

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**

**TELEKOMÜNİKASYON SEKTÖRÜNDE MÜŞTERİ
KAYBI YÖNETİMİ İÇİN META SEZGİSEL TABANLI
KARAR DESTEK SİSTEMİ**

**Hazırlayan
Mihrimah ÖZMEN**

**Danışman
Doç. Dr. Emel KIZILKAYA AYDOĞAN**

Doktora Tezi

**Eylül 2017
KAYSERİ**

**T.C.
ERCIYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ENDÜSTRİ MÜHENDİSLİĞİ ANABİLİM DALI**

**TELEKOMÜNİKASYON SEKTÖRÜNDE MÜŞTERİ
KAYBI YÖNETİMİ İÇİN META SEZGİSEL TABANLI
KARAR DESTEK SİSTEMİ**

(Doktora Tezi)

**Hazırlayan
Mihrimah ÖZMEN**

**Danışman
Doç. Dr. Emel KIZILKAYA AYDOĞAN**

**Bu çalışma; Sanayi Bakanlığı SANTEZ programı tarafından
0777.STZ.2014 kodlu proje ile desteklenmiştir.**

**Eylül 2017
KAYSERİ**

BİLİMSEL ETİĞE UYGUNLUK

Bu çalışmadaki tüm bilgilerin, akademik ve etik kurallara uygun bir şekilde elde edildiğini beyan ederim. Aynı zamanda bu kural ve davranışların gerektirdiği gibi, bu çalışmanın özünde olmayan tüm materyal ve sonuçları tam olarak aktardığımı ve referans gösterdiğimi belirtirim.

Mihrimah ÖZMEN

İmza:

YÖNERGEYE UYGUNLUK

“Kaba Küme İle Telekomünikasyon Sektöründe Müşteri Kaybı Yönetimi İçin Meta Sezgisel Tabanlı Karar Destek Bilişim Sistemi” Doktora tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma Yönergesi 'ne uygun olarak hazırlanmıştır.

Tezi Hazırlayan

Danışman

Mihrimah ÖZMEN

Doç. Dr. Emel KIZILKAYA AYDOĞAN



Endüstri Mühendisliği ABD Başkanı

Prof. Dr. Mithat ZEYDAN

Doç. Dr. Emel KIZILKAYA AYDOĞAN danışmanlığında **Mihrimah ÖZMEN** tarafından hazırlanan “**Telekomünikasyon Sektöründe Müşteri Kaybı Yönetimi İçin Meta Sezgisel Tabanlı Karar Destek Sistemi**” adlı bu çalışma, jürimiz tarafından Erciyes Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalında **Doktora** tezi olarak kabul edilmiştir.

28/09/2017

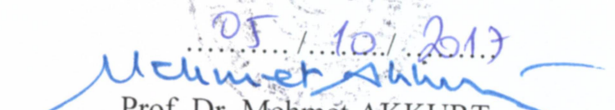
(Tez Savunma Sınav Tarihi)

JÜRİ:

Danışman	:Doç. Dr. Emel KIZILKAYA AYDOĞAN
Üye	:Prof. Dr. Derviş KARABOĞA
Üye	:Doç. Dr. M. Duran TOKSARI
Üye	: Doç. Dr. İbrahim AKGÜN
Üye	: Doç. Dr. Erkan KÖSE

ONAY:

Bu tezin kabulü Enstitü Yönetim Kurulunun 03/10/2017 tarih ve 217/43-31 sayılı kararı ile onaylanmıştır.

05 / 10 / 2017

 Prof. Dr. Mehmet AKKURT

Enstitü Müdürü

ÖNSÖZ / TEŞEKKÜR

Çalışmalarım boyunca farklı bakış açıları ve bilimsel katkılarıyla beni yönlendiren, bana her konuda destek olan, çok değerli sayın hocam Doç. Dr. Emel KIZILKAYA AYDOĞAN' a, yine kıymetli tecrübelerinden yararlandığım değerli hocalarım Prof. Dr. Derviş KARABOĞA, Doç. Dr. M. Duran TOKSARI ve Yrd. Doç. Dr. Yılmaz DELİCE' ye savunma jürimde bulunan hocalarım Doç. Dr. İbrahim AKGÜN ve Doç. Dr. Erkan KÖSE' ye, çalışmalarım süresince sabır göstererek beni daima destekleyen sevgili eşim, çocuklarım ve aileme teşekkürü bir borç bilirim.

