki değişkenlerin farklı birimlerde olması nedeniyle modelin performansını etkileyebilecek tutarsızlıkları gidermeyi amaçlar. Bunun için Min-Max Normalizasyonu (değerlerin $[0, 1]$ aralığına sıkıştırılması) ve Z-Skor Standardizasyonu (verilerin ortalamasının 0, standart sapmasının 1 olacak şekilde yeniden ölçeklendirilmesi) gibi yöntemler kullanılır. Bunun yanı sıra, kategorik değişkenlerin işlenmesi için kullanılan One-Hot Encoding (Tekil Kodlama), bu tür verilerin model tarafından anlaşılabilir hale getirilmesini sağlar. One-Hot Encoding, kategorik bir değişkenin her bir sınıfını yeni bir sütun olarak temsil eder ve bu sütunlara 0 veya 1 değerlerini atar. Veri ayırma, ölçeklendirme ve One-Hot Encoding gibi veri hazırlık adımları, modelleme sürecinin doğruluğunu ve güvenilirliğini artıran temel hazırlık işlemleridir.

3.5.1. One-Hot Encoding İşlemleri

OneHotEncoder, kategorik verileri ikili (binary) vektörlere dönüştürmek için kullanılan bir kodlama tekniğidir. Bu teknik, modelin kategorik değişkenleri anlamasına ve işlemesine yardımcı olur. Her kategori için ayrı bir sütun oluşturur, bu da modelin kategorik ilişkileri öğrenmesini sağlar. Bu yöntem, özellikle doğrusal

modeller ve gradyan tabanlı algoritmaların daha verimli çalışmasını sağlar.

Veri setini tekrar gözden geçirdiğimizde veri setinde bağımlı değişkenimiz hariç 4 adet değişkenin sayısal formatta olmadığı görülmüştür. Eğitim durumunu ifade eden Education değişkeni, medeni durumu ifade eden Marital_Status değişkeni, gelir sınıflarını ifade eden Income_Class değişkeni ve değerli müşteri gruplarını içeren Recency_Category değişkenleri kategorik veriler içermekte olup gerekli dönüşümlerin yapılması gerekmektedir. Her bir gözlem, ilgili kategoriye ait sütunda 1 olarak işaretlenir, diğer sütunlarda ise 0 olarak kalır. Sınıflar arasında sıralama ve öncelik olmadığından one-hot encoding yöntemi ile kategorik değişkenlerin dönüşümü yapılır.

Şekil 37: Veri Setindeki Değişken Tiplerinin Gösterimi

```
# Column Non-Null Count Dtype
---
0 Education 2240 non-null object
1 Marital_Status 2240 non-null object
2 Income 2240 non-null float64
3 Kidhome 2240 non-null int64
4 Teenhome 2240 non-null int64
5 Recency 2240 non-null int64
6 MntWines 2240 non-null float64
7 MntFruits 2240 non-null float64
8 MntMeatProducts 2240 non-null float64
9 MntFishProducts 2240 non-null float64
10 MntSweetProducts 2240 non-null float64
11 MntGoldProds 2240 non-null float64
12 NumDealsPurchases 2240 non-null float64
13 NumWebPurchases 2240 non-null float64
14 NumCatalogPurchases 2240 non-null float64
15 NumStorePurchases 2240 non-null int64
16 NumWebVisitsMonth 2240 non-null float64
17 Complain 2240 non-null int64
18 selectedcampaign 2240 non-null object
19 Age 2240 non-null float64
20 Income_Class 2240 non-null category
21 Recency_Category 2240 non-null object
dtypes: category(1), float64(12), int64(5), object(4)
memory usage: 370.0+ KB
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Şekil 38: Kategorik Değişkenlerin Gösterimi

	Education	Marital_Status	Income_Class	Recency_Category
0	Graduation	Single	Medium	Potansiyel Müşteriler
1	Graduation	Single	Medium	Potansiyel Müşteriler
2	Graduation	Together	High	Aktif Müşteriler
3	Graduation	Together	Medium	Aktif Müşteriler
4	PhD	Married	Medium	Kayıp Müşteriler
...	...	...	...	...
2235	Graduation	Married	High	Potansiyel Müşteriler
2236	PhD	Together	High	Potansiyel Müşteriler
2237	Graduation	Divorced	Medium	Kayıp Müşteriler
2238	Master	Together	High	Aktif Müşteriler
2239	PhD	Married	Medium	Potansiyel Müşteriler

2240 rows x 4 columns

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Değişken dönüşümünü gerçekleştirirken yazılan kod bloğu aşağıdaki gibidir. Bu sayede kategorik verilerimiz her biri bir satırda olacak şekilde veri setine yeni sütunlar eklenmiş ve modelin anlayacağı formata dönüştürülmüştür.

Şekil 39: One-Hot Encoding İşlemi Uygulaması

```
#One-Hot Encoder İşlemleri
encoder = OneHotEncoder(sparse=False)
encoded_data = encoder.fit_transform(df[["Education", "Marital_Status", "Income_Class", "Recency_Category"]]) #İlgili sütunları dönüştürme
encoded_df = pd.DataFrame(encoded_data, columns=encoder.get_feature_names_out(["Education", "Marital_Status", "Income_Class", "Recency_Category"]))
df_encoded = pd.concat([df.drop(["Education", "Marital_Status", "Income_Class", "Recency_Category"], axis=1), encoded_df], axis=1)
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Şekil 40: One-Hot Encoding Uygulaması Sonrası Değişkenlerin Gösterimi

	Education_2n Cycle	Education_Basic	Education_Graduation	Education_Master	Education_PhD	Marital_Status_Absurd	Marital_Status_Alone	Marital_Status_Di
0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
1	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
2	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
3	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
4	0.0	0.0	0.0	0.0	1.0	0.0	0.0	
...	...	...	...	...	...	...	...	...
2235	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
2236	0.0	0.0	0.0	0.0	1.0	0.0	0.0	
2237	0.0	0.0	1.0	0.0	0.0	0.0	0.0	
2238	0.0	0.0	0.0	1.0	0.0	0.0	0.0	
2239	0.0	0.0	0.0	0.0	1.0	0.0	0.0	

2240 rows x 20 columns

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Veri hazırlama sürecinde, One-Hot Encoding işlemleri başarıyla tamamlanmıştır. Ancak, Selected Campaign değişkeni hala object formatında olup, makine öğrenmesi modeli oluşturulmadan önce bu değişkenin uygun bir formata dönüştürülmesi gerekmektedir. Bu dönüşüm yapılmadığı takdirde, özellikle Grid Search CV gibi yöntemlerde hatalar ortaya çıkabilir. Çünkü, çok sınıflı sınıflandırma problemi bağlamında, modelin bu tür bir değişkenle çalışabilmesi için doğru bir formatlama zorunludur. Bu bağlamda, Selected Campaign değişkeni data frame formatına dönüştürme işlemi sonrasında veriler daha düzenli ve analiz edilebilir hale gelmiştir. Makine öğrenmesi modelleri genellikle tablo benzeri yapılarla çalıştığı için, verinin data frame formatında sunulması hem modelin giriş verisi olarak kullanılabilirliği açısından hem de veri işleme adımlarının uygulanabilirliği açısından önemlidir. Dönüşüm işlemleri tamamlandıktan sonra, veriler ölçeklendirme (scaling) adımıyla işlenerek modelleme için hazır hale getirilecektir. Bu süreç, modelin doğruluğunu artırmak ve hatasız bir analiz gerçekleştirmek adına kritik bir aşamadır.

Aşağıdaki kod bloğu yardımıyla çok değişkenli sınıflandırma problemi içeren hedef değişkenim selectedcampaign liste formatına dönüştürüldü.

Şekil 41: Hedef Değişkeni Liste Formatına Çevirmek

```
df_encoded["selectedcampaign"] = df_encoded["selectedcampaign"].apply(  
    lambda x: [int(i) for i in x] |  
)
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Dönüşüm sonrası hedef değişkenim olan selectedcampaign [0,0], [0,1], [1,0] , [1,1] değerlerini içeren liste formatında bir çıktı döndürür. Liste içerisindeki 1. indeks kampanya 1 'i temsil ederken 2. indeks ise kampanya 2' i temsil etmektedir.

Liste formatına dönüştürdüğümüz hedef değişkenimizin data frame formatına dönüştürüldüğü kod bloğu aşağıdaki gibidir.

Şekil 42: Yeni Oluşturulan Değişkenin Encode Edilmiş Data İle Birleştirilmesi

```
new_df = pd.DataFrame(df_encoded["selectedcampaign"].tolist(), columns=[f'campaign_{i+1}' for i in range(2)])
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Şekil 43: Dönüşüm Sonrası Hedef Değişken Gösterimi

	campaign_1	campaign_2
0	0	0
1	0	0
2	0	0
3	0	0
4	0	0
...	...	...

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Label encoder işlemi yapılan değişkenim daha sonra encode edilen veri setinin devamına eklenmiştir. Dönüşüm sonrası oluşturulan selected campaign değişkenlerinden ilki kampanya 1 değişkeninin kabul edilip edilmemesi durumunu 0 ve 1'ler ile temsil ederken diğeri ise kampanya 2'nin kabul edilip edilmediğini gösteriyor. Oluşturulan bu yeni data frame değişkeni orijinal dataya pd.concat() fonksiyonu her biri bir sütunu ifade edecek şekilde eklenmiştir.

3.5.2. Veri Ölçeklendirme ve Eğitim- Test Verisine Ayırma

Modeli ölçeklendirmek tüm özelliklerin aynı aralıkta olmasını sağlayarak her bir özelliğin modele eşit katkı sağlamasına yardımcı olur. Makine öğrenmesi

modellerinde oldukça kritik bir adımdır, çünkü özelliklerin farklı ölçü birimlerinde ve aralıklarda olması, modelin öğrenme sürecinde dengesizliğe neden olabilir. Ölçeklendirilmiş verilerde modelin eğitim süreçleri daha stabil olur ve model daha iyi genelleme yapabilir. Özelliklerin farklı ölçeklerde olması optimizasyon sürecinin dengesiz ve yavaş ilerlemesine yol açabilir. Özellikle bazı algoritmalar, verilerin aynı ölçeklerde olmasını bekler, aksi takdirde bazı özellikler daha fazla ağırlık kazanabilir, bu da modelin doğru öğrenmesini engelleyebilir.

Ölçeklendirilmiş veriler, özelliklerin birbirine eşit şekilde katkı sağlamasına imkan tanır ve modelin daha dengeli bir şekilde öğrenmesini sağlar. Bu da modelin daha hızlı ve verimli eğitim almasına olanak tanır. Örneğin, gradyan tabanlı algoritmalar (lojistik regresyon, sinir ağları) daha stabil ve hızlı çalışırken, mesafe tabanlı yöntemler (KNN, K-means) daha doğru sonuçlar verebilir.

Modeli eğitim ve test veri seti olarak ikiye ayırmak, makine öğrenimi sürecinde modelin gerçek dünyadaki performansını değerlendirebilmek için kritik bir yöntemdir. Bu işlemde, veri seti rastgele seçilerek eğitim ve test seti olarak iki ayrı kısma bölünür. Eğitim seti, modelin öğrenme sürecinde kullanılır ve modelin verilerdeki desenleri ve ilişkileri kavramasını sağlar. Test seti ise modelin doğruluğunu ve genelleme yeteneğini bağımsız bir veri kümesinde sınamak için ayrılır. Bu ayırım, modelin yeni verilerle karşılaştığında nasıl performans göstereceğini anlamamıza yardımcı olur ve aşırı öğrenme (overfitting) veya yetersiz öğrenme (underfitting) gibi problemleri tespit etmeyi mümkün kılar. Eğitim ve test setleri arasındaki bu stratejik bölünme, modelin güvenilirliğini ve genel geçerliliğini sağlamak için temel bir adımdır. Aşağıda veri setinin MinMaxScaler fonksiyonuyla ölçeklenip eğitim ve test veri seti olarak ikiye ayrıldığı kod bloğu yer almaktadır.

Şekil 44: Verilerin Ölçeklendirme ve Eğitim- Test Veri Setlerinin Ayrılması

```
# MinMaxScaler ile bağımsız değişkenleri ölçeklendirme
scaler = MinMaxScaler()

# Bağımsız değişkenler (features)
X = df_encoded.drop(["selectedcampaign"], axis=1)

# Bağımsız değişkenleri ölçeklendiriyoruz
X_scaled = scaler.fit_transform(X)

# Veriyi eğitim ve test olarak ayırıyoruz
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=45)
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Bu kod bloğunda, makine öğrenmesi modeli oluşturulmadan önce veriyi hazırlamak için gerekli ölçeklendirme ve veri ayırma işlemleri gerçekleştirilmiştir. Öncelikle, veri setindeki bağımsız değişkenler (features) ve bağımlı değişkenler (target) belirlenmiştir. Bağımsız değişkenler, modele girdi sağlayan özellikleri temsil ederken, bağımlı değişkenler sınıflandırma ya da regresyon amacıyla modelin tahmin edeceği çıktılarıdır. Bu çalışmada bağımlı değişkenler, kampanya sınıflarını temsil eden iki ayrı sütun olarak tanımlanmıştır (campaign_1 ve campaign_2).

Bağımsız değişkenler, farklı ölçeklerde değerlere sahip olabileceği için MinMaxScaler kullanılarak [0, 1] aralığına ölçeklendirilmiştir. Bu adım, özellikle gradyan tabanlı algoritmaların daha hızlı ve kararlı bir şekilde öğrenme sağlaması için gereklidir. Ölçeklendirme işlemi, df_encoded adlı veri çerçevesinden selectedcampaign değişkeninin çıkarılmasıyla elde edilen bağımsız değişkenlere uygulanmıştır.

Daha sonra, ölçeklendirilmiş bağımsız değişkenler ve bağımlı değişkenler, eğitim ve test alt kümelerine ayrılmıştır. Bu işlem, train_test_split fonksiyonu ile %80 eğitim ve %20 test oranında gerçekleştirilmiş ve işlemin rastgeleliğini kontrol etmek için bir sabitlik parametresi (random_state=45) kullanılmıştır. Eğitim verisi, modelin öğrenme sürecinde kullanılacakken, test verisi modelin genelleme kabiliyetini değerlendirmek için kullanılacaktır. Çoklu çıktı desteğine sahip bir problem olduğu için, bağımlı değişkenler doğrudan iki sütunlu bir formatta bırakılmıştır. Bu sayede, çok sınıflı kampanya tahmini yapmaya yönelik algoritmalar için uygun bir veri yapısı sağlanmıştır.

3.5.3. Hiperparametre Optimizasyonu

Doğru hiper parametreler modelin genelleme yeteneğini artırır ve performans ölçütlerini iyileştirir. Öğrenme oranı arttıkça model daha iyi sonuçlar vereceği ve gerçeğe yakınlacağı için model başarılı olur. Modellerin daha iyi sonuçlar vermesi açısından en iyi hiperparametre değerlerini belirleyip modelleri en optimum değerlerle çalıştırmak model geçerliliği açısından oldukça önemlidir. GridSearch, hiper parametrelerin belirli ve sınırlı bir değer aralığında olduğu durumlarda idealdir. Bu yöntem, tüm olası kombinasyonları sistematik olarak test ederek, her bir

kombinasyonun model performansı üzerindeki etkisini analiz eder. Özellikle küçük veri setleri veya hızlı eğitilebilen modellerle çalışırken, GridSearch'ün tüm kombinasyonları denemesi pratik ve zaman açısından verimli olabilir. Mevcut bir modelin performansını iyileştirmek amacıyla en uygun hiperparametreleri bulmak için GridSearch sıkça tercih edilir. GridSearchCV, belirlenen hiperparametre

aralığında tüm olasılıkları dener ve en iyi performansı sağlayan parametreleri tespit eder. Bu durum aynı zaman da modelin ezberlemesinin (overfit) ya da veriyi hiç anlayamamasının (underfit) önüne geçmek içinde yapılır. Hedef değişkenimiz çok sınıflı sınıflandırma problemine sahip olduğu için birden fazla bağımlı değişkeni aynı anda tahmin etmek istediğimizde modelin parametreleri biraz değişebilir. Bunun nedeni, her bağımlı değişkenin ayrı bir tahmin yapacak olmasıdır. Çoklu sınıflandırma problemlerinde (multi-class classification) parametrelerin çoğu yine aynı olacaktır ancak bazı parametreler özel olarak çoklu çıktı modelleri için uyarlanabilir.

Tez içerisinde kullanılan Random Forest, Desicion Tree, SVC, Xgboost ve CatBoost algoritmaları için Grid Search Cv yöntemi kullanılarak hiperparametre optimizasyonları yapılmıştır. Aşağıda, Random Forest algoritması için kullanılacak hiperparametrelerin seçimi amacıyla yazılmış kod bloğu görülmektedir.

Şekil 45: Random Forest Algoritması için Hiperparametre Optimizasyonu

```
param_grid = {
    'estimator__n_estimators': [50, 100],
    'estimator__max_depth': [10, 20],
    'estimator__min_samples_split': [2, 5]
}

grid_search = GridSearchCV(MultiOutputClassifier(RandomForestClassifier(random_state=45)),
                           param_grid, scoring='accuracy', cv=3)
grid_search.fit(X_train, y_train)
best_model = grid_search.best_estimator_
print(best_model)

MultiOutputClassifier(estimator=RandomForestClassifier(max_depth=10,
                                                         n_estimators=50,
                                                         random_state=45))
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Bu kod bloğunda, RandomForestClassifier modeli ile çoklu bağımlı değişkenlerin tahmin edilmesi amacıyla GridSearchCV kullanılarak hiperparametre

optimizasyonu gerçekleştirilmiştir. İlk olarak, modelin `n_estimators` (ağaç sayısı), `max_depth` (ağaç derinliği) ve `min_samples_split` (bölme için minimum örnek sayısı) gibi hiperparametreleri belirli bir grid üzerinde taranmıştır. `GridSearchCV`, her parametre kombinasyonunu değerlendirerek en iyi sonuçları elde etmek için 3 katlı çapraz doğrulama (cross-validation) kullanmıştır. Eğitim verisi üzerinde yapılan bu işlemin ardından, en iyi parametrelerle eğitilen model `best_estimator_` aracılığıyla seçilmiş ve doğruluk oranı ile değerlendirilmiştir. Bu yöntem, modelin doğruluğunu artırarak daha sağlam tahminler elde edilmesine yardımcı olmuştur. Bu sayede, `RandomForestClassifier` modeli çoklu bağımlı değişkenleri doğru bir şekilde tahmin edebilecek şekilde optimize edilmiştir. Burada `max_depth` değişkenimiz için 10, `n_estimators` değişkenimiz için ve `random_state` değişkenimiz için 45 değerleri en uygun kombinasyonla seçilmiş parametrelerimizdir.

