T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MÜŞTERİ TERK ETME DAVRANIŞI TAHMİNİ VE MÜŞTERİ TUTUNDURMA YÖNTEMLERİ TESPİTİ AMAÇLI VERİ ANALİZİ İŞ SÜRECİ YAZILIM MİMARİSİ

Damla EREK

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Prof. Dr. Mehmet S. AKTAŞ

Ocak , 2023

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MÜŞTERİ TERK ETME DAVRANIŞI TAHMİNİ VE MÜŞTERİ TUTUNDURMA YÖNTEMLERİ TESPİTİ AMAÇLI VERİ ANALİZİ İŞ SÜRECİ YAZILIM MİMARİSİ

Damla EREK tarafından hazırlanan tez çalışması 17.01.2023 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Mehmet S. AKTAŞ
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Mehmet S. AKTAŞ, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Oya KALIPSIZ, Üye
Yıldız Teknik Üniversitesi _____

Prof. Dr. Selim AKYOKUŞ, Üye
İstanbul Medipol Üniversitesi _____

Danışmanım Prof. Dr. Mehmet S. AKTAŞ sorumluluğunda tarafımca hazırlanan Müşteri Terk Etme Davranışı Tahmini ve Müşteri Tutundurma Yöntemleri Tespiti Amaçlı Veri Analizi İş Süreci Yazılım Mimarisi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Damla EREK

İmza

Dedicated to my family
and my best friend

TEŞEKKÜR

Öncelikle bana bu tez ile beraber çalışma fırsatı sunan Sayın Prof. Dr. Mehmet Sıddık AKTAŞ hocama teşekkür ederim. Tez süresince yapmış olduğu desteklerinden ve çıkarmış olduğumuz yayınlardaki katkılarından dolayı sonsuz teşekkür ederim. Ayrıca çalışmamda bana yardımcı olan tüm MNG KARGO çalışanı olan uygulama geliştirici arkadaşlarıma teşekkür ederim. Son olarak, hayatım boyunca benimle kederlenen ve benimle sevinen aileme manevi desteklerinden dolayı çok teşekkür ederim. Bu uzun yolda ailemin sunmuş olduğu desteğin önemini bir kez daha kavramış bulunmaktayım ve bu tezi aileme adıyorum.

Damla EREK

İÇİNDEKİLER

KISALTMA LİSTESİ viii

ŞEKİL LİSTESİ ix

TABLO LİSTESİ x

ÖZET xi

ABSTRACT xii

1 GİRİŞ 1

1.1 Müşteri Terketme Davranışı Tahminini Gerektiren Gereksinimler 2
1.2 Müşterinin Geri Kazanılmasını Gerektiren Gereksinimler 3
1.3 Motivasyon 3
1.4 Tezin İçeriği 4
1.5 Tezin Amacı 4
1.6 Literatür Özeti 5
1.7 Orjinal Katkı 8

2 Genel Bilgi 9

2.1 Terk Etme Davranışı Oranı(Churn) 9
2.1.1 Gönüllü Kayıp 9
2.1.2 Gönülsüz Kayıp 10
2.2 Makine Öğrenmesi 10
2.2.1 Denetimli Öğrenme 10
2.2.2 Denetimsiz Öğrenme 11
2.2.3 Takviyeli Öğrenme 12
2.3 Yapay Sinir Ağları 12
2.4 Aktivasyon Fonksiyonları 13
2.4.1 Popüler Aktivasyon Fonsiyonları 13

3 ARAŞTIRMA PROBLEMİNİN TANIMI 17

3.1 Araştırma Soruları 17

3.1.1 Kargoculuk ve lojistik sektöründe, müşterilerin terketme davranışını en iyi tahmin edilebilecek bir veri analiz süreci aracının yazılım mimarisi nasıl olabilir? 18
3.1.2 Kargo ve lojistik sektöründe müşterinin tutundurma yöntemini en iyi tahmin edilebilecek bir veri analiz süreci aracının yazılım mimarisi nasıl olabilir? 18
3.1.3 Kargoculuk ve lojistik sektöründe, müşteri terketme davranışında kullanılabilecek, müşteriyi terketmeden önce kalmasını sağlayabilecek, tahmin gücü en başarılı öznitelikler nelerdir? 18
3.1.4 Kargo ve lojistik sektöründe müşteri tutundurma yöntemi iş sürecinde hangi algoritmalar kullanılabilir? 18
3.1.5 Kargoculuk ve lojistik sektöründe, müşterinin terketme davranışı tahmininde kullanılmasını önerdiğimiz öznitelikler ile başarılı tahminler yapılabilir mi? 19
3.1.6 Kargoculuk ve lojistik sektöründe, müşterinin tutundurma yöntemi için kullanılmasını önerdiğimiz mimari ile başarılı şekilde müşteriler firmada kalabilir mi? 19

4 ÖNERİLEN YÖNTEM **20**
4.1 Önerilen Terk Etme Davranışı İş Süreci Yazılımı Mimarisi Ana Birimleri 21
4.1.1 Veri Önişleme Modülü 21
4.1.2 Otomatik Boru Hattı(Pipeline) Birimi 24
4.1.3 Hiper-Parametre Optimizasyonu Birimi 31
4.1.4 Bayes Optimizasyonu 31
4.2 Önerilen Tutundurma İş Süreci Yazılımı Mimarisi Ana Birimleri 32
4.2.1 Müşteriyi Elde Tutma Tahmin Modülünün Oluşturulması 34

5 PROTOTİP UYGULAMA **38**
5.1 Teknoloji açısından Prototip Uygulamanın sınıflandırılması 38
5.1.1 Müşteri Terk Etme Eğilimi Tahmin Modeli Verisinin Ön İşleme Modülünde Kullanılacak Teknolojiler 38
5.1.2 Müşteri Terk Etme Eğilimi Tahmin Modeli Olusturulması ve Tahmini ve Müşteri Tutundurma İş Süreci Yazılım Mimarisinde Kullanılacak Teknolojiler 39
5.2 Kullanılan yöntem açısından Prototip Uygulamanın sınıflandırılması . 40
5.2.1 Müşteri Terk Etme Eğilimi Tahmin Modeli Verisinin Ön İşleme Modülünde Kullanılacak Yöntemler 40
5.2.2 Müşteri Terk Etme Eğilimi Tahmin Modeli Oluşturulması ve Tahmininde Kullanılacak Yöntemler 42

5.2.3 Müşteri Tutundurma İş Süreci Yazılım Mimarisinde Kullanılacak Yöntemler 42

6 ÖNERİLEN YÖNTEMİN DEĞERLENDİRİLMESİ 43

6.1 Kullanılacak araçlar 43

6.1.1 Teknoloji ve Altyapı 43

6.1.2 Kontrol Veri Seti 43

6.2 Test - 1: Müşterinin terketme davranışından önce kalmasını sağlayacak durumlar tespitinin tahminde en başarılı sonuçlar verecek özniteliklerin bulunması 44

6.3 Test-2: En başarılı sonuçlar verecek özniteliklerin oluşturulmasından sonra başarı oranlarının verilmesi 44

6.3.1 Duyarlılık(Recall) 45

6.3.2 Kesinlik(Precision) 46

6.3.3 F1 Skor(F1 Score) 46

6.3.4 Doğruluk(Accuarcy) 47

6.4 Test - 3: Müşterinin tutundurma yöntemi tahminde en başarılı sonuçlar verecek yöntemin ve bu yöntemin çalışma süresinin bulunması 48

6.5 Değerlendirme 50

7 SONUÇ VE ÖNERİLER 52

7.1 Araştırma Sorularına Cevaplar 52

7.2 Gelecekte Planlanan Çalışmalar 54

KAYNAKÇA 55

TEZDEN ÜRETİLMİŞ YAYINLAR 60

KISALTMA LİSTESİ

BO	Bayes Optimizasyonu
DBMRRS	Dönem sonundaki MRR sayısı
DBMS	Dönem başındaki müşteri sayısı
DSMRRS	Dönem sonundaki MRR sayısı
DSMS	Dönem sonundaki müşteri sayısı
GBM	Gradyan Artırma Makinesi
GP	Gauss Süreci
HPO	Hiper Parametre Optimizasyonu
LTI	Loss of Term Income
MRR	Vadeli Gelir Kaybı
QLR	Üç Aylık Zarar Oranı
RF	Rasgele Orman
SRM	Sıralı Kural Madencilik
SVM	Destek Vektör Makinesi
UD	Üyelik Değişimleri
YSA	Yapay Sinir Ağları

ŞEKİL LİSTESİ

Şekil 2.1 Sinir Ağları ve Nöron Yapısı 12
Şekil 2.2 Basamak Fonksiyonu [49] 14
Şekil 2.3 Doğrusal (Linear) Fonksiyon 15
Şekil 2.4 Sigmoid Fonksiyonu [50] 15
Şekil 2.5 ReLu Aktivasyon Fonksiyonu [51] 16
Şekil 4.1 Müşteri Kaybı Davranışı Verilerine İlişkin Etiketli Verilerin Dağılımları 22
Şekil 4.2 Kademeli Veri Ön İşleme 23
Şekil 4.3 Automated Pipeline Mimarisi 24
Şekil 4.4 Tek Gizli Katmanlı Sinir Ağı 28
Şekil 4.5 Hiper-Parametre Optimizasyon Birimi 31
Şekil 4.6 Kargo Müşterisinin Tutundurma Tahmini Geri Dönüş Yazılım Mimarisi 32
Şekil 4.7 Kargo ve Lojistik Sektörü Endüstrisinde Müşteriyi Elde Tutma Tahmini için Öznitelikler Oluşturma Süreç Mimarisi 33
Şekil 4.8 Veri Dönüşüm Kontrolü 33
Şekil 4.9 Müşteriyi Elde Tutma Modeli 34
Şekil 4.10 Prefix Span Algoritması 35
Şekil 4.11 Spade algoritması 36
Şekil 4.12 GSP algoritması 36
Şekil 5.1 Kutu Analizi Yöntemi Şeması 41
Şekil 6.1 Duyarlılık için bar grafiği 45
Şekil 6.2 Kesinlik için bar grafiği 46
Şekil 6.3 F1 Score için bar grafiği 47
Şekil 6.4 Doğruluk için bar grafiği 47
Şekil 6.5 Müşterilerin Oluşturabileceği Örnek Tipik Sıralı Eylem 49
Şekil 6.6 Farklı setlerden elde edilen test sonuçlarının karşılaştırılması farklı boyut verileri 49

TABLO LİSTESİ

Tablo 6.1 En Önemli Öznitelikler 44
Tablo 6.2 Karmaşıklık Matrisi 45

ÖZET

Müşteri Terk Etme Davranışı Tahmini ve Müşteri Tutundurma Yöntemleri Tespiti Amaçlı Veri Analizi İş Süreci Yazılım Mimarisi

Damla EREK

Bilgisayar Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

Danışman: Prof. Dr. Mehmet S. AKTAŞ

Müşteri terketme davranışı tahmini son yılların en önemli araştırma konularından biri olup bir hizmet aboneliği iptal etme olasılığı olan müşterileri tespit etmekten oluşur. Son zamanlarda özellikle kargo ve lojistik pazarı hızla büyüyen bir pazar olup şiddetli rekabet durumundan dolayı müşteri terketme tahmini ve yönetimi giderek önemli bir hal almaktadır. Bu nedenle rakabetçi ortamda hayatta kalma ihtiyacını karşılamak için şirketlerin odak noktası müşteri tabanı oluşturmaktan müşterileri elde tutmaya doğru kaymıştır. Rekabetçi ortamda hayatta kalma ihtiyacını karşılamak için müşterilerin elde tutulması önemli olup hangi müşterilerin bir rakip şirkete geçme ihtimalinin yüksek olduğunu bilmek değerlidir. Bir müşterinin gerçek değerini anlamak, şirkete müşteri ilişkileri yönetiminde yardımcı olacaktır. Ayrıca rekabetin yüksek olduğu bu ortamlarda müşterilerin davranışlarının tanınması ve analiz edilmesi gerekmektedir. Bu analizlerin en önemlisi müşteriyi elde tutma ve kaybetme oranıdır. Müşteri terk etme davranışı analizi hizmete yönelik sözleşmelerini iptal etme olasılığı en yüksek olan müşterilerin tahmin etmelerini sağlar. Hizmete yönelik sözleşmelerini iptal etme olasılığı en yüksek olan müşterilerin tahmin edilmesi ve hizmetlerini iptal eden müşterilerin geri kazandırılması sorununu çözmek için, kargo ve lojistik sektörlerinde kayıp analizini yapacak bir veri analizi ve müşteri tutundurma yöntemlerini tahmin edilmesini sağlayacak bir veri analizi iş akışı yazılım mimarisi önerilmektedir. Elde edilen sonuçlar, önerilen veri analizi iş akışının müşteri kayıp analizinde ve müşteri tutundurmasında başarılı olduğunu göstermektedir.

Anahtar Kelimeler: Kayıp analizi, Modelleme, Veri analizi, Kargo, Müşteri elde tutma

ABSTRACT

Data Analysis Business Process Software Architecture for Customer Abandonment Behavior Prediction and Customer Retention Methods Detection

Damla EREK

Department of Computer Engineering
Master of Science Thesis

Supervisor: Prof. Dr. Mehmet S. AKTAŞ

Customer abandonment behavior prediction is one of the most important research topics in recent years and consists of identifying customers who are likely to cancel a service subscription. Recently, especially the cargo and logistics market is a rapidly growing market, and customer abandonment forecasting and management is becoming increasingly important due to the fierce competition situation. Therefore, to meet the need to survive in the competitive environment, the focus of companies has shifted from building a customer base to retaining customers. Customer retention is important to meet the need to survive in a competitive environment, and it is valuable to know which customers are most likely to switch to a competitor. Understanding the true value of a customer will assist the company in customer relationship management. In addition, in these environments where competition is high, the behavior of customers needs to be recognized and analyzed. The most important of these analyzes is customer retention and loss rate. Customer abandonment behavior analysis allows them to predict the customers who are most likely to cancel their contract for the service. To solve the problem of predicting the customers who are most likely to cancel their contracts for the service and recovering the customers who have canceled their services, a data analysis that will analyze the loss in the cargo and logistics industries and a data analysis workflow software architecture that will provide the prediction of customer retention methods is proposed. The results show that the proposed data analysis workflow is successful in customer churn analysis and customer retention.

Keywords: Churn analysis, Modeling, Data analysis, Cargo, Customer retention

1
GİRİŞ

Pazardaki rekabetin artmasıyla şirketler, yeni müşterileri ikna etmek yerine önceki müşterileri elde tutma isteğine girmişlerdir. Çünkü yeni müşteri edinmek veriler toplama ve analiz etme süreçlerinden dolayı yüksek maliyetler gerektirmektedir. Ayrıca yeni müşteriler edinmek onları rakip şirketlerden geçiş yapmak için ikna etme gerektirmektedir. Bu da zaman alır. Bu gibi sebeplerden dolayı bir şirketin mevcut müşterilerini elde tutmaya odaklanması daha mantıklıdır. Müşteriyi elde tutma, sadakat ve müşteri kaybı konusu birçok sektörde ilgi görmektedir. Bir şirket, müşteri kaybından ve elde tutma oranından yola çıkarak gerçekten ne kadar kaybedildiğini hissedecektir. Müşteri kaybı tahminleriyle birlikte müşteri değeri analizi, pazarlama programlarının daha spesifik müşteri gruplarını hedeflemesine yardımcı olacaktır. Kargoculuk ve lojistik sektöründe bulunan şirketler çalıştıkları müşteriler için sözleşmeyle anlaşmalı olarak çalışırlar. Anlaşmalı müşteriler bireysel müşterilerden farklı olarak adreslerinden kargoları alınır. Bu müşterilere özel faturalandırılma uygulanır. Gönderi hazırlama süreçleri daha hızlıdır. Anlaşmalı müşterilerin sözleşmesi bir yıl sürelidir ve istenildiği zaman karşılıklı iptal hakkı bulunmaktadır. Müşterilerin sözleşme süresinde şirketle yaşadığı çeşitli olumsuz durumlar, sözleşmelerini iptal etmesine yol açmaktadır. Bunun sonucunda müşteriler başka şirketlerle çalışmak istemektedir. Bu aşamalarda müşteriler yeni fiyat ve hizmet teklifleriyle eski müşterilerin geri kazanılması için uğraşılmaktadır. Ve bu zorlayıcı bir süreç olmaktadır. Bundan dolayı bu aşamaya gelmeden müşterinin kayıp tahminin yapılması gerekmektedir. Bu müşterilerin tutulabilmesi için hangi tekliflerin yapılabileceğinin belirlenmesine yönelik stratejiler geliştirilecektir. Bu amaçla müşteri memnuniyetsizliğinin nedenleri ve müşterinin geri kazanılması için uygulanacak aksiyonlar detaylı olarak irdelenmelidir.

1.1 Müşteri Terketme Davranışı Tahminini Gerektiren Gereksinimler

Müşteri memnuniyetsiliğinin sebepleri çeşitlilik göstermektedir. Bu çeşitlilikler; personel davranışları, müşteriyi oyalamaya yönelik söz ve davranışlar, hatalı ürün/hizmet, yetkiliye ulaşamama ve fiyat olarak 5 ana başlık altında incelenmektedir.

- **Personel Davranışları:**

  Kargo ve lojistik sektöründe diğer sektörlerde olduğu gibi müşterilere ilgili davranılmak zorundadır. Müşteriye yeteri kadar ilgili davranılmaması müşterinin kaçış davranışı göstermesine yol açmaktadır. Müşterinin önceden kaçış durumu göstereceği tahmin edilip müşteriyle başka bir personelin yardımcı olması veya personele gerekli eğitimlerin verilmesi sağlanmalıdır.

- **Müşteriyi Oyalamaya Yönelik Söz Ve Davranışlar:** Müşteriler için ilginin önemli olduğu kadar güvenirlik ve tutarlılık da önemlidir. Şirket vaatlerini gerçekleştirmesi gerekmektedir. Aksi olduğu durumlarda güven sarsıcı bir ortma oluşur. Bu da müşterinin memnuniyetsiz olmasına ve rakip firmalarla anlaşma yapmasına neden olmaktadır. Müşterinin önceden kaçış durumu tahmin edilip müşterinin bu vaadlerin gerçekleştirileceğine karşı güveninin artması sağlanılması sağlanılmalıdır.

- **Hatalı Hizmet** Pazardaki rekabetin artmasıyla mücadelenin ilk şartı hatasız hizmet sunabilmektedir. Hatalı hizmet sunmak müşteri memnuniyetsizliği yaratabilmektedir. Müşterinin kaçma davranışı tahmin edilip hata telafi edilmelidir.

- **Yetkiliye Ulaşamama:** Müşteriler alacakları hizmetle ilgili yetkili birinden bilgi talep edebilmektedir. Yetkili birisine ulaşmakta sıkıntı çıktığı durumlarda müşteri memnuniyetsizliği yaşanmaktadır. Bu memnuniyetsizliğin ortadan kalkması için müşterinin kaçma durumu önceden tahmin edilmesi gerekmektedir. Müşteri ile yetkili birisinin iletişimi sağlanıp bu memnuniyetsizlik ortadan kalkar.

- **Fiyat:** Müşteri, alacağı ürün ya da hizmet ile ilgili fiyat politikasını önemser. Diğer bir ifadeyle fiyat faktörü satın alma davranışını direkt etkilemese de tüketicinin markaya olan sadakatini etkileyen önemli bir faktördür. Fiyat artışı pazardaki rekabetin artmasıyla müşterilerin rakip firmalarla anlaşma yapmalarını sağlamaktadır. Müşteri kaçma davranışı önceden tahmin edilip müşteriye özel teklifler sunularak müşterinin kaçması engellenmelidir.

1.2 Müşterinin Geri Kazanılmasını Gerektiren Gereksinimler

Müşteriler sözleşme süresinde Bölüm 1.1'de bahsedildiği gibi bazı olumsuz koşullar nedeniyle sözleşmeyi fesedip rakip şirketlerle çalışmak istemesine yol açmaktadır. Yeni müşteri edinmek zaman almakta ve mevcut müşteriyi elde tutmaktan 6-7 kat daha maaliyetli olmaktaktadır. Bu nedenle rakip şirketlerle sözleşme imzalamış müşteriyi birden fazla aksiyonlar uygulanarak geri kazanmak oldukça önemlidir.

1.3 Motivasyon

Giriş bölümünde bahsedilen olumsuzlukların ortadan kaldırılması için, müşteri terketme davranışının tahmini önem taşımaktadır. Müşterilerin sistemde sürekli tutulabilmeleri için hangi tekliflerin yapılabileceğinin belirlenmesine yönelik stratejiler geliştirilmelidir. Bunun için müşteri terk etme davranışını ve müşteri tutundurma yöntemini tahmin edebilecek bir veri analizi iş sürecinin yazılım mimarisi önerilmesi gerekmektedir.

Pazardaki rekabetin artmasıyla şirketler, yeni müşterileri ikna etmek yerine önceki müşterileri elde tutma isteğine girmişlerdir. Çünkü yeni müşteri edinmek veriler toplama ve analiz etme süreçlerinden dolayı yüksek maliyetler gerektirmektedir. Ayrıca yeni müşteriler edinmek onları rakip şirketlerden geçiş yapmak için ikna etme gerektirmektedir. Bu da zaman alır. Bu gibi sebeplerden dolayı bir şirketin mevcut müşterilerini elde tutmaya odaklanması daha mantıklıdır. Müşteriyi elde tutma, sadakat ve müşteri kaybı konusu birçok sektörde ilgi görüyor. Bir şirket, müşteri kaybından ve elde tutma oranından yola çıkarak gerçekten ne kadar kaybedildiğini hissedecektir.

Bölüm 1.1'de açıklanan nedenlerden dolayı, müşteri kaybı ve tutundurma iş süreci yazım mimarisi yapma ihtiyacı artmaktadır. Kargoculuk sektöründe müşteri kaçış eğilimlerinin ve tutundurma yöntemi tahminlenmesine yönelik az sayıda çalışma bulunmaktadır. Kargoculuk sektörüne yönelik yapılan araştırmalar, daha çok müşteri memnuniyeti odaklı olup, memnuniyet ölçümlerine ve değerlendirmelerine yöneliktir. Müşteri kaybı ve tutundurma iş süreci için yazılım mimarisi geliştirmek amaçlanmaktadır. Bundan dolayı bu konu için gerekli olan mimariler ve prototipler araştırılıp, geliştirilmesi gerekmektedir. Müşteri kaybı ve tutundurma yöntemi tahminleriyle birlikte müşteri değeri analizi, pazarlama programlarının daha spesifik müşteri gruplarını hedeflemesine yardımcı olacaktır.

1.4 Tezin İçeriği

Bu tez çalışması, müşteri terk etme davranışını ve müşteri tutundurma yöntemini tahmin edebilecek bir veri analizi iş sürecinin yazılım mimarisi önermektedir. Tezin organizasyon yapısı takip eden şekildedir. Bölüm 1'de müşteri terketme davranışı tahminini gerektiren gereksinimler ve müşterinin geri kazanılmasını gerektiren gereksinimler özetlenmektedir. Bölüm 2'de müşteri terketme davranışı tahmini için kullanılan tanımlar verilmekte ve müşteri terketme davranışı tahmini ve müşteri tutundurma iş süreci yazım mimarisiyle ilgili literatürde bulunan çalışmalar incelenmektedir. Müşteri terketme davranışı ve müşteri tutundurmasında çözülmesi gereken problem ve araştırma soruları Bölüm 3'te verilmektedir. Bölüm 4'de müşteri terketme davranışı tahmini ve müşteri tutundurması için önerilen çalışma yöntemi, önerilen çalışma yöntemi içinde kullanılan prototipin ana birimleri ve uygulamanın gerçekleştirilmesi için kullanılan yazılım mimarisi anlatılmaktadır. Bölüm 5'te önerilen yazılım mimarisinin prototip uygulaması anlatılmaktadır. Bu bölüm prototipte kullanılan teknolojiler ve yöntemleri içermektedir. Bu yazılım mimarisinin gerçekleştirmek için kullanılan teknolojiler, önerilen yazılım mimarisi ve prototip uygulaması kullanılarak, müşteri terketme davranışı tahmini için önerilen yöntemin doğruluk ve müşteri tutundurma yöntemi tahmini için çalışma süresi açısından performans testleri Bölüm 6'da verilmektedir. Bölüm 7'de elde edilen sonuçlar ışığında 3'de verilen araştırma sorularına cevaplar verilmekte ve gelecekte yapılması planlanan çalışmalar anlatılmaktadır.