Ayrıca, bu tez çalışmasına maddi destek veren Sanayi Bakanlığı SANTEZ programına (Proje No: 0777.STZ.2014) teşekkür ederim.

Mihrimah ÖZMEN

Kayseri, Eylül 2017

TELEKOMÜNİKASYON SEKTÖRÜNDE MÜŞTERİ KAYBI YÖNETİMİ İÇİN META SEZGİSEL TABANLI KARAR DESTEK SİSTEMİ

Mihrimah ÖZMEN

Erciyes Üniversitesi, Fen Bilimleri Enstitüsü
Doktora Tezi, Eylül 2017
Danışman: Doç. Dr. Emel KIZILKAYA AYDOĞAN

ÖZET

Telekomünikasyon sektöründe yer alan firmalar, pazar koşulları gereği müşterilerini elde tutabilmek için etkili bir müşteri kaybı yönetimine ihtiyaç duymaktadırlar. Bu ihtiyaca yönelik olarak müşterilerin hizmet aboneliklerini sonlandırma nedenlerini analiz etmede, müşteri kaybı analizi teknikleri kullanılmaktadır. Telekomünikasyon sektöründe büyük miktarlarda müşteri verisi olsa da bunların yararlılığı veri kalitesiyle ilişkilidir. Müşteri veri kümeleri; yüksek boyutlu ve dengesiz olması sebebiyle müşteri kaybı tahmini için arzu edilen performansı elde etmeyi zorlaştırmaktadır. Hazırlanan tezde, sınıflandırmayı zorlaştıran; yüksek boyutluluk, sınıfsal dengesizlik ve asimetrik yanlış sınıflandırma maliyetleri ile başa çıkabilen, maliyet duyarlı entegre bir müşteri kaybı yönetimi yaklaşımı önerilmektedir. Önerilen model, nitelik seçimi ve sınıflandırma aşamalarından oluşmaktadır. Nitelik seçimi aşamasında, öncelikle müşteri kaybı tahmininde etkin sonuçlar elde etmede önem arz eden nitelikler ayrıntılı olarak literatüre kazandırılmış olup, daha sonra bu niteliklere bulanık kaba küme tabanlı FRQR (The fuzzy-rough Quick Reduct) ve maliyet tabanlı NSGA-II (Non-dominated Sorted Genetic Algorithm) nitelik seçim algoritmaları uygulanarak sınıflandırma kalitesini arttıracak indirgenmiş nitelikler elde edilmiştir. Sınıflandırma aşamasında ise, nitelik seçimi yaklaşımlarından elde edilen indirgenmiş veriler, önerilen maliyet duyarlı KKO (Karıncı Kolonisi Optimizasyonu) algoritması ile sınıflandırılmıştır. Geliştirilen müşteri kaybı sınıflandırma modeli, Türkiye'de ilk 100 Bilgi Teknolojileri (BT) şirketleri içerisinde yer alan bir firmanın müşteri verilerine uygulanmıştır. Elde edilen sonuçlar, önerilen modelin müşteri kaybını tahmin etmede 0.99 AUC değeri ile etkin bir performans sergilediğini göstermekte olup, telekomünikasyon sektörü literatürü için de farklı bir bakış açısı kazandırmıştır.

Anahtar Kelimeler: Veri Madenciliği, Kaba Küme, Müşteri Kaybı Yönetimi, Maliyet Duyarlı Öğrenme, Nitelik İndirgeme.