Desicion tree algoritmasında kullanılacak olan hiperparametre seçimi için yazılmış kod bloğu aşağıda görülmektedir.

Şekil 46: Desicion Tree Algoritması için Hiperparametre Optimizasyonu

```
param_grid = {
    'criterion': ['gini', 'entropy'],
    'max_depth': [3, 4, 5, 6],
    'min_samples_split': [5, 10, 15],
    'min_samples_leaf': [3, 5, 7],
    'max_features': ['sqrt', 'log2', None],
    'ccp_alpha': [0.05, 0.1, 0.2]
}

# Modeli oluşturma ve GridSearchCV ile parametreleri optimize etme
dt = DecisionTreeClassifier(random_state=42)
grid_search = GridSearchCV(estimator=dt, param_grid=param_grid, cv=10, scoring='accuracy', n_jobs=-1, verbose=1)
grid_search.fit(X_train, y_train)

# En iyi parametreleri yazdırma
print("Best parameters found: ", grid_search.best_params_)
```

Best parameters found: {'ccp_alpha': 0.05, 'criterion': 'entropy', 'max_depth': 3, 'max_features': None, 'min_samples_leaf': 3, 'min_samples_split': 5}

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Öncelikle, hiperparametreler bir sözlük yapısında tanımlanmıştır. Bu parametreler arasında, bölme ölçütü olarak kullanılan `criterion` (gini veya entropy), karar ağacının maksimum derinliğini sınırlandıran `max_depth` (3, 5, 7, 10), bir düğümün dallanabilmesi için gereken minimum örnek sayısını ifade eden `min_samples_split`, yaprak düğümünde bulunması gereken minimum örnek sayısını belirten `min_samples_leaf`, bölme sırasında dikkate alınacak maksimum özellik

sayısını tanımlayan max_features (sqrt, log2 veya sınırsız) ve modelin karmaşıklığını kontrol ederek daha iyi genelleme sağlamayı amaçlayan ccp_alpha parametresi yer almaktadır. Bu parametre çoklu sınıflandırma problemlerinde daha dengeli ve geliştirilebilir dallanma yapıları oluşturmak için idealdir.

Bu hiperparametrelerin optimize edilmesi için GridSearchCV yöntemi, 5 katlı çapraz doğrulama (cv=5) ile uygulanmıştır. Ayrıca, işlem süresinin optimize edilmesi için tüm işlemci çekirdeklerinin kullanılması sağlanmıştır (n_jobs=-1). Optimizasyon işlemi sırasında model, eğitim verileri üzerinde eğitilmiş ve en iyi hiperparametre kombinasyonu belirlenmiştir. Son olarak, en iyi parametreler grid_search.best_params_ ile elde edilmiş ve bu değerler modelin performansını artırmak için kullanılmıştır. Bu yöntem, karar ağacının daha etkili ve genellenebilir bir şekilde tahmin yapmasını sağlamıştır.

Bu işlem sonucunda ccp_alpha 0.05, criterion entropy, max_depth 3, min_sample_leaf 3 ve min_sample_split 5 olacak şekilde en uygun kombinasyonla seçilmiş parametrelerimizdir.

Support Vector Classifier algoritmasında kullanılacak olan hiperparametre seçimi için yazılmış kod bloğu aşağıda görülmektedir.

Şekil 47: Support Vector Classifier Algoritması için Hiperparametre Optimizasyonu

```
param_grid = {
    'svc__estimator__C': [0.1, 2, 1, 10],
    'svc__estimator__gamma': ['scale', 'auto'],
    'svc__estimator__kernel': ['rbf'],
}

# 4. GridSearchCV ile parametreleri optimize edelim (K-fold cross-validation kullanıyoruz)
grid_search = GridSearchCV(estimator=svc_pipeline, param_grid=param_grid, cv=10, scoring='accuracy', n_jobs=-1, verbose=1)
grid_search.fit(X_train, y_train)

# 5. En iyi parametreyi yazdırma
print("Best parameters found: ", grid_search.best_params_)
```

```
Best parameters found: {'svc__estimator__C': 2, 'svc__estimator__gamma': 'scale', 'svc__estimator__kernel': 'rbf'}
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Belirlenen hiperparametreler arasında, modelin karar sınırlarını kontrol eden C (0.1, 2, 1 ve 10), kernel fonksiyonu için gamma değeri (scale veya auto) ve kernel türü olarak rbf (radial basis function) bulunmaktadır. Optimizasyon işlemi, 10 katlı çapraz doğrulama (cv=10) yöntemiyle gerçekleştirilmiş ve modelin performansı

doğruluk metriği (accuracy) kullanılarak değerlendirilmiştir. İşlem süresinin daha verimli hale getirilmesi için tüm işlemci çekirdeklerinin kullanımı sağlanmıştır (n_jobs=-1). Eğitim verisi üzerinde uygulanan bu süreç sonucunda, en iyi hiperparametre kombinasyonu grid_search.best_params_ ile elde edilmiş ve bu parametreler, modelin doğruluğunu artırmak için kullanılmıştır. Bu yöntem, SVC modelinin genelleme kabiliyetini geliştirmeyi amaçlamaktadır.

İşlem sonucunda svc__estimator__c 2, svc__estimator__gamma 'scale', svc__estimator__kernel 'rbf' olacak şekilde en uygun kombinasyonla seçilmiş parametrelerimizdir.

CatBoost algoritmasında kullanılacak olan hiperparametre seçimi için yazılmış kod bloğu aşağıda görülmektedir.

Şekil 48: CatBoost Algoritması için Hiperparametre Optimizasyonu

```
# CatBoost modelini oluşturuyoruz
catboost = CatBoostClassifier(learning_rate=0.1, iterations=500, depth=6, verbose=0)

# Hiperparametreler için bir grid tanımlıyoruz
param_grid = {
    'estimator__learning_rate': [0.01, 0.05, 0.1],
    'estimator__depth': [4, 6, 8],
    'estimator__iterations': [500, 1000]
}

# OneVsRestClassifier ile CatBoost'u sarıyoruz
ovr_catboost = OneVsRestClassifier(catboost)

# GridSearchCV ile hiperparametre optimizasyonu yapıyoruz
grid_search = GridSearchCV(ovr_catboost, param_grid, cv=5, n_jobs=-1)
grid_search.fit(X_train, y_train_bin)

# En iyi parametreleri yazdırıyoruz
print(f"Best parameters: {grid_search.best_params_}")

Best parameters: {'estimator__depth': 4, 'estimator__iterations': 500, 'estimator__learning_rate': 0.01}
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Model, çok sınıflı bir problem yapısına uygun hale getirilmek amacıyla OneVsRestClassifier ile birleştirilmiştir. Bu yöntem, her bir sınıf için bağımsız bir model eğiterek çok sınıflı sınıflandırma problemini çözmeyi amaçlamaktadır.

CatBoost modeline yönelik olarak, öğrenme oranını kontrol eden learning_rate (0.01, 0.05, 0.1), ağacın maksimum derinliğini belirleyen depth (4, 6, 8) ve yineleme sayısını ifade eden iterations (500, 1000) hiperparametreleri

değerlendirilmiştir. Bu hiperparametre kombinasyonlarının performansı, GridSearchCV yöntemi kullanılarak 5 katlı çapraz doğrulama (cv=5) ile optimize edilmiştir. Optimizasyon sırasında işlem süresi ve hesaplama verimliliği açısından tüm işlemci çekirdekleri kullanılmıştır (n_jobs=-1).

Eğitim verisi üzerinde uygulanan bu süreç sonucunda, en iyi hiperparametre kombinasyonu grid_search.best_params_ ile elde edilmiş ve modelin doğruluğunu artırmak için bu parametreler kullanılmıştır. Bu yöntem, CatBoost modelinin genelleme kabiliyetini artırmayı ve çok sınıflı sınıflandırma problemlerinde daha yüksek performans sağlamayı hedeflemektedir.

İşlem sonucunda estimator__depth 4, estimator__iterations 500 ve estimator__learning_rate 0.01 olacak şekilde en uygun kombinasyonla seçilmiş parametrelerimizdir.

Bu çalışmada, XGBoost modelinin varsayılan hiperparametrelerle elde ettiği sonuçlar yüksek doğruluk oranlarına ulaştığı için hiperparametre optimizasyonuna gerek duyulmamıştır. Varsayılan ayarlarla yapılan modelleme sürecinde, XGBoost'un hızlı hesaplama kabiliyeti ve genelleme yeteneği sayesinde, sınıflandırma performansının tatmin edici düzeyde olduğu gözlemlenmiştir. Bu nedenle, hiperparametre arama işlemi yapılmadan model doğrudan kullanılmış ve sonuçlar değerlendirilmiştir. Bu yaklaşım ,hem hesaplama maliyetlerini azaltmış hem de modelleme sürecini hızlandırmıştır.

3.5.4. Makine Öğrenmesi Modeli Oluşturulması ve Performans Ölçütü

3.5.4.1. Random Forest Algoritması

Belirlenen hiperparametreler ile makine öğrenme algoritması oluşturulur ve algoritmaların başarı performansları incelenir. Random Forest algoritmasının temel prensipleri, ensemble yapısı, bootstrap örnekleme, rastgele özellik seçimi ve tahminlerin birleştirilmesi üzerine kuruludur. Algoritma, belirli sayıda karar ağacından oluşan bir topluluk öğrenmesi modeli (ensemble) oluşturur. Her ağaç, orijinal veri setinden rastgele seçilen bir alt küme (bootstrap örnekleme) ile eğitilir. Eğitim sırasında birçok karar ağacı oluşturularak ve oluşan bireysel ağaçların ortalama tahminini çıkararak çalışır. Bu yöntem, her ağacın farklı bir veri alt kümesi üzerinde

eğitim almasını sağlayarak modelin genelleme yeteneğini artırır. Ayrıca, her bir düğümde tüm özellikler arasından rastgele seçilen bir alt küme kullanılarak en iyi bölünme noktası belirlenir. Bu da ağaçların birbirinden bağımsız olmasını sağlar ve korelasyonları azaltarak modelin performansını iyileştirir.

Random Forest, hem ikili (binary) hem de çok sınıflı (multi-class) sınıflandırma problemlerinde etkili bir şekilde kullanılabilen, güçlü ve esnek bir topluluk öğrenme algoritmasıdır.

Çok sınıflı sınıflandırma problemlerinde Random Forest, her karar ağacının her sınıfa ait olasılık tahminlerini hesaplamasına dayanır. Algoritma, bir örneğin hangi sınıfa ait olduğunu belirlemek için tüm ağaçlardan gelen tahminlerin çoğunluk oylamasını alır. Bu yaklaşım, veri setindeki sınıflar arasında dengesizlik olsa bile genellikle iyi performans gösterir. Ayrıca, her sınıfa ait tahmin olasılıklarını birleştirerek sınıf olasılık dağılımlarını da çıkartabilir, bu da karar eşiklerini özelleştirmeyi mümkün kılar. Random Forest, veri setindeki sınıf sayısından bağımsız olarak, genellikle sağlam, doğru ve yorumlanabilir sonuçlar üreten bir algoritmadır. Özellikle, karmaşık veri yapılarında ve yüksek boyutlu veri setlerinde etkin bir seçimdir.

Random Forest algoritması çok etiketli bir sınıflandırma problemi için modeli eğitme, tahmin yapma ve performans değerlendirme işlemlerini kapsamaktadır. Bu süreçte model, veri setindeki bağımsız değişkenler ile bağımlı değişkenler arasındaki ilişkiyi öğrenir. Random Forest algoritması, dengesiz sınıf dağılımını dengelemek için `class_weight='balanced'` parametresi ile uygulanmıştır. Modelin doğruluğu, precision, recall, F1-Score, Confusion Matrix gibi klasik metriklerle ölçülmüş, ayrıca Matthews Correlation Coefficient, Hamming Loss ve Precision-Recall AUC gibi çok etiketli sınıflandırmaya özel metriklerle desteklenmiştir. Sınıf dengesizliği göz önüne alınarak ağırlıklı ortalama yöntemi kullanılmıştır. Modelin performans metriklerinin çıktısı aşağıdaki gibidir.

Şekil 49: Random Forest Algoritması için Performans Metrikleri Çıktısı

General Metrics:		Matthews Correlation Coefficient (MCC) for each label:		
Metric	Value	Label	Metric	Value
Accuracy	0.993304	Label 0	MCC	0.986809
F1 Score (Weighted)	0.979212	Label 1	MCC	0.965927
Precision (Weighted)	1			
Recall (Weighted)	0.959459			
Jaccard Score (Weighted)	0.959459			
Hamming Loss	0.00334821			
Train Score	1			
Test Score	0.993304			

Precision-Recall AUC for each label:		
Label	Metric	Value
Label 0	Precision-Recall AUC	0.989211
Label 1	Precision-Recall AUC	0.970982

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Accuracy değeri, modelin genel olarak doğru tahmin yapma oranını gösterir. %99.33 modelin çok başarılı olduğunu ve test veri setindeki çoğu örneği doğru sınıflandırdığını göstermektedir.

F1 Score (weighted) değeri, precision (kesinlik) ve recall (duyarlılık) değerlerinin harmonik ortalamasını temsil eder. Ağırlıklı ortalama yöntemi kullanıldığı için sınıf dengesizlikleri hesaba katılmıştır. Bu skor modelin dengeli bir performans sergilediğini ortaya koymaktadır. Bu değer modelin genel performansını yansıtır. 1.0'a yakın olması, modelin dengeli ve başarılı tahminler yaptığını gösterir.

Precision (Weighted) kesinlik değeri, modelin yaptığı pozitif tahminlerin ne kadarının doğru olduğunu gösterir. Ağırlıklı ortalama ile sınıflar arasında hesaplanan bu değer, modelin pozitif tahminlerde %100 doğruluk sağladığını ifade etmektedir. Bu sonuç, modelin aşırı güvenli tahminlerde bulunduğunu gösterir.

Recall (Weighted) duyarlılık değeri, modelin gerçek pozitif örneklerin ne kadarını doğru tahmin ettiğini ölçer. Ağırlıklı ortalama kullanıldığında %95.94 gibi oldukça yüksek bir değer elde edilmiştir. Bu sonuç, modelin tüm sınıflardaki pozitif örneklerin yaklaşık %96 'sını başarılı bir şekilde yakaladığını ifade eder.

Jaccard Score (Weighted) değeri, Jaccard skoru, gerçek ve tahmin edilen pozitif kümeler arasındaki benzerliği ölçer. Yüksek bir değere sahiptir, bu da sınıflandırma performansının başarılı olduğunu kanıtlar. Jaccard skor (%95.94),

modelin pozitif sınıflandırmalarda tutarlı ve başarılı olduğunu göstermektedir.

Hamming Loss değeri, yanlış sınıflandırılan etiket oranını ölçer. Bu değer ne kadar düşüğe modelin başarısı o kadar yüksektir. %0.33 gibi düşük bir değer, çok etiketli sınıflandırmada yapılan hataların oldukça az olduğunu kanıtlamaktadır.

MCC (Matthews Correlation Coefficient) değerleri çok sınıflı sınıflandırma problemlerinde önemli bir metriktir. Sınıflar arası uyumu ifade eder. Tablodan da anlaşılacağı gibi %96 ve %98 oranları sınıflar arasındaki uyumun çok iyi olduğunu göstermektedir.

Train Score ve Test Score değeri, modelin eğitim ve test veri setlerinde gösterdiği başarıyı ölçer. Eğitim skorunun 1.0 olması, modelin eğitim setindeki tüm örnekleri doğru tahmin ettiğini gösterir. Test skorunun 0.993304 olması ise modelin test veri setinde de yüksek doğrulukla performans gösterdiğini ortaya koymaktadır. Bu durum, modelin aşırı öğrenme (overfitting) eğiliminde olmadığını ve genelleştirilebilirliğinin yüksek olduğunu gösterir.

Özetle, accuracy ve precision değerlerinin çok yüksek olması, modelin büyük oranda doğru ve güvenilir tahminler yaptığını, recall ve F1 Score'un yüksek olması, modelin her sınıfta pozitif örnekleri büyük bir başarıyla tahmin ettiğini ve hamming loss değerinin düşük olması, modelin çok etiketli sınıflandırmada hatalı tahmin oranının son derece az olduğunu gösterir. Rastgele orman algoritması bu çok etiketli sınıflandırma probleminde üstün performans sergilemiş ve doğruluğu oldukça yüksek bir model oluşturmuştur.

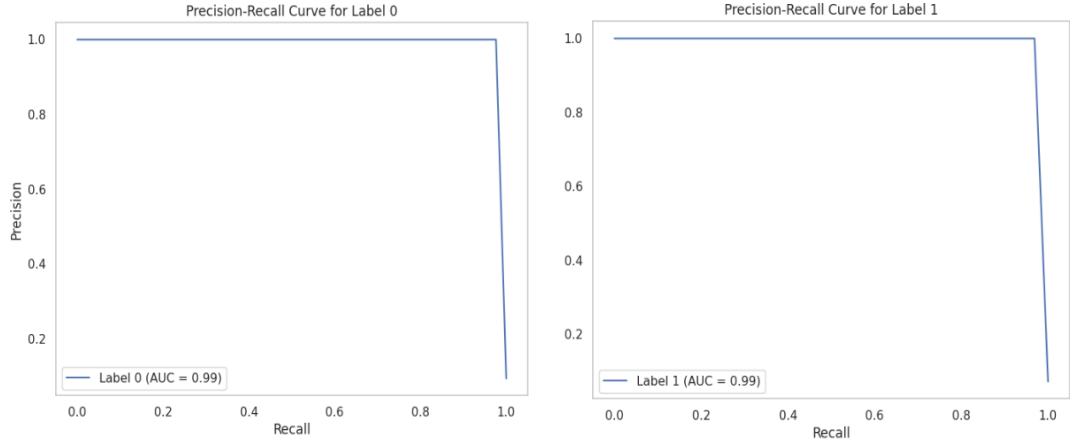
Sınıfların performansları ayrı ayrı incelenmek istenirse aşağıdaki classification report çıktısından incelenebilir.