1.5 Tezin Amacı

Pazardaki rekabetin artmasıyla şirketler, yeni müşterileri ikna etmek yerine önceki müşterileri elde tutma isteğine girmişlerdir. Çünkü yeni müşteri edinmek veriler toplama ve analiz etme süreçlerinden dolayı yüksek maliyetler gerektirmektedir. Ayrıca yeni müşteriler edinmek onları rakip şirketlerden geçiş yapmak için ikna etme gerektirmektedir. Bu da zaman alır. Bu gibi sebeplerden dolayı bir şirketin mevcut müşterilerini elde tutmaya odaklanması daha mantıklıdır. Müşterilerin sistemde sürekli tutulabilmeleri için hangi tekliflerin yapılabileceğinin belirlenmesine yönelik stratejiler geliştirilmelidir. Bunun için müşteri terk etme davranışını ve müşteri tutundurma yöntemini tahmin edebilecek bir veri analizi iş sürecinin yazılım mimarisi önerilmesi gerekmektedir.

1.6 Literatür Özeti

Literatürde, kargoculuk sektöründe müşteri terketme eğilimlerinin tahminlenmesine ve müşteri tutundurma yöntemi tahminlenmesine dayalı olarak az sayıda tespit yöntemi önerilmektedir. Müşteri terketme davranışı tahmini kullanılan algoritmaya göre çeşitlilik göstermektedir: Rasgele Orman, Lojistik Regresyon, Destek Vektör Makineleri, Karar Ağaçları, NaiveBayes. Kargoculuk sektörüne yönelik yapılan araştırmalar, daha çok müşteri memnuniyeti odaklı olup, memnuniyet ölçümlerine ve değerlendirmelerine yöneliktir. Literatür incelendiğinde, müşteri terk etme davranışı ve tutundurma analizi çalışmalarının, yoğunluklu olarak kullanıldığı sektörlerin bankacılık ve telekom sektörleri olduğunu görmekteyiz. Bu tezde kargoculuk sektöründe müşteri terketme davranışının tahmini için makine öğrenmesi algoritma yöntemleri detaylı olarak irdelenerek, özellikle hiperparametrik optimizasyonu yapısı üzerinde, teknolojideki en son gelişmeleri yansıtan bir değerlendirme yapılmaktadır. Ayrıca müşteri tutundurma iş süreci yazım mimarisi için sıralı kural madenciliği algoritmaları kullanılarak performansı incelenmektedir.

2012 yılında Bingquan Huang ve diğerleri tarafından yürütülen telekomünikasyonda müşteri kaybını tahmin etme çalışması [1] 2 aylık altı ay Henley 4 aylık doğru çağrı ayrıntıları, hat bilgileri, fatura ve ödeme bilgileri, hesap bilgileri, demografik profiller, hizmet siparişler, şikayet bilgileri dahil olmak üzere sabit hat müşteri kaybını tahmin etmek için yeni bir dizi özellik tanıtılmıştır ve bu özellikler için yedi tahmin yöntemiyle karşılaştırmalı deneyler yapılmıştır. Deneylerin sonucu, yeni özelliklerin öncekilerden daha başarılı olduğunu göstermektedir. Ning Lu ve diğerleri tarafından Boosting Kullanan Telekom Sektöründe Müşteri Kaybı Tahmin Modeli [2] çalışmasında müşteri kayıp tahmini modeli için boosting algoritması önermekte olup bu algoritma tarafından müşterileri iki kümeye ayırmaktadır. Bunun sonucu olarak yüksek riskli bir müşteri kümesi tespit edilmiştir. lojistik regresyon sınıflandırma algoritması kullanılarak her küme üzerinde bir kayıp modeli oluşturulmuştur. Sonuçlar tek bir lojistik regresyon modeliyle karşılaştırıldığında boosting algoritmasının kayıp verilerinin daha iyi ayrılmasını sağladığını göstermektedir. Bu tezde müşteri kaybını tahmin etme çalışması [1] ve Boosting Kullanan Telekom Sektöründe Müşteri Kaybı Tahmin Modeli [2] çalışmalarından farklı olarak hiperparametrik optimizasyon yöntemi ile oluşturulan pipeline sonrasında gözetimli makine ögrenmesi yöntemiyle modeller oluşturulmuş ve müşteri terk etme davranışı tahmininde başarısı irdelenmiştir.

2015 yılında T. Vafeiadis ve diğerleri tarafından müşteri için makine öğrenimi tekniklerinin karşılaştırılması kayıp tahmini müşteri için makine öğrenimi tekniklerinin karşılaştırılması kayıp tahmini [3] çalışmasında tüm modeller

popüler, kamuya açık bir veri kümesi üzerinde çapraz doğrulama kullanılarak uygulandı ve değerlendirilmiş olup Boosting'in sunduğu performans iyileştirmesi sağlanmıştır. En iyi genel sınıflandırıcı, AdaBoost'u neredeyse %97 doğruluk ve %84'ün üzerinde F-ölçüm ile kullanan SVM-POLY idi. Bu tezde müşteri için makine öğrenimi tekniklerinin karşılaştırılması kayıp tahmini müşteri için makine öğrenimi tekniklerinin karşılaştırılması kayıp tahmini [3] çalışmasından farklı olarak Lojistik Regresyon, Karar Ağaçları, Yapay Sinir Ağları, Gradyan Arttırma Ağaçları Sınıflandırıcısı (Gradient-Boosted Trees Classifier), Doğrusal Destek Vektör Sınıflandırıcısı (Linear SVC), Rasgele Orman ve FM Sınıflandırıcısı olmak üzere yedi tane sınıflandırıcı makine öğrenmesi algoritması kullanılmıştır.

2017 yılında Chen ve arkadaşları tarafından yürütülen konteyner taşımacılığı endüstrisi için hizmet özellikleri ile müşteriyi elde tutma olasılığı arasındaki ilişkiyi araştırmak için beta regresyonunun kullanılması [4] çalışmasında konteyner taşıma endüstrisinde verilen hizmet ve müşterinin sadakati arasındaki bağlantılar incelenmiş olup pazarın bölümlere ayrılması ile ilgili bir analiz çalışması üzerinde durulmuştur. Bunun sonucunda bölünmüş bir değişken olarak hizmet kalitesi, genel modeli farklı işlevsel ilişkilere sahip beş alt modele ayırmış olup fiyat ve indirim, kişisel satış ve müşteri ilişkisinin nitelikleri, müşteriyi elde tutma olasılığı üzerinde önemli etkiye sahip ,tatmin edici fiyat ve indirimler, müşteriyi elde tutma olasılığını desteklemek için gerekli bir nitelik ve tatmin edici kişisel satış, müşteriyi elde tutma olasılığını artırmak için en önemli özellik olduğunun sonuçlarına ulaşılmıştır. Bu tezde konteyner taşımacılığı endüstrisi için hizmet özellikleri ile müşteriyi elde tutma olasılığı arasındaki ilişkiyi araştırmak için beta regresyonunun kullanılması [4] çalışmasından farklı olarak müşteri tutundurma yöntemi tahmini yapılmıştır.

2018 yılında Arno De Caigny ve diğerleri tarafından lojistik regresyon ve karar ağaçlarına dayalı müşteri kayıp tahmini için yeni bir hibrit sınıflandırma algoritması [5] çalışmasında yeni bir hibrit algoritma olan logit yaprak modeli önerilmektedir. Karar kuralları kullanılarak müşteri segmentleri belirlenir ve bu ağacın her yaprağı için bir model oluşturulur. Bu yeni hibrit yaklaşım, tahmine dayalı performans ve anlaşılabilirlik açısından karar ağaçları, lojistik regresyon, rastgele ormanlar ve lojistik model ağaçları ile karşılaştırıldığında tahmin için daha iyi performans gösterdiği görülmektedir. Bu tezde lojistik regresyon ve karar ağaçlarına dayalı müşteri kayıp tahmini için yeni bir hibrit sınıflandırma algoritması [5] çalışmasından farklı olarak öznitelikler kümelenip sanal öznitelikler oluşturulur ve makine öğrenmesi modeli için en başarılı sonuçlar verecek öznitelik haline getirilir. Böylece müşteri kayıp tahmininde daha başarılı sonuçlar elde edilmektedir.

Jose tarafından yapılan denetimli makine öğrenimi kullanarak uygulama tabanlı

bir işletmede müşteriyi elde tutmayı tahmin etme [6] çalışmasında, yazılım sektöründeki müşteri verileri analiz edilmiş ve bu analiz sonucunda elde tutulan müşterileri doğru bir şekilde tahmin edebilecek bir makine öğrenmesi modelinin geliştirilmesi üzerinde durulmuştur. Jyh-Fu Jeng ve Bailey tarafından yapılan mobil telekomünikasyonda müşteri tutma stratejilerinin değerlendirilmesi: Hibrit mcdm yaklaşımı [7] çalışmasında telekom sektöründe müşteriyi elde tutmanın temel motivasyonları ve aralarındaki ilişkiler incelenmiş ve müşterilerin ilgili reklamlara ilişkin algılanan değerini değerlendirmek için bir değerlendirme çalışması yapılmıştır. Chu ve arkadaşları, müşteriyi elde tutma problemleriyle başa çıkmak için hibritleştirilmiş bir mimari önerilmiştir. Bu araştırma [8] kapsamında XBoost makine öğrenmesi sınıflandırıcısı kullanılarak özel bir bankanın müşteri elde tutma modeli oluşturulmuştur. Bu tezde denetimli makine öğrenimi kullanarak uygulama tabanlı bir işletmede müşteriyi elde tutmayı tahmin etme [6] , mobil telekomünikasyonda müşteri tutma stratejilerinin değerlendirilmesi: Hibrit mcdm yaklaşımı [7] ve müşteriyi elde tutmak için hibrit bir veri madenciliği modeline doğru [8] çalışmalarından farklı olarak Spade, GSP ve Prefix Span olmak üzere yedi tane sıralı kural madenciliği algoritması kullanılmıştır.

AL-Shatnwai ve Faris tarafından bu çalışmada [9] kullanılan sınıflandırıcı, dengesiz veri seti üzerindeki performansını iyileştirmek için farklı aşırı örnekleme yöntemleri ile birleştirildiğinde test edilmiştir. Genetik teknolojinin en yüksek performansı sunduğu sonucuna varılmıştır. Eiben ve arkadaşları, genetik programlama, kaba veri analizi ve lojistik regresyon analizi kullanarak bir müşteri tutma modeli oluşturmuş ve müşteri davranışını etkileyen faktörler hakkında fikir sahibi olmuştur. Araştırma [10], müşterinin firma ile ilişkisi sona erdikten sonra pazarlama sektörü üzerine yapılmıştır. Sim ve arkadaşları, konaklama endüstrisini kullanarak müşteri memnuniyetinin ve elde tutmanın öncüllerini ve sonuçlarını değerlendirmeye yönelik bir yaklaşım önermektedir [11]. Zhao ve arkadaşları tarafından telekom sektöründeki veriler üzerinde MBBTree algoritmasına dayalı müşteri elde tutma modelleri oluşturularak etkin sonuçlar elde edilmiştir [12]. Karp tarafından yapılan [13] çalışmada müşteri elde tutma analizi için lojistik regresyon modeli kullanılarak tahminler yapılmıştır. Bu çalışmada [14], Sabbeh telekomünikasyon endüstrisinde diskriminant analizi, karar ağaçları (CART), örnek tabanlı öğrenme (k-en yakın komşular), destek vektör makineleri, lojistik regresyon için bir tahmin sorununu ele aldı. Çalışmanın bulguları, rastgele orman ve ADA artışının her ikisinin de en yüksek doğruluk seviyelerine ulaştığını göstermektedir. Bu tezde bu çalışmadan farklı olarak Spade, GSP ve Prefix Span olmak üzere yedi tane sıralı kural madenciliği algoritması kullanılarak müşteri tutundurma yöntemi tahmini yapımlması sağlanmıştır.

Günümüzde müşteri terketme davranışının tahmini ve tutundurma iş süreci yazım

mimarisi ağırlıklı olarak Yapay zeka(AI), Makine öğrenmesi(ML) Derin öğrenme(DL) ve veri madenciliği üzerinedir.

Bu tez kapsamında, kargoculuk sektörüne yönelik terk etme davranışı tahminine yönelik bir iş süreci geliştirilmektedir. Literatürde, dağıtık programlama ve web servislere dayalı iş süreçlerinin bulunduğu gözlemlenmektedir [15–27]. Ancak, bu tez kapsamında, bu çalışmalardan farklı olarak, veri analizi ve veriden anlam çıkartma süreçleri üzerinde yoğunlaşılmaktadır ve dağıtık programlama konuları, gelecekte üzerinde çalışılacak konular olarak belirlenmiştir. Literatür incelendiğinde, farklı veri tanımlama yöntemleri üzerinde çalışmalar yapıldığı gözlemlenmektedir [28–32]. Bu tez kapsamında, bunlardan farklı olarak, veri tanımlama yöntemleri üzerinde değil, veri analizi süreçleri üzerinde yoğunlaşılmıştır. Tasarlanan iş sürecinin faydalı olduğunu göstermek adına geliştirilen yazılımın bir prototip uygulaması geliştirilmiştir. Literatürde, yazılımların, kalitesinin belirlenmesi üzerinde çalışmalar olduğu gözlemlenmektedir [33–40]. Bu araştırma kapsamında, yazılım kalitesi kapsam dışında bırakılmıştır.

1.7 Orjinal Katkı

Bu tezde makine öğrenmesi ile geliştirilmiş yazılımlar için müşteri terketme tahmininin nasıl yapılacağı ve sıralı kurallı madencilik algoritmaları kullanılarak müşteri tutundurma yöntermi tahmininin nasıl gerçekleştirileceği problemi üzerinde durulmaktadır. Bu araştırmanın özgün katkısı, makine öğrenmesi ve sıralı kurallı madencilik algoritmaları metriklerini kullanarak python programlama diliyle ile yazılmış projelerde müşteri kaçış davranışı ve tutundurması yöntemi tahmini için kolayca eklenebilir bir yöntem geliştirmek ve terketme ve tutundurma davranışı bilimsel araştırma alanına katkı sağlamaktır.

2
Genel Bilgi

2.1 Terk Etme Davranışı Oranı(Churn)

Müşteri kayıp analizi, bir müşterinin bir ürün veya hizmetten ayrılma olasılığı üzerine yapılan analitik çalışma olarak tanımlanabilir. En basit tanımıyla müşterilerin rekabet nedeniyle firmayı tercih etmekten vazgeçmesi anlamına gelmektedir [41]. Amaç, müşterinin ürün veya hizmetten ayrılmadan önce bu durumu tespit etmek ve ardından önleyici bazı faaliyetleri gerçekleştirmektir [42]. Örneğin 5 ayın başlangıcında 5000 müşteriniz varsa ve 5 ay süresinde bu müşterilerinin 10 tanesi hizmeti veya ürünü almayı bıraktıysa terk etme davranışı oranı 10/5000 yani %2'e karşılık gelir. Her şirket terk etme davranışı oranı %0 olmasını ister. Çünkü yeni bir müşteri elde etmek genellikle müşteriyi elde tutmaktan daha maliyetlidir. Müşterilerin şirketin ürün veya hizmetini kullanmayı bırakma nedenleri, müşteri memnuniyetinin nasıl artırılacağı, kaçış davranışı sergileyecek müşteri davranışlarının tahmin edilip edilemeyeceği ve buna göre teklifler yapılıp yapılamayacağı öğrenilerek terk etme oranı azaltılabilir.

Kayıp analizinin hesabı Denklem-2.1'deki gibidir.

$$QLR = \frac{DBMS - DSMS}{DBMS} \tag{2.1}$$

Gelir analizinin hesabı Denklem-2.2 gibidir.

$$QLR = \frac{DBMRRS - DSRRS - UD}{DBMRRS} \tag{2.2}$$

2.1.1 Gönüllü Kayıp

Gönüllü kayıp, konukların işletmeden serbestçe ayrılıp farklı işletmelerin hizmetlerinden yararlanma durumudur. Örnek vermek gerekirse, bir bankanın müşterisi, hesabını başka bir bankaya taşırsa, misafir kaybının serbestçe geçtiği

söylenebilir. Bu genellikle müşterinin kendisi için daha karlı olduğuna inandığı koşullar altında gerçekleşir [43].

2.1.2 Gönülsüz Kayıp

Müşterinin kendi tercihleri dışında gelişen ve genellikle çevresel koşullardan etkilenen bir kayıp durumudur. Örnek vermek gerekirse, müşterinin başka bir ülkeye taşınması, ölüm veya sağlık sorunları gibi sebeplerden dolayı olabilir. Genellikle, müşteri kaybı analizi, istatistiksel modeller veya veri madenciliği çalışmaları için geliştirilen araçlar, benzer istem dışı kayıpları göz ardı eder. Göz ardı edilmesinin en büyük nedeni, misafir kaybının sıklıkla gerekli olmasıdır. azami müşteri kaybı analizleri, istem dışı kayıpların istatistiksel tahminlerini içerir, ancak bu araçlar genellikle oranı bir zaman serisi olarak tahmin etme eğilimindedir [43].

2.2 Makine Öğrenmesi

Makine öğrenimi, bilgisayarların açıkça programlanmadan işlerin nasıl yapılacağını bulma sürecidir. Belirli görevlerin nasıl yürütüleceğini öğrenmek için verileri kullanan bilgisayarları ifade eder. Bilgisayarlara tahsis edilen temel görevler için eldeki sorunu çözmek için gereken tüm adımları nasıl gerçekleştireceğini makineye bildiren algoritmalar oluşturmak mümkündür; bilgisayar tarafında öğrenmeye gerek yoktur. Daha karmaşık etkinliklerde gerekli insan algoritmalarını manuel olarak oluşturmak zor olabilir. Pratikte, her adımı seçmek için insan programcılara güvenmektense, bilgisayarın kendi algoritmasını geliştirmesine yardımcı olmak daha verimli olabilir [44].

Makine öğrenimi, bilgisayarları, tam olarak uygun bir çözümün bulunmadığı görevleri yürütmek üzere eğitmek için bir dizi teknik kullanır. Birden fazla olası yanıt olduğunda, bir strateji doğru yanıtlardan bazılarını geçerli olarak işaretlemektir. Bu bilgi daha sonra bilgisayarın doğru yanıtlara ulaşmak için kullandığı algoritmaları oluşturmak için kullanılabilir. Örneğin, MNIST el yazısı rakamları koleksiyonu, bir sayısal karakter tanıma sistemini eğitmek için sıklıkla kullanılır [44].

2.2.1 Denetimli Öğrenme

Denetimli algoritmalarda sınıflar önceden belirlenir. Bu sınıflar, insan tarafından tanımlanan sonlu bir küme şeklinde oluşturulur, bu da pratikte belirli bir veri bölümünün bu sınıflandırmalarla etiketleneceği anlamına gelir. Makine öğrenimi algoritmasının görevi, kalıpları bulmak ve matematiksel modeller oluşturmaktır. Bu

modeller daha sonra, verilerin kendisindeki varyans ölçümleriyle ilişkili olarak tahmin kapasitesine dayalı olarak değerlendirilir [45].

İki ana denetimli model arasında fark önemlidir: sınıflandırma modelleri (sınıflandırıcılar) ve regresyon modelleri. Regresyon modelleri, girdi uzayını gerçek değer alanına eşler. Sınıflandırıcılar, giriş uzayını önceden tanımlanmış sınıflara eşler. Sınıflandırıcıları temsil etmek için pek çok alternatif vardır. Örneğin, destek vektör makineleri, karar ağaçları vs. Tekniğin hem Veri Madenciliği hem de uygulamalarında diğer alanlarda büyük etkisi olduğundan, sınıflandırmadaki ilerlemenin potansiyel faydaları çok fazladır [45].

Denetimli öğrenmenin zorlukları aşağıdaki gibidir [46].

- Denetimli öğrenme modellerini doğru bir şekilde yapılandırmak için belirli düzeyde uzmanlık gerekebilir.
- Denetimli öğrenme modellerinin eğitimi çok zaman alabilir.Veri kümeleri, yüksek insan hatası olasılığına sahip olabilir ve bu da algoritmaların yanlış öğrenmesine neden olur.
- Denetimsiz öğrenme modellerinden farklı olarak, Denetimli öğrenme verileri kendi başına kümeleyemez veya sınıflandıramaz.

2.2.2 Denetimsiz Öğrenme

Denetimsiz öğrenme algoritmalarına sınıflandırmalar sağlanmaz. Denetimsiz öğrenmenin ana görevi, sınıflandırma etiketlerini otomatik olarak geliştirmektir. Bu algoritmalar, kategorize edilip edilemeyeceklerini belirlemek ve bir grup oluşturup oluşturamayacaklarını belirlemek için veri parçaları arasındaki benzerliği araştırıyor. Bu gruplar sözde kümeler olarak adlandırılır ve kümeleme makine öğrenimi tekniklerinin tüm ailesini temsil ederler. Bu denetimsiz sınıflandırmada (küme analizi), makine kümelerin nasıl gruplandırıldığını bilmez. Küme analizini kullanarak kendimizi şaşırtmak için daha büyük bir potansiyel var. Bu nedenle, küme analizi, birçok makale arasındaki ilişkilerin araştırılması için çok umut verici bir araçtır [45].

Gözetimsiz öğrenmenin zorlukları aşağıdaki gibidir.

- Gözetimsiz verilerilerin çıktılarının doğruluğu belirsiz olabilir.
- Sonuçları doğrulamak için etiketlenmiş veri setleri olmadığı için denetimsiz öğrenme çıktılarının doğruluğunu kontrol etmek zordur. Mühendislerin ve

veri bilimcilerin, gözetimli öğrenmeye göre gözetimsiz öğrenme sonuçlarını yorumlamak ve etiketlemek için daha fazla zamana ihtiyacı vardır.

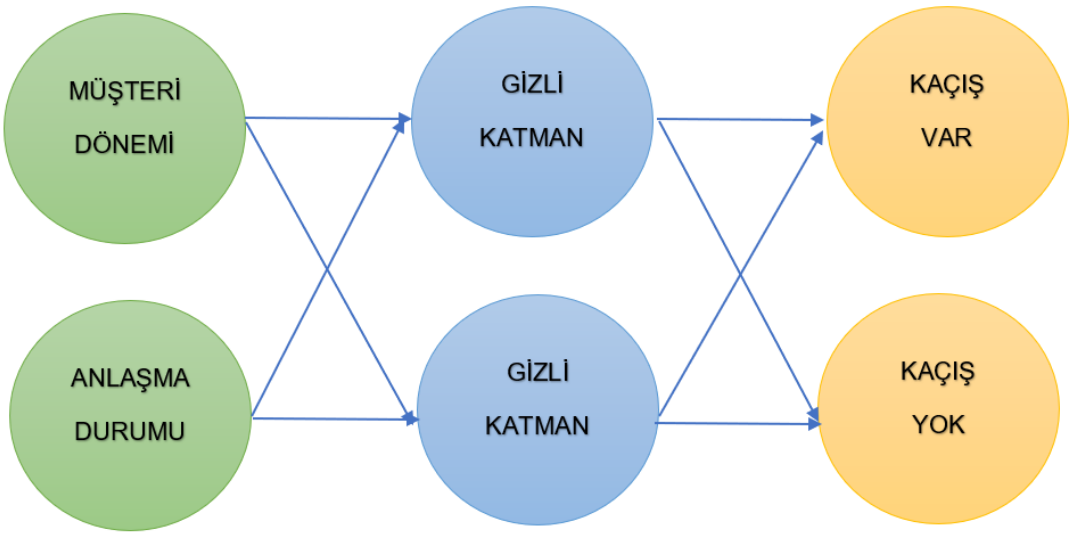
- Gözetimsiz bir sistemin, sonuçlarına nasıl veya neden ulaştığına dair kavrayış eksikliği mevcuttur.