META-HEURISTIC BASED DECISION SUPPORT SYSTEM FOR CUSTOMER CHURN MANAGEMENT IN TELECOMMUNICATIONS SECTOR

Mihrimah ÖZMEN

Erciyes University, Graduate School of Natural and Applied Sciences

Ph.d. Thesis, September 2017

Supervisor: Assoc. Prof. Dr. Emel KIZILKAYA AYDOĞAN

ABSTRACT

The telecommunications sector companies need effective customer churn management to keep customers due to the market conditions. Therefore; customer churn analysis techniques are used to analyze the reasons for cancellation of subscriptions. In the telecommunications sector, even though there are large amounts of customer data, their usefulness is limited by data quality and relevance. The characteristics of telecommunications datasets such as; high dimensionality and imbalanced are making it difficult to achieve the desired performance for churn prediction. In this thesis, an integrated cost sensitive customer churn management approach is proposed, which can cope with high dimensional, class imbalance and asymmetric misclassification costs. The proposed model consists of feature selection and classification stages. In the stage of feature selection, the features that are important to obtain effective results in predicting customer churn have been elaborated in detail, and then these features have been applied to Fuzzy Rough Quick Reduct (FRQR) and cost-based NSGA-II (Non-dominated Sorted Genetic Algorithm) feature selection algorithms to obtain reduced features that increase the classification quality. In the classification stage, the reduced data obtained from the feature selection approaches are classified by the proposed cost sensitive ACO (Ant Colony Optimization) algorithm. The developed customer churn model has been applied to the customer data of a company in the first 100 Information Technology (IT) companies in Turkey. The results show that the proposed model performs effectively with AUC values of 0.99 to predict the churners and it gives a different perspective to the literature of the telecommunication sector.

Keywords: Data Mining, Rough Set, Churn Management, High Dimension, Cost Sensitive Learning and Feature Selection

İÇİNDEKİLER

TELEKOMÜNİKASYON SEKTÖRÜNDE MÜŞTERİ KAYBI YÖNETİMİ İÇİN META SEZGİSEL TABANLI KARAR DESTEK SİSTEMİ

BİLİMSEL ETİĞE UYGUNLUK	i
YÖNERGEYE UYGUNLUK.....	ii
KABUL VE ONAY	iii
ÖNSÖZ / TEŞEKKÜR	iv
ÖZET	v
ABSTRACT.....	vi
İÇİNDEKİLER	vii
KISALTMALAR	xi
TABLolar LİSTESİ.....	xii
ŞEKİLLER LİSTESİ	xiii

1. BÖLÜM

GİRİŞ

1.1. Tezin Kapsamı ve Önemi	1
1.2. Tezin Amacı ve Yöntemi	2
1.3. Tezin Literatüre Katkısı	3
1.4. Tezin Bölümleri	3

2. BÖLÜM

TELEKOMÜNİKASYON SEKTÖRÜNDE MÜŞTERİ KAYBI YÖNETİMİ ve VERİ MADENCİLİĞİ UYGULAMALARI

2.1. Giriş	5
2.2. Telekomünikasyon Sektörü.....	5
2.3. Telekomünikasyon Sektöründe Veri Madenciliği Uygulamaları	8

2.3.1. Telekomünikasyon Sektöründe Pazarlamada Alanında Veri Madenciliği Uygulamaları	9
2.3.2. Telekomünikasyon Sektöründe Dolandırıcılık Tespitinde Veri Madenciliği Uygulamaları	10
2.3.3. Telekomünikasyon Sektöründe Ağ Yönetiminde Veri Madenciliği Uygulamaları	11
2.4. Telekomünikasyon Sektöründe Müşteri Kaybı Yönetimi.....	12
2.4.1. Müşteri Kaybı.....	12
2.4.2. Müşteri Kaybının Sınıflandırılması.....	18
2.4.2.1. İstemeyerek Ayrılan Müşteriler	20
2.4.2.2. İsteyerek Ayrılan Müşteriler	20
2.4.3. Telekomünikasyon Sektöründe Müşteri Kaybı Yönetimi Alanında Veri Madenciliği Uygulamaları	26
2.4.4. Diğer Sektörler ile Telekomünikasyon Sektörünün Müşteri Kaybı Yönetimi Açısından Değerlendirilmesi.....	32

3. BÖLÜM

VERİ MADENCİLİĞİ

3.1. Giriş	37
3.2. Veri Madenciliği	38
3.3. Veri Madenciliği Tekniklerinin Sınıflandırılması.....	40
3.3.1. Denetimli Öğrenme: Sınıflandırma ve Regresyon	40
3.3.2. Denetimsiz Öğrenme: Kümeleme ve Örüntü Bulma	41
3.4. Veri Madenciliğinin Görevleri ve Teknikleri	41
3.4.1. Sınıflandırma ve Tahmin	42
3.4.2. Kümeleme.....	43
3.4.3. Birliktelik Analizi	44
3.4.4. Veri Madenciliğinde Nitelik İndirgeme Yaklaşımları	45