Şekil 50: Random Forest Algoritması Classification Report Çıktısı

Classification Report:				
	precision	recall	f1-score	support
0	1.00	0.98	0.99	42
1	1.00	0.94	0.97	32
micro avg	1.00	0.96	0.98	74
macro avg	1.00	0.96	0.98	74
weighted avg	1.00	0.96	0.98	74
samples avg	0.13	0.13	0.13	74

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Şekil 51: Kesinlik-Duyarlılık Grafiği

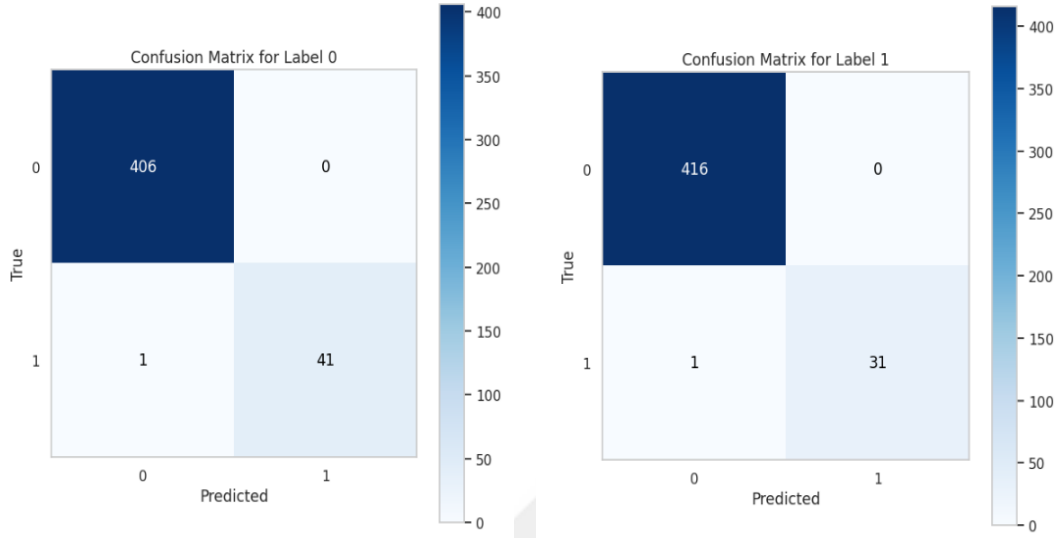


Kaynak: Araştırmacı tarafından oluşturulmuştur.

Precision-Recall eğrileri, modelin sınıflandırma performansını özellikle sınıf dengesizliği olan durumlarda değerlendirmek için kullanılmaktadır. Precision, tahminlerin ne kadarının doğru olduğunu Recall ise gerçekte doğru olanların ne kadarının yakalandığını ölçer.

Label 0, birinci kampanyayı Label 1 ise ikinci kampanyayı temsil eder. Kampanya 1 ve kampanya 2 için eğri boyunca precision değeri büyük ölçüde 1.0 seviyesinde ilerlemekte ve recall değeri 1.0'a ulaştığında hafif bir düşüş göstermektedir. Eğrinin altında kalan alan (AUC) 0.99 olarak hesaplanmıştır. Bu da modelin iki sınıfta da mükemmel yakın performans gösterdiğini ifade etmektedir. Çok etiketli sınıflandırma probleminde rastgele orman modelinin yüksek performans sergilediğini ve hem kesinlik hem de duyarlılık açısından tutarlı sonuçlar verdiğini doğrulamaktadır. Model, her iki etiket için de sınıf dengesizliklerinden etkilenmeden başarılı tahminler yapabilmektedir.

Şekil 52: Karmaşıklık Matrisi (Confusion Matrix)



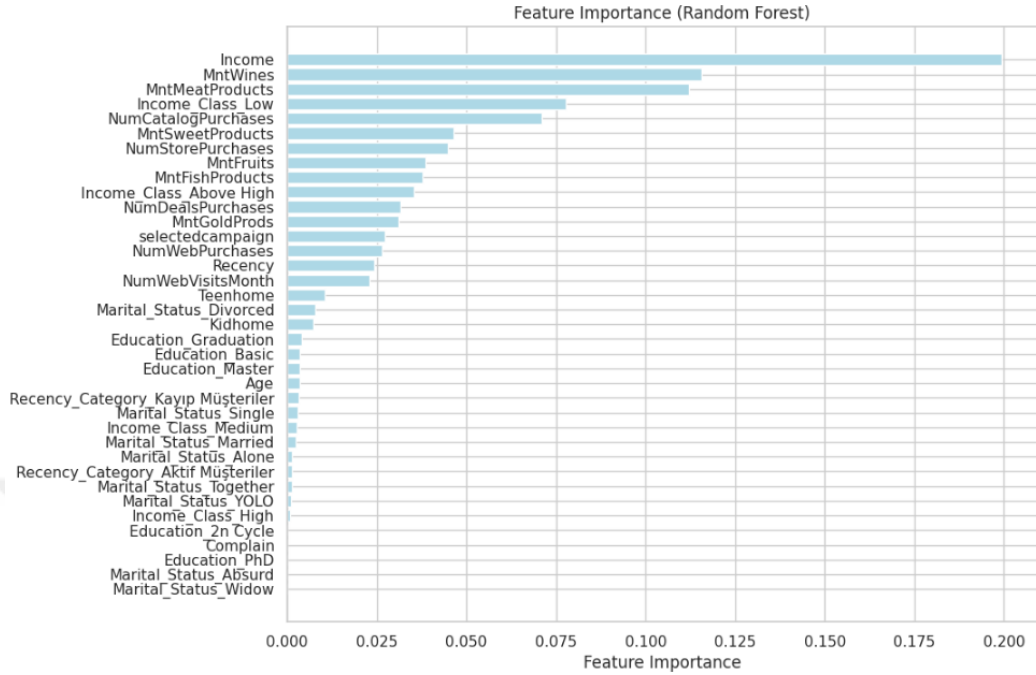
Kaynak: Araştırmacı tarafından oluşturulmuştur.

Model, birinci kampanya (Label 0) için gerçekte kampanyanın kabul edilmeme durumunu 406 örnekte doğru bir şekilde 0 olarak tahmin etmiştir. Model sınıfı 0 iken yanlışlıkla 1 olarak tahmin edilen örnek yoktur. Gerçekte kampanyanın kabul edilme durumunu ifade eden 1 sınıfında 41 örneği doğru bir şekilde 1 olarak tahmin etmiştir. Model, gerçekte kabul edilen sınıfı temsil eden kampanyayı 1 örnekte yanlışlıkla 0 olarak tahmin etmiştir.

Model, ikinci kampanya (Label 1) için gerçekte kampanyanın kabul edilmeme durumunu 406 örnekte doğru bir şekilde 0 olarak tahmin etmiştir. Model sınıfı gerçekte kampanyanın kabul edilmeme durumunu temsil ederken yanlışlıkla 1 olarak tahmin edilen örnek yoktur. Gerçekte kampanyanın kabul edilme durumunu ifade eden 1 sınıfında 41 örneği doğru bir şekilde 1 olarak tahmin etmiştir. Model, gerçekte kabul edilen sınıfı temsil eden kampanyayı 1 örnekte yanlışlıkla 0 olarak tahmin etmiştir.

Bu sonuçlar modelin kampanya 1 ve kampanya 2 için yüksek doğruluk ve güvenilir performans sunduğunu göstermektedir.

Şekil 53: Özellik Önem Düzeyi Analizi



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Bu grafik, Random Forest modeli kullanılarak elde edilen öznelik önem düzeylerini göstermektedir. Öznelik önem düzeyleri, modelin tahmin yaparken hangi değişkenlere daha fazla ağırlık verdiğini belirlemektedir.

Modelde en yüksek önem düzeyine sahip Income (gelir) değişkenidir. Bu durum, gelir seviyesinin kampanya seçimini etkileyen en belirleyici faktörlerden biri olduğunu göstermektedir.

Şarap ve et ürünlerine yapılan harcamaların kampanya seçimini belirlemede diğer önemli değişkenler olduğunu belirtir. Bu da müşterilerin harcama alışkanlıklarının önemli bir etkiye sahip olduğunu gösterir.

Income_Class_Low değişkeni düşük gelir sınıfında yer almanın, kampanya seçimi üzerinde etkili bir faktör olduğunu gösteren bir diğer önemli faktördür.

Katalog üzerinden yapılan alışverişler de, müşterilerin kampanyalara verdiği tepkiyi etkileyen bir faktör olduğu görülmektedir. Mağazada yapılan alışveriş sayısının, tatlı ve meyve ürünlerine yapılan harcamaların da belirli bir etkisi gözlemlenmiştir.

Evde yaşayan çocuk sayısı, eğitim durumu, müşteri etkileşim durumu gibi değişkenler modelde düşük önem düzeyine sahiptir.

Şikayet durumunu ifade eden complain değişkeninin kampanya seçiminde belirgin bir etkisi görülmemektedir.

Bu sonuçlar, müşteri harcamaları ve gelir seviyesinin kampanya seçimini belirlemede en kritik faktörler olduğunu ortaya koymaktadır.

3.5.4.2. Desicion Tree Algoritması

Karar ağacı algoritması, veri madenciliği ve makine öğrenmesi alanlarında kullanılan ve verileri sınıflandırmak veya tahmin yapmak için dallanma yapısı kullanan bir algoritmadır. Bu algoritma, veri setindeki karar noktalarını belirleyerek veriyi çeşitli dallara ayırır. Her bir düğüm, bir özelliğin belirli bir değeri üzerindeki karar noktasını temsil ederken, dallar bu karara göre verinin nasıl bölüneceğini gösterir. En uçtaki yaprak düğümler ise sınıf etiketlerini veya tahmin edilen değerleri belirtir. Karar ağacı algoritması hem sınıflandırma hem de regresyon problemlerinde etkili olup, modelin anlaşılması ve yorumlanması kolaydır. Ancak, aşırı öğrenme (overfitting) gibi sorunlarla karşılaşabileceği için dikkatli bir şekilde budanması (pruning) gerekebilir. Ağaç yapısı, girdi özelliklerinin değeri temel alınarak veri kümesini alt kümelerle ayırır, dallar oluşturur ve yaprak düğümlere ulaştığında tahmin edilen değerlerle biter. Çok değişkenli sınıflandırma problemlerinde karar ağaçları, veriyi farklı değişkenler üzerinden bölerek sınıfları tahminlemeye çalışır. Model, her düğümde bir değişken seçer ve bu değişkene bağlı olarak bir eşik değeri belirler. Bölünme işlemi birden fazla sınıfa göre yapılır. Bu süreç, her iç düğümde devam eder ve model en homojen (en saf) bölünmeyi hedefler. Bölünme kriteri olarak Gini, Bilgi Kazancı veya Entropi gibi yöntemler kullanılır.

Karar ağacı algoritmasının anlaşılması ve görselleştirilmesi kolaydır. Çok az veri ön işleme gerektirir. Sayısal ve kategorik veriler üzerinde oldukça güzel performansa sahiptir. Ancak, aşırı uyuma meyillidir.

Bu algoritma özelinde aşırı uyum (overfitting) probleminin önüne geçmek için hiperparametre optimizasyonu yapılmıştır. En uygun parametrelerle model çalıştırılmış çıktı sonuçları ve performans metrikleri incelenmiştir.

Şekil 54: Desicion Tree Algoritması için Performans Metrikleri Çıktısı

Metric	Value	Metric	Value
Accuracy	0.970982	MCC for Label 0	0.817957
F1 Score (Weighted)	0.896079	MCC for Label 1	1
Jaccard Score (Weighted)	0.824324	Precision-Recall AUC for Label 0	0.859747
Hamming Loss	0.0145089	Precision-Recall AUC for Label 1	1
Training Score	0.97433	Training Score	0.97433
Test Score	0.970982	Test Score	0.970982

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Modelin performans metrikleri çok sınıflı sınıflandırma problemi kapsamında ve sınıflar arası dağılımın dengeli olmadığı bilinerek yorumlanmalıdır.

Sınıflar arası dağılımı dengelemek amacıyla F1 score ve jaccard score metrikleri ağırlıklandırılmış performans metrikleri olarak çalıştırılır.

Accuracy değeri %97,09 çıkmıştır. Bu oran modelin test verisi üzerindeki doğruluğunun yüksek olduğunu göstermektedir.

Doğruluk metriği, toplam doğru tahminlerin tüm gözlemlere oranını ifade eder. Ancak sınıflar arasında dengesizlik (imbalance) olduğunda yanıltıcı olabilir.

F1 score (weighted) değeri, modelin hem precision (kesinlik) hem de recall (duyarlılık) metriklerini dikkate alarak ortalama performansını gösterir. F1 skorunun yüksek olması, sınıflar arasında dengeli performans elde edildiğini ifade eder.

Jaccard score (weighted) değeri tahmin doğruluğunu ve tahminlerinin %82'sinin örtüştüğünü gösterir. Bu skorun yüksek olması, modelin başarılı bir sınıflandırma yaptığını destekler.

Hamming Loss değerinin %1.4 olması yanlış sınıflandırılan örneklerin oranını verir. Bu değerin 0'a yakın olması hatalı tahminlerin çok az olduğunu göstermektedir.

MCC değeri sınıflar arasındaki uyumu ölçen güçlü bir metriktir. Çok sınıflı sınıflandırma problemlerin ve dengesiz veri dağılımı söz konusu olan sınıflar için güçlü bir metriktir.

Kampanya 2 için mükemmel bir sonuç vermesi dikkat çekmekte olup

kampanya 1 için ise %81 uyum göstermiştir. Uyumun kampanya 1 özelinde mükemmel olması overfitting durumu olabileceğini düşündürdüğünden diğer metrikler de incelenmelidir.

Eğitim ve test skorlarının birbirine çok yakın olması, modelin overfitting (aşırı öğrenme) yapmadığını ve veriye iyi genelleme yaptığını gösterse de diğer performans metriklerinden precision ve recall değerleri incelenmelidir. Dengeli olmayan veri setlerinde bu metrikler daha bilgilendirici olabilir.

Şekil 55: Precision – Recall- F1 Score Çıktısı

	precision	recall	f1-score	support
0	1.00	0.69	0.82	42
1	1.00	1.00	1.00	32

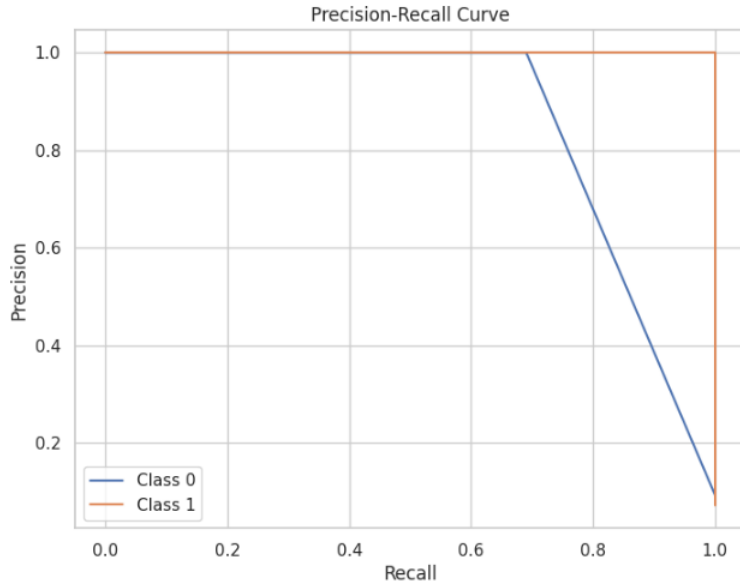
Kaynak: Araştırmacı tarafından oluşturulmuştur.

Bu, modelin kampanya 2 için tüm örnekleri doğru tahmin ettiğini gösterir. Ancak bu mükemmel sonuç genelleme sorunu olduğuna işaret edebilir.

Kampanya 1 için precision değeri 1'dir. Bu durum kampanyanın kabul edildiği halde kabul edilmemiş gibi tahminlenen negatif sınıf örneklerini ifade eden örneklerin hepsinin doğru tahmin edildiğini gösterir. Recall değerinin 0.69 olması ise negatif sınıftaki örneklerin sadece %69'unu doğru tahmin ettiğini gösterir. Bu modelin bazı sınıfları atladığını gösterir.

Kampanya 2 için mükemmel performans varken, Kampanya 1 için daha zayıf bir performans var. Bu durum, modelin Kampanya 2 'ye aşırı uyum sağladığını (overfitting) ve Kampanya 1'i yeterince öğrenemediğini gösteriyor. Bu durum aşağıda precision-recall curve grafiğinde de açıkça görülmektedir. Model kampanya 2 eğrisi için başarılı bir grafik oluşturmuş ancak kampanya 1 için aynı başarıyı gösterememiştir.

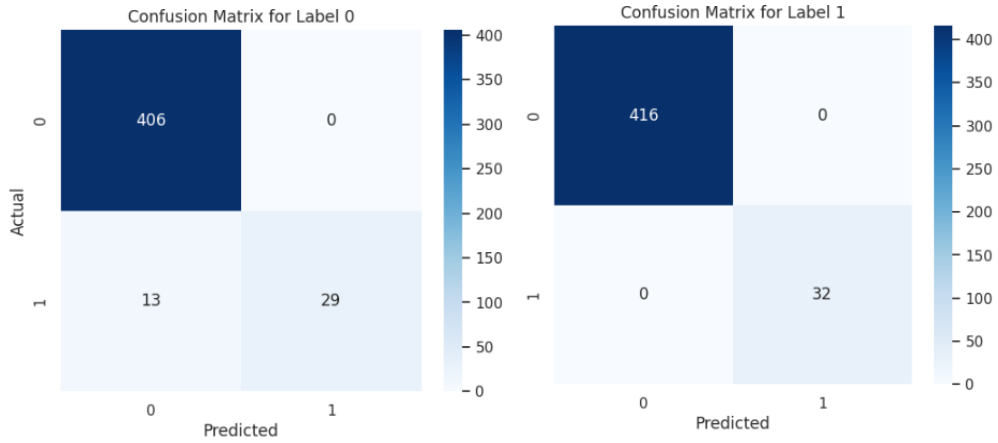
Şekil 56: Kesinlik - Duyarlılık Grafiği



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Modelin ne kadar örneği doğru tahmin ettiğini aşağıda confusion matrix çıktısında görülmektedir.