2.2.3 Takviyeli Öğrenme

Bulunduğu ortamı algılayan ve kendi başına kararlar alabilen bir sistemin, hedefine ulaşabilmesinde doğru kararlar almayı nasıl öğrenebileceğini gösterir. Bu yöntem robotik, oyun programlama, hastalık teşhisi ve fabrika otomasyonu gibi alanlarda sıklıkla kullanılır [46].

2.3 Yapay Sinir Ağları

Yapay sinir ağları (YSA), insan beyninin bilgi işleme tekniğinden esinlenerek geliştirilmiş bir bilgi işlem teknolojisidir. YSA ile basit biyolojik sinir sisteminin çalışma şekli taklit edilir. Yani biyolojik nöron hücrelerinin ve bu hücrelerin birbirleri ile arasında kurduğu sinaptik bağın dijital olarak modellenmesidir [47]. Bir sinir ağı girdi olarak aldığı sayıları mekanik bir sezgi, etiket veya kümeleme yöntemiyle çıktılara dönüştürür.Sinir ağlarının müşterinin kaçışının olup olmadığını anlayabilmesi için bu verilerin dönüştürülmesi gereklidir.



Şekil 2.1 Sinir Ağları ve Nöron Yapısı

Şekil-2.1 bir sinir ağına örnektir. Bu ağ girdi olarak müşteri ile ilgili özellikleri almaktadır. Müşteri özelliklerinden müşteri kaçışının bulunup bulunmadığını ayırt

edecek bir şekilde eğitilmiştir. Sarı ile gösterilen ara katman nöronları müşteride bulunan özellikler almaktadır. Örneğin bir nöron müşteri dönemi ve anlaşma durumu değerleri alınıp belirli oranda karıştırarak müşterinin kaçış davranışının bulunup bulunmadığı ihtimalini ortaya koyar. Örnek bir nöron çıktı hesaplaması Denklem 2.3 'de verilmiştir. B(bias) değeri nöronun sapma değeridir ve çıktıya eklenir.

$$GN_{1cikti} = k1 * w1 + k2 * w2 + b1 \tag{2.3}$$

2.4 Aktivasyon Fonksiyonları

Derecesi birden fazla olan ve çizildiğinde bir eğriliği olan bu fonksiyonlar, Doğrusal olmayan fonksiyonlar olarak bilinir. Herhangi bir veriyi ve girdileri çıktılara eşleyen herhangi bir keyfi karmaşık işlevi öğrenmek, temsil etmek ve işlemek için bir sinir ağı gereklidir. Yapay Sinir Ağları, Evrensel İşlev Yaklaşımlayıcıları(Universal Function Approximators) olarak da bilinir; bu, kendilerine sağlanan herhangi bir işlevi hesaplayabilecekleri ve öğrenebilecekleri anlamına gelir. Akla gelebilecek herhangi bir süreç, Sinir Ağlarında işlevsel bir hesaplama olarak temsil edilebilir [48].

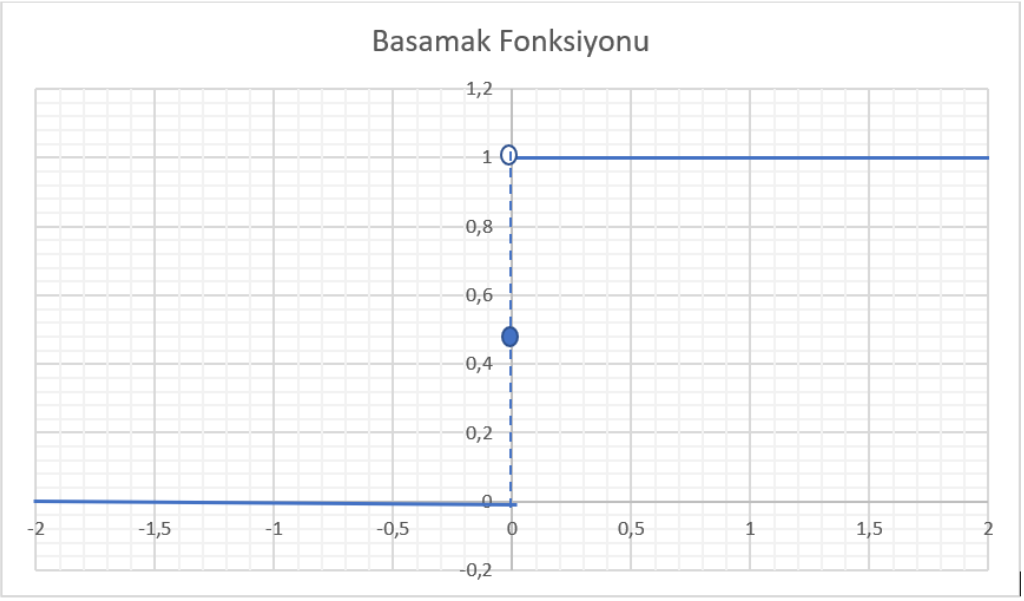
Bu nedenle, ağı dinamik hale getirmek için bir etkinleştirme işlevi uygulamamız ve ona verilerden karmaşık ve karmaşık bilgileri çıkarma ve girdi ile çıktı arasında doğrusal olmayan kıvrımlı rastgele işlevsel eşlemeleri temsil etme yeteneği eklememiz gerekir. Dolayısıyla, ağa doğrusal olmayan aktivasyon fonksiyonları yardımıyla doğrusal olmayanlığı ekleyerek, girdilerden çıktılara doğrusal olmayan eşlemeler elde edebiliyoruz. Bir aktivasyon fonksiyonunun önemli bir özelliği, ağırlıklara göre hataları veya kayıpları hesaplamak ve sonunda hataları azaltmak için Gradient Descend veya başka herhangi bir optimizasyon tekniğini kullanarak ağırlıkları optimize etmek için geri yayılım optimizasyon stratejisini uygulayabilmemiz için türevlenebilir olması gerektiğidir. dünya bilgileri verilir [48].

2.4.1 Popüler Aktivasyon Fonsiyonları

Net girdiler, bir nöral ağ yapısındaki en önemli birimlerdir ve aktivasyon işlevi veya eşik işlevi veya bir skaler dönüşüm olan transfer işlevi olarak adlandırılan işlevi uygulayarak birimin aktivasyonu olarak bilinen bir çıktı sonucuna dönüştürülür [48].

Genel olarak kullanılan popüler aktivasyon fonksiyonları basamak fonksiyonu, doğrusal fonksiyonu, sigmoid fonksiyonu, relu aktivasyon fonksiyonu ve softmax aktivasyon fonksiyonudur. Bunlar aşağıda kısacak özetlenmektrdir.

Basamak Fonksiyonu



Şekil 2.2 Basamak Fonksiyonu [49]

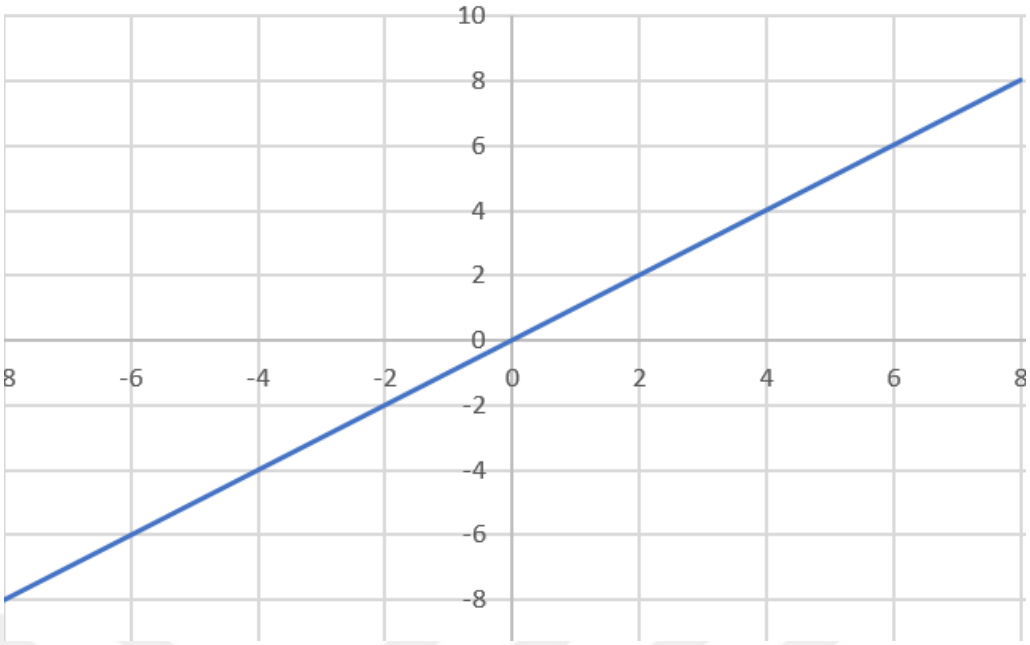
İkili değer alan bir fonksiyondur, belli bir aralığın üstünü pozitif, altını ise negatif olarak ele alır ve ikili sınıflayıcı olarak kullanılır. Örneğin bir soruya cevap vermek istiyor olabiliriz. Var, yok, doğru veya yanlış tarzı cevapların aktivasyon fonksiyonu ikili adım fonksiyonu olabilir. Bu yüzden genellikle çıkış katmanlarında tercih edilir. Gizli katmanlarda türevi öğrenme değeri temsil etmediği için kullanılması tavsiye edilmez. Basamak fonksiyonunun görünüşü Şekil- 2.2'deki gibidir. İkili adım fonksiyonunun alçalma fonksiyonu bir sayının türevi 0 olduğundan sabit bir sayı olan 0'dır.

Doğrusal (Linear) Fonksiyon

Bir dizi aktivasyon değeri üretir ve bunlar basamak fonksiyonundaki gibi ikili değerler değildir. Kesinlikle bir kaç nöronu birbirine bağlamaya izin verir. Nöronların aktive edilmesi formülü Denklem 2.4'deki gibidir.

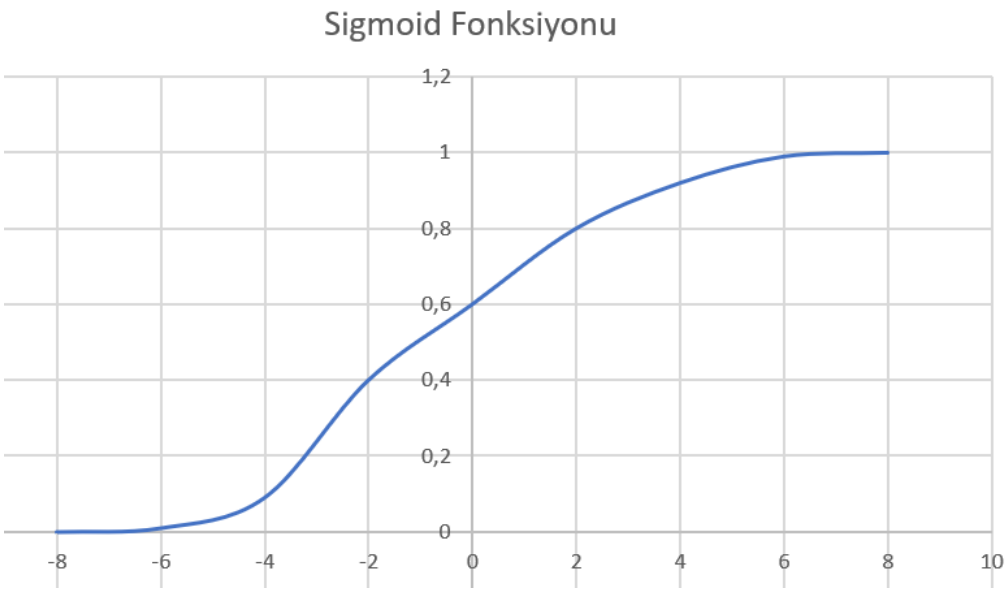
$$f(x) = ax + b \tag{2.4}$$

Doğrusal fonksiyonların hata fonksiyonu, türevi alındığında sabit bir sayı çıktığı için nöron eğitmede kullanılamaz. Hata sabit bir değer çıkacağından dolayı doğru nöronlar da hatalı gözükecektir. Linear fonksiyonunun görünüşü Şekil- 2.3'deki gibidir.



Şekil 2.3 Doğrusal (Linear) Fonksiyon

Sigmoid Fonksiyon



Şekil 2.4 Sigmoid Fonksiyonu [50]

Doğrusal aktivasyon fonksiyonu geri yayılma için uygun olmadığından yerine daha sonra Sigmoid fonksiyonu gelmiştir. Sigmoid fonksiyonunun görünüşü 2.4'deki gibidir [46]. Bu fonksiyonun en büyük avantajı doğrusal olmamasıdır. Derin öğrenme için doğrusal sonraki en uygun fonksiyondur. Hesaplanması Denklem 2.5'deki gösterildiği gibidir.

$$\sigma(z) = \frac{1}{1+e^{-z}} \tag{2.5}$$

ReLu Aktivasyon Fonksiyonu

Grafiksel gösterimi 2.5'deki gibidir. ReLU günümüzde en çok kullanılan aktivasyon fonksiyonudur. En önemli artısı nöronun çıktısının negatif değerlere düştüğünde aktifleştirilmemesidir. Böylece sinir ağındaki nöronların bir miktarı aktif olarak çıktıya etki eder ve daha başarılı bir sinir ağına sahip olunmaktadır [46]. Hesaplanması Denklem 2.6'deki gibidir.

$$Relu(z) = max(0, z) \tag{2.6}$$



Şekil 2.5 ReLu Aktivasyon Fonksiyonu [51]

SoftMax Aktivasyon Fonksiyonu:

Softmax kümeleme işlemlerinde kullanılmaktadır. Daha önce incelediğimiz sınıflandırma aktivasyon fonksiyonu ikili sonuç veren bir fonksiyondu. Daha fazla durumu içeren sınıflandırmalar için ise Softmax fonksiyonu kullanılmaktadır. Softmax fonksiyonu tüm çıktıları toplamlarına bölerek normalleştirme işlemini de yapmaktadır [46]. Hesaplanması Denklem 2.7'deki gibidir.

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad for\ i = 1, 2, \ldots, K \tag{2.7}$$

3
ARAŞTIRMA PROBLEMİNİN TANIMI

Bölüm 2'de verilen literatür taraması diğer sektörlerde müşteri terketme davranışının tespit edilebildiğini ancak kargoculuk sektöründe makine öğrenme algoritması kullanılarak müşterinin terketme tahmini tahmini en doğru sonuçlarla tespit edemediğini göstermektedir. Ayrıca müşteri tutundurması daha çok müşteri memnuniyetini ölçülüp değerlendirilmektedir ve sıralı kurallı madencilik algoritması kullanılarak müşteri tutundurma iş süreci yazım mimarisi en doğru sonuçlarla yapılamamaktadır.

Kargoculuk sektörü müşterinin kargosunu A noktasından B noktasına ulaşıncaya kadar, güvenle taşımak ve teslim etmekten sorumludur. Bu sektörde, anlaşmalı müşterilere şirketlerinin ihtiyaçlarına göre indirim uygulanabilir, özel tekliflerde bulunulabilinir. Anlaşmalı müşteriler bir kargo firması için önem teşkil etmektedir. Her bir anlaşmalı müşteri kayıbı, firma için çok önemlidir. Kargo firmaları; anlaşmalı müşterilerinin davranışlarını izleyerek henüz müşterileri kendilerini terk etmeden, müşterinin memnuniyetsizliğini tahmin edebilmelidir. Bu sayede, müşteriyi firmalarında tutmak için gerekli pazarlama aktivitelerini gerçekleştirebilirler. Müşteri kaçış davranışlarınının makine öğrenmesine dayalı tahmin yöntemleri ile tahmin edebilme yeteneği, kargoculuk alanında öncelikli bir konuma gelmiştir.

3.1 Araştırma Soruları

Müşterinin terketme davranışından önce kalmasını sağlayacak durumların tespiti ve terketme davranışı gösteren veya gösterme ihtimalinin olan müşterilerinin şirkette kalabilmesi için hangi aksiyonların gerçekleştirilmesi gerektiğinin tespiti için aşağıdaki araştırma sorularının cevapları aranması gerekmektedir.

3.1.1 Kargoculuk ve lojistik sektöründe, müşterilerin terketme davranışını en iyi tahmin edilebilecek bir veri analiz süreci aracının yazılım mimarisi nasıl olabilir?

Sistem performansı, güvenlik, değiştirilebilirlik ve donanım üzerinde yürütme gibi daha sonra değiştirilmesi zor olan yazılım kararlarının tümü yazılım mimarisini oluşturmaktadır. Kargoculuk ve lojistik sektöründe müşterilerin terketme davranışını en iyi tahmin edilebilecek bir veri analiz süreci aracının hangi yazılım mimarisini kullanılacağı bilinmemektedir. Bu araştırma ile Kargoculuk ve lojistik sektöründe müşterilerin terketme davranışını en iyi tahmin edilebilecek bir veri analiz süreci aracında kullanılabilecek yazılım mimarisinin kullanılabilirliğini ortaya koymak için bir protatip geliştirilmiştir.

3.1.2 Kargo ve lojistik sektöründe müşterinin tutundurma yöntemini en iyi tahmin edilebilecek bir veri analiz süreci aracının yazılım mimarisi nasıl olabilir?

Kargo ve lojistik sektöründe müşteri tutundurma yöntemini en doğru şekilde tahmin edebilen veri analiz süreci aracının hangi yazılım mimarisini kullanılacağı bilinmemektedir. Bu araştırma sonucunda kargo ve lojistik sektöründe müşteri tutundurma yöntemi öngörmek için yazılım mimarisinin bir veri analizi süreç aracında ne kadar iyi kullanılabileceğini gösteren bir prototip oluşturulmuştur.

3.1.3 Kargoculuk ve lojistik sektöründe, müşteri terketme davranışında kullanılabilecek, müşteriyi terketmeden önce kalmasını sağlayabilecek, tahmin gücü en başarılı öznitelikler nelerdir?

Müşterinin terketme davranışından önce kalmasını sağlayacak durumlar tespitinin tahmini yapılırken bazı öznitelikler başarılı sonuç verirken bazı öznitelikler daha az başarılı sonuç vermektedir. Bu çalışmayla hangi özniteliklerin müşterinin terketme davranışından önce kalmasını sağlayacak durumlar tespitinin tahmini için daha önemli olduğu ortaya konulmaktadır.

3.1.4 Kargo ve lojistik sektöründe müşteri tutundurma yöntemi iş sürecinde hangi algoritmalar kullanılabilir?

Bazı sıralı madencilik algoritmaları, müşterinin şirkette kalması için uygulanacak olan eylemlerin tahminleri yapılırken etkili sonuçlar verirken, bazıları daha az başarılı sonuçlar vermektedir. Tutundurma iş süreci yazılım mimarisi için hangi sıralı kural madenciliği stratejilerinin daha önemli olduğu bu araştırma ile ortaya çıkmaktadır.

3.1.5 Kargoculuk ve lojistik sektöründe, müşterinin terketme davranışı tahmininde kullanılmasını önerdiğimiz öznitelikler ile başarılı tahminler yapılabilir mi?

Müşterinin terketme davranışından önce kalmasını sağlayacak durumlar tespitinin tahmininde makine öğrenme algoritmaları kullanılarak başarılı sonuç verip vermediği araştırılacaktır.

3.1.6 Kargoculuk ve lojistik sektöründe, müşterinin tutundurma yöntemi için kullanılmasını önerdiğimiz mimari ile başarılı şekilde müşteriler firmada kalabilir mi?

Müşterinin şirkette kalması için uygulanacak olan eylemlerin tahmininde sıralı kurallı madencilik algoritmaları kullanılarak başarılı sonuç verip vermediği araştırılacaktır.

4
ÖNERİLEN YÖNTEM

Bu tez kapsamında, anlaşmalı müşterinin terk etme davranış bilgileri ve bu müşteriye ait olan kargo gönderim bilgileri periyodik olarak veri kaynağı üzerinden depolanmıştır. Bu depolanan verilerden haraketle müşterinin bilgileri ve anlaşmalı müşterilerin kargo gönderim kayıtları, müşterinin terketme davranışları bilgileri toplanarak veritabanı oluşturulmuştur. Veri önişleme adımlarından geçirilerek matris veri seti haline getirilmiştir. Ayrıca pazarlama departmanı tarafından birden çok tipte müşteriyle birçok kanaldan iletişime geçip çeşitli aksiyonlar gerçekleştirilmiştir. Bu aksiyonların müşterilerin sözleşmesinin devam etttimesi üzerindeki etkisi incelenerek müşteri tutundurma verilerin oluşmasını sağlanmıştır. Bu toplanan pazarlama departmanı günlük verileri sıralı kural madenciliği algoritmalarının okuyabileceği standart bir giriş formatına dönüştürerek veri seti haline getirilmiştir.

Kargo ve lojistik sektörlerine yönelik olarak, müşteri terk etme davranışını ve müşteri tutundurma yöntemini tahmin edebilecek bir veri analizi iş sürecinin yazılım mimarisi önerilmektedir. Önerilen yazılım mimarisinin prototip bir uygulaması geliştirilmiştir. Müşteri terk etme davranışı verileri ve tutundurma iş süreci yazılım mimarisi için kullanılan srm verileri farklı aşamalardan geçmektedir. Bu aşamaların sonucunda, her bir müşteri için terk etme eğilimi olup olmadığı ve müşterinin şirkette sözleşmesinin devam edebilmesi için hangi tutundurma yöntemini kullanılması gerektiğini tahmin edilebilmektedir.

Müşteri terk etme davranışı tahmini yazılım mimarisinin prototip ayarları Sistem Yöneticisi tarafından ayarlanmış olup, optimize edilecek her bir hiperparametre için değer aralığı tanımlar. Değerler prototip için önemlidir. Kabul edilebilir ve doğru şekilde tanımlanmladır. Parametre yapılandırılma sırasında optimizasyon tekniği de dikkate alınmalıdır. Dikkate alınması gereken diğer faktörler, her otomatik veri hattı(automated pipeline) için maksimum yineleme sayısının yanı sıra doğruluk eşiğini de içerir. Müşteri terketme davranışı ile ilgili gerçekleştirdiğimiz yazılımın mimarisi Bölüm 4.1'de anlatılmaktadır.

Müşteri tutundurma iş süreci yazılım mimarisinin prototip uygulaması farklı veri boyutlarında test edilir. Veri boyutları prototip için ehemmiyetlidir. Seçilen veri boyutlarındaki verilerin testleri sırasında sıralı kural madencilik algoritmaları göz önünde bulundurulmalıdır. Göz önünde bulundurulması gereken diğer faktör ise çalışma süresidir. Müşteri tutundurma iş süreci yazılım mimarisinin prototip uygulamanın çalışma süresi açısından performans değerlendirilmesi yapılmıştır. Müşteri tutundurma modeli ile ilgili gerçekleştirdiğimiz yazılımın mimarisi Bölüm 4.2'de anlatılmaktadır.