4. BÖLÜM

ÖNERİLEN MÜŞTERİ KAYBI TAHMİN MODELİ

4.1. Giriş	47
4.2. Veri Madenciliğinde Kaba Küme Teorisi Yaklaşımları.....	48
4.2.1. Kaba Küme Teorisinde Nitelik Seçimi Yaklaşımı.....	49
4.2.2. FRQR Nitelik Seçimi Algoritması.....	51
4.3. Maliyet Duyarlı Veri Madenciliği Yaklaşımları.....	53
4.4. Veri Madenciliğinde Meta Sezgisel Yaklaşımlar	58
4.4.1. Veri Madenciliğinde Çok Amaçlı Meta Sezgisel Yaklaşımlar	60
4.4.2. Nitelik İndirgeme ve Sınıflandırma için Çok Amaçlı Meta Sezgisel Yaklaşımlar	62
4.5. NSGA-II	63
4.5.1. Literatürde NSGA-II ile Nitelik Seçim Yaklaşımları	70
4.5.2. Önerilen Çok Amaçlı ve Maliyet Duyarlı NSGA-II Nitelik Seçim Yaklaşımı.....	71
4.6. KKO	72
4.6.1. Çok Amaçlı KKO Algoritması	74
4.6.2. Literatürde KKO Sınıflandırma Algoritması.....	75
4.6.3. Değerlendirme Ölçütleri	76
4.6.4. Önerilen Çok Amaçlı ve Maliyet Duyarlı KKO Sınıflandırma Yaklaşımı	77

5. BÖLÜM

ÖNERİLEN MÜŞTERİ KAYBI TAHMİN MODELİNİN UYGULAMASI

5.1. Giriş	79
5.2. Veri Hazırlama ve Örneklemeye	79
5.3. Müşteri Veri Seti Nitelikleri.....	81
5.4. Önerilen Müşteri Kaybı Tahmin Modeli	85
5.5. Önerilen Modelin Müşteri Kaybı Tahminine Uygulaması	94

6. BÖLÜM

SONUÇ.....	102
KAYNAKLAR	103
ÖZGEÇMİŞ.....	118



KISALTMALAR

BT	: Bilgi Teknolojileri
ETL	: Extraction-Transformation-Loading
FRQR	: The fuzzy-rough Quick Reduct
KKO	: Karınca Kolonisi Optimizasyonu
NCL	: Komşuluk Temizleme Kuralı
NSGA-II	: Non-dominated Sorted Genetic Algorithm
ROS	: Rastgele Üst Örneklem
RUS	: Rastgele Alt Örneklem
SL-SMT	: Güvenlik Seviyesi SMOTE
SMT	: Sentetik Azınlık Üst Örneklem Tekniği
SMT+ENN	: SMT+Düzenlenmiş en Yakın Komşu
SVM	: Destek vektör makinesi
TL	: Tomek Links

TABLOLAR LİSTESİ

Tablo 2.1.	Telekomünikasyon sektöründe, müşteri kaybı yönetiminde veri madenciliği uygulamaları.	30
Tablo 4.1.	Çok amaçlı meta sezgisel yaklaşımlarda nitelik seçimi alanında çalışmalar	70
Tablo 4.2.	İkili sınıflandırma problemi için karışıklık matrisi.....	76
Tablo 5.1.-a	Müşteri Niteliklerinin Tanımları.....	82
Tablo 5.1.-b	Fatura ve Ödeme Niteliklerinin Tanımları.....	83
Tablo 5.1.-c	Arama Niteliklerinin Tanımları	83
Tablo 5.2.	Müşteri Kaybı Tahmininde Uygulanan Maliyet Matrisi	97
Tablo 5.3.	Deneylerde Kullanılan Yöntemler	99
Tablo 5.4.	Telekomünikasyon Müşteri Veri Setinin Deney Sonuçları	100