Şekil 57: Karmaşıklık Matrisi (Confusion Matrix)



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Modelin performansı iki ayrı karmaşıklık matrisi ile değerlendirilmiştir. Kampanya 1'i temsil eden (Label 0) matriste, model 406 doğru tahminle kampanyaya katılmayan müşterileri doğru bir şekilde belirlemiştir. Ayrıca, 29 müşteriyi doğru bir şekilde kampanyaya katılmış olarak tahmin ederken, 13 müşteriyi yanlışlıkla

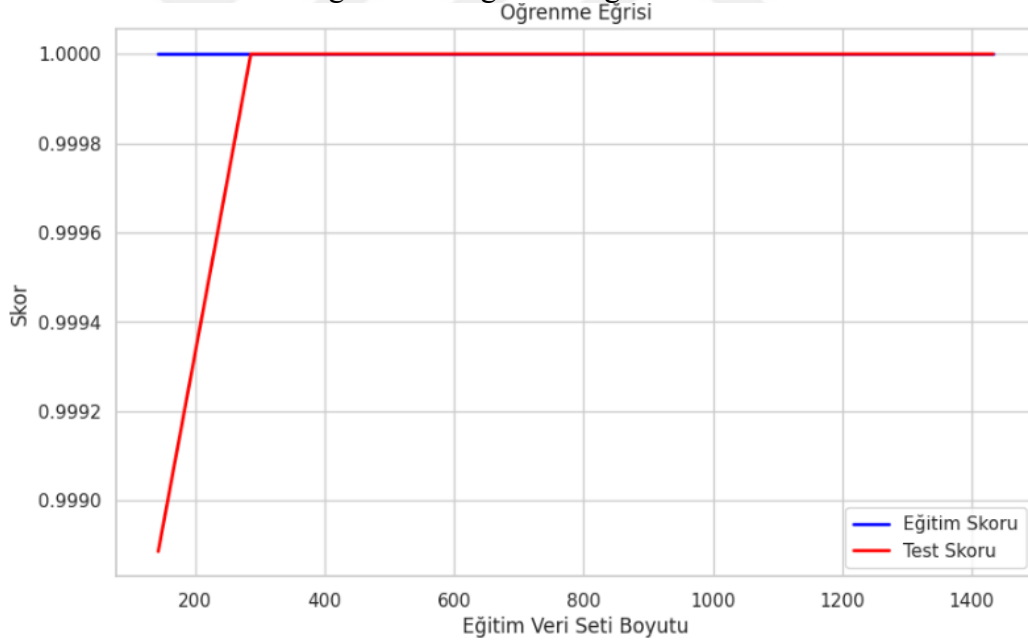
kampanyaya katılmamış olarak tahmin etmiştir. Yanlış pozitif tahmin bulunmamaktadır, yani model kampanyaya katılmayan müşterileri yanlışlıkla katılmış olarak tahmin etmemiştir.

Kampanya 2'yi temsil eden (Label 1) matriste ise model mükemmel bir performans sergilemiştir. Toplamda 416 müşterinin kampanyaya katılmadığını, 32 müşterinin ise katıldığını doğru bir şekilde tahmin etmiş ve hiçbir yanlış tahminde bulunmamıştır. Genel olarak model, Kampanya 2'de hatasız çalışırken, Kampanya 1 için küçük bir hata oranı göstermiştir.

Bu durum, Kampanya 1 tahmin doğruluğunu iyileştirmeye yönelik çalışmalar yapılabileceğini ortaya koymaktadır.

Öğrenme eğrisi ile modelin veri setini ezberlediği (overfit) görülmektedir.

Şekil 58: Decision Tree Algoritması Öğrenme Eğrisi



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Bu grafikte, modelin eğitim ve test performansını gösteren öğrenme eğrisi incelenmiştir. Grafikte, eğitim skoru (mavi) ve test skoru (kırmızı) değerlerinin oldukça kısa bir eğitim veri seti boyutundan sonra 1.0 seviyesine ulaştığı ve sabitlendiği görülmektedir. Bu durum, modelin eğitim verisinde ve test verisinde mükemmel bir performans sergilediğine işaret etmektedir. Ancak, skorların tam

olarak 1.0 değerine ulaşması, modelin veriyi ezberlemiş olabileceğini ve aşırı öğrenme (overfitting) sorunuyla karşı karşıya kaldığını göstermektedir. Aşırı öğrenme durumunda model, eğitim verisindeki örnekleri çok iyi öğrendiği halde yeni ve gerçek dünya verilerinde genelleme yapma yetisini kaybedebilir. Bu nedenle, modelin performansını daha güvenilir hale getirmek için aşırı öğrenmeyi azaltıcı yöntemler uygulanabilir.

Overfitting, gerçek dünyadaki birçok projede karşılaşılan bir sorundur. Bu sorunun önüne geçebilmek için tekrar hiperparametre optimiasyonu yapıp yeni değerler ile model çalıştırılmıştır. Bu, modelin karmaşıklığını azaltarak ezberleme durumunun önüne geçmeye çalışılır. 5 katlı ve 10 katlı çapraz doğrulama yapılarak modelin doğrulama verisiyle tekrar test edilip ezberleme probleminin önüne geçilmeye çalışılmıştır. Veri setinde sınıflar arası veri dengesizliği söz konusu olduğu için değişkenler tekrar ağırlıklandırılarak veri seti dengeli hale getirilmeye çalışılmıştır. Veri setinin 2240 gözlemden oluşmakta olduğu göz önünde bulundurularak smote yöntemiyle veri artırma işlemleri uygulanmaya çalışılmıştır. Bu yöntemlerle overfitting kısmen giderildi ancak veri setinin sınırlılığı nedeniyle tamamen çözülemedi. Bu durum ileride daha büyük veri setleriyle yapılacak çalışmalarda ele alınabilir.

3.5.4.3. XgBoost Algoritması

XGBoost (eXtreme Gradient Boosting), denetimli öğrenme algoritmaları arasında güçlü bir araç olarak bilinir ve özellikle regresyon ve sınıflandırma problemlerinde geniş bir uygulama alanına sahiptir. XGBoost, boosting (güçlendirme) yöntemine dayalı bir algoritmadır ve temel olarak zayıf öğrencileri (genellikle karar ağaçları) ardışık olarak eğitir. Her yeni ağaç, önceki ağaçların hatalarını düzeltmeye çalışarak modelin genel doğruluğunu artırmaya yönelik olarak eğitim yapılır. Bu süreçte, her bir ağacın katkısı, ağırlıklı olarak hatalı sınıflandırmaları düzelten veri noktalarına verilir, böylece modelin performansı zamanla iyileşir.

XGBoost'un çalışma prensibi, gradient boosting yaklaşımına dayalıdır. Bu, her yeni ağacın önceki modelin hatalarını minimize etmek için gradyan düşüşü (gradient descent) kullanarak öğrenmesini sağlar. Bu sayede, model her iterasyonda

daha iyi sonuçlar üretir ve overfitting (aşırı uyum) riski azalır. Ayrıca, XGBoost'un hız ve performans açısından oldukça verimli olması, onu çok büyük veri setlerinde bile tercih edilen bir algoritma haline getirmiştir. Paralel işlem yetenekleri, bellek yönetimi ve özelleştirilmiş loss fonksiyonları gibi optimizasyonlar da modelin başarısını artıran unsurlardandır.

Çok sınıflı sınıflandırma problemlerindeki başarısı, XGBoost'un her bir sınıf için bağımsız karar ağaçları oluşturması ve çoklu etiketli sınıflandırma sorunlarında dahi yüksek doğruluk oranları sağlamasıyla kendini gösterir. Model, her sınıf için ayrı ayrı ağaçlar üreterek sınıflar arası etkileşimleri ve sınıflandırma hatalarını başarılı bir şekilde öğrenir. XGBoost, hem doğruluk hem de hız açısından, çok sınıflı problemler için genellikle güçlü bir performans sergileyerek, özellikle karmaşık veri setlerinde ve yüksek boyutlu problemlerde etkin sonuçlar elde edilmesine olanak tanır.

XGBoost, dengesiz veri setlerinde yüksek performans gösteren bir algoritmadır. Bu tür veri setlerinde genellikle bir sınıf çok fazla örnek içerirken, diğer sınıflar daha az örneğe sahiptir. Bu dengesizlik, modelin nadir sınıfları doğru bir şekilde sınıflandırma yeteneğini olumsuz etkileyebilir. Doğru parametre ayarları ve ön işleme teknikleri ile, nadir sınıfların doğru bir şekilde sınıflandırılmasını sağlayarak, modelin genel başarımını arttırılabilir. Bununla birlikte, dengesiz veri setleriyle çalışırken, sadece doğruluk gibi basit metriklerle değil, daha kapsamlı değerlendirme metrikleri ile modelin başarısını ölçmek önemlidir.

XgBoost algoritması çalıştırılırken hiperparametre ayarlarının yapılmasına gerek duyulmamıştır. Model en basit haliyle veri dengesizlikleri giderildikten sonra gayet başarılı bir şekilde çalışmaktadır. Amaç hiperparametre optimizasyonu yapmadan da modelin çok etiketli sınıflandırma problemlerinde başarılı olduğunu göstermektir.

İlk olarak dengesiz veri setleri için Smote (Synthetic Minority Over-sampling Technique) yöntemiyle verinin yeniden örnekleme yapılmıştır. İlk başta her etiket için ayrı ayrı SMOTE uygulanmış, ardından tüm etiketlerin yeniden örneklenmiş olan sonuçları birleştirilmiştir. Smote, veri setindeki azınlık sınıfın örneklerini çoğaltarak veri dengesizliğini gidermeye çalışır. Bu adımlar, veri setindeki etiket dengesizliğini gidererek modelin her etiket üzerinde daha dengeli tahminler

yapmasını sağlar.