4.1 Önerilen Terk Etme Davranışı İş Süreci Yazılımı Mimarisi Ana Birimleri

Önerilen yazılım mimarisi veri önişleme, otomatik veri hattı(automated pipeline), hiperparametre optimizasyonu ve model seçim olmak üzere dört bölümden oluşmaktadır.

4.1.1 Veri Önişleme Modülü

Veri önişleme modülü, veri özniteliklerinin oluşturulması, etiketli verinin oluşturulması ve müşteri terketme davranış verisinin işlenmesi olmak üzere üç bölümden oluşmaktadır.

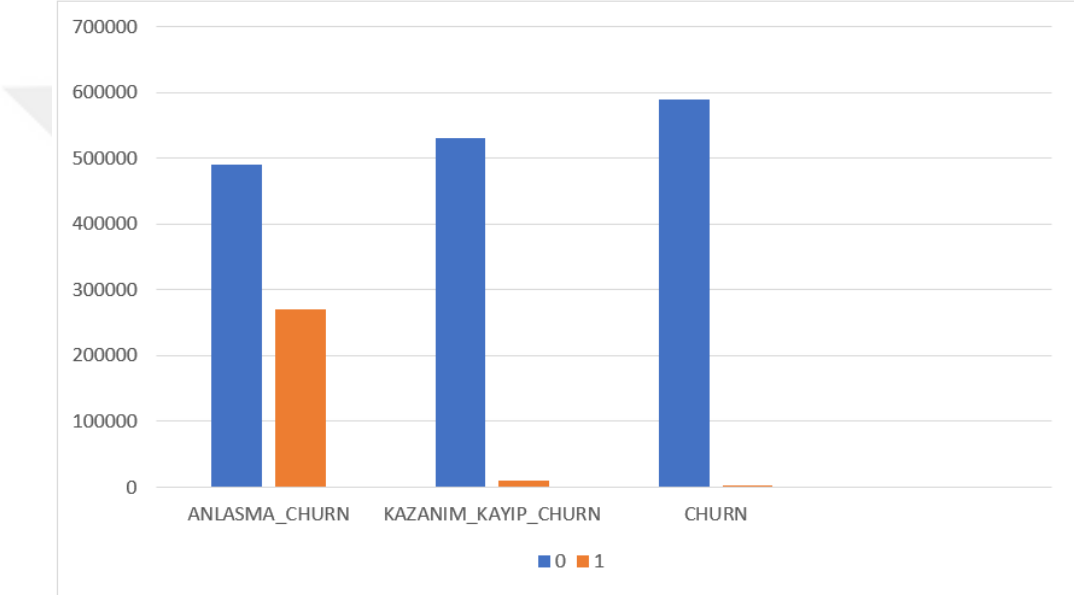
4.1.1.1 Veri Özniteliklerin Oluşturulması

Kargoculuk sektöründe müşteri kavramı kargo taşıma hizmetinin kapsamının ve koşullarının karşılıklı imza altına alındığı müşteriler ve sadece kargo gönderimi ihtiyacı durumunda gönderi yapılacak müşteriler olmak üzere ikiye ayrılır.

Müşteri terketme davranışlarının tahmini için 120 öznitelik bulunmaktadır. Anlaşmalı müşterilerin kendi firmasına ait öznitelikler, anlaşmalı müşterilerin kayıt edildiği lokasyon ile ilgili öznitelikler, müşterinin kargo gönderimi ve teslimi sırasında tercih ettiği öznitelikler, anlaşmalı müşterinin anlaşma öznitelikleri ve anlaşmalı müşterinin kargo gönderim öznitelikleri olmak üzere beş kategoriye ayrılmaktadır. Anlaşmalı müşterilerin kendi firmasına ait özniteliklerde firma türü ve sektörü gibi öznitelikler içermektedir. Müşterinin bulunduğu firma türünün vb kaçma eğilimi ile bağlantı olup olmadığı sorununa cevap alabilecek öznitelikler seçilmiştir. Anlaşmalı müşterilerin kayıt edildiği lokasyon ile ilgili özniteliklerde müşterinin hangi lokasyonun müşterisi olduğuna dair öznitelikler içermektedir. Müşterinin kargo gönderimi ve teslimi sırasında tercih ettikleri özniteliklerde müşterilerin kargo gönderimlerinin hangi

durumlarda olması gerektiğine karar veren öznitelikler içermektedir. Örneğin kargo gönderimi ne zaman yapılacak, kargo gönderimi sırasında fark kesilip kesilmeyecek mi sorularına cevap veren öznitelikler bulundurmaktadır. Anlaşmalı müşterinin anlaşma özniteliklerinde anlaşmanın bulunduğu lokasyon anlaşma türü gibi öznitelikleri bulundurur. Anlaşmalı müşterinin kargo gönderim özniteliklerinde müşterilerin bir dönemde gönderdiği kargo adetleri, tutarları, sms ücreti gibi öznitelikler bulunmaktadır. Örneğin artan kargo ücretleri dolayısıyla bir terketme davranışı bulunup bulunmadığını araştırır.

4.1.1.2 Etiketli Verinin Oluşturulması



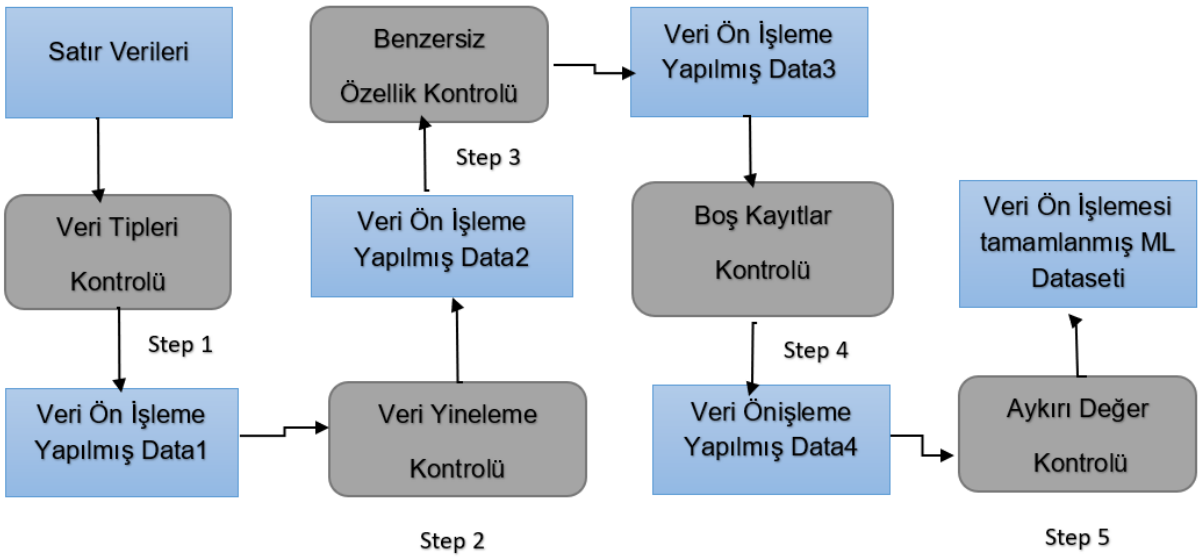
Şekil 4.1 Müşteri Kaybı Davranışı Verilerine İlişkin Etiketli Verilerin Dağılımları

Müşterinin terk etme davranışı verisi oluşturulurken Kazanım_Kayıp_Etiket, Aktif_Pasif_Etiketi ve Anlaşma _Aktif _Pasif _Etiketi olmak üzere üç farklı öznitelik etiketi eklenmiştir. Kazanım_Kayıp_Etiketinde kargoculuk sektöründe kayıp olup olmadığı ve müşterinin kazanılıp kazanılmadığı müşteri ilişkileri iş birimi tarafından oluşturulan kurallar baz alınarak belirlenebilmektedir. Bu etiket belirlenen müşterinin müşterinin kazanılıp kazanılmadığı belirten etikettir. Aktif _Pasif Etiketi müşterinin kazanılıp kazanılmadığı ile ilgili verilerde müşterinin aktif olup olmaması baz alınarak oluşturulmuştur. Anlaşma _Aktif _Pasif _Etiketinde bir müşterinin anlaşmayı bitirdikten sonra bir süre geçip tekrar anlaşma yapma ihtiyacı duyabilir. Zaman etkileşim verilerinden haraketle bu etiket oluşturulmuştur. Müşterinin, bir dönem içerisinden aktif anlaşması olup olmadığı kontrol edilir, aktif anlaşması yoksa terk etti olarak işaretlenmesi yapılmıştır. Veri üzerinde etiket verilerin dağılımı Şekil-4.1 'de gösterilmiştir. Veri üzerindeki etiket dağımlarından yola çıkarak makine öğrenmesi

algoritmalarının daha başarılı tahminlerde bulunması için Anlaşma _Aktif _Pasif _Etiketi hedef etiket belirlenmesine karar verilmiştir.

4.1.1.3 Müşteri Terketme Davranış Verisinin İşlenmesi

Müşteri terketme davranışları verisi, veri analiz aşamalarından geçerek, makine öğrenme algoritmaları için başarılı sonucu verecek şekildeki son halini almıştır. Veriyi işlemeden önce ham data 120 öznitelik ve 591662 satır bulundururken, veri işlendikten sonra makine öğrenmesi algoritmaları için kullanılacak son dataset 103 öznitelik ve 563423 satır bulundurmaktadır. Aşağıdaki Şekil- 4.2'de detaylı anlatılmıştır.



Şekil 4.2 Kademeli Veri Ön İşleme

Veri önişleme adımları 4.2'de gösterilmiş olup 5 adımdan oluşmaktadır. Birinci adım olan data tipleri kontrol adımında özniteliklerin data tipleri kontrol edilmelidir. Eğer düzeltilmesi gerektiği durumlarda data tiplerin düzeltilmesi sağlanmalıdır. Örneğin, müşteri terk etme davranışı verisinde, "MUST_NO" özniteliği kategorik iken numeric tipine dönüştürülmüştür.

Veri önişlemenin ikinci adımı tekrarlayan kayıt kontrolüdür. Bu adımda bir kaydın birden fazla tekrar etmesi durumları incelenecektir. Bu gibi durumlar makine öğrenmesi algoritmaları için sorun oluşturulmaktadır. Müşteri verisi incelenerek mükerrer kayıtların tek kayıt haline getirilmesi sağlanmıştır.

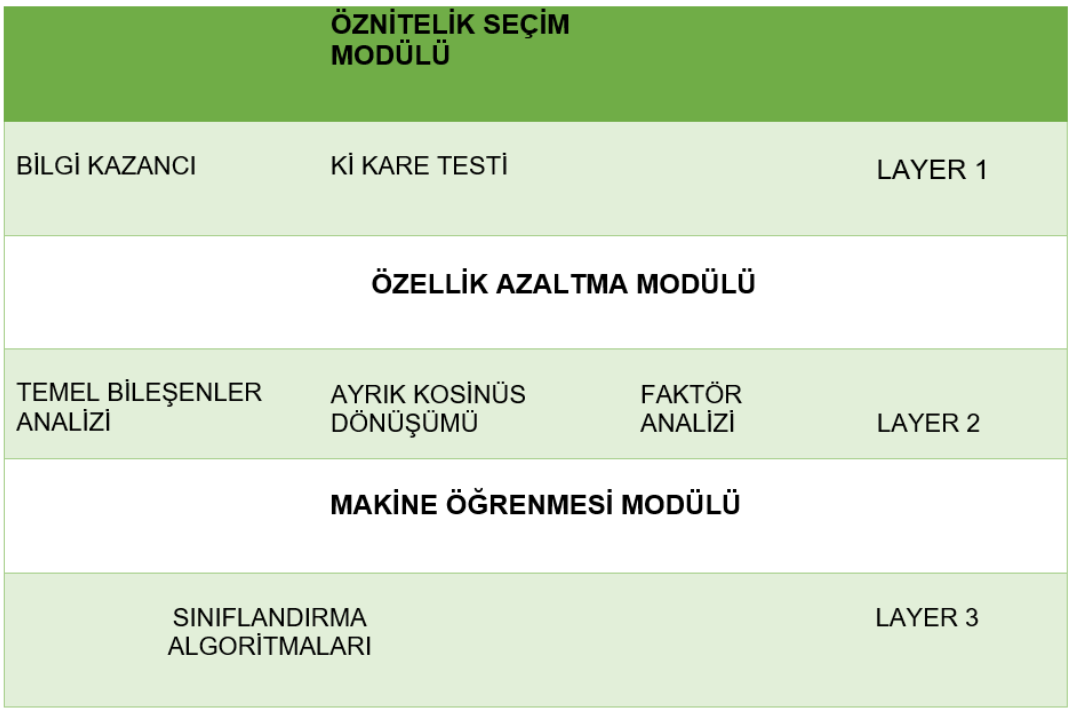
Veri önişlemenin üçüncü aşaması benzersiz öznitelikler kontrolüdür. Benzersiz olan öznitelikler model oluşturmakta etkisizdir, modelleri yanlış değerlendirip, yanlış

sonuçlar vermeye sebep olabilirler. Data tipleri kontrolleri sırasında düzelttiğimiz "MUST _NO" buna bir örnektir. Benzersiz öznitelikler araştırılıp Müşteri terk etme davranışı verisinden çıkarılmıştır.

Veri önişlemenin dördüncü adımı boş kayıtlar kontrolüdür. Bu aşamada boş kayıtlarla ne yapılacağı her kayıt için ayrı ayrı düşünmelidir. Ve özniteliğin içerdiği bilgi ve önemlilik derecesine göre yapılması gereken işleme karar verilmelidir. Boş kayıt ile karşılaşılması durumunda eksik değerleri sahip satır ve sütunlar silinebilir, özniteliğin ortalama, medyan veya mod değerini boş alana yazılabilir, benzersiz bir kategori atanabilir, eksik değerler tahmin edilebilir veya eksik değerleri destekleyen algoritmalar kullanılabilir. Bu çalışmada boş kayıtların aynı zamanda benzersiz kayıtlar olduğu farkedilmiştir. Ve bu kayıtların veri setinden çıkarılması sağlanmıştır.

Veri önişlemenin son adımı aykırı değer kontrolüdür. Öznitelik içeriğinin geri kalanından oldukça farklı olan gözlemlere aykırı değer denir. Aykırı değerler ortalamanın çok aşağısında yada çok yukarısında kalır.Aykırı değerler hassastır onları silmek veri bütünlüğünü bozup veri kaybına yol açabilir. Bu araştırmada aykırı değerler box-plot analizi yöntemi ile veri setinden çıkartılmıştır.

4.1.2 Otomatik Boru Hattı(Pipeline) Birimi



Şekil 4.3 Automated Pipeline Mimarisi

Otomatik Boru Hattı(Pipeline) Birimi terk etme davranışı iş süreci yazılımı mimarisi uygulamasının ana parçasıdır. Otomatik Boru Hattı(Pipeline) üç katmandan

oluşmaktadır. Bu katmanlar üç modül ile yönetilmektedir. İlk katman özellik seçimi modülü olup bilgi kazanımı(information gain) ve ki-kare testi algoritmalarını içerir. İkinci katman ise özellik azaltma modülüdür. Temel bileşenler analizi, ayrık kosinüs dönüşümü, ve faktör analizi algoritmaları katmanlarını ihtiva eder. Son katman olan makine öğrenmesi katmanı olmakta ve Lojistik Regresyon,Karar Ağaçları, Yapay Sinir Ağları, Gradyan Arttırma Ağaçları Sınıflandırıcısı (Gradient-Boosted Trees Classifier), Doğrusal SVC, Rasgele Orman Sınıflandırıcısı ve FM Sınıflandırıcısı algoritmalarından oluşmaktadır. Her katmandaki algoritma seçimi Hiperparametre Optimizasyon Birimi tarafından yapılır. Belirli yineleme için the Otomatik Boru Hattı(Pipeline) her katmandaki bir bölümü etkinleştirebilir. Şekil-4.3'de pipeline yapısı görülmektedir.

4.1.2.1 Öznitelik Seçim Modülü

Veri kümesi içerisinde hangi özniteliklerin sonuç üzerinde etkili olduğunu belirlemek için parça parça değerlendirmeye alma işlemi kullanılır. Bu çalışmada özniteliklerin veri üzerindeki önemini hesaplanmış ve iki tane öznitelik seçim algoritması kullanılmıştır. Bunlardan ilki information gain algoritmasıdır. İnformation gain bir özniteliğin bir sınıf hakkında ne kadar bilgi sakladığının ölçütü olarak kullanılır.

Bilgi Kazancı(Information Gain) hesaplanmasında Entropi kullanılır. Entropi fiziksel bir sistemin düzensizliğinin ölçüsüdür ve beklenmeyen bir olayın ortaya çıkma olasılığını gösterir. Entropi hesabı Denklem 4.1'de verilmektedir.

$$= H(p_1, p_{2,\ldots\ldots p_m}) = -\sum_{k=1}^{m} p_i \log_2 p_i \tag{4.1}$$

Burada m entropisi hesaplanacak durum sayısıdır ve p_i i durumunun olasılığıdır.

Entropideki azalma miktarı kazanç olarak belirtilir. Entropi azalmasına en çok sebebiyet veren öznitelik en çok kazanç sağlayan özniteliktir. Information Gain hesabı Denklem 4.2'de verilmektedir.

$$Gain(S,B) = Entropy(S) - \sum_{v \varepsilon Values(B)} \frac{|S_V|}{S} Entropy(S_v) \tag{4.2}$$

Gain(S,A) B niteliğinin S sistemine göre kazancını ifade eder. Sistemin entropisinden özniteliğin entropi değerinin çıkarılmasıyla elde edilmektedir. Diğer algoritma olan Ki-kare testleri kategorik veri dağılımlarıyla ilgili hipotezleri test etme yollarını bize sunan bir grup önem testidir.

Ki-Kare test hesaplanması Denklem 4.3 verilmektedir. C tahminin hesaplanmasına giren bağımsız bilgi parçalarının sayısıdır. O ölçülen değerdir ve E beklenen değerdir.

$$X_c^2 = -\sum \frac{(O_i - E_i)^2}{E_i} Entropy(S_v) \tag{4.3}$$

4.1.2.2 Öznitelik Azaltma Modülü

Bu modülde öznitelik sayısının fazla olduğu koşullarda öznitelikler kümeleyip sanal bir öznitelik oluşuturulur ve makine öğrenmesi modeli için en başarılı sonuçlar verecek öznitelik haline getirilir. Bu modülde dört tane öznitelik azaltma algoritması kullanılmaktadır. Bunlardan ilki faktör analizi algoritmasıdır. Öznitelikler arasındaki karşılıklı ilişkili olanlar bir kategoriye toplanır ve daha bu özniteliklerin büyük sayılarını daha küçük faktör kümelerine indirger. İkinci algoritma Temel Bileşenler Analizi algoritmasıdır. Temel Bileşenler Analizi algoritmasının temel amacı yüksek boyutlu verilerde verilerdeki genel özellikleri bularak boyut sayısının azfaltılmasını sağlar ve verilerin sıkıştırılması sağlanarak bilgi kaybının en aza indirilmesi için kullanılır. Bunu varyansı yüksek korele olmayan yeni değişkenler oluşturarak yapar. Son algoritma ise Ayrık Kosinüs Dönüşümü algoritmasıdır. Ayrık Kosinüs Dönüşümü algoritması farklı frekanslarda salınan kosinüs fonksiyonlarının toplamı cinsinden sonlu bir veri noktaları dizisini ifade eder. Bu algoritmayla artıklık sorunu ortadan kalkar.

4.1.2.3 Veri Setinde Eğitim ve Test Veri Setinin Ayırımı

Kullanılan müşteri verisi sınıflandırma algoritmasının başarılı sonuçlar verebilmesi için %70 eğitim ve %30 test verisi olmak üzere ayrılmıştır. Veri seti bir buçuk yıllık dönemsel müşteri terketme davranışı verisi kullanılarak oluşturulmuştur. Müşterinin terketme eğilimi, bir önceki aşamada oluşturulan tahmin modeli kullanılarak tahmin edilir. Müşteriyi terk etme eğilimi, terketme eğilimi var terketme eğilimi yok olmak üzere ikili bir karar kullanılarak hesaplanır.

4.1.2.4 Makine Öğrenmesi Algoritmaları Modülü

Seçilen öznitelikler yardımıyla farklı sınıflandırıcı algoritmaları kullanılarak müşteri terk etme davraşının etiketlenmesi sağlanmaktadır. Lojistik Regresyon, Karar Ağaçları, Yapay Sinir Ağları, Gradyan Arttırma Ağaçları Sınıflandırıcısı (Gradient-Boosted Trees Classifier), Doğrusal Destek Vektör Sınıflandırıcısı (Linear SVC), Rasgele Orman ve FM Sınıflandırıcısı olmak üzere yedi tane sınıflandırıcı makine öğrenmesi algoritması

kullanılmıştır.

1. **Lojistik Regresyon**

   Bağımlı değişken, sürekli bir değişken olmayıp, olabilecek veya olmayabilecek bir durum veya belirli bir sınıflandırmadaki bir kategori olduğunda, binom (ikili) lojistik regresyon kullanılır [52]. Logistic Regression algoritması bağımlı değişkenin tahmini değerlerini olasılık olarak hesaplar ve olasılık kurallarına uygun sınıflandırır. Lojistik regresyon sürekli ve/veya kategorik değişkenlerin sınıflandırılması ve sonucunun tahmin etmesinde kullanılmaktadır. Normal regresyon modelinin aksine, lojistik regresyondaki bağımlı değişken tipik olarak ikilidir: q olasılıkla 1 değerini veya 1-q olasılıkla 0 değerini alabilir [53]. Basit lineer regresyon denklemi Denkelm 4.4 verilmektedir.

$$p(x) = \beta_0 + \beta_1 X \tag{4.4}$$

   Bu denklemin sağ tarafında bağımsız değişken X, bağımsız değişkenin katsayısı 1'dir. X'deki bir birimlik değişim p(x) değerini $beta_1$ oranında etkilemektedir. Sonuç olarak, sabit bir değişken oluşturulmuştur. Lojistik fonksiyonunun amacı, bağımsız bir değişkenin bir kategoriye dahil edilip edilemeyeceğini göstermektir. Lojistik fonksiyonu denklemin sağ tarafına uygulanmaktadır ve lineer regresyon denkleminin sol tarafında 0 ile 1 arasında bir değer elde edilmektedir.

$$p(x) = \frac{e^{\beta_0+\beta_1 x}}{1 + e^{\beta_0+\beta_1 x}} \tag{4.5}$$

   eşitlik düzenlenirse

$$\frac{p(X)}{1-p(X)} = e^{\beta_0+\beta_1 x} \tag{4.6}$$

   π(x) lojistik fonksiyonunun logit dönüşümü Denklem 4.7'deki gibidir.

$$g(x) = \ln\left[\frac{p(x)}{1-p(x)}\right] = \beta_0 + \beta_1 x \tag{4.7}$$

   Bu eşitliğin sol tarafı logit olarak bilinir. Doğrusal regresyondan farklı olarak, $beta_1$, p(X)'deki X'in her değerini aynı oranda değiştiremez. Değişim ise şimdi X'in mevcut değeri tarafından belirlenecek. p(X) ve x arasındaki ilişki artık doğrusal değildir. Lineer model uyumunda en küçük kareleri doğrusal olmayan modellerin uyumunda ise genelde maximum likelihood yaklaşımı kullanılmaktadır. Amaç, bir veri kümesinden seçilen bir elemanın β_0 ve β_1,

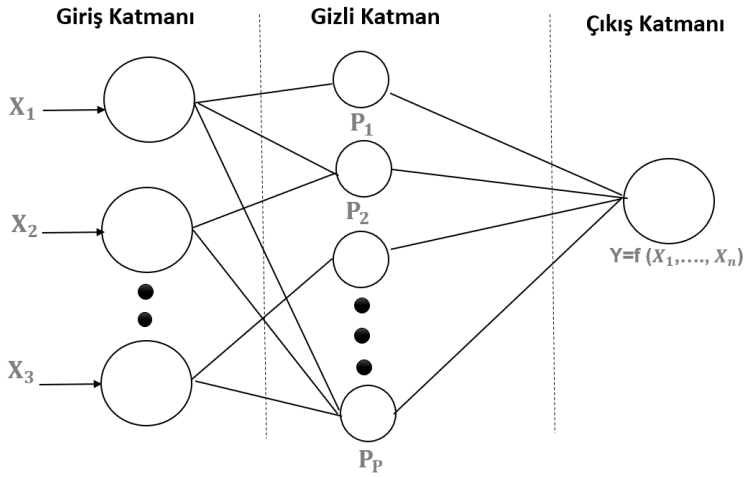
β_2,β_3.... değerlerini ve ayrıca elemanın hangi kümeye dahil edileceğini tahmin etmektir.