ŞEKİLLER LİSTESİ

Şekil 2.1.	Yeni teknoloji adaptasyon eğrisi.....	15
Şekil 2.2.	Telekomünikasyon sektörü yeni teknoloji adaptasyon eğrisi	16
Şekil 2.3.	Büyüme Döneminde Pazar Payı	17
Şekil 2.4.	Olgunlaşma Aşamasında Pazar Payı.....	18
Şekil 2.5.	Müşteri kaybının sınıflandırılması.....	19
Şekil 2.6.	Müşteri ayrılma nedenleri	22
Şekil 3.1.	Bilgi keşfinin adımları	39
Şekil 3.2.	Veri madenciliği sürecine katkı sağlayan alanlar	40
Şekil 3.3.	Veri madenciliğinin görevleri	41
Şekil 3.4.	Kümeleme Analizi	43
Şekil 3.5.	Kümeleme Yöntemleri.....	44
Şekil 3.6.	Boyut indirgeme yöntemlerinin sınıflandırılması.....	45
Şekil 3.7.	Nitelik Seçimi Yaklaşımları.....	46
Şekil 4.1.	Kaba Kümenin Gösterimi	48
Şekil 4.2.	FRQR Algoritması	53
Şekil 4.3.	İki amaçlı bir problem için NSGA sıralama mekanizması	64
Şekil 4.4.	NSGA akış şeması	64
Şekil 4.5.	Hızlı bastırılmamış sıralama prosedürü	66
Şekil 4.6.	Yakınlık mesafesi hesaplama.....	67
Şekil 4.7.	NSGA-II sıralama akış şeması.....	68
Şekil 4.8.	NSGA-II akış şeması	69
Şekil 4.9.	Gerçek bir karıncanın davranışına örnek	72
Şekil 4.10.	KKO Akış Şeması (Gezgin Satıcı Problemi için).....	73
Şekil 4.11.	KKO sınıflandırma algoritması	75
Şekil 5.1.	Bir müşteri ayrılma eğiliminde mi değil mi?	81
Şekil 5.2.	Önerilen Müşteri Kaybı Tahmin Modeli Yapısı.....	88
Şekil 5.3.	Üyelik fonksiyonu.....	90
Şekil 5.4.	Önerilen müşteri kaybı tahmin modelinin gerçekleşme adımları.....	94

Şekil 5.5.	NSGA-II ve FRQR + KKO ile elde edilen Pareto Çözümler	96
Şekil 5.6.	NSGA-II + KKO ile elde edilen Pareto Çözümler	96
Şekil 5.7.	FRQR + KKO ile elde edilen Pareto Çözümler.....	97
Şekil 5.8.	Telekomünikasyon Müşteri Veri Setinin Deney Sonuçları Grafik Gösterimi	101



1. BÖLÜM

GİRİŞ

1.1. Tezin Kapsamı ve Önemi

Telekomünikasyon sektörü özellikle son on yılda kaydettiği, internet ve gelişmiş iletişim ürünlerindeki gelişmelerle, yaşam tarzında belirgin değişikliklere sebep olmuştur. Bu açıdan bakıldığında, ekonomik gelişmelerin en önemli göstergelerinden birisi olan telekomünikasyon, stratejik bir sektör haline gelmiştir. Bu sektörü anlayabilen ve analiz edebilen ekonomiler gelecek ekonomisinde de söz sahibi olabileceklerdir. Tüm bu gelişmeleri takip eden ülkeler son zamanlarda telekomünikasyon sektöründeki özelliklerini üst seviyelere taşımak için yarışmaktadırlar [1]. Bu teknolojik rekabet, sektördeki rekabeti stratejik bir savaşa dönüştürmüştür. Bu konunun önemi nedeniyle, akademisyenler, müşteri kaybının nedenlerini ve sonuçlarını incelemek üzere kapsamlı birçok araştırma yürütmüşlerdir [2-11]. Bu araştırmalara dayanarak, veri madenciliği literatürüne birçok tahmin modeli önerisi kazandırılmıştır [12-22].