Şekil 59: Smote Yöntemi Uygulama Çıktısı

```
Resampling for Label 0...
Resampling for Label 1...
X_train_res shape: (3322, 39)
y_train_res shape: (1792, 2)
X_train_res shape: (3334, 39)
y_train_res shape: (3334,)

X_train_res shape: (3334, 39)
y_train_res_full shape: (3334, 2)
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

İlk yeniden örnekleme işleminden sonra `X_train_res` veri setinin 3322 örnek ve 37 özellikten oluştuğunu, ancak `y_train_res` etiket setinin 1792 örnekten oluştuğunu gösteriyor. İlk adımda, her etiket için SMOTE uygulandığında, etiketlerin boyutları arasında uyumsuzluklar oluşmuş. Ancak daha sonra bu uyumsuzluğu çözmek için gerekli işlem yapılmış ve sonunda her etiket için eşleşen boyutlar elde edilmiştir. Bu, SMOTE uygulandıktan sonra `X_train_res` veri setinin yeniden örneklenerek 3334 örnekle genişlediğini gösteriyor. Ancak `y_train_res` etiket setinin boyutunun, her etiketin tek boyutlu hale gelmesiyle (yani 1D array olarak) değiştiği görülüyor. Son olarak, her iki etiket için örnekleme işlemi tamamlandığında, `X_train_res` veri setinin 3334 örnekle tamamlandığı ve `y_train_res_full` etiket setinin her iki etiket için 3334 örnek içerdiği belirtiliyor. Bu, etiket setlerinin boyutlarının düzgün bir şekilde eşitlendiğini ve her etiketin yeniden örneklenmiş verileri doğru şekilde içerdiğini gösteriyor.

Yapılan işlemler sonrasında oluşturulan XgBoost algoritmasına ait çıktı sonuçları aşağıda yer almaktadır.

Şekil 60: XgBoost Algoritması için Performans Metrikleri Çıktısı

Metric	Value	Metric	Value
Accuracy	0.907046	MCC for Label 0	0.914177
Precision (Micro)	0.917557	Precision-Recall AUC for Label 0	0.960065
Recall (Micro)	0.977236	MCC for Label 1	0.884893
F1-Score (Micro)	0.946457	Precision-Recall AUC for Label 1	0.94506
Jaccard Score (Micro)	0.898356	Training Score	0.998875
Hamming Loss	0.0509745	Test Score	0.907046

Kaynak: Araştırmacı tarafından oluşturulmuştur.

XGBoost algoritması ile yapılan modelin performansı çeşitli metriklerle değerlendirilmiştir. "Micro" terimi, genellikle çok sınıflı (multiclass) ve dengesiz veri setlerinde kullanılan bir değerlendirme yaklaşımını ifade eder. Özellikle, Precision, Recall, F1-Score gibi metriklerin mikro (micro) versiyonları, tüm sınıfların performansını birleştirerek tek bir sayısal değer verir. Bu yaklaşım, her sınıfın etkisini eşit olarak değerlendirir ve her örneğin doğru veya yanlış sınıflandırılmasını toplar.

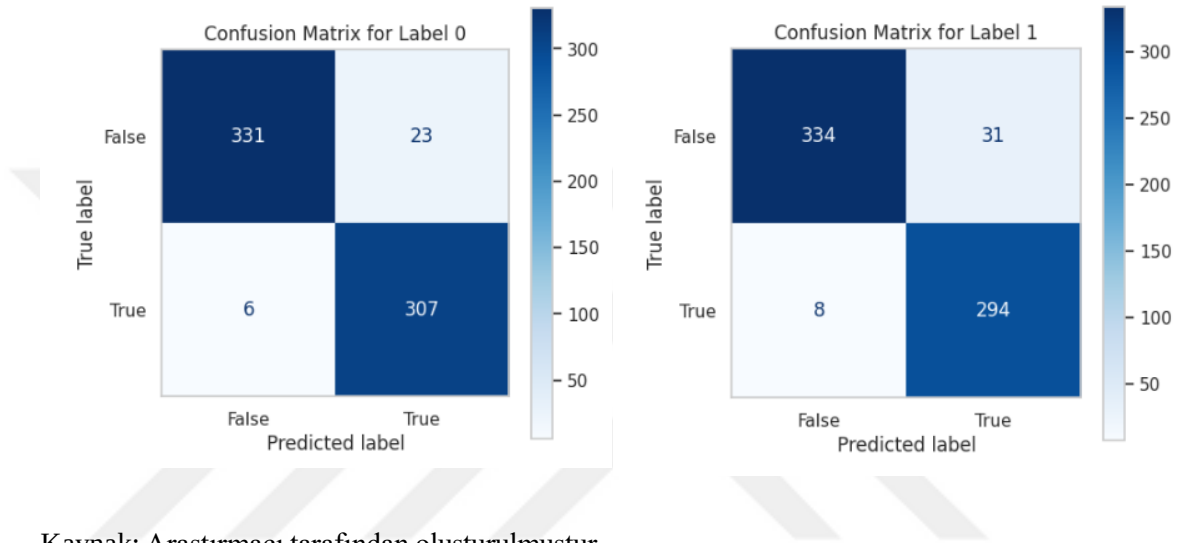
Modelin genel doğruluk oranı (Accuracy) %90.7 olarak elde edilmiştir. Precision (Micro) değeri ise %91.8'dir, bu da modelin sınıflandırma doğruluğunun yüksek olduğunu göstermektedir. Recall (Micro) değeri %97.7, bu da modelin pozitif sınıfı doğru bir şekilde tahmin etme oranının oldukça yüksek olduğunu ifade etmektedir. F1-Score (Micro) değeri %94.6 ile modelin doğruluk (precision) ve recall oranlarını dengeleyerek güçlü bir genel performans sergilediğini göstermektedir.

Jaccard Score (Micro) %89.8 olarak belirlenmiştir, bu da modelin sınıflar arasındaki benzerliği iyi bir şekilde yakaladığını gösterirken, Hamming Loss %5.1 ile düşük bir hata oranı olduğunu belirtmektedir.

Ayrıca, etiketlere göre modelin performansı daha ayrıntılı bir şekilde değerlendirilmiştir. Kampanya 1 için MCC değeri 0.914, Precision-Recall AUC değeri ise 0.960'dır, bu da modelin bu sınıfta oldukça başarılı olduğunu göstermektedir. Kampanya 2 için MCC değeri 0.885, Precision-Recall AUC değeri

ise 0.945'tir, bu da modelin ikinci sınıfta da güçlü performans sergilediğini göstermektedir. Eğitim verisi üzerindeki başarı oranı %99.9 iken, test verisi üzerindeki başarı oranı ise %90.7 olarak kaydedilmiştir. Bu sonuçlar, modelin test verisindeki genel performansının yüksek olduğunu ve aşırı öğrenme (overfitting) sorunundan kaçınıldığını göstermektedir.

Şekil 61: XgBoost için Karmaşıklık Matrisi (Confusion Matrix)



Kaynak: Araştırmacı tarafından oluşturulmuştur.

İlk matris, Kampanya 1 için modelin katılım tahminlerinin büyük ölçüde doğru olduğunu gösteriyor. Doğru pozitif (TP) değerinin 307 ve doğru negatif (TN) değerinin 331 olması, modelin katılım gösteren müşterileri doğru bir şekilde tanıyıp, katılım göstermeyenleri de doğru şekilde sınıflandırdığı anlamına gelir. Ancak, modelin yanlış pozitif (FP) 23 ve yanlış negatif (FN) 6 olması, bazı katılım göstermeyen müşterilerin katılım gösterecek olarak yanlış sınıflandırıldığını ve bazı katılım gösterenlerin de katılım göstermeyen olarak yanlış etiketlendiğini gösteriyor. Bu hata oranı, modelin genellikle doğru tahminler yapsa da, katılım gösterecek müşterileri kaçırma ve katılım göstermeyecek müşterileri yanlışlıkla katılım gösteriyor olarak etiketleme riskinin az da olsa olduğunu gösterir.

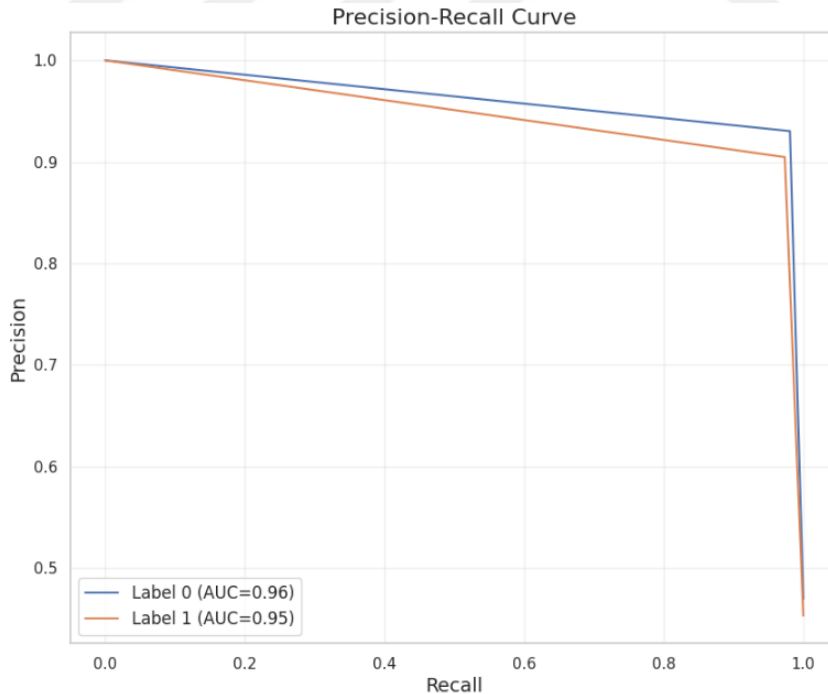
İkinci matris, Kampanya 2 için modelin performansının biraz daha düşük olduğunu gösteriyor. Doğru pozitif (TP) 294 ve doğru negatif (TN) 334 olmakla birlikte, yanlış pozitif (FP) 31 ve yanlış negatif (FN) 8 oranları daha yüksektir. Bu durum, modelin katılım gösteren bazı müşterileri yanlışlıkla katılım göstermeyen

olarak etiketlediğini ve katılım göstermeyen bazı müşterileri ise yanlışlıkla katılım gösterecek olarak sınıflandırdığını gösteriyor. Özellikle yanlış pozitif (FP) sayısının daha yüksek olması, modelin katılım göstermeyen kişileri katılım gösteriyor olarak yanlış sınıflandırması ve böylece pazarlama kampanyalarını yanlış kişilere yönlendirmesi riski taşıdığını ifade eder.

Kampanya 1 için modelin katılım tahminlerinde daha doğru sonuçlar verdiği, ancak Kampanya 2'de katılım durumunu tahmin etme noktasında daha fazla hata yaptığı söylenebilir. Bu fark, her kampanyanın hedef kitlesi ve müşteri davranışlarıyla ilgili farklı özellikler gösterebileceğini ve modelin her iki kampanya için de optimize edilmesi gerektiğini gösteriyor.

Hiperparametre optimizasyonu yapılmadığı göz önünde bulunarak temel modelden çıkan sonuçlar modelin çok sınıflandırma problemlerinde ve dengesiz veri setlerinde başarılı sonuçlar verdiği gözlemlenmiştir. Hiperparametre ayarlamaları yapılarak modelin daha iyi bir hale getirilebileceği unutulmamalıdır. Sonraki çalışmalarda dikkate alınarak iyileştirilebilir.

Şekil 62: Kesinlik – Duyarlılık Grafiği



Kaynak: Araştırmacı tarafından oluşturulmuştur.

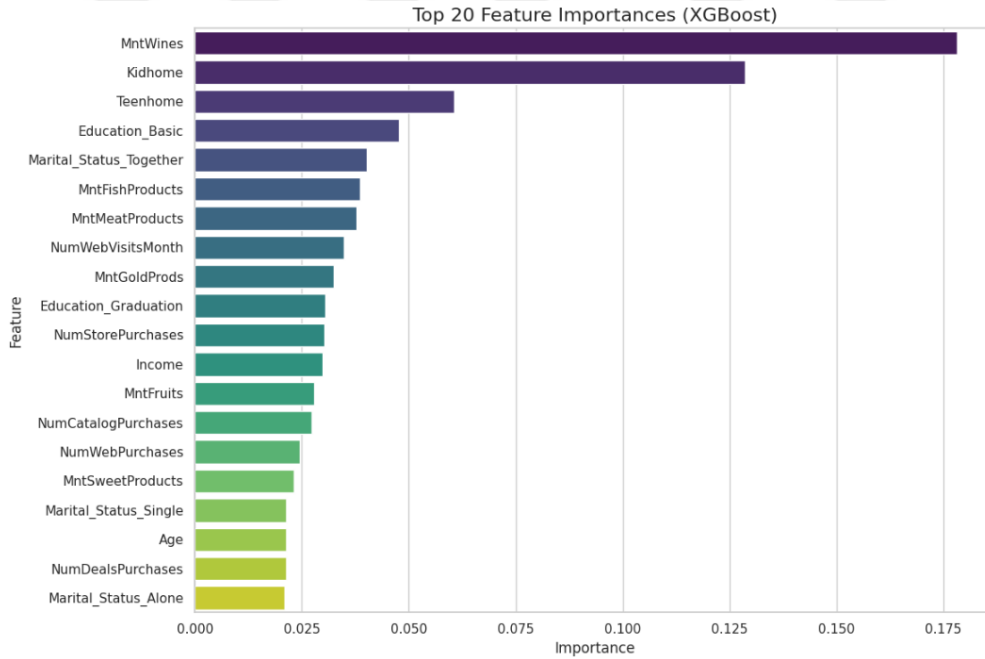
Kampanya 1 için AUC (Eğri Altındaki Alan) değeri 0.96 olarak hesaplanmıştır. Kampanya 2 için AUC değeri 0.95 olarak hesaplanmıştır. Grafik, modelin hem Precision (Kesinlik) hem de Recall (Duyarlılık) açısından oldukça yüksek performans gösterdiğini ortaya koymaktadır.

Precision değeri pozitif tahminlerin ne kadar doğru olduğunu recall değeri ise gerçekte pozitif olan örneklerin ne kadarının doğru olduğunu gösterir. Bu iki değer yüksedikçe eğri altında kalan alan (AUC) yükselir. Eğri altında kalan alanın değeri 1'e yaklaştıkça modelin performansı artar.

Görüldüğü gibi, Kampanya 1 için Precision-Recall eğrisi Kampanya 2'ye kıyasla biraz daha iyi performans göstermektedir. Ancak her iki sınıf için de modelin başarısı oldukça yüksektir.

Bu grafik, modelin iki farklı etiket üzerinde tutarlı ve başarılı performans sergilediğini kanıtlamaktadır.

Şekil 63: Öznitelik Önem Düzeyi Analizi



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Bu grafik, XGBoost algoritması kullanılarak oluşturulan modelin, tahmin yaparken hangi özelliklere ne kadar önem verdiğini gösteren "özellik önem

sıralaması" (Feature Importance) analizidir. Yatay ekseninde özelliklerin önem derecesi (Importance) bulunurken, dikey ekseninde modele dahil edilen özellik isimleri yer almaktadır. Modelin performansını en fazla etkileyen özellik şarap harcamaları olup, bu özellik diğerlerine kıyasla açık bir farkla en yüksek katkıyı sağlamaktadır. İkinci sırada evdeki çocuk sayısı, üçüncü sırada ise evdeki genç sayısı yer alarak, demografik faktörlerin de kampanya seçiminde önemli bir rol oynadığını göstermektedir.

Bunun yanı sıra, eğitim düzeyi değişkenlerinden temel eğitim ve mezuniyet düzeyi gibi faktörler sıralamada dikkat çekmektedir. Aynı şekilde, harcama kalıplarını temsil eden balık ürünleri harcamaları, et ürünleri harcamaları, altın ürünleri harcamaları ve meyve harcamaları gibi değişkenler de modeldeki karar süreçlerine kampanya seçimine katkıda bulunmaktadır.

Ek olarak, web ziyaret sayısı ve alışveriş sayıları gibi tüketici davranışını gösteren değişkenler de modelin önemli bulduğu özellikler arasındadır. Daha az önem verilen özellikler arasında ise yaş, indirimli ürün alışveriş sayısı ve medeni durumu ifade eden "Marital_Status_Alone" ve "Marital_Status_Single" gibi değişkenler yer almaktadır.

Bu analiz, modelin harcama alışkanlıkları, demografik özellikler ve tüketici davranışlarını güçlü bir şekilde dikkate aldığını ortaya koymaktadır.

3.5.4.4. CatBoost Algoritması

CatBoost, birden fazla sınıf içeren hedef değişkenlerle çok sınıflı sınıflandırma problemlerinde üstün performans sunan bir gradient boosting algoritmasıdır. Özellikle kategorik veriler üzerinde yüksek performans gösterir. Bu tür problemler için, her sınıfa ait olasılık tahminlerini optimize etmek amacıyla softmax loss fonksiyonunu kullanır. Model, her veri örneği için sınıflar arasındaki sınırları öğrenir ve tahminlerini olasılık dağılımı olarak sunar. Bir müşterinin belirli bir kampanyayı seçme ihtimali değerlendirildiğinde, CatBoost her kampanya için ayrı bir olasılık hesaplayarak en yüksek olasılıklı sınıfı tahmin eder. Kategorik verilerin işlenmesinde, sıralama istatistikleriyle hedef değişkenin ilişkisini dikkate alır ve bu süreçte overfitting riskini azaltır. Ayrıca, ordered boosting yöntemi

sayesinde, veri sızıntısını önleyerek modelin daha iyi genellenmesini sağlar. Bu yenilikçi yaklaşımlar, CatBoost'u çok sınıflı sınıflandırma problemleri için güçlü ve güvenilir bir seçenek haline getirir.

Çok sınıflı sınıflandırma için CatBoost, her bir sınıfın olasılığını tahmin etmek için her veri noktasına ait sınıfın doğru olma olasılığını hesaplar. Her sınıf için ayrı bir tahmin yapmaz. Bunun yerine, tüm sınıflar arasındaki ilişkiyi aynı anda öğrenmek için karar ağaçlarını kullanır.

Kategorik özellikler, CatBoost'un sıralama yöntemiyle otomatik olarak işlenir. Eksik veriler varsa CatBoost bunları otomatik olarak işler veya doldurur. Ağaç yapıları, sınıf sınırlarını öğrenmek için optimize edilir.

CatBoost algoritmasının veri ön işleme adımlarına ihtiyaç duymadan ham verilerle modellerin çalıştırabildiği bilinmektedir. Fakat algoritma çalıştırıldığında veri boyutlarıyla ilgili hatalar verdiği için veri ön işleme adımları gerçekleştirilerek veri boyutlarının birbirine eşitlenmesi için smote yöntemiyle veri artırma yoluna gidilmiştir. Eşitlenen bu veriler için hedef değişkenin hala dengesiz bir dağılım oluşturduğu bilinmektedir. Model çalıştırılıp çıktılar analiz edildiğinde modelin precision ve recal değerlerinin 1 olduğu dikkat çekmektedir.

Şekil 64: Overfit Olmuş Performans Metrikleri

	precision	recall	f1-score	support
0.0	1.00	1.00	1.00	356
1.0	1.00	1.00	1.00	311
accuracy			1.00	667
macro avg	1.00	1.00	1.00	667
weighted avg	1.00	1.00	1.00	667

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Diğer metrikler incelenerek modelin ezberlediği kanaatine varılmıştır. Bu durumun önüne geçmek için CatBoost L2 regularizasyonu kullanılmıştır. L2 regularizasyonu, modelin karar sınırlarını belirleyen parametreleri (örneğin, ağaç yapısındaki dallanmalar) cezalandırır. Bunu, kayıp fonksiyonuna (loss function) parametrelerin karesini ekleyerek yapar. Böylece büyük ağırlık değerlerinin etkisi sınırlanır ve modelin daha basit, genellenebilir bir yapıya sahip olması sağlanır. Bu

yöntem, modelin aşırı öğrenme riskini azaltarak daha dengeli bir yapı oluşturmasını sağlamıştır. Model performansını daha güvenilir bir şekilde değerlendirebilmek için k-fold (5) çapraz doğrulama yapılmıştır. StratifiedKFold yöntemi ile veri alt kümelerinde sınıf dağılımlarının korunması sağlanmış, böylece modelin farklı veri bölümlerindeki başarısı ölçülmüştür. Bu süreçler, modelin genel doğruluğunu artırmak ve çeşitli sınıflardaki başarısını değerlendirmek amacıyla uygulanmıştır. Tüm bu işlemler uygulanırken Catbosst modeli diğer modellere göre çok hızlı çıktılar üretmiştir. Yapılan işlemler sonucunda CatBoost algoritmasının çıktısı aşağıdaki gibidir.

Şekil 65: CatBoost Algoritması için Performans Metrikleri Çıktısı

```
=== K-Fold CV Results ===
```

Metric	Label 0 Average	Label 1 Average
precision	0.930146	0.923678
recall	0.928013	0.921114
f1-score	0.927918	0.921007
Jaccard Index	0.869772	0.858356
Hamming Loss	0.0719874	0.0788862
Matthews Correlation Coefficient (MCC)	0.858154	0.84474

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Precision (kesinlik) değeri, Kampanya 1 (Label 0) için %93 değeri, modelin bu kampanyada doğru tahmin oranının oldukça yüksek olduğunu göstermektedir. Kampanya 2 (Label 1) için %92 değeri ise benzer şekilde yüksek, ancak Kampanya 1'e kıyasla biraz daha düşük bir kesinlik sunmaktadır.

Recall (duyarlılık) değeri, Kampanya 1 (Label 0) için recall değeri %92 olarak hesaplanmış, bu da modelin gerçek pozitifleri başarılı bir şekilde bulduğunu gösteriyor. Kampanya 2 (Label 1) için recall değeri %92, bu kampanyada modelin performansının yine yüksek olduğunu ancak Kampanya 1'den bir miktar düşük kaldığını gösteriyor.

F1 score değeri, precision ve recall değerlerinin harmonik ortalamasıdır ve modelin genel sınıflandırma başarısını özetler. Kampanya 1 ve Kampanya 2 için de küçük bir farkla %92 olarak hesaplanmıştır. Bu değerler modelin her iki kampanyada

da dengeli ve güçlü bir performans gösterdiğini doğrulamaktadır. Kampanya 1’de Jaccard Index’in biraz daha yüksek olması, bu kampanyada modelin tahminlerinin daha net olduğunu göstermektedir.

Hamming Loss değeri, Kampanya 1 için %7.19, Kampanya 2 için %7.8 değerleri, modelin hata oranının düşük olduğunu ve genel olarak güvenilir tahminler yaptığını gösteriyor. Kampanya 2’de hata oranı Kampanya 1’e göre bir miktar daha yüksektir.

MCC, modelin tüm sınıflar için genel doğruluğunu ölçen ve özellikle dengesiz veri kümelerinde etkili bir metriktir. Kampanya 1 için %85, Kampanya 2 için %84 olarak hesaplanmıştır. Bu değerler, modelin her iki kampanyada da tutarlı ve güçlü bir performans sergilediğini ancak Kampanya 1’de bir miktar daha iyi olduğunu göstermektedir.

Şekil 66: CatBoost Algoritması için Sınıflandırma Çıktısı

Classification Report for Label 0 - Fold 5:				
	precision	recall	f1-score	support
0.0	0.98	0.86	0.92	333
1.0	0.88	0.98	0.93	333
accuracy			0.92	666
macro avg	0.93	0.92	0.92	666
weighted avg	0.93	0.92	0.92	666

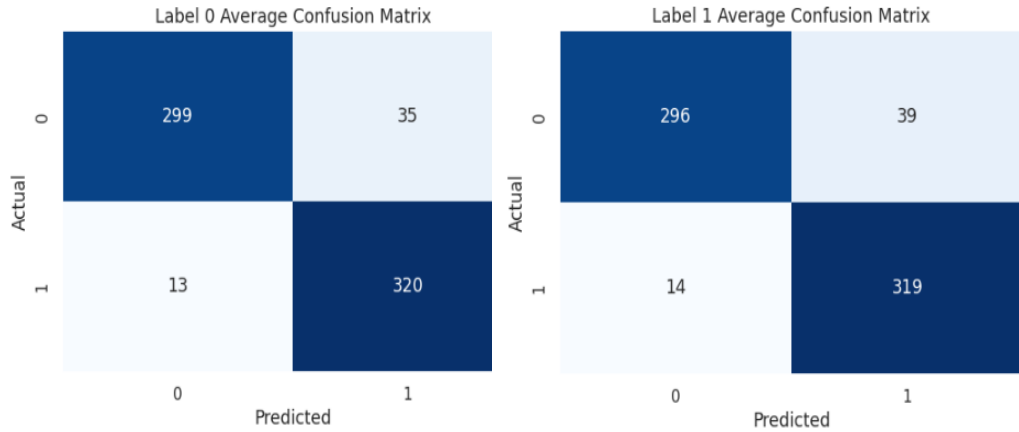
Classification Report for Label 1 - Fold 5:				
	precision	recall	f1-score	support
0.0	0.96	0.87	0.91	326
1.0	0.88	0.96	0.92	340
accuracy			0.91	666
macro avg	0.92	0.91	0.91	666
weighted avg	0.92	0.91	0.91	666

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Bu çıktı dan hem sınıf bazında hem de etiket bazında modelin başarısı incelenebilir. Overfit olmayan model çıktısı gayet başarılı tahminlerde bulunduğunu F1 score metriği ile gösteriyor.

Model iki kampanya için de oldukça başarılı sonuçlar elde etmiştir. Kampanya 1’de performansın biraz daha yüksek olması, bu kampanya verilerinde modelin daha iyi öğrenmiş olabileceğine işaret etmektedir.

Şekil 67: Karmaşıklık Matrisi(Confusion Matrix)

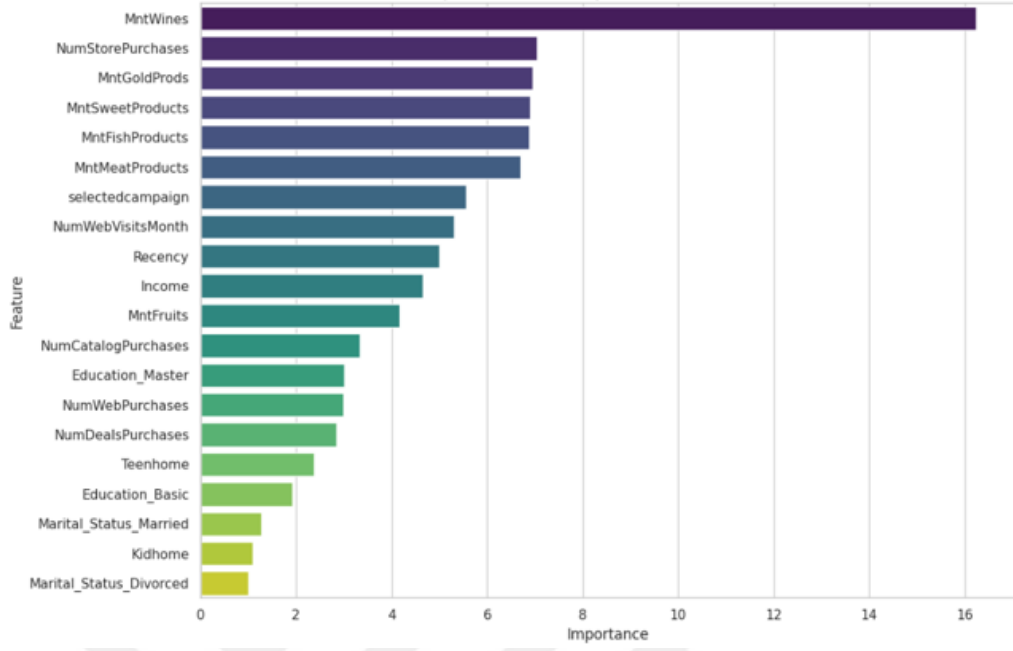


Kaynak: Araştırmacı tarafından oluşturulmuştur.

Birinci matris, Kampanya 1 için doğru pozitif (TP) değerinin 320 ve doğru negatif (TN) değerinin 229 olması, modelin katılım gösteren müşterileri doğru bir şekilde tanıyıp, katılım göstermeyenleri de doğru şekilde sınıflandırdığı anlamına gelir. Modelin yanlış pozitif (FP) 35 ve yanlış negatif (FN) 13 olması, bazı katılım göstermeyen müşterilerin katılım gösterecek olarak yanlış sınıflandırıldığını ve bazı katılım gösterenlerin de katılım göstermeyen olarak yanlış etiketlendiğini gösteriyor. Bu hata oranı, modelin genellikle doğru tahminler yapsa da, katılım gösterecek müşterileri kaçırma ve katılım göstermeyecek müşterileri yanlışlıkla katılım gösteriyor olarak etiketleme riskinin az da olsa olduğunu gösterir.

İkinci matris, Kampanya 2 için modelin performansının biraz daha düşük olduğunu gösteriyor. Doğru pozitif (TP) 296 ve doğru negatif (TN) 319 olmakla birlikte, yanlış pozitif (FP) 39 ve yanlış negatif (FN) 14 oranları daha yüksektir. Bu durum, modelin katılım gösteren bazı müşterileri yanlışlıkla katılım göstermeyen olarak etiketlediğini ve katılım göstermeyen bazı müşterileri ise yanlışlıkla katılım gösterecek olarak sınıflandırdığını gösteriyor. Özellikle yanlış pozitif (FP) sayısının daha yüksek olması, modelin katılım göstermeyen kişileri katılım gösteriyor olarak yanlış sınıflandırması ve böylece pazarlama kampanyasını yanlış kişilere yönlendirmesi riski taşıdığını ifade eder.

Şekil 68: Özellik Önem Düzeyi Analizi



Kaynak: Araştırmacı tarafından oluşturulmuştur.

Grafikte CatBoost modeli için en önemli 20 değişkenin önem sıralaması yer almaktadır. Modelin en güçlü belirleyicisi, müşterilerin şarap harcaması (MntWines) olurken, bunu mağazadan satın alma sayısı, altın ürünler ve tatlı ürünler gibi harcama kalemleri takip etmiştir. Bu, müşterilerin alışveriş tercihlerini anlamada tüketim kalemlerinin kilit rol oynadığını göstermektedir. Dijital davranışlar da dikkat çekici bir şekilde modele katkı sağlamıştır; aylık web sitesi ziyaret sayısı, kampanya katılımı ve internetten yapılan alışverişler, müşterilerin dijital etkileşimlerinin satın alma davranışları üzerindeki etkisini ortaya koymaktadır.

Demografik değişkenler de modelin performansına önemli katkılar sağlamıştır. Gelir seviyesi, eğitim durumu (yüksek lisans ve temel eğitim düzeyleri), medeni durum (evli ve boşanmış), evdeki çocuk ve genç sayısı gibi faktörler, müşterilerin alışveriş alışkanlıklarını etkileyen temel özellikler olarak öne çıkmıştır. Bunun yanında, et ve balık ürünlerine yönelik harcamalar gibi diğer tüketim kalemleri de müşterilerin satın alma davranışlarını anlamada önemli bir rol oynamıştır. Genel olarak, bu bulgular, hem müşterilerin harcama alışkanlıklarının hem de demografik ve davranışsal özelliklerinin modelin tahmin gücü üzerinde kritik bir etkiye sahip

olduğunu ortaya koymaktadır. Bu sonuçlar, hedefe yönelik pazarlama stratejileri geliştirmek ve müşteri segmentasyonu yapmak için değerli iç görüler sunmaktadır.

Özetle, CatBoost, diğer gradient boosting algoritmalarına kıyasla daha doğru tahminler sağladığı bilindiğine göre model üzerinde biraz daha oynamalar yapılırsa daha yüksek tahmin sonuçlarına ulaşılabilir.

3.5.4.5. Support Vector Machine

Destek vektör makineleri (SVM), hem sınıflandırma hem de regresyon problemlerinde kullanılan güçlü bir makine öğrenmesi algoritmasıdır. SVM, veriyi yüksek boyutlu bir uzaya projekte ederek, farklı sınıfları ayıran en iyi hiper düzlemi bulmayı amaçlar. Bu hiper düzlem, sınıflar arasındaki marjini maksimize eder, bu da modelin genelleme yeteneğini artırır. SVM'nin bir diğer avantajı, karmaşık veri kümelerinde bile etkili performans gösterebilmesi ve doğrusal olmayan problemlerde çekirdek fonksiyonları kullanarak başarılı sonuçlar elde edebilmesidir. Ancak, büyük veri setlerinde hesaplama maliyeti yüksek olabilir ve modelin Hiperparametre ayarlamaları dikkatli bir şekilde yapılmalıdır. Destek vektör makineleri yüksek boyutlu uzaylarda oldukça etkilidir. Özellik sayısının fazla olduğu durumlarda iyi çalışır. Aşırı uyuma karşı dayanıklıdır. Ancak, özellikle büyük veri kümeleri için hesaplama açısından biraz yoğun olabilir. Parametrelerin dikkatlice ayarlanmasını ve kernel seçimini gerektirir. Çok sınıflı sınıflandırma problemlerinde SVM doğrudan birden fazla sınıflı aynı anda ayırmak yerine, genellikle ikili sınıflandırma algoritmasını genişleterek çalışır. Dengesiz sınıflandırma, bir sınıfın çok daha fazla örneğe sahip olması ve diğer sınıfların çok az sayıda temsil edilmesi durumudur. SVM'nin, marjı maksimize etme amacı, dengesiz veri kümelerinde küçük sınıfı göz ardı etmesine neden olabilir. Dengesiz sınıflandırma problemlerinde SVM, sınıf ağırlıklarını ayarlayarak ve hiperparametre ayarlamaları yaparak güvenli sonuçlar verebilir.

Şekil 69: Destek Vektör Makinesi Algoritması için Sınıflandırma Çıktısı (k=5)