2. **Karar Ağaçları**

   Karar ağacı, bilinen veri sınıflarından öğrenmek için endüktif yöntemi kullanan bir karar yapısıdır. Karar ağacı, büyük hacimli verileri yönetilebilir gruplara ayırmak için basit karar verme prosedürlerini kullanan bir öğrenme yöntemidir. Sonuç grubunun üyeleri, her başarılı ayrılığın bir sonucu olarak daha benzer hale gelir. Okunması kolay, veritabanlarına dahil edilmesi kolay ve güvenilir olduğu için tanımlayıcı ve öngörücü özelliklere sahip karar ağacı en çok kullanılan sınıflandırma yöntemlerinden biridir [54]. Karar ağacı bir karar düğümü ve birden fazla dallanma içerebilmekte olup kategorik ve sayısal verilerden oluşabilir.

3. **Yapay Sinir Ağları**

   Yapay sinir ağı (YSA), sinir hücreleri adı verilen çok sayıda basit düğümden oluşan karmaşık bir ağdır. YSA, insan beyin dokusu ve sinir sistemi üzerine yapılan son teknoloji biyoloji araştırmalarına dayanarak geliştirilmiştir ve insan beynindeki bilgi işleme sinirsel aktivitesini modellemek için kullanılabilir [55]. Bir YSA'daki bilgi işleme düğümlerinin topolojik yapıları, verileri paralel bir şekilde dağıtmak için kullanılır. Doğrusal olmayan aktarım işlevlerinin birleşimleri, girdi eşlemeleri ve tahmini çıktı yanıtları elde etmek için kullanılır. Herhangi bir matematiksel model geliştirmeden, bulanık, doğrusal olmayan ve gürültü içeren verileri işlemek için geçmiş deneyimleri, sinir hücrelerini, belleği ve ilişkilendirmeyi kullanarak yapay sinir ağlarının eğitim algoritmalarını incelemek için kendi kendine uyarlanabilir bilgi örüntü tanıma özelliğini kullanabiliriz [56].



Şekil 4.4 Tek Gizli Katmanlı Sinir Ağı

Sinir ağları tek katmanlı yapay sinir ağları ve çok katmanlı katmanlı sinir ağları olmak üzere ikiye ayrılmaktadır. Çok katmanlı yapay sinir ağları dış ortamdan veya diğer nöronlardan toplanan veriler kullanılarak ağırlıklardan oluşan girişler, toplama fonksiyonu, aktivasyon fonksiyonu ve çıktılardan oluşmaktadır. Giriş ve çıkış katmanları arasında, sinir ağı bir veya daha fazla gizli nöron katmanına sahiptir. Bu tür ara katmanlar doğrudan dışarıdan girdi almadığından, bunlar gizli katmanlar olarak bilinir ve bunların içine gömülü düğümler gizli düğümler olarak bilinir. Çok sayıda gizli katmana ve birden fazla girişe sahip olabilir. Sorunun akışına bağlı olarak, gizli katman sayısı artırılabilir veya azaltılabilir. Sorun, yapısına bağlı olarak birden çok işlev kullanılarak ele alınabilir ve daha sonra gizli katmanı kullanarak çıktı katmanına gönderilebilir. Bir gizli katmanlı yapay sinir ağı Şekil-4.4'de gösterilmiştir.

Gizli birimin aktivasyon fonksiyonlarını Denklem 4.8'deki yazılmış ikinci bir işlem elemanı katmanı kullanılıp dönüştülerek, gizli katman ağ çıktısı olan çok katmanlı yapay sinir ağları oluşturulabilir. Burada D, L ve J sırasıyla giriş, gizli ve çıkış katmanındaki toplam birim sayısıdır ve f bir aktivasyon fonksiyonudur.

$$output_{net}(j) = f\left(\sum_{n=1}^{L} w_{jl} f\left(\sum_{n=1}^{L} w_{li} x_i\right)\right) \tag{4.8}$$

4. **Gradyan Arttırma Agaçları Sınıflandırıcısı**

   Gradyan Arttırma veya GBM, hem regresyon hem de sınıflandırma problemleri için çalışan bir başka topluluk makine öğrenme algoritmasıdır. GBM, güçlü bir öğrenici oluşturmak için birtakım zayıf öğrenicileri birleştirerek, yükseltme tekniğini kullanır. Temel öğrenici olarak kullanılan regresyon ağaçları, her biri seri halinde olan bir ağaç önceki ağaç tarafından hesaplanan hatalar üzerine kuruludur.

5. **Doğrusal Destek Vektör Sınıflandırıcısı**

   SVM, bir girdi alanındaki doğrusal olmayan problemleri yüksek boyutlu bir özellik alanındaki doğrusal problemlerle eşleştirerek matematiksel veri sınıflandırmasına izin verir [57]. Giriş vektörlerinin yüksek boyutlu özellik alanına doğrusal olmayan eşlemesi sayesinde SVM, optimal ayırma hiper düzlemleri olarak adlandırılan ikili bir sınıflandırıcı oluşturur. SVM, karar işlevini tahmin etmek için doğrusal bir model oluşturmak üzere doğrusal olmayan sınıf kenarlıklarını ve destek vektörlerini kullanır. SVM, verileri doğrusal olarak ayrılırsa, verileri hatasız ve hiper düzlem ile en yakın eğitim noktaları arasındaki maksimum mesafeye bölen en uygun hiper düzlem için doğrusal makineleri eğitir. Destek vektörleri, en uygun ayırma hiper düzlemine

en yakın olan eğitim konumlarıdır. İkili sınıf sınırlarını tanımlamak için tüm ek eğitim örnekleri önemsizdir. SVM, veriler doğrusal olarak ayrılmadığında genel senaryolarda eğitim kümesindeki hata sayısını en aza indiren bir hiper düzlemi tanımlamak için doğrusal olmayan makineler kullanır [58].

Linear SVC ve SVM fark, SVC'nin libsvm'de uygulandığı sırada Linear SVC'nin liblinear açısından uygulanmasıdır. Linear SVC'nin penalties ve kayıp fonksiyonlarının seçiminde daha fazla esnekliğe sahip olmasının nedeni budur. Ayrıca çok sayıda örneğe daha iyi ölçeklenir.

6. **Rasgele Orman Algoritması**

   RF bir topluluk öğrenme yöntemidir. Birbirinden farklı olarak kurulan sınıflama ve regresyon karar ağaçları (CART) karar ormanı topluluğunu oluşturur. Karar ormanı oluşumu sırasında elde edilen sonuçlar bir araya getirilerek en son tahmin yapılır. RF yönteminde ağaçlar, seçilen bootstrap örneklemleri ve her düğüm ayrımında rasgele seçilen m adet tahminci ile oluşturulur. m adet tahmincinin toplam tahminci sayısından oldukça küçük olmasına dikkat edilir. Oluşturulan her bir karar ağacı en geniş haliyle bırakılır ve budanmaz. Sınıflandırma için ağaçlar; her yaprak düğümü sadece bir sınıfın üyelerini içerecek şekilde oluşturulurlar. Regresyon için ise; yaprak düğümde az sayıda birim kalana kadar ağaçlar bölünmeye devam ederler [59].

   Random Forest Algoritmasının Avantajları: [60]

   - Birçok uygulama türü için doğru tahminler sağlar.
   - Eğitim veri setine göre her bir özelliğin önemini ölçebilir.
   - Numuneler arasındaki ikili yakınlık, eğitim veri seti ile ölçülebilir.

   Random Forest Algoritmasının Dezavantajları [60]:

   - Farklı düzey sayılarına sahip kategorik değişkenler içeren veriler için, rastgele ormanlar daha fazla düzeye sahip öznitelikler lehine önyargılıdır.
   - Veriler, çıktı için benzer alaka düzeyine sahip ilişkili özellik gruplarını içeriyorsa, daha küçük gruplar daha büyük gruplara göre tercih edilir. Karar ağacı yapısının bağlı olduğu faktörler şunlardır [60]:
     - Her düğümde kullanılacak kararın şekli.
     - Her yaprakta kullanılacak tahmin edici türü.
     - Her düğümde optimize etmek için bölme hedefi.
     - Ağaçlara rastgelelik enjekte etme yöntemi.

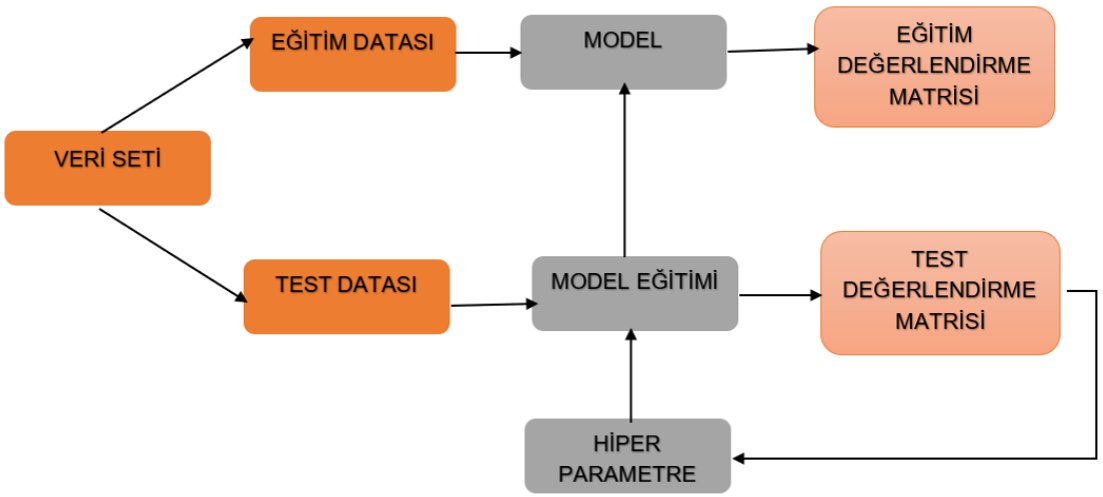
Analiz edilecek veri, veri önişleme adımlarından geçirilerek hazırlanılır. Algoritma herbir örnek için karar ağacı oluşturur ve her karar ağacının tahmini bir değer sonucu olur. Sınıflandırma problemi için mode regression problem için medyan kullanılmak üzere tahmin sonucu oluşan her değer için bir oylama gerçekleştirilir. En çok oylanan değeri seçerek sonuç oluşturulur.

7. **FM Sınıflandırıcısı**

   FM sınıflandırma modeli, dereceli azalma yöntemi ile çözülebilen lojistik kaybı kullanır ve L2 gibi düzenlileştirme terimleri genellikle aşırı öğrenmeyi önlemek için kayıp fonksiyonuna eklenir.

4.1.3 Hiper-Parametre Optimizasyonu Birimi

Hiperparametre Optimizasyonu Birimi, Otomatik Boru Hattı(Pipeline) Biriminin akıllı hale getirilmesi için önemli bir rol oynar. Her katmanın algoritmik bloklarının seçimi, bir optimizasyon yaklaşımına dayanarak akıllıca ve doğru olmalıdır. Bu çalışmada Terk etme davranışı iş süreci yazılımı mimarisi uygulaması için Bayes Optimizasyon Algoritmasını kullanılmaktadır. Hiper-Parametre Ayarlama(Tuning) Şekil-4.5 gösterilmektedir.



Şekil 4.5 Hiper-Parametre Optimizasyon Birimi

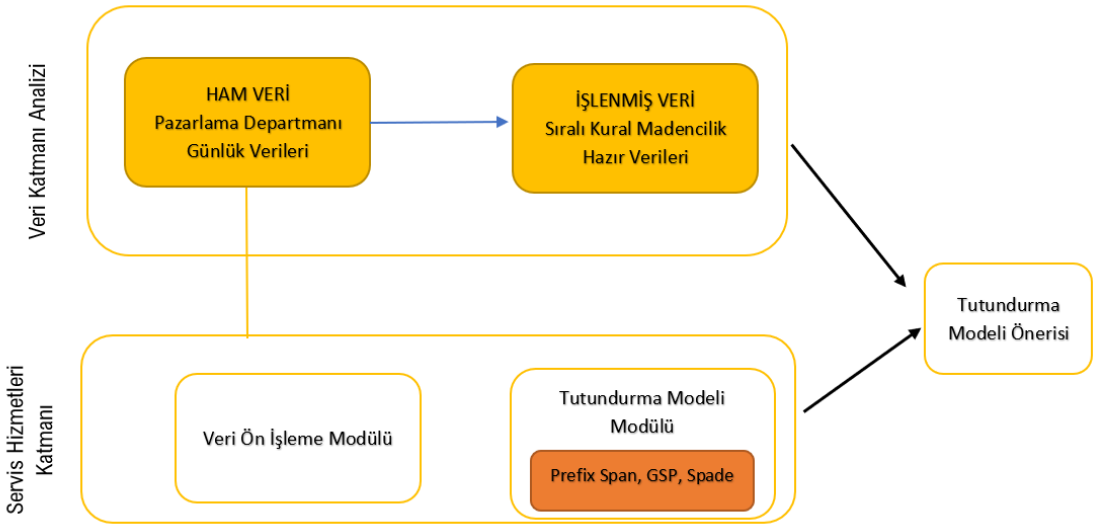
4.1.4 Bayes Optimizasyonu

BO, neredeyse tüm küresel optimizasyon türleri için tipik bir yöntemdir ve daha fazla veriyle daha az hata yapmayı amaçlamaktadır. Bayes optimizasyonundaki birçok gelişme yeni kazanım fonksiyonlarında, modellerde ve paralelleştirme şemalarında HPO'ya kolayca uygulanabilmektedir. Kullanıcıların hiperparametrelerin dağılımı

hakkında ön bilgiye sahip olmaları gerekmemesi ve arkasında olasılık fikrine dayanmasından dolayı rastgele arama ve ızgara arama yöntemlerinden daha verimli sonuçlar elde etmektedir.

BO algoritmasında [61] birçok olasılık modeli kullanılabilir, ancak Gauss Süreci (Gauss Process, GP) [62] çoğunlukla tercih edilmektedir. GP, tahminlerde belirsizliğin ölçülmesini sağlayan BO için tercih edilen modeldir. Parametrik olmayan bir modeldir ve parametre sayısı sadece giriş noktalarına bağlıdır. Bununla birlikte, GP'nin bazı dezavantajları da vardır. Örneğin, BO teorisiyle birlikte anlaşılması kavramsal olarak zordur. Ayrıca, yüksek boyutlu veya çok sayıda veri noktası ile zayıf ölçeklenebilirliği önemli bir sorundur [63].

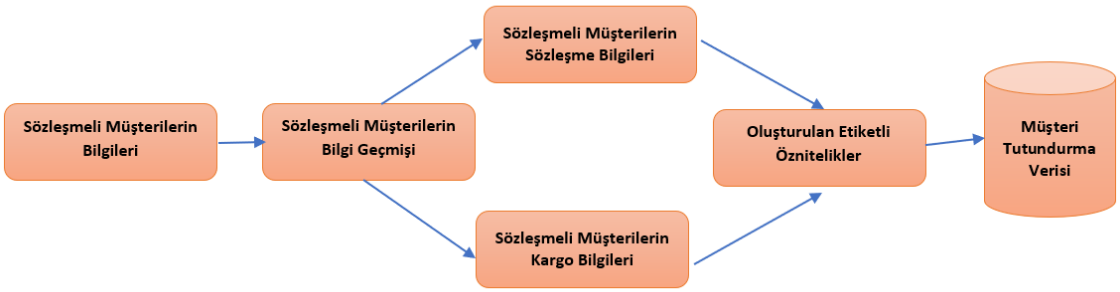
4.2 Önerilen Tutundurma İş Süreci Yazılımı Mimarisi Ana Birimleri



Şekil 4.6 Kargo Müşterisinin Tutundurma Tahmini Geri Dönüş Yazılım Mimarisi

Bu araştırma kapsamında, kargo müşteri verileri üzerinde, müşteri tutundurma iş süreci için yazılım mimarisi önerilmektedir. Önerilen mimari, Şekil-4.6'de gösterilmektedir. Müşteri tutundurma modeli girdisi olan ham verilerin yazılım mimarisi için uygun hale dönüştürülme süreci, Veri Ön İşleme alt başlığında, detaylı anlatılmaktadır. Önerdiğimiz tutundurma iş süreci yazılım mimarisinde, müşteri tutundurma analizi gerçekleştirilirken izlenilen iş süreci Şekil-4.7' de anlatılmaktadır.

Şekil-4.7 de görülebileceği üzere, müşteri tutundurma verileri farklı aşamalardan geçmektedir. Bu aşamalar pazarlama departmanı günlük verileri oluşturulması, müşteri tutundurma yöntemini günlük verileri oluşturulması ve müşteri ön işlenmesi aşaması ve müşteri tutundurma tahmin yöntemlerinin oluşturulmasıdır.



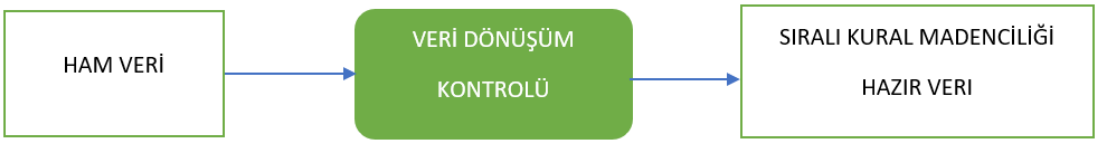
Şekil 4.7 Kargo ve Lojistik Sektörü Endüstrisinde Müşteriyi Elde Tutma Tahmini için Öznitelikler Oluşturma Süreç Mimarisi

Bu aşamaların sonucunda, müşterilere hangi sıralı tutundurma yöntemi uygulanması gerektiği tahmin edilmektedir. Önerdiğimiz yazılım mimarisi ve veri analizi iş süreci ile, bu bildirinin "Araştırma Soruları" bölümündeki 2. soruya yanıt veriyoruz. Önerdiğimiz mimarinin kullanılabilirliğini ortaya koymak için, bu araştırma kapsamında bir protatip geliştirilmiştir. Bu bildirinin Bölüm 5'de yer alan "Prototip Uygulama" ve Bölüm 6'de yer alan "Önerilen Yöntemin Değerlendirilmesi" kısımlarında, protatip uygulamayı açıklıyoruz. Bunun yanı sıra, uygulamanın çalışma süresi açısından performansını inceliyoruz. Aşağıda bu temel veri analizi süreçlerinin detayları verilmektedir.

4.2.0.1 Müşteri Tutundurma Verilerinin İşlenmesi

Müşteri tutundurma verisi, veri analiz aşamalarından geçerek, sıralı kural algoritması madenciliği algoritmaları kullanabilecek şekildeki nihai halini almıştır.

4.2.0.2 Verilerin Dönüşüm Kontrolü



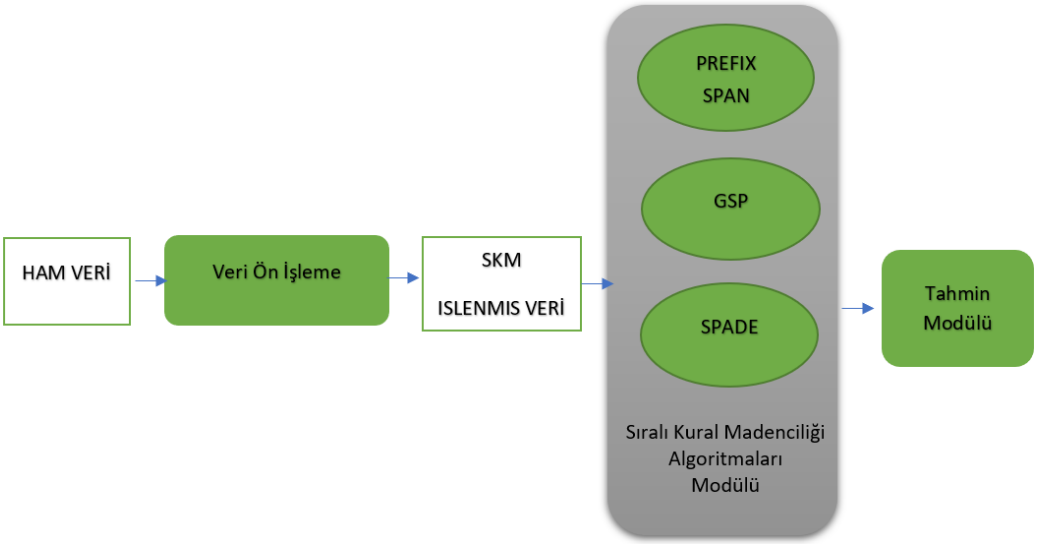
Şekil 4.8 Veri Dönüşüm Kontrolü

Şekil-4.8 veri önişleme adımını göstermektedir. Pazarlama departmanı günlük verisetinde 17 olay (müşteri ara, müşteriye SMS, müşteriye e-posta gönderme, müşteriye yerinde ziyaret ve indirim teklifi vb.) ve 989.818 satır bulunmaktadır. Bu her bir satırdaki sıralı kurallar tutundurma yöntemlerini oluşturmaktadır. Sıralı kural

madencilik algoritmalarında kullanılması için numeric tipinde satırlar içiren srm hazır data veriseti haline dönüştürülmesi gereklidir.

4.2.1 Müşteriyi Elde Tutma Tahmin Modülünün Oluşturulması

Şekil-4.9, müşteriyi elde tutma yöntemi tahmin sürecini göstermektedir. Şekil-4.9 modülleri, yani sıralı kural madenciliği algoritmalarını ve tahmin modüllerini görselleştirir. Bu modüller aşağıda ayrıntılı olarak açıklanmıştır.



Şekil 4.9 Müşteriyi Elde Tutma Modeli

4.2.1.1 Sıralı Kural Madenciliği Algoritmaları Modülü

Seçilen öznitelikler, farklı sınıflandırıcı algoritmalara verilerek, sonuçlar doğrultusunda müşteri tutundurma yöntemlerin frekanslarına göre sıralanması sağlanmıştır. Prefix span, spade ve gsp algoritmaları bu çalışmada kullanılan sıralı kural madenciliği algoritmalarıdır. Önerilen sıralı kural madenciliği algoritmaları ile bu makalenin "Araştırma Soruları" bölümündeki 5. soru cevaplanmıştır.