Araştırmacılar özellikle telekomünikasyon sektöründe müşteri kaybı tahmini için: destek vektör makineleri [12,18,20,22], sinir ağları [6,7,14,19,21], karar ağaçları ve lojistik regresyon [12,23], naïve bayes ve bayesian ağları, neuro-fuzzy, çok katmanlı sinir ağları, sezgisel algoritmalar, kaba küme gibi birçok veri madenciliği yöntemlerinden yararlanmıştır.

Türkiye'de müşteri kaybı analizi için literatürde birçok araştırma da yürütülmüştür. Gürsoy, U.T.S. [24], Türkiye'deki önemli bir telekomünikasyon şirketinin müşteri verilerini kullanarak, özel kampanyalar teklif edebilmek için ayrılmak isteyen

müşterileri lojistik regresyon ve karar ağaçları tekniklerinden yararlanarak tahmin etmişlerdir. Pınar Kisioglu ve Y. Ilker Topcu [25], ayrılma eğiliminde olan müşterilerin davranışlarını belirleyebilmek için “Bayesian Ağları” nı kullanarak bir model önermişlerdir. Bu modelde Türkiye’de faaliyet gösteren bir telekomünikasyon şirketinin müşteri verilerinden yararlanılmışlardır. Veri kümesi 534 ayrılan olmak üzere 2000 müşteri ve 23 nitelikten oluşmaktadır. Veride yapılan işlemlerden sonra nitelikler 9’a indirgenmiştir. Karahoca, A., ve Karahoca, D. [11], Türkiye’ deki bir GSM operatörünün ayrılmak isteyen müşterilerini ANFIS yöntemi ile tahmin etmişlerdir.

Telekomünikasyon sektörü muazzam boyutlarda veriye sahip olduğundan dolayı tahmin modellerini eğitmek ve değerlendirmek zorlaşmaktadır. Telekomünikasyon müşterilerinin büyük boyutlardaki verilerini faydalı bilgiye dönüştürebilmek için nitelik indirgeme mekanizması sınıflandırma işlemlerine eşlik etmiştir. Bu nedenle bu çalışmada, nitelik indirgeme ve sınıflandırma algoritmasının entegre olduğu bir model önerilmiştir.

1.2. Tezin Amacı ve Yöntemi

Hazırlanan tezde, telekomünikasyon sektörünün müşteri beklentilerini ve davranışlarını daha iyi anlamasına ve bunun için stratejiler geliştirmesine yardımcı olabilmek için etkili bir müşteri kaybı tahmin modeli önerilmesi amaçlanmıştır. Telekomünikasyon müşteri verisine özgü yeni bir müşteri kaybı tahmin modeli geliştirilmiş ve model yeni niteliklerle zenginleştirilmiştir. Önerilen model Türkiye'nin ilk 100 BT firmasından biri olan bir telekomünikasyon şirketi müşteri verilerine uygulanmıştır. Böylelikle, telekomünikasyon sektörü müşteri yapısına uygun bir sistem kurarak, rekabet gücünü daha da iyi hale getirebilmek için sektöre katkıda bulunulmaya çalışılmıştır.

Müşteri verilerinin dengesiz dağılımı ve yüksek boyutlu özelliği nedeniyle maliyet duyarlı entegre bir model geliştirilmiştir. Önerilen model nitelik seçimi ve sınıflandırma olmak üzere iki ana aşamadan oluşmaktadır. Nitelik seçimi aşamasında, FRQR ve maliyet duyarlı NSGA-II’ yi temel alan 2 ayrı yaklaşım uygulanmıştır. Sınıflandırma aşamasında ise çok amaçlı ve maliyet duyarlı KKO sınıflandırma algoritması uygulanmıştır. Nitelik seçimi aşaması müşteri verisinin büyük boyutlu nitelikleri ile başa çıkmaya çalışırken, aynı zamanda maliyet duyarlı KKO sınıflandırıcı da dengesiz

dağılıma sahip veri ile başa çıkmaya çalışmaktadır. Çok amaçlı olarak geliştirilen NSGA-II temelli nitelik seçimi algoritması iki amacı minimize etmeyi amaçlamaktadır: nitelik alt kümesi eleman sayısı ve yanlış sınıflandırma maliyeti. Bu aşamada elde edilen nitelik alt kümeleri modelin başarısına katkı sağlamıştır. Ayrıca, yine çok amaçlı olarak geliştirilen KKO sınıflandırıcı aynı zamanda; AUC değerini maksimize ederken yanlış sınıflandırma maliyetini de minimize etmeyi amaçlamaktadır. Önerilen entegre model geçmiş müşteri davranışlarından müşteri kaybı tahmini yaparak; firmaya müşteri kayıp oranını düşürmede yardımcı olması, müşteri davranış ve ihtiyaçlarını daha iyi anlaması amaçlanmaktadır.