```
Fitting 5 folds for each of 12 candidates, totalling 60 fits
Best parameters found: {'estimator__C': 0.1, 'estimator__gamma': 'scale', 'estimator__kernel': 'linear'}
SVC Classifier Results:
F1 Score (weighted): 1.0
Jaccard Score (weighted): 1.0
Accuracy Score: 1.0

Classification Report:

              precision    recall  f1-score   support

     0           1.00       1.00       1.00        42
     1           1.00       1.00       1.00        32

   micro avg       1.00       1.00       1.00       74
   macro avg       1.00       1.00       1.00       74
  weighted avg       1.00       1.00       1.00       74
   samples avg       0.14       0.14       0.14       74

Hamming Loss: 0.0
Precision-Recall AUC for label 0: 1.0
Precision-Recall AUC for label 1: 1.0
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Svm algoritması 5 katlı çapraz doğrulama ve hiperparametre optimizasyonu yapılarak çalıştırılmış fakat performans sonuçlarından da anlaşılacağı gibi model veriyi ezberlemiş overfit olmuştur. Modelin doğruluğunu yansıtan metriklerin hepsi %100 dür. Yani model veriyi mükemmel derecede anlamış ve veriyi ezberlemiştir. Eğitim verisini ezberlediği gibi test verisinde hiçbir veriyi anlayamayacaktır.

K-Katlı çapraz doğrulamayı 10 yaparak ve bazı parametreleri değiştirerek yeniden çalıştıracamız. OneVsRestClassifier kullanarak çok sınıflı sınıflandırma problemlerinde SVM nin başarısının yükseltilmesi hedeflenmektedir. Aşağıda hiperparametre ayarlamaları yapılmış SVM modeli çıktısı gösterilmektedir.

Şekil 70: Destek Vektör Makinesi Algoritması için Sınıflandırma Çıktısı (k=10)

```
Fitting 10 folds for each of 8 candidates, totalling 80 fits
Best parameters found: {'svc__estimator__C': 2, 'svc__estimator__gamma': 'scale', 'svc__estimator__kernel': 'rbf'}
SVC Classifier Results:
F1 Score (weighted): 1.0
Jaccard Score (weighted): 1.0
Accuracy Score: 1.0
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	42
1	1.00	1.00	1.00	32
micro avg	1.00	1.00	1.00	74
macro avg	1.00	1.00	1.00	74
weighted avg	1.00	1.00	1.00	74
samples avg	0.14	0.14	0.14	74