1. **Prefix Span**

   PrefixSpan algoritmasının yürütme prosedürü şu şekilde açıklanmaktadır: İlk olarak, PrefixSpan algoritması, verilen minimum desteğe göre 1-uzunluk sıralı örüntüler olan tüm sık öğeleri bulmak için sıralı veri tabanını tarar ve ardından dizi veri tabanı, bu sık öğelere göre farklı alt kümelere bölünür; burada her alt küme, sık 1-diziye göre dizi veritabanının izdüşümüdür. Sıralı kalıpların

Girdi: Bir dizi veri tabanı D_S, ve minimum destek eşiği min_sup

Çıktı: Sıralı modelin eksiksiz seti

Metot: Çağır PrefixSpan ($\langle\rangle$,0, D_S)

Alt Program: PrefixSpan (α,l,S/α)

Parametreler: α sıralı model; l:α'nın uzunluğu; S/α öngörülen veri tabanı,

Eğer $\alpha \neq \langle\rangle$ aksi takdirde, dizi veri tabanı S

Method:

1. Önce S/α 'yı tara, sık öğeler kümesini bulun β öyle ki
   a. β, sıralı bir model oluşturmak için α'nın son elemanına monte edilebilir; veya
   b. Sıralı bir model oluşturmak için α 'ya β eklenebilir
2. Her bir sık kullanılan öge için β, sıralı bir model $\overline{\alpha}$ oluşturmak için bunu α 'ye ekleyin ve $\overline{\alpha}$ çıktısını alın
3. Herbir $\overline{\alpha}$ için, $\overline{\alpha}$ öngörülen veritabanı S/α 'yi oluştur ve PrefixSpan ($\overline{\alpha}$,l+1,$S/\overline{\alpha}$) çağırın

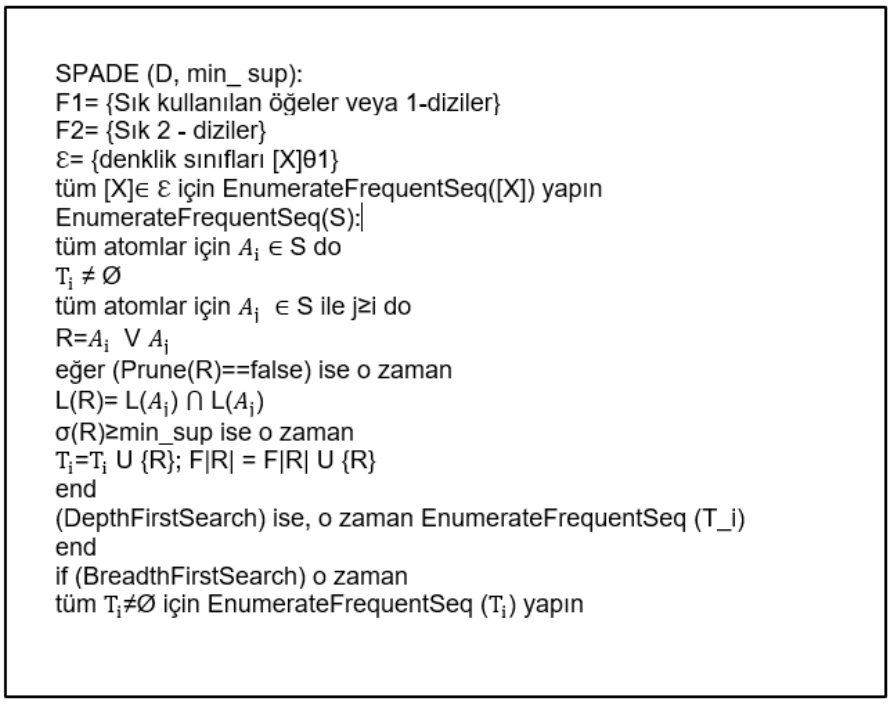
Şekil 4.10 Prefix Span Algoritması

alt kümeleri, karşılık gelen öngörülen veritabanları kümesi oluşturularak çıkarılır. Bu öngörülen veritabanlarından, PrefixSpan algoritması, sık 1-diziler kullanarak, karşılık gelen aynı ön ek ile sık 2-dizileri oluşturmaya devam eder. Her sık k-dizisi için PrefixSpan algoritması, sık (k + 1)-dizilerini yinelemeli olarak araştırmak için öngörülen bir veritabanı oluşturur. Bu tanımlara dayanarak, PrefixSpan'in algoritması Şekil-4.10'de sunulmaktadır [64].

2. **Spade**

SPADE, verimli kafes arama teknikleri(lattice search techniques) ve basit birleştirmeler(simple joins) kullanarak sık dizileri bulmak için önerilen bir algoritmadır. SPADE, orjinal problemi, her bir sınıfın bağımsız olarak çözülebilmesi için sık dizilerdeki denklik sınıflarını kullanarak daha küçük alt problemlere ayrıştırır. SPADE genellikle yalnızca üç veritabanı taraması yapar, ilki sık 1-diziler için, ikincisi sık 2-diziler için ve bir tane daha tüm diğer sık dizileri oluşturmak için. 2 dizi desteği mevcutsa, yalnızca bir tarama gerekir [65].

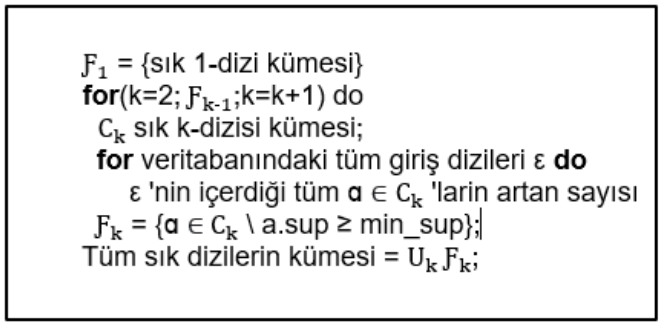
SPADE, GSP algoritmasını iki kat ve önceden hesaplanmış 2 dizi desteğiyle bir büyüklük sırasına göre geride bırakır. Aynı zamanda, girdi dizisi sayısı, girdi dizisi başına olay sayısı, olay boyutu ve potansiyel maksimum sık olay ve dizilerin boyutu gibi bir dizi parametreye göre mükemmel ölçek büyütme



Şekil 4.11 Spade algoritması

özelliklerine sahiptir [66]. Spade algoritması Şekil-4.11'te gösterilmiştir.

3. **Gsp**



Şekil 4.12 GSP algoritması

GSP, veritabanı üzerinden birden çok geçiş yapmaktadır. İlk geçişte, tüm tek öğeler (1- diziler) sayılır. Sık kullanılan maddelerden bir dizi aday 2 dizisi oluşturulur. Desteklerini toplamak için başka bir geçiş yapılır. Sık 2 dizileri, aday 3 dizilerini oluşturmak için kullanılır ve bu işlem, daha sık diziler bulunana kadar tekrarlanır. Algoritmada iki ana adım vardır. [66]. GSP algoritması Şekil-4.12'te gösterilmiştir.

1. Aday Üretimi: Sık $(k-1)$ diziler F_{k-1} kümesi verildiğinde, bir sonraki geçiş için adaylar, F_{k-1} 'in kendisiyle birleştirilerek oluşturulur. Bir budama aşaması,

alt dizilerinden en az biri sık olmayan herhangi bir diziyi ortadan kaldırır. Hızlı sayım için, aday diziler bir hash ağacında saklanır.

2. Destek Sayımı: Bir girdi dizisi ε'de yer alan tüm adayları bulmak için, kavramsal olarak ε'nin tüm k alt dizileri oluşturulur. Bu tür her alt dizi için, hash ağacında bir arama yapılır. Karma ağaçtaki bir aday alt diziyle eşleşirse, sayısı artırılır.

4.2.1.2 Tahminleme Modülü

İş sürecindeki son adım veri analizidir. Elde edilen bulguları müşteri elde tutma verilerine uygulayarak müşterinin elde tutulması için hangi sıralı elde tutma yönteminin tercih edilmesi gerektiğini gösterir. Müşteri tutundurma yöntemi tahmini, sıralı kural şeklinde (ara/mesaj/..vb) tahminlenmesi yapılmaktadır.

5
PROTOTİP UYGULAMA

Tez kapsamı süresince terk etme davranışı ve tutundurma iş süreci yazılım mimarisinin prototip uygulmasının yapılabilmesi için oluşturulan modüllerde birden çok altyapı ve yazılım dili kullanılmıştır.

Tez süresince terk etme davranışı ve tutundurma iş süreci yazılım mimarisi üzerinde, performans ve maliyetten kazanım sağlayacak teknoloji ve programla dilleri kullanılarak uygulamalar geliştirilmiştir. Uygulama geliştirilme sürecinde oluşurulan modüllerde çalışmaya uygun programlama dili, performans, maliyet, geliştirilebilirlik gibi unsurlar göz önünde bulundurularak testler yapılmıştır.

5.1 Teknoloji açısından Prototip Uygulamanın sınıflandırılması

5.1.1 Müşteri Terk Etme Eğilimi Tahmin Modeli Verisinin Ön İşleme Modülünde Kullanılacak Teknolojiler

Müşteri terketme eğilimi tahmin modeli ve müşteri tutundurma verisinin ön işlemisinin gerçeklenebilmesi için Spyder 4.1.4 kullanılmıştır ve başarılı bir şekilde önişleme gerçekleştirilmiştir.

5.1.1.1 Spyder Anaconda platformu

Spyder, Anaconda Navigatorün geliştirildiği ortam olup Python dilinde kod yazılmasını sağlayan açık kaynaklı basit arayüzlü bir ortamdır. Spyder, bilim adamları, veri analistleri ve mühendisler için tasarlanmıştır, ancak öğrenciler ve Python geliştiricileriyle bağımsız olarak çalışan herkes tarafından da kullanılabilir. Spyder, gelişmiş düzenleme, etkileşimli test, hata ayıklama ve iç gözlem yeteneklerine sahip sağlam bir etkileşimli Python programlama ve IPython uyumluluğu ve NumPy, SciPy ve matplotlib gibi popüler Python kitaplıkları sayesinde sayısal bir hesaplama ortamıdır.

Spyderın kullanımı kolaydı, hata yapılan yeri ve hata yapılma sebebini gösterir. Spyder

fonksiyonları tamamlar ve pythondaki işlemler otomatik olarak renklendirilerek yazılır, bu hata yapma ihtimalimizi en aza indirir. Bu gibi sebeplerden dolayı müşteri terk etme eğilimi tahmin modeli verisinin ön İşleme modülünde Spyder platformu tercih edilmiştir [67].

5.1.2 Müşteri Terk Etme Eğilimi Tahmin Modeli Olusturulması ve Tahmini ve Müşteri Tutundurma İş Süreci Yazılım Mimarisinde Kullanılacak Teknolojiler

Müşteri terk etme eğilimi tahmini modeli oluşturulması ve tahmini ve müşteri tutundurma iş süreci yazım mimarisi modullerin ölçeklenebilirlik testi için büyük veri işleme platformu olan Apache Spark veri işleme motorunun spark3.0.0 versiyonu kullanılmıştır. Hiperparametre optimizasyonu ve tahmininde ise PyCharm 2020.2 versiyonunda geliştirildi.

5.1.2.1 Pycharm platformu

Pycharm Python dili için bilgisayar programlasında kullanılan ve Window,MacOs ve Linux sürümleri ile çapraz bir platformdur. Akıllı bir kod editörü, akıllı kod navigasyonu, hızlı ve güvenli yeniden düzenleme ile birlikte gelir. PyCharm, hata ayıklama, test etme, profil oluşturma, dağıtımlar, uzaktan geliştirme ve veritabanı araçları gibi özelliklerle entegredir [68].

Pycharm Platformunun Avantajları:

- Otomatik kod tamamlama, hata tespiti, hızlı sabitleme vb. konularında onlara yardımcı olan geliştiricilere akıllı bir platform sağlar.
- Birçok maliyet tasarrufu faktörünü artırarak multiple framework desteği sağlar.
- Platformlar arası geliştirme gibi zengin bir özelliği destekler, böylece geliştiriciler farklı platformlarda da bir komut dosyası yazabilir.
- PyCharm ayrıca özelleştirilebilir arayüzün iyi bir özelliğiyle birlikte gelir ve bu da üretkenliği artırır.

Pycharm Platformunun Dezavantajları:

- PyCharm, istemciye sağladığı özellikler ve araçlar göz önüne alındığında biraz pahalı bir araçtır.

- İlk kurulum zordur ve bazen arada kalabilir fakat kurulum adımları takip edilerek kolayca halledilebilir.

5.1.2.2 Apache Spark kütüphanesi

Apache Spark, büyük veri kümeleri üzerinde paralel olarak işlem yapılmasını sağlayan, Scala dili ile geliştirilmiş açık kaynak kodlu kütüphanedir. Sk bazlı çalışma yapısına sahip olan "MapReduce"un oluşturduğu performans maliyetlerinin çözümüyle ortaya çıkan Spark, bellek içi veri işleme özelliğiyle büyük veri uygulamalarında Apache Hadoop'tan daha hızlı çalışabilmektedir. Dolayısıyla verileri işlerken diskten veri okumadan veya diske veri yazmadan bellekte tutmasıyla Spark'ın analitik motorunun daha hızlı işlemler yapabildiğini söylemek mümkündür. Bellek içi veri işleme, Spark'ın temel programlama özeti olan "RDD (Resilient Distributed Datasets)"ler ile yapılmaktadır. RDD'ler elemanlar üzerinde paralel olarak işlemlerin yapılmasını sağlayan, makineler arasında bölünmüş veri koleksiyonlarıdır. RDD'ler oluşturularak diskteki veriler, geçici belleğe taşınmaktadır. Bu RDD'ler üzerinde yapılacak senaryolara göre sorgular yazılarak veriler işlenmektedir [67].

5.2 Kullanılan yöntem açısından Prototip Uygulamanın sınıflandırılması

5.2.1 Müşteri Terk Etme Eğilimi Tahmin Modeli Verisinin Ön İşleme Modülünde Kullanılacak Yöntemler

Müşteri terk etme eğilimi tahmin modeli verisinin ön işleme modülünde veri önişlemesi yapılırken kendi içerisinde 5 farklı yöntem kullanılmaktadır. Bu yöntemler Data Tipleri Kontrolü, Benzersiz Öznitelikler Kontrolü, Boş Kayıtlar Kontrolü, Aykırı Değer Kontrolü ve Tekrarlayan Kayıt Kontrolüdür. Bu beş farklı veri ön işleme yöntemleri sonucu olarak veriyi işlemeden önce ham data 120 öznitelik ve 591662 satır barındıran veri seti, veri işlendikten sonra makine öğrenmesi algoritmaları uygulanacak nihai dataset 103 öznitelik ve 563423 satır haline gelmektedir. Bu veri seti daha doğru sonuçlar vermektedir.

5.2.1.1 Data Tip Kontrolü için Kullanılacak AsType Yöntemi

Bazı durumlarda sayısal değerler string, string değerler de sayısal değer olarak gözükebilmektedir. Bu durum hatalara yol açmaktadır. Bu hatanın önlenebilmesi için astype methodu kullanılarak veri istenilen veri tipine dönüştürülür. Müşteri terk etme eğilimi tahmin modeli verisinin ön işleme modülünde astype methodu kullanılarak

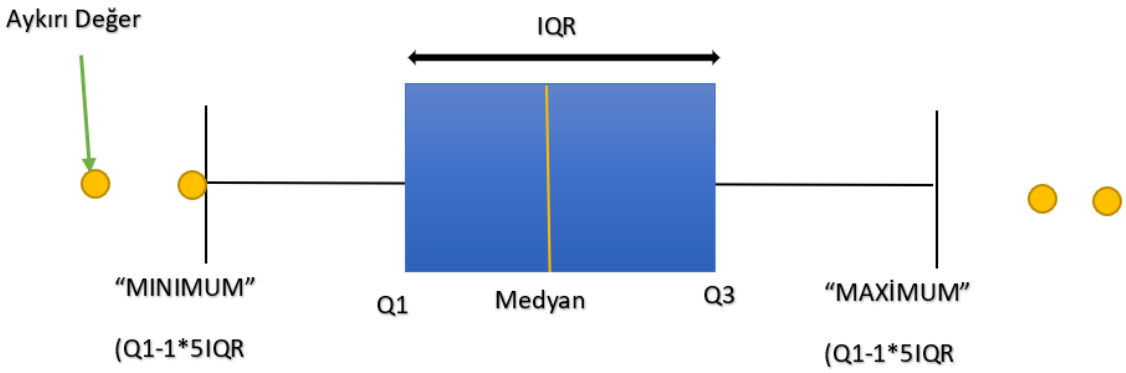
bazı kategorik değişkenler sayısal değere, bazı sayısal değişkenler kategorik değişkene dönüştürülmüştür.

5.2.1.2 Boş Kayıt Kontrolü Kontrolü için Kullanılacak Isnull Yöntemi

Veri setinde bazı özelliklere ait kayıtlar boş olabilmekte bu müşteri terketme davranışı tahmininde başarılı sonuçlar almayı engellemektedir. Isnull methodu ile veri setinde kaç tane boş kayıt bulunduğunu belirlenmiştir. Bu boş kayıtlar veri setinden silinerek başarılı sonuçlar alınması sağlanmıştır.

5.2.1.3 Aykırı Kontrolü Kontrolü için Kullanılacak Kutu Analizi Yöntemi

Bir veri setindeki gözlemlerin geri kalanından büyük ölçüde farklı olan herhangi bir veri noktası aykırı değer olarak isimlendirilmektedir. Aykırı değerler tespit edilmeli ve bu aykırı değerin yok edilmesi için bir takım işlemler yapılmalıdır. Müşteri terketme davranışı tahmininde kullanılacak verisetinde aykırı değerler tespit edilmiştir. Bu aykırı değerlerin yok edilmesi için kutu analizi yöntemi kullanılmıştır. Kutu Analizi için gerekli parametreler Şekil-5.1 gösterilmiştir.



Şekil 5.1 Kutu Analizi Yöntemi Şeması

5.2.1.4 Boş Kayıt Kontrolü Kontrolü için Kullanılacak Duplicated Yöntemi

Bazı durumlarda bazı kayıtlar tekrarlanabilir durumdadır. Hataya yol açtığından dolayı bu kayıtların bulunup silinmesi gerekmektedir. Müşteri terk etme eğilimi tahmin modeli verisinin ön işleme modülünde duplicated methodu kullanılarak tekrar eden kayıtlar bulunmakta ve drop duplicated methoduyla da tekrar eden kayıtlar silinmektedir.

5.2.2 Müşteri Terk Etme Eğilimi Tahmin Modeli Oluşturulması ve Tahmininde Kullanılacak Yöntemler

Müşteri terk etme eğilimi tahmin modeli oluşturulması ve tahmininde bayes optimizasyonu ve makine öğrenme algoritmaları kullanılmıştır. Bu Bölüm-4.1.3 ve Bölüm-4.1.2.3'de detaylıca anlatılmıştır.

5.2.3 Müşteri Tutundurma İş Süreci Yazılım Mimarisinde Kullanılacak Yöntemler

Müşteri tutundurma iş süreci yazılım mimarisinde sıralı madencilik algoritmaları kullanılmıştır. Bu Bölüm-4.3.1.1'de detaylıca anlatılmıştır.

6
ÖNERİLEN YÖNTEMİN DEĞERLENDİRİLMESİ

6.1 Kullanılacak araçlar

Apache Spark, büyük veri kümelerinin paralel olarak işlenmesini sağlayan Scala programlama dilinde yazılmış açık kaynaklı bir kütüphanedir. Spark, mapreduce'un disk tabanlı çalışma yapısından kaynaklanan performans yükünü ortadan kaldıran bellek içi veri işleme özelliği sayesinde büyük veri uygulamalarında Apache Hadoop'tan daha hızlı çalışabilir. Sonuç olarak, Spark'ın analitik motorunun, verileri işlerken diske erişmek veya diske yazmak yerine bellekte depolayarak daha hızlı işlemler gerçekleştirebileceğini iddia etmek mantıklıdır. Ölçeklenebilirlik testi için büyük bir veri işleme platformu olan Apache Spark 3.0.0 veri işleme motoru kullanıldı.

6.1.1 Teknoloji ve Altyapı

Spyder, Python geliştirme için Python tabanlı açık kaynaklı bir Ide'dir, PyCharm ise bilgisayar programlama, özellikle Python dili için entegre bir geliştirme ortamıdır. Müşteri terk etme davranışı verileri ve müşteri tutundurma verilerinin önceden işlemesi için Spider sürüm 4.1.4'te , hiperparametre optimizasyonu ve tahmini ve müşteri tutundurma iş süreci yazılım mimarisi geliştirmek için PyCharm sürüm 2020.2 kullanıldı. 64 GB RAM ve iki işlemci çekirdeği olan bir sunucuda testler çalıştırıldı.

6.1.2 Kontrol Veri Seti

Müşteri terk etme davranışının ve müşteri tutundurma iş süreci yazılım mimarisinde kullanılan veri kümesi, çeşitli müşteri ilişkileri veritabanlarından gelen bilgilerin birleştirilmesiyle oluşturulmuştur. 2019'dan 2020'ye kadar yaklaşık bir buçuk yıllık veri bulunmaktadır. Bu veriler 76364 sözleşmeli tüketici hakkında bilgiler içermektedir.

Müşteri terketme davranışı tahminin için periyodik olarak üretilen veri koleksiyonunda Eylül Ocak 2019 - Eylül 2020 tarihleri arasında toplam 563423

verisi satırı bulunmaktadır. Veri koleksiyonunda toplam 120 öznitelik bulundu.

Müşteri tutundurma iş süreci yazılım mimarisinde için kullanılan veri setinde ise 17 olay (müşteri ara, müşteriye SMS, müşteriye e-posta gönderme, müşteriye yerinde ziyaret ve indirim teklifi vb.) ve 989.818 satır bulunmaktadır.

6.2 Test - 1: Müşterinin terketme davranışından önce kalmasını sağlayacak durumlar tespitinin tahminde en başarılı sonuçlar verecek özniteliklerin bulunması

Geliştirilen uygulama içerisinde en yüksek tahmin gücüne sahip öznitelikler bilgi kazancı metriğine bakılarak karar verilmiştir. Tüm özniteliklerin özniteliklerin bilgi kazancı değerleri hesaplanmıştır. Ve tüm özniteliklerin bilgi kazancı değerleri karşılaştırılmıştır. Bu hesaplama yöntemi, Bölüm 4.4.1'de detaylı olarak anlatılmaktadır. En yüksek tahmin gücüne sahip özniteliklerin bazıları Tablo 6.1'de gösterilmektedir.