1.3. Tezin Literatüre Katkısı

Hazırlanan tez hem önerilen model hem de uygulaması açısından literatüre katkı sağlamıştır. Birçok meta sezgisel yöntem, çok amaçlı sınıflandırma yaklaşımı olarak literatürde uygulanmıştır. KKO sınıflandırma algoritmasının da performansını görebilmek için bu tezde çok amaçlı ve maliyet duyarlı olarak geliştirilmiştir. Bunun yanı sıra NSGA-II temelli nitelik seçimi yaklaşımları literatürde yer almaktadır ancak maliyet duyarlı NSGA-II nitelik seçimi yaklaşımı, bu çalışmada ilk kez yanlış sınıflandırma maliyeti uygunluk fonksiyonu ile uygulanmıştır. Önerilen müşteri kaybı modeli bu özellikleri ile orijinalliğini ortaya koymaktadır.

Hazırlanan tezin uygulaması açısından da literatüre birçok katkısı yer almaktadır. Öncelikle önerilen modelin uygulandığı müşteri veri setinde yer alan niteliklerin bir kısmı tez kapsamında konusunda uzman kişilerden de görüş alınarak belirlenmiş ve literatüre kazandırılmıştır. Bunun yanı sıra müşteri veri seti literatürdeki benzer çalışmalara nazaran daha detaylı ve geniş nitelikler içermektedir. Ayrıca FRQR nitelik seçimi yaklaşımı, KKO sınıflandırma algoritması ve çok amaçlı, maliyet duyarlı meta sezgisel sınıflandırma yaklaşımları ise telekomünikasyon sektörü müşteri kaybı analizinde ilk kez uygulanmıştır.

1.4. Tezin Bölümleri

Bir sonraki bölümde, telekomünikasyon sektöründe müşteri kaybı analizi hakkında bilgi verilmiş ve literatürde yapılan çalışmalar özetlenmiştir. Bölüm 3' te veri madenciliği ve

literatürde yer alan veri madenciliği yaklaşımları incelenmiştir. Bölüm 4’ te önerilen algoritma ve sonraki bölümde de önerilen algoritmanın müşteri kaybı yönetiminde uygulaması yer almaktadır. Son bölümde ise tez çalışmasında elde edilen sonuçlara yer verilerek ileride yapılabilecek çalışmalar ile ilgili öneriler sunulmuştur.



2. BÖLÜM

TELEKOMÜNİKASYON SEKTÖRÜNDE MÜŞTERİ KAYBI YÖNETİMİ ve VERİ MADENCİLİĞİ UYGULAMALARI

2.1. Giriş

Müşteri kaybı, dünya genelinde telekomünikasyon sektöründe birçok geçerli sebepten ötürü en önemli konulardan biridir. Bu konu hakkında sektöre yeni giren veya mevcut firmaların her birinin farklı görüşleri olsa da ortak görüş müşteri kaybının olumsuz sonuçları olduğudur. Hâlbuki başka bir bakış açısıyla bakıldığında aslında müşteri kaybını fırsata çevirmek mümkün olabilir. Müşteri kaybı, telekomünikasyon sektörünün bir gerçeğidir ve firmalar bu durumla başa çıkabilmelidirler. Bu durumun, gelirlerinde düşmeye ve pazarlama giderlerinde artışa sebep olarak firmaları önemli ölçüde etkilediği bilinmektedir. Ancak iyi olan şu ki, sektördeki bütün rakipler aynı pazar koşullarına sahiptir. Bu noktada, ya sektördeki bütün firmalar şimdiye kadar yapmakta olduklarına devam edip mevcut pazar paylarını koruyabilir ya da avantaja çevrilebilir. Bu durumu avantaja çevirebilmek için; firmanın sektördeki yerinde, kârlılık stratejilerinde ve şirketin organizasyon yapısında radikal değişimler düşünülebilir. Açıkçası, müşteri kaybı firmalara işlerini daha iyi yapabilmeleri için kendilerini gözden geçirme fırsatı sunar. Bu sayede, firmalar müşterilerini ve isteklerini daha yakından tanıyabilirler.