```
Precision-Recall AUC for label 0: 1.0
Precision-Recall AUC for label 1: 1.0
Hamming Loss: 0.0
Eğitim Skoru: 1.0
Test Skoru: 1.0
```

Kaynak: Araştırmacı tarafından oluşturulmuştur.

10 katlı çapraz doğrulama yapılmış ve hiperparametreleri birden fazla kez değiştirilerek çalıştırılmıştır. Model sonuçlarında görüldüğü gibi modelin verilere aşırı uyumu devam etmektedir.

Model tüm değişkenleri mükemmel derecede anlamış ve %100 doğru performans göstermektedir.

Hamming loss değeri 0'dır. Yani model tüm sınıflardaki tüm etiketleri doğru tahmin etmiştir.

Kampanya seçimini barındıran hedef değişkenimiz çok değişkenli sınıflandırma problemi kapsamında sınıflar arasındaki etiketlerin dengesiz dağılımı durumu göz önüne alınarak değerlendirilmelidir.

Bu sebeple smote yöntemi kullanılarak veri seti çoğaltılarak dengeli hale getirilmeye çalışılmış ve model sonuçları incelenmiştir.

Aşağıda smote yöntemi kullanılarak yeniden örneklenen modelin çıktısı görülmektedir.

Şekil 71: Destek Vektör Makinesi Algoritması için Performans Metrikleri

	precision	recall	f1-score	support
0	0.14	0.14	0.14	42
1	0.03	0.03	0.03	29
micro avg	0.09	0.10	0.10	71
macro avg	0.09	0.09	0.09	71
weighted avg	0.10	0.10	0.10	71
samples avg	0.01	0.01	0.01	71

Jaccard Score for Label 0: 0.07692307692307693

Jaccard Score for Label 1: 0.016666666666666666

Kaynak: Araştırmacı tarafından oluşturulmuştur.

Sınıflandırma raporu, modelin dengesiz bir veri seti üzerinde düşük performans sergilediğini göstermektedir. Sınıf 0 için precision, recall ve F1-skoru 0.14 iken, sınıf 1 için bu değerler oldukça düşüktür (0.03). Bu durum, modelin özellikle azınlık sınıfını (sınıf 1) doğru bir şekilde tahmin edemediğini ve dengesizlikten ciddi şekilde etkilendiğini ortaya koymaktadır. Ayrıca, mikro, makro ve ağırlıklı ortalamalar da oldukça düşük (0.09-0.10 aralığında) olup, genel sınıflandırma performansının zayıf olduğunu göstermektedir. Bu durumda SVM modelinin çok sınıflı sınıflandırma problemlerinde dengesiz dağılım söz konusu olduğu için başarısız olduğunu söyleyebiliriz.

SONUÇ

Pazar sektöründe faaliyet gösteren firmalar müşteri memnuniyetini artırma ve rekabet avantajı elde etme amacıyla pazar araştırmaları ve veri analizi sonuçlarına dayalı stratejiler benimsemelidir. Özellikle, kişiselleştirilmiş fiyatlandırma stratejileri, müşterilere daha cazip teklifler sunma ve onların alışveriş deneyimlerini geliştirme konusundaki önemli bir araçtır. Bu stratejiler, müşterilerin satın alma davranışlarını ve tercihlerini derinlemesine anlamak için makine öğrenmesi algoritmalarının sağladığı verileri kullanır. Müşteri davranışlarını ve geçmiş satın alma verilerini analiz ederek, işletmelerin belirli ürünleri paketler halinde sunması ve bu ürünlerin fiyatlarını kişiselleştirmesi, müşteri memnuniyetini artırmada etkili bir yöntemdir. Bu tür paket teklifler, müşterilere daha değerli ve uygun fiyatlı alışveriş seçenekleri sunar, dolayısıyla müşteri sadakatini güçlendirir ve satışları artırabilir. Kampanya seçimlerinde şirketlerin müşterilerin ihtiyaçlarını belirleyip doğru kampanyalar sunarak satış hacmini artırma gayreti göstermekle birlikte müşterilerin satın alabileceği fiyatlara karşı hassasiyet gösterip müşteri sadakati elde edebilirler.

Makine öğrenmesi algoritmaları, bu süreçte önemli bir rol oynar. Rastgele Ormanlar, XgBoost ve CatBoost gibi algoritmalar, kampanya kişiselleştirilmesinde yüksek doğruluk ve etkili sonuçlar sağlar. Bu algoritmaların kullanımı, müşteri verilerini analiz ederek en uygun fiyatları belirlemeye ve pazarlama kampanyalarını optimize etmeye yardımcı olur. Veri madenciliği ve makine öğrenmesi tekniklerinin entegrasyonu, işletmelere müşteri ihtiyaçlarını daha iyi anlama ve pazarlama çabalarını kişiselleştirme yeteneği kazandırır. Bu stratejiler, pazarın taleplerine daha iyi yanıt verme ve uzun vadeli başarıyı destekleme potansiyelini artırır. İşletmeler, bu teknikleri kullanarak müşteri memnuniyetini yükseltmekte ve rekabet avantajını güçlendirmekte önemli bir adım atmış olurlar.

Bu çalışmada, kampanya katılım durumunu tahmin etmek amacıyla çok etiketli bir sınıflandırma problemi ele alınmış ağaç tabanlı yöntemler ve SVM algoritması kullanılarak kampanya tahmini yapılmıştır. Çalışma içerisinde kullanılan algoritmalar Random Forest, Desicion Tree, XgBoost, CatBoost ve SVM algoritmasıdır. Model başarıları değerlendirilerek şirketlerin kişiye özel kampanyalar yapmayı destekleyen ve kullanım önerisi sunarak şirketlerin yapmayı hedeflediği

kampanyalarda doğru tahminlere ulaşmaları açısından yol göstermektedir.

Random Forest algoritması kullanılarak geliştirilen model, dengesiz sınıf dağılımını dengelemek için `class_weight='balanced'` parametresi ile eğitilmiş ve doğruluk, precision, recall, F1-Score gibi temel metriklerin yanı sıra Matthews Correlation Coefficient (MCC), Hamming Loss, ve Precision-Recall AUC gibi çok etiketli sınıflandırmaya özgü metriklerle değerlendirilmiştir.

Elde edilen sonuçlar, modelin yüksek doğruluk (%99.33), güçlü F1-Score (%95.94) ve düşük Hamming Loss (%0.33) ile dengeli ve başarılı bir performans sergilediğini göstermektedir. Öznitelik önem düzeyi analizi, kampanya katılımını etkileyen en belirleyici faktörlerin gelir seviyesi, şarap ve et harcamaları, ve katalog alışverişleri olduğunu ortaya koymuştur. Buna karşın, eğitim durumu, evdeki çocuk sayısı ve şikayet durumu gibi değişkenlerin daha düşük bir etkiye sahip olduğu görülmüştür.

Random Forest algoritmasının kampanya katılım durumunu doğru ve güvenilir bir şekilde tahmin edebildiği ve özellikle müşteri harcamaları ile gelir seviyesinin kampanya katılımını belirlemede kritik öneme sahip olduğu belirlenmiştir.

Kampanya katılım durumunu tahmin etmek için Decision Tree algoritması kullanılmış ve model performansı hiperparametre optimizasyonu, çapraz doğrulama ve veri artırma teknikleriyle iyileştirilmiştir. Model, dengesiz sınıf dağılımına rağmen yüksek doğruluk (%97,09), ağırlıklı F1 skoru ve Jaccard skoru ile başarılı bir performans sergilemiştir. Ancak, model Kampanya 2'de mükemmel sonuçlar verirken, Kampanya 1 için daha düşük doğruluk ve recall değerleri nedeniyle sınırlı performans göstermiştir.

Overfitting sorununu önlemek amacıyla, sınıf dengesizliğini gidermek için smote yöntemiyle veri artırma yapılmış, değişkenler ağırlıklandırılmış ve hiperparametre optimizasyonu ile model yeniden çalıştırılmıştır. Öğrenme eğrisi, modelin eğitim verisini ezberleyerek overfitting yaşadığını ancak çapraz doğrulama ve dengeli veri seti kullanımıyla bu sorunun kısmen giderildiğini göstermektedir. Gelecekte, daha geniş veri setleriyle yapılacak çalışmaların modelin genelleme kabiliyetini artıracakı öngörülmektedir.

XGBoost algoritması kullanılarak iki farklı kampanya için müşteri katılım durumu analiz edilmiştir. Çalışma, dengesiz veri setlerini dengelemek amacıyla SMOTE yöntemiyle yeniden örnekleme işlemi gerçekleştirmiş ve elde edilen dengeli veri seti üzerinde XGBoost algoritması uygulanmıştır. Model, hiperparametre optimizasyonu yapılmaksızın, dengesiz veri setlerinde ve çok etiketli sınıflandırma problemlerinde başarılı sonuçlar vermiştir.

Modelin performansı; doğruluk, precision, recall, F1-score, Jaccard skor ve Hamming Loss gibi metriklerle değerlendirilmiştir. Elde edilen sonuçlara göre model, test verisinde %90.7 doğruluk oranı sağlamış, mikro düzeyde %91.8 precision, %97.7 recall ve %94.6 F1-score değerleriyle güçlü bir performans sergilemiştir. Ayrıca, MCC ve Precision-Recall AUC değerleri, her iki kampanya için de modelin yüksek başarı elde ettiğini göstermiştir.

Kampanya 1'de model, doğru sınıflandırma oranının daha yüksek olduğu ve katılım durumunu daha doğru tahmin ettiği görülmüştür. Kampanya 2'de ise katılım göstermeyen kişilerin yanlış sınıflandırılma oranı biraz daha yüksek çıkmıştır. Bu durum, her kampanyanın farklı özelliklere sahip müşteri gruplarına hitap ettiğini ve modele özgü optimizasyon gerektirdiğini göstermektedir.

Öznitelik önem sıralaması analizine göre, tahminlerde en etkili faktörler arasında şarap harcamaları, evdeki çocuk ve genç sayısı, eğitim düzeyi, balık, et ve meyve gibi harcamalar yer almaktadır. Bu durum, tüketici harcama alışkanlıklarının ve demografik özelliklerin kampanya katılımını belirlemede önemli bir rol oynadığını ortaya koymaktadır.

XGBoost algoritmasının çok etiketli sınıflandırma problemleri ve dengesiz veri setlerinde başarılı bir şekilde uygulanabileceği ve hiperparametre optimizasyonu ile daha iyi sonuçlar elde edilebileceği görülmüştür.

Kullanılan bir diğer algoritma müşterilerin kampanya katılım durumlarını sınıflandırmak amacıyla CatBoost algoritmasıdır. CatBoost, özellikle kategorik verilerle çalışmada üstün performans göstermesi ve veri sızıntısını önleyen ordered boosting yöntemiyle öne çıkmaktadır. Model, çok sınıflı sınıflandırma probleminde her sınıfa ait olasılık tahminlerini optimize etmiş ve hem Kampanya 1 hem de Kampanya 2 için başarılı tahminler üretmiştir.

Model performansı, precision, recall, F1 score, Hamming Loss ve MCC gibi metriklerle değerlendirilmiştir. Kampanya 1 için F1 score %92, MCC %85 olarak hesaplanmış; Kampanya 2’de ise bu değerler sırasıyla %92 ve %84 olarak bulunmuştur. Precision ve recall değerleri de benzer şekilde yüksek olup, modelin hem doğru tahmin oranı hem de gerçek pozitifleri bulmadaki başarısını ortaya koymaktadır.

Çalışmada, veri dengesizliklerinin smote yöntemiyle giderilmesi, overfitting riskini azaltmak için L2 regularizasyonun uygulanması ve StratifiedKFold çapraz doğrulama yöntemiyle modelin genellenebilirliğinin değerlendirilmesi gibi yöntemler kullanılmıştır. Ayrıca, değişken önem analizi, müşterilerin harcama alışkanlıkları ve demografik özelliklerinin kampanya katılım davranışlarında belirleyici olduğunu göstermiştir.

CatBoost algoritması, kampanya katılım durumunun tahmininde etkili ve güvenilir bir yöntem olarak öne çıkmıştır. Modelin tahmin gücü, pazarlama stratejilerinin geliştirilmesi ve müşteri segmentasyonu için değerli içgörüler sunmaktadır.

Kampanya katılım durumlarının sınıflandırılmasında kullanılan bir diğer algoritma Destek Vektör Makineleri (SVM)’dir.. Model, dengesiz sınıflandırma problemleri için sınıf ağırlıkları ayarlanarak ve hiperparametre optimizasyonları uygulanarak çalıştırılmıştır. Ancak ilk denemelerde, modelin %100 doğrulukla sonuç verdiği, dolayısıyla veriyi ezberleyerek overfit olduğu tespit edilmiştir. Bu durum, modelin test verilerinde yetersiz performans gösterebileceğini ortaya koymuştur.

Modelin performansını artırmak amacıyla 10 katlı çapraz doğrulama ve smote yöntemiyle veri setinin dengelenmesi gibi ek yöntemler uygulanmıştır. Dengeli veri setiyle yeniden eğitilen modelin sonuçları incelendiğinde, azınlık sınıflar için precision, recall ve F1 score değerlerinin oldukça düşük olduğu görülmüştür. Özellikle azınlık sınıfını doğru tahmin edemeyen model, genel sınıflandırma performansı açısından zayıf bir sonuç sergilemiştir.

Sonuç olarak, bu çalışmada SVM algoritmasının çok sınıflı sınıflandırma problemlerinde dengesiz veri setleriyle çalışırken performansının sınırlı olduğu ve smote gibi yöntemlerle denge sağlanmasına rağmen azınlık sınıflar üzerindeki başarısının yetersiz kaldığı belirlenmiştir. Bu durum, dengesiz veri setleriyle

çalışmak için alternatif algoritmaların veya farklı dengeleme stratejilerinin değerlendirilmesi gerektiğini göstermektedir.

Müşteri ilişkileri yönetimi (CRM) kapsamında kampanya katılım durumu analiz edilerek farklı makine öğrenimi algoritmalarının performansları değerlendirilmiştir. CRM'in en önemli hedeflerinden biri, müşterilere doğru zamanda, doğru kanaldan ve doğru mesajla ulaşmaktır. Bu doğrultuda kampanya yönetimi, müşterilerin özelliklerine ve ihtiyaçlarına uygun hedefli pazarlama stratejilerinin geliştirilmesinde kritik bir rol oynar.

CRM ve kampanya yönetimi açısından bu analizler, müşteri segmentasyonu ve hedefli pazarlama stratejileri geliştirilmesi için önemli ipuçları sunmaktadır. Özellikle, müşteri davranışlarını anlamak ve kampanya katılımını artırmak amacıyla veri odaklı yaklaşımların önemi bir kez daha vurgulanmıştır. CatBoost'un güçlü performansı, pazarlama kampanyalarının etkili bir şekilde optimize edilmesine olanak sağlamış ve müşterilere daha kişiselleştirilmiş deneyimler sunulmasını desteklemiştir. Gelecekte, bu tür modellerin farklı veri setlerinde daha da geliştirilmesi ve diğer algoritmalarla entegrasyonu, CRM uygulamalarında daha yüksek başarı elde edilmesine katkı sağlayacaktır.

Bu çalışmada kullanılan algoritmaların her biri, CRM (Müşteri İlişkileri Yönetimi) kampanya yönetiminde farklı avantajlar sunmaktadır. Random Forest algoritması, yüksek doğruluk ve güçlü F1-Score sonuçlarıyla dengesiz veri setlerinde güçlü bir performans sergilemiştir. Müşteri harcama alışkanlıkları ve demografik verilerle yapılan segmentasyonlar sayesinde kişiye özel kampanya stratejileri geliştirmede oldukça etkili olabilir.

XGBoost algoritması da yüksek doğruluk oranı ve güçlü performansı ile dikkat çekmektedir. Büyük ve karmaşık veri setlerinde başarılı sonuçlar verir ve hiperparametre optimizasyonu ile daha iyi sonuçlar elde edilebilir. Bu özellikleri, CRM sistemlerinde müşterilerin kampanya katılım durumunu tahmin etmek ve çoklu kampanyalar için özelleştirilmiş stratejiler geliştirmek adına avantaj sağlar.

CatBoost ise özellikle kategorik verilerle çalışmada yüksek başarı gösterdiği için, pazarlama ve müşteri segmentasyonu gibi konularda önemli bir yer tutabilir. Model, her iki kampanya için de başarılı tahminler yaparak, tüketici harcama alışkanlıkları ve demografik özellikler gibi faktörleri dikkate alarak etkili içgörüler

sunabilir.

Karar Ağaçları (Decision Tree), daha anlaşılır ve açıklanabilir sonuçlar sunduğundan, özellikle kullanıcı dostu raporlar ve basit karar destek sistemleri için tercih edilebilir. Ancak, daha karmaşık veri setlerinde XGBoost veya Random Forest kadar etkili olmayabilir. Son olarak, SVM (Destek Vektör Makineleri) algoritması, başlangıçta overfitting sorunları yaşasa da, doğru dengeleme yöntemleriyle iyileştirilebilir. Ancak, azınlık sınıflar üzerinde düşük performans gösterdiği için, daha dengeli veri setleriyle ve alternatif algoritmalarla çalışılması önerilir.

Genel olarak, CRM kampanya yönetiminde en güçlü algoritmalar arasında Random Forest, XGBoost ve CatBoost öne çıkmaktadır. Bu algoritmalar, kişiselleştirilmiş kampanya önerileri geliştirmek, müşteri segmentasyonu yapmak ve veriye dayalı stratejiler oluşturmak için son derece uygundur. SVM ise dengesiz veri setleriyle çalışırken sınırlı başarı gösterdiği için, kullanımı daha dikkatli yapılmalıdır.

Bu çalışma, şirketlere kişiselleştirilmiş kampanyaları optimize etme konusunda değerli bir rehber sunmakta ve çeşitli makine öğrenmesi algoritmalarının etkinliğini değerlendirmektedir. Şirketler, bu bulguları kullanarak müşteri verilerini daha iyi analiz edebilir, müşteri ilişkilerini daha kolay yöneterek müşterilerini yakından tanır ve rekabet avantajı elde edebilirler. Gelecekte, bu çalışmanın ışığında, yeni algoritmaların ve derin öğrenme yöntemlerinin araştırılması, veri çeşitliliğinin artırılması gibi gelişmeler yapılabilir.

KAYNAKÇA

Acı, M. ve Doğansoy, G.A. (2022). *Demand forecasting for e-retail sector using machine learning and deep learning methods*. Journal of the Faculty of Engineering and Architecture of Gazi University, 37 (3): 1325-1339.

Adalı, G. K. (2022). *Measuring the attitudes of governmental policies and the public Towards the COVID-19 pandemic*. In Handbook of research on interdisciplinary perspectives on the threats and impacts of pandemics (pp. 163-187). IGI Global.

Adegbola, M. D., Adegbola, A. E., Amajuoyi, P., Benjamin, L. B., & Adeusi, K. B. (2024). *Leveraging financial incentives for enhanced diversity: A review and new models*. International Journal of Applied Research in Social Sciences, 6(5), 1037-1047.

Akinsulire, A. A., Idemudia, C., Okwandu, A. C., & Iwuanyanwu, O. (2024). *Dynamic financial modeling and feasibility studies for affordable housing policies: A conceptual synthesis*. International Journal of Advanced Economics, 6(7), 288-305.

Akpunar, E. N. (2017). *Türkiye'de Elektronik Dış Ticaretin Gelişimi ve İstihdam İlişkisi*. Harran Education Journal, 18-32.

Alabay, Y. M. (2010). *Geleneksel Pazarlamadan Yeni Pazarlama Yaklaşımlarına Geçiş Süreci*. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 213-235.

Alankuş, Z. (2021). *COVID-19'UN Gölgesinde Dijital Pazarlama Ve Yeni Yaklaşımlar*. Uluslararası Halkla İlişkiler ve Reklam Çalışmaları Dergisi, 4(1), 94-125.

Atalay, M. ve Çelik, E. (2017). *Büyük veri analizinde yapay zekâ ve makine*

öğrenmesi uygulamaları-artificial intelligence and machine learning applications in big data analysis. Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 9(22), 155-172.

Ay, D. (2009). *Veri Madenciliği ve Apriori Algoritması ile Sipermarket Analizi*. Yayınlanmamış Yüksek Lisans Tezi. Sakarya: Sakarya Üniversitesi FBE.

Aylak, B. L., Oral, O., Yazıcı, K., “Yapay Zeka ve Makine Öğrenmesi Tekniklerinin Lojistik Sektöründe Kullanımı” El-Cezerî Fen ve Mühendislik Dergisi 2021, 8 (1); 74-93

Bach, M. P., Juković, , Dumići, K., & Šarlija, N. (2013). *Business Client Segmentation in Banking Using Self-Organizing Maps*. South East European Journal of Economics and Business, 32-41. doi:10.2478/jeb-2013

Bartuğ, B.B., Bulut, V. (2022). *Müzik Veri Setinin Analizi ve Sınıflandırma Algoritmaları Kullanılarak Şarkı Türü Tahminleme Çalışması*. Avrupa Bilim ve Teknoloji Dergisi, (40), 143-150.

Bayram Arlı, N. (2021). Büyük Veri ve Veri Madenciliği. İçinde Z.

Sabuncuoğlu, K. Kaymaz ve M. Paksoy (Ed.), *İşletme Yönetiminde Dijital Yaklaşımlar* (1. Baskı, ss. 293-319). Dora.

Beliner, E. (2015). *Müşteri İlişkileri Yönetiminde İstatistiğin Yeri ve Bir Uygulama*. Yayınlanmamış Yüksek Lisans Tezi. Ankara: Hacettepe Üniversitesi

Bonaccorso, G. (2017). *Machine Learning Algorithms*, Birmingham: Pacht Publishing Ltd, 2017. pp.360

Bozkurt, A. H. ve Yalçın, N. (2024). *Topluluk Öğrenmesi Algoritmaları Kullanarak Amazon Yemek Yorumları Üzerine Duygu Analizi*. Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi, 11(1), 128-139. <https://doi.org/10.35193/bseufbd.1300732>

Breiman, L. (2001). *Random forests*. *Machine learning*, 45, 5- 32. AbdelMoniem, M., Gasser: M., El-Mahallawy, M. S., Fakhr, M. W. ve Soliman, A. (2019). Enhanced NOMA system using adaptive coding and modulation based on LSTM neural network channel estimation. *Applied Sciences*, 9(15), 3022. preprint arXiv:1412.6980.

Bulunmaz, B. (2016). *Gelişen teknolojiyle birlikte değişen pazarlama yöntemleri ve dijital pazarlama*. Trt Akademi, 1(2), 348-365.

Cemalcılar, İ. (1987). “*Pazarlama karması (4P) kavramında yeni gelişmeler*”. *Pazarlama Dünyası Dergisi*, 4, 23-24.

Cemalcılar, İ. (2008). “*Pazarlama Nedir?*”. *Eskişehir İktisadi ve Ticari İlimler Akademisi Dergisi*, 1(2), 63-72.

Chang, W., Park, J. E., & Chaik: (2010). “*How Does CRM Technology Transform into Organizational Performance? A Mediating Role of Marketing Capability*”, *Journal of Business Research*, 63(8): 849-855.