Tablo 6.1 En Önemli Öznitelikler

Bilgi Kazancı	Öznitelik
0.41997081682535475	ANLASMA _FL _DURUM
0.3838990425225518	MUSTERI _DONEM
0.3838990425225518	OPERASYON _DONEM
0.28628699982933736	FL_ANA _ANLASMAYI _KULLAN
0.2662831003955598	RF_ANL_BOLGE _KODU
0.21630655612641525	LU_ANLASMA_TUR
0.20808995897344718	LU_SEKTOR
0.19796376116814485	LU_ANLASMA_KATEGORI
0.1090970817988397	FL_MERKEZI_ODEME
0.08266661315292767	FL_AMBALAJ_SATIS_YOK
0.06976202709416013	RF_SUBE_KOD
0.06676972990616103	RF_SEGMENT_KOD

6.3 Test-2: En başarılı sonuçlar verecek özniteliklerin oluşturulmasından sonra başarı oranlarının verilmesi

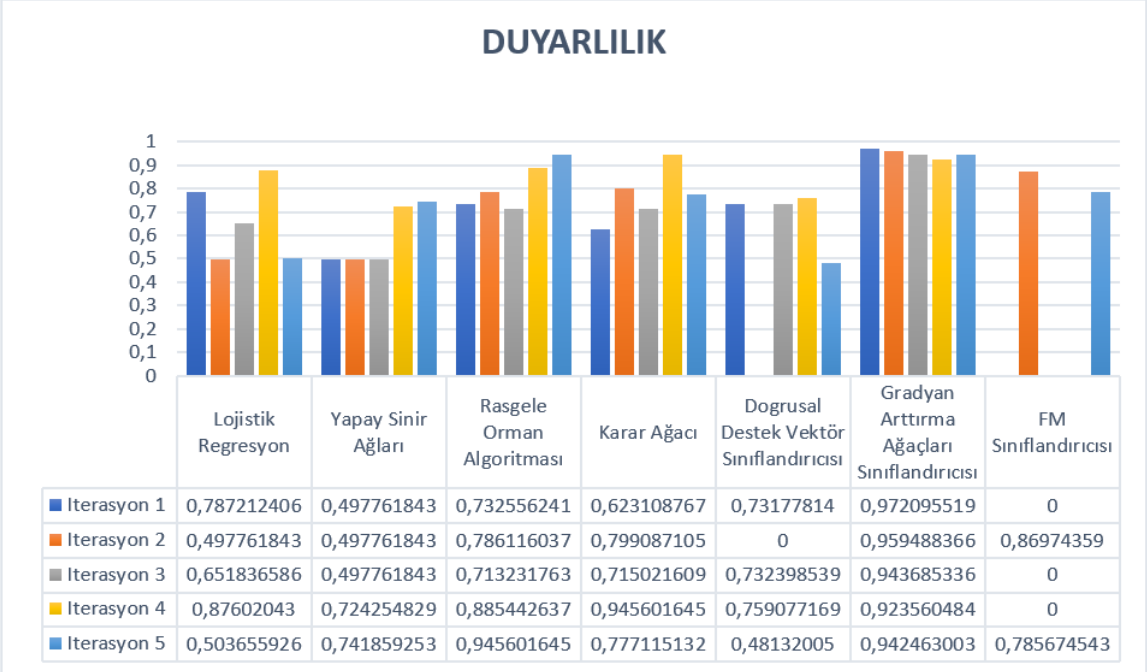
Bu bölümde, önerilen yöntemin doğruluğunu tespit yönteminde kullanılacak algoritmalarla test edilecektir. İlk olarak öncesi veri ön işleme aşamaları uygulanarak, verinin makine öğrenmesine hazır hale gelmesi sağlanmıştır. Davranışı matrisi veriseti üzerinde çalıştırılan, müşteri terk etme eğilimi tahmin modeli oluşturma süreci sonrasında modeller oluşturulmuş ve müşteri terk etme davranışı tahmininde başarısı

irdelenmiştir. Hangi algoritmanın daha başarılı olduğunun tespiti için karmaşıklık matrisi hesaplanmalıdır. Karmaşıklık matrisi kullanılarak oluşturulan ve doğruluk, kesinlik, duyarlılık ve f1 scoredan oluşan dört test metriği bulunmaktadır. Bu metriklerin hesaplanması oluşturulan karmaşıklık matrisi Tablo 6.2'de verilmiştir.

Tablo 6.2 Karmaşıklık Matrisi

	Gerçek Pozitif	Gerçek Negatif
Tahmini Pozitif	Gerçek Pozitif(TP)	Yanlış Pozitif(FP)
Tahmini Negatif	Yanlış Negatif (FN)	Doğru Negatif(TN)

6.3.1 Duyarlılık(Recall)



	Lojistik Regresyon	Yapay Sinir Ağları	Rasgele Orman Algoritması	Karar Ağacı	Dogrusal Destek Vektör Sınıflandırıcısı	Gradyan Arttırma Ağaçları Sınıflandırıcısı	FM Sınıflandırıcısı
Iterasyon 1	0,787212406	0,497761843	0,732556241	0,623108767	0,73177814	0,972095519	0
Iterasyon 2	0,497761843	0,497761843	0,786116037	0,799087105	0	0,959488366	0,86974359
Iterasyon 3	0,651836586	0,497761843	0,713231763	0,715021609	0,732398539	0,943685336	0
Iterasyon 4	0,87602043	0,724254829	0,885442637	0,945601645	0,759077169	0,923560484	0
Iterasyon 5	0,503655926	0,741859253	0,945601645	0,777115132	0,48132005	0,942463003	0,785674543

Şekil 6.1 Duyarlılık için bar grafiği

Duyarlılık(Recall) Positive olarak tahmin edilmesi gereken durumların kaç tanesinin positive olarak tahmin ettiğimizi gösteren bir metriktir. Duyarlılık değeri FP oranının yüksek olduğu durumlarda değerlendirme için yardımcı olan bir metriktir. Duyarlılık değerinin yüksek olması gerekir. Duyarlılık hesabı aşağıda verilmiştir.

$$Duyarlilik = \frac{TP}{TP + FN} \tag{6.1}$$

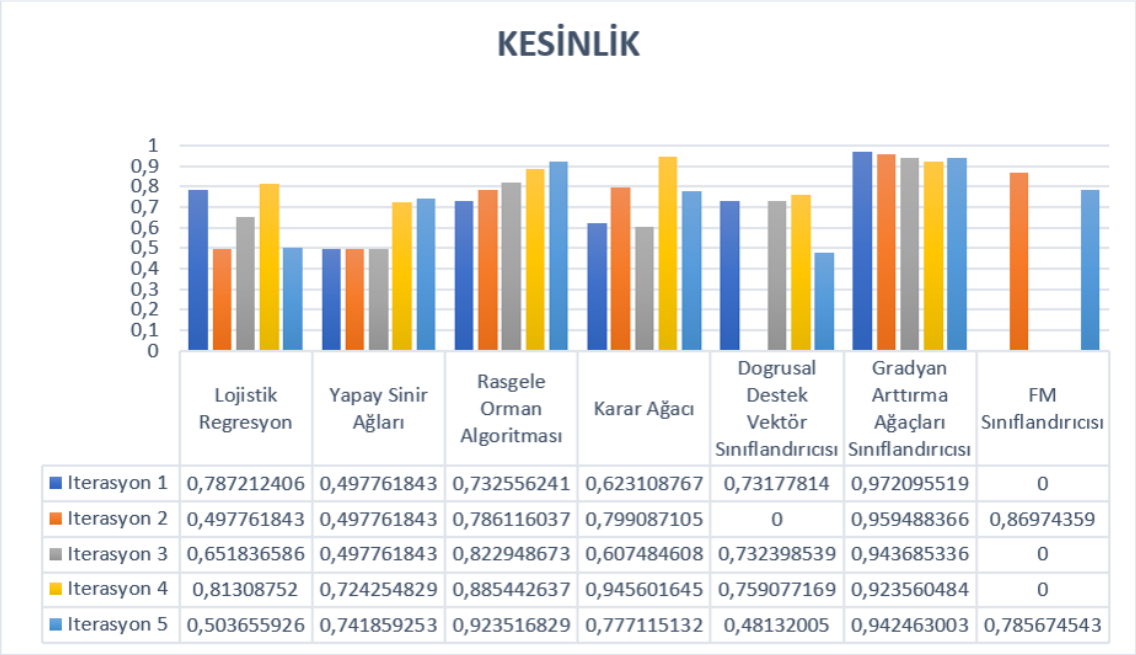
Kullanılan bu hesaplamayla başarılı sonuç için kullanılan algoritmaların duyarlılık değerleri Şekil- 6.1'de verilmiştir.

6.3.2 Kesinlik(Precision)

Kesinlik(Precision) değeri False Positive tahminlemenin yüksek olduğu durumlarda önemli olup hangi modelin kullanılacağı konusunda önemli bilgi vermektedir. Kesinlik değerinin yüksek olması gerekir. Kesinlik hesabı aşağıda verilmiştir.

$$Kesinlik = \frac{TP}{TP + FP} \tag{6.2}$$

Kullanılan bu hesaplamayla başarılı sonuç için kullanılan algoritmaların Kesinlik değerleri Şekil 6.2'de verilmiştir.



	Lojistik Regresyon	Yapay Sinir Ağları	Rasgele Orman Algoritması	Karar Ağacı	Dogrusal Destek Vektör Sınıflandırıcısı	Gradyan Arttırma Ağaçları Sınıflandırıcısı	FM Sınıflandırıcısı
Iterasyon 1	0,787212406	0,497761843	0,732556241	0,623108767	0,73177814	0,972095519	0
Iterasyon 2	0,497761843	0,497761843	0,786116037	0,799087105	0	0,959488366	0,86974359
Iterasyon 3	0,651836586	0,497761843	0,822948673	0,607484608	0,732398539	0,943685336	0
Iterasyon 4	0,81308752	0,724254829	0,885442637	0,945601645	0,759077169	0,923560484	0
Iterasyon 5	0,503655926	0,741859253	0,923516829	0,777115132	0,48132005	0,942463003	0,785674543

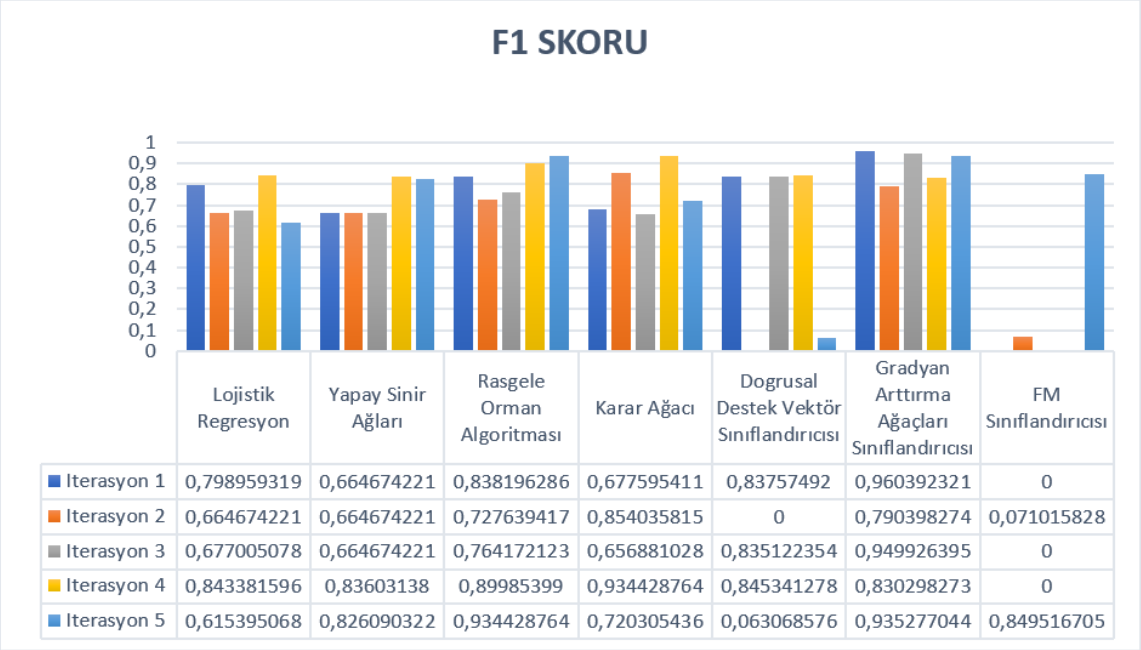
Şekil 6.2 Kesinlik için bar grafiği

6.3.3 F1 Skor(F1 Score)

F1 Score kesinlik ve duyarlılık değerlerinin harmonik ortalamasıdır. Doğruluk yerine F1 Score kullanılmasının temel nedeni eşit dağılmayan veri kümelerinde doğru model seçimi yapmaktır. Tüm hata maaliyetlerini içerecek bir ölçme metriğidir. F1 Score hesaplanması için öncelikle duyarlılık ve kesinlik değerlerinin hesaplanması gerekmektedir. F1 Score hesabı aşağıda verilmiştir.

$$F1Score = 2 \cdot \frac{Duyarlilik \cdot Kesinlik}{Duyarlilik + Kesinlik} \tag{6.3}$$

Kullanılan bu hesaplamayla başarılı sonuç için kullanılan algoritmaların F1 Score

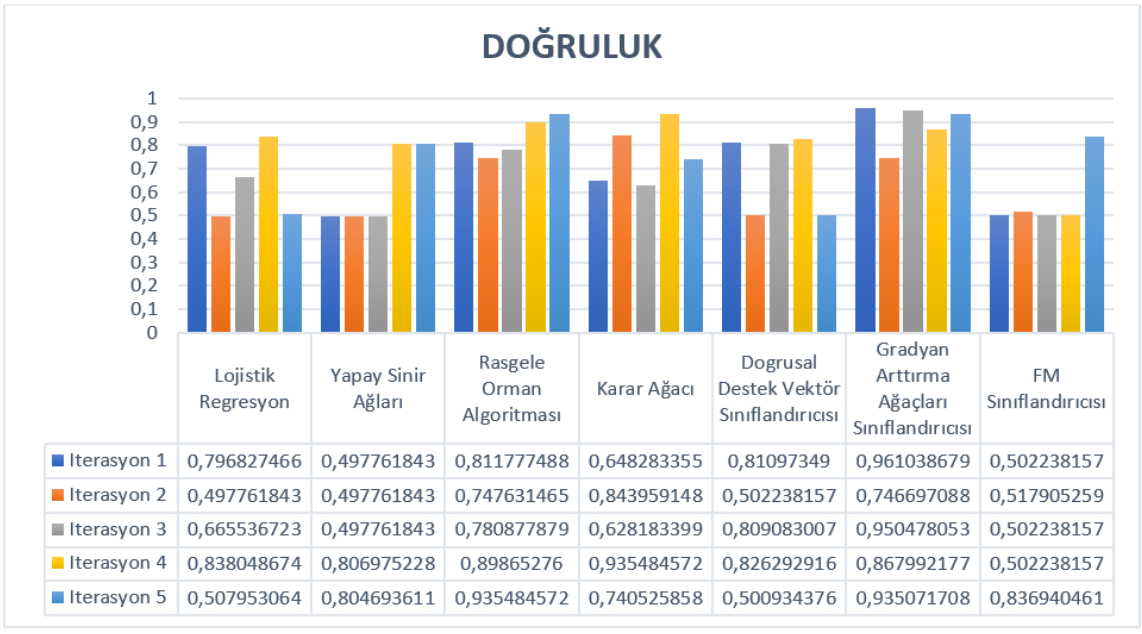


	Lojistik Regresyon	Yapay Sinir Ağları	Rasgele Orman Algoritması	Karar Ağacı	Dogrusal Destek Vektör Sınıflandırıcısı	Gradyan Arttırma Ağaçları Sınıflandırıcısı	FM Sınıflandırıcısı
Iterasyon 1	0,798959319	0,664674221	0,838196286	0,677595411	0,83757492	0,960392321	0
Iterasyon 2	0,664674221	0,664674221	0,727639417	0,854035815	0	0,790398274	0,071015828
Iterasyon 3	0,677005078	0,664674221	0,764172123	0,656881028	0,835122354	0,949926395	0
Iterasyon 4	0,843381596	0,83603138	0,89985399	0,934428764	0,845341278	0,830298273	0
Iterasyon 5	0,615395068	0,826090322	0,934428764	0,720305436	0,063068576	0,935277044	0,849516705

Şekil 6.3 F1 Score için bar grafiği

değerleri Figure 6.3'de verilmiştir.

6.3.4 Doğruluk(Accuarcy)



	Lojistik Regresyon	Yapay Sinir Ağları	Rasgele Orman Algoritması	Karar Ağacı	Dogrusal Destek Vektör Sınıflandırıcısı	Gradyan Arttırma Ağaçları Sınıflandırıcısı	FM Sınıflandırıcısı
Iterasyon 1	0,796827466	0,497761843	0,811777488	0,648283355	0,81097349	0,961038679	0,502238157
Iterasyon 2	0,497761843	0,497761843	0,747631465	0,843959148	0,502238157	0,746697088	0,517905259
Iterasyon 3	0,665536723	0,497761843	0,780877879	0,628183399	0,809083007	0,950478053	0,502238157
Iterasyon 4	0,838048674	0,806975228	0,89865276	0,935484572	0,826292916	0,867992177	0,502238157
Iterasyon 5	0,507953064	0,804693611	0,935484572	0,740525858	0,500934376	0,935071708	0,836940461

Şekil 6.4 Doğruluk için bar grafiği

Doğruluk(Accuarcy) toplam doğru tahmin ettiğimiz alanları ifade etmektedir. Doğruluk hesabı aşağıda verilmiştir.

$$Dogruluk = \frac{TP}{TP + FP + TN + FN} \tag{6.4}$$

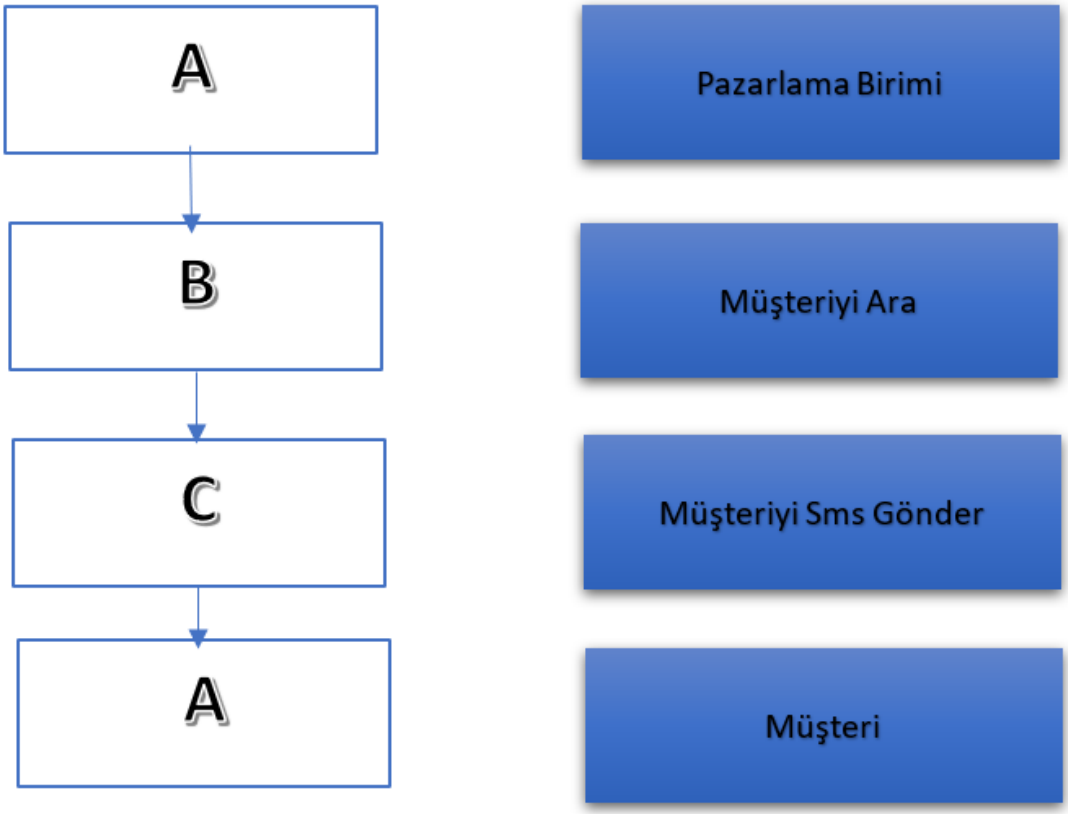
Kullanılan bu hesaplamayla başarılı sonuç için kullanılan algoritmaların doğruluk değerleri Şekil 6.4'de verilmiştir.

6.4 Test - 3: Müşterinin tutundurma yöntemi tahminde en başarılı sonuçlar verecek yöntemin ve bu yöntemin çalışma süresinin bulunması

Metodoloji bölümünde belirtilen sistem mimarisini uygulamak ve performansını değerlendirmek için bir prototip uygulama geliştirilmiştir. Geliştirilen sistem, pazarlama departmanından elde edilen veriler üzerinde çalışmaktadır. Bu sistem, müşteriyi aramak, müşteriye SMS göndermek, müşteriyi elinde tutmak için müşteriye e-posta göndermek gibi işlemleri içermektedir. Pazarlama departmanı, birçok kanal aracılığıyla birden fazla müşteri türü ile iletişim kurarak verileri oluşturmuştur. Veri Ön İşleme Modülü, toplanan bu pazarlama departmanı günlük verilerini, sıralı kural madenciliği algoritmalarının okuyabileceği standart bir giriş formatına dönüştürmek için geliştirilmiştir. Pazarlama departmanı günlük verileri, müşteri tutundurma yöntemi tahmin modülü tarafından toplu olarak işlenir.

Müşteri modellerinin sıralı davranışının elde edildiği yöntem, müşteri tutundurma yöntemi tahmin modülünün en önemli bileşenidir. Bu araştırmanın nihai amacı, müşterilerin sözleşme iptallerini azaltmaktır. İlk olarak, müşterinin türünden bağımsız olarak, bu hedefi gerçekleştirmek için müşteriye uygulanan eylem verilerinden sıralı ilişkilendirme kuralları türetilir. Müşterinin aranmasından kendisine uygulanan promosyona kadar olan adımlar, müşteri tutma yöntemi tahminleme modülünde birer birer kararlaştırılır. Aşağıdaki resim, müşteriyi elde tutma yöntemi tahmin modülünün tanımlayabileceği örnek bir dizi faaliyeti göstermektedir.

Bir pazarlama biriminin yaptıkları Şekil-6.5'te gösterilmektedir. Müşteriyle iletişim kuran pazarlama biriminin, müşteriyle temasa geçmeye çalıştığı gösterilmektedir. Bu eylemler dizisi sistem tarafından kaydedilir. Bu şekilde elde edilen tüm tüketicilerin davranışları kontrol edilir ve müşteri tutundurma yöntemi tahmin modülünde karşılaştırılabilir dizilerin frekansları belirlenerek frekans değerlerine göre sıralanır. Müşteriyi elde tutma yöntemi bu şekilde hesaplanır. Bir müşteri tutundurma yöntemi tahmin modülü geliştirmek için Apache Spark kullanıldı. txt formatındaki metin dosyası, SRM modülü tarafından üretilen tüm kuralları içerir. Müşteriyi tutundurma yöntemi tahmin modülü ile proje kapsamında oluşturulan sıralı



Şekil 6.5 Müşterilerin Oluşturabileceği Örnek Tipik Sıralı Eylem

ilişkilendirme kurallarının elde edilmesine yönelik algoritmaların oluşturulmasında Python dili kullanılmaktadır. Projenin müşteri tutundurma yöntemi tahmin modülü, Python'da yazılan ve veriler üzerinde Apache Spark tarafından yürütülen sıralı ilişkilendirme kuralları algoritmalarını oluşturmak için kullanılır.

ALGORİTMA	3 MB	13 MB	23 MB
PREFİX SPAN	1,3687218s Std. 0,0674	9,09405192s Std. 0,4820	24,36332334s Std. 6,5769
GSP	0,76856754s Std. 0,1623	1,69631232s Std. 0,1318	1,0052s Std. 0,05

Şekil 6.6 Farklı setlerden elde edilen test sonuçlarının karşılaştırılması farklı boyut verileri

Müşteri tutundurma yöntemi tahmin modülünün performansı, farklı veri boyutları üzerinde test edilmiştir. Spark tarafından geliştirilen modül, artan veri boyutuna bağlı olarak çalışma süreleri kayıt altına alınarak değerlendirilmektedir. Testler, 64 GB RAM ve iki işlemci çekirdeğine sahip bir sunucu üzerinde gerçekleştirilmiştir. SRM algoritmaları modülünün sağladığı üç farklı algoritma için testler tekrarlanır. En başarılı algoritma testi sonuçları Şekil-6.6'de gösterilmektedir. SRM algoritmaları modülünün müşteriyi elde tutma davranışı verileri yoluyla sıralı ilişkilendirme

kuralları elde etme hızı test edilerek veri boyutu katlanarak 1 MB'tan 25 MB'a çıkarılmıştır.