Chen, T., Guestrin C., *Xgboost: A scalable tree boosting system*, In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, San Francisco California / USA, August 13-17, 2016, pp: 785-794.

Chen, I. J., & Popovich, K. (2003). “*Understanding Customer Relationship Management (CRM)*”, *Business Process Management Journal*, 9 (5): 672-688.

Chollet, F. (2020). *Derin öğrenme ile yapay zekâ: Python ve Keras uygulamaları* (İ. B. Aksoy, Çev.). Buzdağı Yayınevi.

Cortes, C. ve Vapnik, V. (1995). *Support-vector networks*. *Machine learning*, 20, 273-297.

Çakır, N. S. (2024). *Müşteri İlişkileri Platformu ile Yapay Zeka Entegrasyonu*.

Çiçek, R. (2015). *Pazarlamada Yeni Yaklaşımlar*

M. Akın ve M. Toksarı (Ed.). *Dünyada ve Türkiyede E-Ticaret Potansiyeli ve Geleceği Çağdaş* (283-298). Edirne: Paradigma Akademi Yayınları.

D. Mienye and Y. Sun, (2022), "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," **IEEE Access**, vol. 10, pp. 99129- 99149.

Demir, F. O. ve Kırdar, Y. (2007). *Müşteri ilişkileri yönetimi: CRM*. *Review of Social, Economic & Business Studies*, 8, 293-308.

Demirel, Y. (2006). *Müşteri İlişkileri Yönetimi ve Bilgi Paylaşımı*. İstanbul: IQ Yayıncılık., ss.51

Dhar, V. (2016). *The Future of Artificial Intelligence*. *Big Data*, 4(1), 5-9.
<https://doi.org/https://doi.org/10.1089/big.2016.29004.vda>

Dilmegani, C. (2022). *Image Recognition: In-depth Guide*.
<https://research.aimultiple.com/image-recognition/#marketing> (23.01.2025)

Dixon, P. (2020). *Hemen Her Şeyin Geleceği: Tüm İş Dallarını ve Yaşamımızı Etkileyecek Küresel Değişimler*. M. M. Kemaloğlu (Çev.). İstanbul: Türkiye İş Bankası Kültür Yayınları.

Doyle, P., (2003) “Değer Temelli Pazarlama: Şirketinizi Büyütmek Ve Hissedar Değeri Yaratmak İçin Pazarlama Stratejileri”, MediaCAT, İstanbul, Çeviren:Gülfidan Barış

Duran, C. (2021). *Yapay Zekâ Temelli Pazarlama. Geleneksel Pazarlamanın Sonu mu? N. Ö. İyigün ve M. K. Yılmaz (Ed.), Yapay Zekâ Güncel Yaklaşımlar ve Uygulamalar içinde (s. 31-50). İstanbul: Beta Kitap.*

ElKordy, M. (2014) *The Impact of CRM Capability Dimensions on Organizational Performance*. European Journal of Business and Social Sciences, 2, 128-146.

Erboğa, M., 2023. *Makine öğrenmesi süreçleri*. <https://medium.com/academy-team/makine-%C3%B6%C4%9Frenmesi-s%C3%BCre%C3%A7leri-dokuz-ad%C4%B1mda-ba%C5%9Far%C4%B1ya-giden-yol-c34dab36dc26>, (10.11.2023)

Erciş, M. S. (2010). *Pazarlama iletişiminde temel yaklaşımlar*. Ankara: Nobel Yayın Dağıtım.

Erçetin, C. (2015). *Elektronik Ticarete Tüketicilerin Satın Alma Davranış ve Tercihlerini Etkileyen Unsurlar: E-Ticaret Siteleri Üzerine Bir Çalışma*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: Doğu Üniversitesi SBE.

Erdoğan, Z. (2014). *Pazarlamanın Konusu, Kapsamı, Gelişimi ve Diğer Sosyal Bilimlerle ilişkisi*. Pazarlama, (3-24). Bırsa: Ekin Yayınevi.

Euronews (2020). *Amazon Artan Talebi Karşılama İçin 100 Bin Personel İşe Alacak*. Erişim adresi <https://tr.euronews.com/2020/03/16/amazon-koronavirus-sebebiyle-artan-talebi-karsilamak-icin-100-bin-personel-ise-alacak>, Erişim Tarihi: 30/11/2020.

European Commission. (2018). *The European Commission's High-Level Expert Group On Artificial Intelligence. A Definition Of AI: Main Capabilities And Scientific Disciplines.*

https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

European Commission. (2018). 237 *Final Communication From The Commission To The European Parliament, The European Council, The Council, The European Economic And Social Committee And The Committee Of The Regions Artificial Intelligence For Europe.*

Ezerce, A. (2008). *Müşteri İlişkileri Yönetimi (CRM) ve Veri Madenciliği (Data Mining): Tekstil Sektöründe Bir Uygulama.* Yayınlanmamış Yüksek Lisans Tezi. İstanbul: İstanbul Üniversitesi SBE.

Fatouretchi, M. (2019). *The art of Crm: Proven strategies for modern customer relationship management.* Packt Publishing.

Ganioğlu, B., Aktepe, A., Ersöz, S., & Tebrizcik: (2024). *Denetimsiz Öğrenme İle E-Ticaret Sektöründe Faaliyet Gösteren Bir İşletmede Müşteri Segment Analizi Ve Uygulaması.* International Journal of Engineering Research and Development, 16(2), 880-894.

Gedik, G. (2020). *Robotlara karşı gerçek kişilerin korunması gerekliliği ve robot vergisi önerisi.* Marmara Üniversitesi Hukuk Fakültesi Hukuk Araştırmaları Dergisi, 26(1), 24-48

Géron, A. (2022). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow.* " O'Reilly Media, Inc."

Gıl-Gomez, H., Guerola-Navarro, V., Oltra-Badenes, R., & Lozano Quilis, J. A. (2020). “Customer Relationship Management: Digital Transformation and Sustainable Business Model Innovation”, *Economic Research-Ekonomska Istraživanja*, 1-18. doi: 10.1080/1331677X.2019.1676283

Goldenberg, B. J. (2015). *The Definitive Guide to Social Crm: Maximizing customer relationships with social media to gain market insights, customers, and profits*. Pearson.

Gödekmerdan, L., ve Deniz, A. (2010). *Müşteri İlişkileri Yönetimi: Otobüs Firmaları Üzerine Bir Araştırma*. Dergipark, 241-258.

Gökalp, Ö. M. (2022). *Makine öğrenmesi*. Gazi Üniversitesi, Gazi Bilişim Enstitüsü, Adli Bilişim Bölümü.

Gökmen, Ş. (2014). *Müşteri İlişkileri Yönetiminde Bir Araç Olarak Veri Madenciliği ve Perakende Sektöründe Bir Uygulama*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: Yıldız Teknik Üniversitesi SBE.

Gökşin, E. (2017). *Dijital Pazarlama Temelleri*. İstanbul: Abaküs Kitap Yayınları.

Görentaş, M. B. (2023). *Yapay Zekâ Yöntemleriyle Uyuşmazlık Mahkemesi Kararlarının Tahmini*.

Güngör Naz, D. ve Yardımcı İzmen, U. (2019), <https://stratejico.com/yapay-zeka-kavrami-ve-makine-ogrenme-uygulamalari>, Erişim Tarihi: 02.12.2019

Güven, H. ve Ayvaz Güven, E. T. (2023). *Yapay Zekâ Uygulamalarının E-Ticarette Kullanımı*. *International Journal of Management and Administration*, 7(13), 69-94. <https://doi.org/10.29064/ijma.1194949>

Hacıfendiğolu, Ş. ve Çolular, N. (2008). *İlişkisel Pazarlamada Güven Unsuru ve Otel İşletmelerinde Uygulama*. Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi. 16(2). 107-126.

Han, J. ve Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers

Johnson, M., Smith, R. ve Brown, L. (2010). *Understanding Customer Behavior through Market Basket Analysis*. Journal of Business & Economics Research (JBER), 8(8), 67-72

Jordan, M. I. ve Mitchell, T. M. (2015). *Machine learning: Trends, perspectives, and prospects*. Science. 349(6245): 255-260.

Kağnıcıoğlu, C. H. (2002). *Ürün tasarımı kalite fonksiyon yayılımı*. Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 21(1), 177-188.

Karahan Adalı, G. (2022). *Denetimli Makine Öğrenmesi Aşamaları. İçinde N. Bayram Arlı, S. Gürsakal ve M. Engin (Ed.), Denetimli Makine Öğrenmesi Algoritmaları R ve Python Uygulamaları* (1. Baskı, ss. 11-25). Nobel.

Karakuş, C. (2023). “*Makine Öğrenmesi Ders Notları*”, 2023, pp.15-80

Kaya, F. (t.y.) <http://www.furkankarakaya.com/crm-musteri-iliskileri-yonetimi-genel-anlatim/> (11 Ocak 2018).

Kayhan, L. (2016). *Türkiye’deki E-ticaretin Ekonomideki Yeri ve Geleceği*. Yayınlanmamış Yüksek Lisans Tezi. Manisa: Manisa Celal Bayar Üniversitesi.

Kazançoğlu, İ., Görgün, K., & Serdönmez, C. (2018). *Endüstriyel pazarlarda müşteri edinmede ERP-CRM entegrasyonunun önemi*. Uluslararası İktisadi ve İdari İncelemeler Dergisi, 119-134.

Kestane, S.Ü. (2024). *CRM (Customer Relationship Management) Kavramı Üzerine Yapılan Çalışmaların Vosviewer İle Bibliyometrik Analizi*. Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 33(1), 311-329.

Kılıç, H., Atalay, E. ve Yurtsever, A. E. (2019). *Büyük veri (Bigdata) ve müşteri ilişkileri yönetimi (CRM) işbirliğinin pazarlama iletişimi stratejilerindeki rolü: büyük ölçekli özel bir banka örneği*. Stratejik ve Sosyal Araştırmalar Dergisi, 3(2), 289-310.

Kingsford, C. ve Salzberg, S.L. (2008). *What are Decision Trees?*. Nature biotechnology, 26(9), 1011-1013.

Kolay, H. (2016). *Etkin Müşteri İlişkileri Yönetiminin Müşteri Memnuniyetine Yansımaları "Garanti Bankası Müşteri Hizmetleri"*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: Beykent Üniversitesi SBE.

Kotler, P. (1997). *Marketing management: Analysis, planning, implementation and control*. 9th Edition. New Jersey: Prentice Hall.

Kotler, P. (2000). *Pazarlama Yönetimi (10. Baskı)*. (N. Muallimoğlu, çev.). İstanbul: Beta Basım Yayım Dağıtım A.Ş.

Kotler, P. (215). *Kotler ve Pazarlama*. İstanbul: Aura Kitapları.

Kozlu, C. ve Karaca, H. (2019). *Uluslararası Pazarlama – Dijitalleşen Dünyada Yeni Yaklaşımlar*. İstanbul: Türkiye İş Bankası Kültür Yayınları.

Kumar, V., & Reinartz, W. (2012). *Customer relationship management: Concept, strategy, and Tools*. Springer.

Kumar, V., & Reinartz, W. J. (2018). *Customer Relationship Management*. Springer Texts in Business and Economics.

Kurnu, M. *Müşteri ilişkileri yönetimi stratejilerinin müşteri memnuniyetine etkisi: Perakendecilik sektöründe faaliyette bulunan süpermarket işletmelerinde bir uygulama* (Master's thesis, Sosyal Bilimler Enstitüsü).

Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W., (2005). *Applied Linear Statistical Models*, Cilt 103. McGraw-Hill Irwin, New York, 2005.

Landgren, I. (2020). <https://towardsdatascience.com/a-beginners-guide-to-deep-learning-ed41ac75c4e5>, (24.03.2020)

Learning, S.S. (2006). *Semi-supervised learning*. CSZ2006. html, 5, 2.

LeCun, Y., Bengio, Y. ve Hinton, G. (2015). *Deep learning*. nature, 521(7553), 436-444.

Liang, Y., Cai, Z., Yu, J., Han, Q. ve Li, Y. (2018). *Deep learning-based inference of private information using embedded sensors in smart devices*. IEEE Network, 32(4), 8-14.

Liaw, A., & Wiener, M. (2002). *Classification and Regression by randomForest*. R News, 2(3), 18-22.

B. Lutkevich, and E. Burns, “*What is natural language processing? An introduction to NLP*,” Techtarget.com, 02- Mar-2023. [Online].

Mahesh, B. (2020). *Machine Learning Algorithms -A Review*. *International Journal of Science and Research (IJSR)*, 9(1), 381-386. <https://doi.org/10.21275/ART20203995>

Mountrakis, G., Im, J. ve Ogole, C. (2011). *Support Vector Machines in remote sensing: A review*. *ISPRS journal of photogrammetry and remote sensing*, 66(3), 247-259

Mucuk, İ. (1999). *Pazarlama ilkeleri*. 11. Basım. İstanbul: Türkmen Kitabevi.

Nair, K., Gupta, R. (2021). *Application of AI technology in modern digital marketing environment*, *World Journal of Entrepreneurship, Management and Sustainable Development*, 17(3), 318-328

Ngai, E. W. (2005). *Customer relationship management research (1992-2002) An academic literature review and classification*. *Marketing intelligence & planning*, 23(6), 582-605.

Odabaşı, Y. ve Oyman, M. (2005). *Pazarlama İletişimi Yönetimi*. 13. Baskı. İstanbul: MediaCat Kitapları.

Odabaşı, Y. (2005). *Müşteri İlişkileri Yönetimi*. İstanbul: Sistem Yayıncılık.

Odonkor, T. N., Urefe, O., Agu, E. E., ve Obeng, S.(2024). *Building resilience in small businesses through effective relationship management and stakeholder engagement*. *International Journal of Management & Entrepreneurship Research*, 6(8), 2507-2532.

Olaleye, D.S., Oloye, A.C., Akinloye, A.O. ve Akinwande, O.T., (2024). *Advancing Green Communications: The Role of Radio Frequency Engineering in Sustainable Infrastructure Design*. *International Journal of Latest Technology in Engineering, Management & Applied Science (IJLTEMAS)*, 13(5), p.113. DOI:

10.51583/IJLTEMAS.2024.130511.

Orijji, O., ve Joel, O. S. (2024). *Integrating accounting models with supply chain management in the aerospace industry: A strategic approach to enhancing efficiency and reducing costs in the US*. World Journal of Advanced Research and Reviews, 21(3), 1476-1489.

Özay, M. (2024), *Pazarlama Karması Nedir? Elemanları (4P) Nelerdir?*, <https://www.dopinger.com/tr/blog/pazarlama-karmasi-nedir>, (09.02.2024)

Özdemir, E. (2018). *E-ticarete müşteri ilişkileri yönetimi: Hazır giyim sektöründe bir firmada analitik CRM uygulaması* (Master's thesis, Sosyal Bilimler Enstitüsü).

Özen, N. S., Saraç, S. ve Koyuncu, M. (2021). *COVID-19 vakalarının makine öğrenmesi algoritmaları ile tahmini: Amerika Birleşik Devletleri örneği*. Avrupa Bilim ve Teknoloji Dergisi, (22), 134-139.

Özen, Z. (2022). *Yapay Sinir Ağları. İçinde N. Bayram Arlı, S. Gürsakal, & M. Engin (Ed.), Denetimli Makine Öğrenmesi Algoritmaları R ve Python Uygulamaları* (1. Baskı, ss. 215-262). Nobel

Özkan, Y. (2016). *Veri Madenciliği Yönetleri*. İstanbul: Papatya Yayıncılık Eğitim.

Öztemel, E. (2003). *Yapay sinir ağları*. PapatyaYayincilik, İstanbul.

Öztürk, A.S., (2008). *Hizmet Pazarlaması (8. Baskı)*, Ekin Basım, Bursa, 2008.

Öztürk, M. (2021). *Yapay Zekâ ile Veriye Dayalı Pazarlama*. N. Ö. İyigün ve M. K. Yılmaz (Ed.), *Yapay Zekâ Güncel Yaklaşımlar ve Uygulamalar içinde* (s. 51-67). İstanbul: Beta Kitap.,

Özüpek, E.(2024)., *Drought Prediction Based On Meteorological Data Using Ensemble Machine Learning Techniques*. Yüksek Lisans Tezi, Gebze Teknik Üniversitesi pp: 14 -pp:16

Pirim, H. (2006). *YAPAY ZEKA*. Journal of Yasar University, 1(1), 81-93.

Pulat M., Deveci Kocakoç İ. (2021) *Türkiye’de Makine Öğrenmesi ve Karar Ağaçları Alanında Yayınlanmış Tezlerin Bibliyometrik Analizi*, Yönetim ve Ekonomi Dergisi, 28(2), 287-308

Ramu, P., Thananjayan, P., Acar, E. ve diğerleri. (2022) *Yapısal ve çok disiplinli optimizasyonda makine öğrenimi tekniklerinin incelenmesi*. Struct Multidisc Optim 65, 266 (2022). <https://doi.org/10.1007/s00158-022-03369-9>

Rîpa, I. A. (2022). *Customer relationship management in the international context – theoretical and practical considerations*. Proceedings of the International Conference on Business Excellence, 16(1), 1060-1072.

Rupnik, R., Kukar, M., Vracar, P., Kosir, D., Pevec, D. ve Bosnic, D. (2019). *AgroDSS: A decision support system for agriculture and farming*. Computers and Electronics in Agriculture, (161): 260-271.

SAP. (2021). *Was ist Maschinelles Lernen?* <https://www.sap.com/https://www.sap.com/germany/insights/what-is-machine-learning.htm>, 12.06.2024.

Sezer, E., Çelik, A. E. ve Ayaz, T. B. (2022). *Kampanya Başarısı Tahmininde Öznitelik Seçimi*. 9th International Conference on Future Learning and Informatics

Shanthi, R. (2013). *Customer Relationship Management*. MJP Publishers.

Sheth, J. N., Parvatiyar, A., & Shainesh, G. (2001). *Customer relationship management: Emerging concepts, tools, and applications*. Tata McGrawHill Pub. Co.

Sumathi, S. ve Sivanandam, S., (2006). *Introduction to Data Mining and Its Applications*. Tamil Nadu: Springer Science & Business Media.

Sutton, R. S. ve Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Swift, R. S. (2001). *Accelerating customer relationships: Using Crm and relationship technologies*. Prentice Hall PTR.

Şahin, E. ve Kaya, F. (2019). *Pazarlamada Yeni Dönem: Endüstri 4.0, Yapay Zeka ve Akıllı Asistanlar*. Ankara: Çizgi Yayınevi.

Şahin, Ü. (2011). *Customer segmentation for campaign management by using fuzzy c-means* (Master's thesis, Fen Bilimleri Enstitüsü).

Şengül, O. (2019). *2 Satte A'dan Z'ye Dijital Pazarlama*. İstanbul: Ceres Yayınları.

Şirzad, N., & Turancı, E. (2019). *Dijital Halkla İlişkiler ve Müşteri İlişkileri Yönetimi: Şikayet Yönetimi Bağlamında Memnuniyet Unsurlarının İncelenmesi*. Akdeniz İletişim, (31).

Şükrüoğlu, B. (2008). *Türkiye'de Müşteri İlişkileri Yönetimi (CRM) Uygulayan İşletmelerde Çağrı Merkezi Kullanımının Müşteri Memnuniyetine Etkisi ve Bir Saha Çalışması* (Master's thesis, Marmara Üniversitesi (Turkey)).

Tang, S.(2024). *The box office prediction model based on the optimized XGBoost algorithm in the context of film marketing and distribution*. PLOS ONE, 19(1), e0309227. <https://doi.org/10.1371/journal.pone.0309227>

Tassemedo, A. (2016). *Müşteri İlişkileri Yönetimi Uygulamalarının Müşteri Memnuniyeti Ve Sadakatine Etkileri: Telekomünikasyon Sektöründe Bir Uygulama*. Yayınlanmamış Yüksek Lisans Tezi. İzmir: Ege Üniversitesi SBE.

Tekel, S. (2006). *Sigortacılıkta müşteri ilişkileri yönetimi yaklaşımının önemi ve bir branş uygulaması* (Master's thesis, Marmara Üniversitesi (Turkey)).

Tenekecioğlu, B. (2007). *Pazarlama yönetimi* (No. 1478). Anadolu Üniversitesi.

Tengilimoğlu, D., (2015). *Sağlık Hizmetlerinde Pazarlama Karması Elemanları ve Özellikleri*, Ankara Üniversitesi SBF Dergisi, 55(1), 2015, 187-202.

Tseng, S.M. (2016). “*Knowledge Management Capability, Customer Relationship Management, and Service Quality*”, Journal of Enterprise Information Management, 29(2): 202-221

Ucha, B. D., Ajayi, F. A., ve Olawale, O. (2024). *The evolution of HR practices: An analytical review of trends in the USA and Nigeria*. International Journal of Science and Research Archive, 12(1), 940-957.

Urefe, O., Odonkor, T. N., & Agu, E. E. (2024). *Methodologies and best practices for audit and compliance in governmental financial management*. Finance & Accounting Research Journal, 6(8), 1391-1402.

Uyar, S., (2024). *İkinci Bölüm İçsel Pazarlama*. Sağlıkta Güncel Pazarlama Yöntemleri Evrimi, 17.

Ünal, A ve Bozkurt, M. (2015). *Meslek Yüksek Okulları İçin Pazarlama, M. Bozkurt (Ed.). Genel Olarak Pazarlama Kavramı ve Pazarlama Yaklaşımlarındaki Değişimler* (2-16). Ankara: Paradigma Akademi Yayınları.

Üner, T. (2011). *Pazarlamada Müşteri İlişkileri Yönetimi (MİY) ve E-MİY Analizlerinin Değerlendirilmesi*. Dokuz Eylül Üniversitesi SBE Dergisi, 87-104.

Verma, S. Sharma, S., Deb, S. , Maitra, D. (2021). *Artificial Intelligence in marketing: Systematic review and future research direction*. International Journal of Information Management Data Insight, 1(1)

Yaman, Y. (2021) *Müşteri Yolculuğu Analitiğini Makine Öğrenmesi Modeliyle Tahminleme*. YBS Ansiklopedisi., Kırıkkale Üniversitesi

Yakar, M., ve Baycık, S. (2001). *Çağdaş Müşteri Hizmetleri ve Pazarlama Teknikleri*. Ankara: Tmmob Makina Mühendisleri Odası.

Yasak, M. (2017). *Müşteri İlişkileri Yönetiminin (Crm'nin) Satışa Etkilerinin Servqual Modeli ile İncelenmesi Ve Otomotiv Sektöründe Bir Uygulama*. Yayınlanmamış Yüksek Lisans Tezi. Aksaray: Aksaray Üniversitesi SBE.

Yıldırım, E. (2012). *Müşteri İlişkileri Yönetiminde Bilgi Yönetimi Ve Teknoloji Desteği İle Crm Analizi*. Yüksek Lisans Tezi. İstanbul: Haliç Üniversitesi SBE.

Yiğit, D. (2017, Mayıs). *Veri Madenciliğinin Müşteri İlişkileri Yönetiminde Kullanımı Ve Bir Uygulama*. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: Galatasaray Üniversitesi SBE.

Yurdakul, N. B. (2006). *Bütünleşik Pazarlama İletişimi Ölçümleme Süreci*. Ankara: Nobel Yayın Dağıtım

Zheng, S. , Fan, J., Yu, F., Feng, B., Lou, B., Zou, Q., & Liang, T. (2020). *Viral load dynamics and disease severity in patients infected with SARS-CoV-2 in Zhejiang province, China, January-March 2020: retrospective cohort study*. *bmj*, 369.Springer. <https://doi.org/10.2200/S00196ED1V01Y200906AIM006>

Zhou, Z.-H. (2021). *Machine Learning*. Springer Singapore. <https://doi.org/10.1007/978-981-15-1967-3>

Zhu, X. ve Goldberg, A. B. (2009). *Introduction to Semi-Supervised Learning* (C. 3).

<https://www.kaggle.com/datasets/rodsaldanha/arketing-campaign>, Erişim Tarihi: 26.06.2022