Her bir veri boyutu için bu işlem beşer kez yapılarak elde edilen sonuçların ortalaması ve standart sapması kaydedildi. Oluşturulan modülün yüksek performansla çalıştığı, müşteri tutundurma yöntemi tahmin modülünün gün aşırı çalıştığından anlaşılabilir. Her belirli algoritmanın sonuçları ve ölçeklenebilirliği Şekil- 6.6'de gösterilmektedir.

6.5 Değerlendirme

Bu bölümde, önerilen müşteri terketme davranışının tahmini ve müşteri tutundurma yöntemi tahmininin ne oranda başarılı olduğu irdelenmektedir. Müşteri terketme davranışının tahmininin değerlendirme aşamasında algoritmalardan alınan sonuçların f1 score ve doğruluk değerleri karşılaştırılarak başarı oranları tespit edimiştir. Bölüm 6.2'de yapılan testler sonucunda en yüksek bilgi kazancına sahip özniteliğin MUSTERI_DONEM olduğu belirlenmiştir. MUSTERI_takip eden iyi öznitelik ise 6.1'de gösterildiği gibi ANLASMA_FL_özniteliğidir. Sınıflandırıcı algoritmalar olarak; Lojistik Regresyon,Karar Ağaçları, Yapay Sinir Ağları, Gradyan Arttırma Ağaçları Sınıflandırıcısı (Gradient-Boosted Trees Classifier), Doğrusal SVC,Rasgele Orman Sınıflandırıcısı ve FM Sınıflandırıcısı algoritmaları kullanılmıştır. Veri seti üzerinde, sınıflandırıcılar tarafından oluşturulmuş modellerin performansı, doğruluk metriği açısından incelenmiştir.

Bölüm 6.3'de bayes optimizasyon metodu kullanılarak hangi algoritmaların en uygun parametrelerle en başarılı sonuçlar verdiği araştırılmaktadır. Bunun için Şekil- 6.1'de en yüksek duyarlılık oranına sahip algoritmanın Karar Ağacı ve GBT Sınıflandırıcısı algoritması, Şekil-6.2'de en yüksek kesinlik değerinin Rasgele Orman Sınıflandırıcısı ve GBT Sınıflandırıcısı algoritması,Şekil-6.3'de en yüksek f1 score değerinin Karar Ağacı ve GBT Sınıflandırıcısı algoritması, Şekil-6.4'de en yüksek doğruluk değerinin GBT Sınıflandırıcısı ve Rasgele Orman Sınıflandırıcısı olduğu görülmektedir. Yapılan tüm bu testlerin sonucunda, en yüksek doğruluk oranını veren algoritmanın GBT Sınıflandırıcısı algortiması olduğu görülmüştür. Belirlenen öznitelik ve algoritmalar ile müşteri terketme tahmininde %96 oranında başarı sağlamaktadır.

Müşteri tutundurma tahmininin değerlendirme aşamasında algoritmaların çalışma süreleri karşılaştırılarak başarı oranları tespit edimiştir. Sıralı madencilik algoritmalar olarak; Prefix Span,Spade ve Gsp algoritmaları kullanılmıştır. Veri seti üzerinde, algoritmalar tarafından sıralı modellerin performansı, çalışma süresi metriği açısından incelenmiştir.

Bölüm 6.4'de hangi algoritmaların en başarılı sonuçlar verdiği araştırılmaktadır. SRM algoritmaları modülünün perfomansı farklı veri boyutlarında test edilmiştir. Şekil-6.6'de tüm veri boyutlarında en düşük standart sapma ve ortalama çalışma süresine sahip algoritmanın GSP algoritması olduğu görülmüştür.

7
SONUÇ VE ÖNERİLER

Bu çalışma kapsamında, kargo ve lojistik endüstrisinde müşteri terketme davranışını tahmin etme ve sorununa cevap vermek ve müşteri tutundurma yöntemi tahmin probleminin çözümü için bir yazılım mimarisi önerisi yapılmıştır. Büyük veri işleme ve analiz teknolojilerini kullanarak, önerilen yazılım mimarisi, kargo ve lojistik endüstrisindeki müşteri terk etme davranışlarının ve müşteri tutma modelinin tahmin edilmesine olanak tanır. Araştırmanın değerlendirilmesinde kargo sektöründe yer alan MNG Kargo şirketinden müşteri davranış veri setleri ve müşteri tutundurma veri setleri kullanılmıştır. Önerilen yazılım mimarisi için Prototip yazılım oluşturuldu. Protatıp yazılımının müşteri terkeme davranışı ve müşteri tutundurma yöntemi tahmin etme yeteneği araştırılmıştır. Elde edilen sonuçlar, önerilen yazılım mimarisinin, başarılı tahminler yapılabilmesine olanak sağlayan, kullanılabilir bir mimari olduğunu göstermektedir. Elde edilen sonuçlar, kargoculuk müşterilerine ait davranış verileri üzerinde, müşteri terk etme davranışı tahmininde, GBT Classifier sınıflandırıcınısının ve müşteri tutundurma yöntemi tahmininde GSP sınıflandırıcısının başarılı sonuçlar verdiğini göstermektedir.

7.1 Araştırma Sorularına Cevaplar

Bölüm-3.1'de ortaya konulan araştırma soruları, Bölüm 6'da yapılan testler sonucunda bulunan bulgular ile takip eden kısımlarda yanıtlanmaktadır.

Kargoculuk ve lojistik sektöründe, müşterilerin terketme davranışını en iyi tahmin edilebilecek bir veri analiz süreci aracının yazılım mimarisi nasıl olabilir?

Bu çalışma kapsamında, kargo sektöründe karşılaşılan müşteri terk etme davranışının tahmini probleminin çözümü amacıyla bir yazılım mimarisi ve optimizasyonu yapısı ile oluşturulan pipeline önerisi yapılmaktadır. Önerilen pipeline mimarisi; büyük veri işleme ve analizi teknolojilerini kullanarak, kargo müşteri terk etme davranışları verileri üzerinde, bir veri analizi iş sürecinin gerçekleştirilebilmesine olanak sağlamaktadır. Belirlenen öznitelik vektörlerinin değerleri her bir anlaşmalı

müşteri için dönemsel olarak hesaplanmış ve tüm anlaşmalı müşteri verilerini temsil eden müşteri terk etme davranışı matrisi verisi oluşturulmuştur. Geliştirilen prototip uygulama kapsamında, makine öğrenmesi uygulama öncesi ilk ön işleme ve makine öğrenmesi öncesi ikinci veri ön işleme aşamaları uygulanarak, verinin makine öğrenmesine hazır hale gelmesi sağlanmıştır. Müşteri terk etme davranışı matrisi üzerinde çalıştırılan hiperparametrik optimizasyonu yapısı ile oluşturulan pipeline sonrasında gözetimli makine öğrenmesi yöntemiyle modeller oluşturulmuş ve müşteri terk etme davranışı tahmininde başarısı irdelenmiştir. Bölüm 4.1'de en iyi tahmin edilebilecek bir veri analiz süreci aracının yazılım mimarisi anlatılmaktadır.

Kargoculuk ve lojistik sektöründe, müşteri terketme davranışında kullanılabilecek, müşteriyi terketmeden önce kalmasını sağlayabilecek, tahmin gücü en başarılı öznitelikler nelerdir?

Bu tez kapsamında müşteri terketme davranışının tahmini için 120 tane öznitelik önerilmektedir. Bölüm 6.2'de verilen testler sonucunda seçilmiş olan özniteliklerin %96'in üzerinde doğruluk oranına ulaştığı gösterilmektedir.

Kargoculuk ve lojistik sektöründe, müşterinin terketme davranışı tahmininde kullanılmasını önerdiğimiz öznitelikler ile başarılı tahminler yapılabilir mi?

Geliştirilen yazılım mimarisiyle ile yapılan testler sonucunda, müşteri terketme davranışının tahmini için başarılı tahminler yapılacağı kanıtlanmıştır. Bölüm **??**'de Test-2 sonucunda kalıtım, rasgele orman, karar ağaçları ve GBT sınıflandırıcısı algoritmalarının %90'in üzerinde doğruluk oranı ile en başarılı sonuçları verdiği görülmektedir.

Kargo ve lojistik sektöründe müşterinin tutundurma yöntemini en iyi tahmin edilebilecek bir veri analiz süreci aracının yazılım mimarisi nasıl olabilir?

Bu tez kapsamında, kargo sektöründe karşılaşılan müşteri tutundurma yöntemi tahmini probleminin çözümü için yazılım mimarisi önerilmektedir. Önerilen yazılım mimarisi; büyük veri işleme ve analizi teknolojilerini kullanarak, kargo müşteri tutundurma verileri üzerinde, bir veri analizi iş sürecinin gerçekleştirebilmesine olanak sağlamaktadır. Müşteri tutundurma veri, müşterinin terketme durumunda kalmasını sağlayacak 17 olaydan oluşan sıralı kurallardan meydana gelmektedir. Herbir sıralı kural bir yöntemi oluşturmaktadır. Geliştirilen prototip uygulama kapsamında, sıralı madencilik algoritmaları uygulama öncesi ilk veri önişleme aşamaları uygulanarak verinin sıralı kural madenciliği algoritmalarına hazır hale gelmesi sağlanmıştır. Müşteri tutundurma yöntemleri üzerinde çalıştırılan algoritmalarla tutundurma yöntemlerin frekanslarına göre sıralanması sağlanmıştır.

Böylece müşteriyi elde tutmak için uygulanabilecek en doğru yöntemler belirlenmiş ve müşteri tutundurma iş süreci yazım mimarisinde sistemin performansı irdelenmiştir. Bölüm 4.2'de en iyi tahmin edilebilecek bir veri analiz süreci aracının yazılım mimarisi anlatılmaktadır

Kargo ve lojistik sektöründe tutundurma yöntemi iş sürecinde hangi algoritmalar kullanılabilir ve önerilen mimari ile başarılı tahminler yapılabilir mi?

Bu tez kapsamında müşteri tutundurma iş süreci yazım mimarisi için 3 tane algoritma önerilmektedir. Bölüm 6.4'de TEST-3 testleri sonucunda seçilmiş olan algoritmalardan GSP algoritmasının daha başarılı tahminler yaptığı gösterilmektedir.

7.2 Gelecekte Planlanan Çalışmalar

Bu çalışmanın gelecek adımlarında her gün eklenen yeni müşteriler için sürekli akan verileri işleyebilecek, veri işleme ve analizi platformu geliştirilecektir. Bunun yanı sıra, farklı hiperparetmetre optimizasyonu algoritmaları da kullanılarak, sistemin başarısı irdelenecektir. Ayrıca iki, dört ve sekiz düğümlü farklı nodelar kullanılarak sistemin başarısı incelenecektir. Müşteri tutundurma yöntemi tahmin modelinde yeni bir algoritma kullanılacak ve Spade ve GSP ile karşılaştırılacaktır.

KAYNAKÇA

[1] B. Huang, M. T. Kechadi, B. Buckley, "Customer churn prediction in telecommunications," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1414–1425, 2012.

[2] N. Lu, H. Lin, J. Lu, G. Zhang, "A customer churn prediction model in telecom industry using boosting," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1659–1665, 2012.

[3] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, K. C. Chatzisavvas, "A comparison of machine learning techniques for customer churn prediction," *Simulation Modelling Practice and Theory*, vol. 55, pp. 1–9, 2015.

[4] K. K. Chen, R.-H. Chiu, C.-T. Chang, "Using beta regression to explore the relationship between service attributes and likelihood of customer retention for the container shipping industry," *Transportation Research Part E: Logistics and Transportation Review*, vol. 104, pp. 1–16, 2017.

[5] A. De Caigny, K. Coussement, K. W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *European Journal of Operational Research*, vol. 269, no. 2, pp. 760–772, 2018.

[6] J. Jose, "Predicting customer retention of an app-based business using supervised machine learning," 2019.

[7] D. J.-F. Jeng, T. Bailey, "Assessing customer retention strategies in mobile telecommunications: Hybrid mcdm approach," *Management Decision*, 2012.

[8] B.-H. Chu, M.-S. Tsai, C.-S. Ho, "Toward a hybrid data mining model for customer retention," *Knowledge-Based Systems*, vol. 20, no. 8, pp. 703–718, 2007.

[9] A. M. AL-Shatnwai, M. Faris, "Predicting customer retention using xgboost and balancing methods," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 7, 2020.

[10] A. Eiben, T. Euverman, W. Kowalczyk, E. Peelen, F. Slisser, J. Wesseling, "Comparing adaptive and traditional techniques for direct marketing," in *Proceedings of the 4th European Congress on Intelligent Techniques and Soft Computing*, 1996, pp. 434–437.

[11] J. Sim, B. Mak, D. Jones, "A model of customer satisfaction and retention for hotels," *Journal of Quality Assurance in Hospitality & Tourism*, vol. 7, no. 3, pp. 1–23, 2006.

[12] Y. Zhao, Y.-C. Cao, X.-Q. Pan, Y. Lu, "Mbbctree algorithm for telecom customer retention model," in *2007 International Conference on Wavelet Analysis and Pattern Recognition*, IEEE, vol. 3, 2007, pp. 1026–1029.

[13] A. H. Karp, "Using logistic regression to predict customer retention," in *Proceedings of the Eleventh Northeast SAS Users Group Conference. http://www. lexjansen. com/nesug/nesug98/solu/p095. pdf*, 1998.

[14] S. F. Sabbeh, "Machine-learning techniques for customer retention: A comparative study," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 2, 2018.

[15] M. Aktas, et al., "Implementing geographical information system grid services to support computational geophysics in a service-oriented environment," in *NASA Earth-Sun System Technology Conference 2005*, 2005.

[16] M. Pierce, et al., "The quakesim project: Web services for managing geophysical data and applications," *Earthquakes: Simulations, Sources and Tsunamis, 635-651*, 2008.

[17] M. A. Nacar, et al., "Vlab: Collaborative grid services and portals to support computational material science," *Concurrency and Computation: Practice and Experience 19 (12), 1717-1728*, 2007.

[18] B. Dundar, et al., "A big data processing framework for self-healing ınternet of things applications," *2021 IEEE International Conference on Big Data (Big Data), 2353-2361*, 2021.

[19] M. S. Aktas, et al., "Implementing geographical information system grid services to support computational geophysics in a service-oriented environment," *NASA Earth-Sun System Technology Conference, University of Maryland, Adelphi, Maryland, 2005*, 2005.

[20] M. S. Aktas, G. C. Fox, M. Pierce, S. Oh, "Xml metadata services," *Concurrency and Computation: Practice and Experience 20 (7), 801-823*, 2008.

[21] M. S. Aktas, M. Pierce, "High-performance hybrid information service architecture," *Concurrency and Computation: Practice and Experience 22 (15), 2095-2123*, 2010.

[22] G. C. Fox, M. S. Aktas, et al., "Real time streaming data grid applications," *Distributed Cooperative Laboratories: Networking, Instrumentation, and Measurements, 253-267, Springer*, 2006.

[23] A. Tufek, et al., "Provenance collection platform for the weather research and forecasting model," *2018 14th International Conference on Semantics, Knowledge and Grids*, 2018.

[24] M. S. Aktas, et al., "Fault tolerant high performance ınformation services for dynamic collections of grid and web services," *Future Generation Computer Systems 23 (3), 317-337*, 2007.

[25] G. Aydin, et al., "Building and applying geographical information system grids," *Concurrency and Computation: Practice and Experience 20 (14), 1653-1695*, 2008.

[26] G. Aydin, et al., "Algorithms and the grid," *Computing and Visualization in Science 12, 115-124*, 2009.

[27] M. Aktas, et al., “Iservo: Implementing the ınternational solid earth research virtual observatory by integrating computational grid and geographical information web services,” *Computational Earthquake Physics: Simulations, Analysis and Infrastructure, Part II, 2281-2296*, 2007.

[28] Y. Uygun, et al., “On the large-scale graph data processing for user ınterface testing in big data science projects,” *2020 IEEE International Conference on Big Data (Big Data), 2049-2056*, 2020.

[29] E. Olmezogullari, M. S. Aktas, “Pattern2vec: Representation of clickstream data sequences for learning user navigational behavior,” *Concurrency and Computation: Practice and Experience 34 (9)*, 2022.

[30] E. Olmezogullari, et al., “Representation of click-stream datasequences for learning user navigational behavior by using embeddings,” *2020 IEEE International Conference on Big Data (Big Data), 3173-3179*, 2020.

[31] M. J. Baeth, M. S. Aktas, “An approach to custom privacy policy violation detection problems using big social provenance data,” *Concurrency and Computation: Practice and Experience 30 (21)*, 2018.

[32] M. J. Baeth, et al., “Detecting misinformation in social networks using provenance data,” *2017 13th International Conference on Semantics, Knowledge and Grids*, 2017.

[33] E. Guveyi, M. Aktas, O. Kalipsiz, “Human factor on software quality: A systematic literature review,” in *ICCSA 2020*, 2020.

[34] M. Kapdan, M. Aktas, M. Yigit, “On the structural code clone detection problem: A survey and software metric based approach,” in *ICCSA 2014*, 2014.

[35] M. Uzun-Per, A. B. Can, A. V. Gurel, M. S. Aktas, “Big data testing framework for recommendation systems in e-science and e-commerce domains,” in *IEEE Big Data 2021*, 2021.

[36] M. S. Aktas, M. Kapdan, “Structural code clone detection methodology using software metrics,” in *IJSEKE vol 26 issue 2*, 2016.

[37] M. Sahinoglu, K. Incki, M. Aktas, “Mobile application verification: A systematic mapping study,” in *ICCSA 2015*, 2015.

[38] M. Oz, et al., “On the use of generative deep learning approaches for generating hidden test scripts,” *International Journal of Software Engineering and Knowledge Engineering 31 (10), 1447-1468*, 2021.

[39] R. Oguz, et al., “Extracting information from large scale graph data: Case study on automated ui testing,” *Euro-Par 2021*, 2021.

[40] I. Erdem, et al., “Test script generation based on hidden markov models learning from user browsing behaviorsg,” *2021 IEEE International Conference on Big Data (Big Data)*, 2021.

[41] D. Nettleton, *Commercial data mining: processing, analysis and modeling for predictive analytics projects*. Elsevier, 2014.

[42] O. Çelik, U. O. Osmanoglu, “Comparing to techniques used in customer churn analysis,” *Journal of Multidisciplinary Developments*, vol. 4, no. 1, pp. 30–38, 2019.

[43] S. E. Seker, "Müşteri kayıp analizi (customer churn analysis)," *YBS Ansiklopedi*, vol. 3, no. 1, pp. 26–29, 2016.

[44] E. Alpaydin, *Introduction to machine learning*. MIT press, 2020.

[45] V. Nasteski, "An overview of the supervised machine learning methods," *Horizons. b*, vol. 4, pp. 51–62, 2017.

[46] Ç. Elmas, *Yapay Zeka Uygulamaları, Yapay Sinir Ağları – Bulanık Mantık–Genetik Algoritma*. Seçkin Yayinevi, 2012.

[47] Ç. Elmas, "Yapay zeka uygulamaları (yapay sinir ağları-bulanık mantık-genetik algoritma)(2. baskı)," *Seçkin Yayıncılık*, 2011.

[48] S. Sharma, S. Sharma, A. Athaiya, "Activation functions in neural networks," *towards data science*, vol. 6, no. 12, pp. 310–316, 2017.

[49] E. W. Weisstein, "Heaviside step function," *https://mathworld. wolfram. com/*, 2002.

[50] J. Han, C. Moraga, "The influence of the sigmoid function parameters on the speed of backpropagation learning," in *International workshop on artificial neural networks*, Springer, 1995, pp. 195–201.

[51] A. D. Rasamoelina, F. Adjailia, P. Sinčák, "A review of activation function for artificial neural network," in *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, IEEE, 2020, pp. 281–286.

[52] J. Cramer, "The logit model: An ıntroduction," *Edward Arnold*, 1991.

[53] T. Mutanen, "Customer churn analysis–a case study," *Journal of Product and Brand Management*, vol. 14, no. 1, pp. 4–13, 2006.

[54] A. S. Albayrak, S. K. Yilmaz, "Verİ madencİlİğİ: Karar ağacı algorİtmaları ve İmkb verİlerİ üzerİne bİr uygulama.," *Suleyman Demirel University Journal of Faculty of Economics & Administrative Sciences*, vol. 14, no. 1, 2009.

[55] A. P. Markopoulos, D. E. Manolakos, N. M. Vaxevanidis, "Artificial neural network models for the prediction of surface roughness in electrical discharge machining," *Journal of intelligent Manufacturing*, vol. 19, no. 3, pp. 283–292, 2008.

[56] P. Ramesh, J. Jeba Emilyn, V. Vijayakumar, "Hybrid artificial neural networks using customer churn prediction," *Wireless Personal Communications*, vol. 124, no. 2, pp. 1695–1709, 2022.

[57] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*, Springer, 1998, pp. 137–142.

[58] K.-S. Shin, T. S. Lee, H.-j. Kim, "An application of support vector machines in bankruptcy prediction model," *Expert systems with applications*, vol. 28, no. 1, pp. 127–135, 2005.

[59] A. Cutler, D. R. Cutler, J. R. Stevens, "Tree-based methods," in *High-Dimensional Data Analysis in Cancer Research*, Springer, 2009, pp. 1–19.

[60] T. Prajwala, "A comparative study on decision tree and random forest using r tool," *International journal of advanced research in computer and communication engineering*, vol. 4, no. 1, pp. 196–199, 2015.

[61] K. Eggensperger *et al.*, "Towards an empirical foundation for assessing bayesian optimization of hyperparameters," in *NIPS workshop on Bayesian Optimization in Theory and Practice*, vol. 10, 2013.

[62] C. E. Rasmussen, "Gaussian processes in machine learning," in *Summer school on machine learning*, Springer, 2003, pp. 63–71.

[63] M. Feurer, F. Hutter, "Hyperparameter optimization," in *Automated machine learning*, Springer, Cham, 2019, pp. 3–33.

[64] G. Aloysius, D. Binu, "An approach to products placement in supermarkets using prefixspan algorithm," *Journal of King Saud University-Computer and Information Sciences*, vol. 25, no. 1, pp. 77–87, 2013.

[65] Q. Zhao, S. S. Bhowmick, "Sequential pattern mining: A survey," *ITechnical Report CAIS Nayang Technological University Singapore*, vol. 1, no. 26, p. 135, 2003.

[66] M. J. Zaki, "Spade: An efficient algorithm for mining frequent sequences," *Machine learning*, vol. 42, no. 1, pp. 31–60, 2001.

[67] G. B. V. ve İleri Analitik Danışmanı. "Apache spark nedir, kurulumu ve uygulama örneği." (2022), [Online]. Available: `https://www.gtech.com.tr/apache-spark/` (visited on 06/17/2022).

[68] Bilginç. "Pycharm hakkında bİlİnmeyenler." (2022), [Online]. Available: `https://bilginc.com/tr/blog/1322/pycharm-hakkinda-bilinmeyenler` (visited on 06/17/2022).

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. D. Erek, M. Faydalı, and M. S. Aktas, “On the data analysis workflow for predicting retention model in cargo and logistics sectors: Case study,” in The Eurasia Proceedings of Science, Technology, Engineering Mathematics, IConTech, 2022, pp. 1–6.

2. G. O. Sahinkaya, D. Erek, H. Yaman, and M. S. Aktas, “On the data analysis workflow for predicting customer churn behavior in cargo and logistics sectors: Case study,” in 2021 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), IEEE, 2021, pp. 1–6.