2.2. Telekomünikasyon Sektörü

Hem toplum hem de ekonominin geneli için vazgeçilmez bir yer tutan bilgi ve iletişim sektörü, özellikle son on yıllık süreçte önemli gelişmeler kaydetmiştir. Teknolojik gelişmeler daha çok telekomünikasyon şebekelerinin alt yapısında gerçekleşmiştir. Bu gelişmelerle sektör; telefon ve kablo TV operatörleri, internet servis sağlayıcıları,

kablosuz iletişim, uydu operatörleri ve mobil iletişimi de kapsayan çoklu hizmet sağlayıcıları tarafından şekillendirilmiştir.

Telekomünikasyon sektöründeki gelişmelerin hızlanması; ürün çeşitliliği ile penetrasyon oranının ve dolayısıyla da sektörün ekonomideki ağırlığının artmasını sağlamıştır. Sektördeki büyüme ülke ekonomilerinin üzerinde seyrederek günümüzde ekonomik altyapıların temelini oluşturmaktadır. Ayrıca telekomünikasyon sektörü hem diğer sektörleri desteklemesi hem de ekonomik büyümeye yaptığı katkılardan dolayı hayati bir önem taşımaktadır. Ekonomik gelişmenin en büyük itici güçlerinden biri olan bu sektörde ülkeler ve bölgeler arasında; eğitim, kalifiye iş gücü, dış ticaret hacmi, teşvik ve ulusal stratejiler bakımından farklılıklar gözlenmektedir [1].

Telekomünikasyon sektörünün, ekonomik büyümenin yanı sıra sosyal/kültürel kalkınma üzerindeki etkisi ise bütün kesimlerce kabul görmektedir. İnsanların hem iş hem de günlük hayatının içinde önemli yer tutan bu sektör, günümüzün en önemli sektörlerinden birisi haline gelmiştir. Özellikle internet ve ileri iletişim ürünlerinin hayatımıza girmesiyle de yaşam tarzı dikkat çekici bir şekilde değişmiştir [26]. Bilgiye erişimi sınırsız hale getiren yeni gelişmelerle bu sektör, bilgi toplumuna ve bilgi temelli ekonomiye geçiş için önemli bir ekonomik değer ifade eden stratejik bir sektör konumundadır.

Ekonomik gelişmelerin en önemli göstergelerinden biri olan geleceğin sektörünü yani telekomünikasyonu anlayabilen ve buna ayak uydurabilen ülkeler geleceğin ekonomisinde de söz sahibi olabileceklerdir. Bu minvalde, son zamanlarda ülkeler telekomünikasyon kalitelerini üst seviyelere taşımada birbiriyle yarışmaktadır [1]. Bu teknolojik rekabet de sektördeki rekabeti stratejik bir savaşa çevirmiştir.

Tüm dünya genelinde telekomünikasyon sektörü sürekli bir değişim ve gelişim içindedir. U.S. 1996 yılı telekomünikasyon yasasında yerel ve uzun mesafe arama, kablolu TV, televizyon yayını ve kablosuz hizmetlerindeki kotaları koyan devlet kurallarını sona erdirmiştir. Avrupa'da, sektörde liberalleşme ile İngiltere, İsveç ve Finlandiya'daki pazarlar açılmıştır. 1998 yılında 15 Avrupa ülkesi de diğer ülkelerin etkisiyle liberalleşmiştir. Gelişmekte olan Şili, Malezya ve Peru gibi ülkeler de devlet

telefon tekellerini ortadan kaldırmışlardır. Bu yeniden yapılanma sayesinde, yeni pazar fırsatları ortaya çıkmıştır [27].

İnternet ve sosyal medya kullanımı Türkiye’de de yaygınlaşmıştır. Günümüzde Türkiye başlıca sosyal medya kullanıcısıdır ve nüfusunun yüzde 45’i (35 milyon) internete erişebilmektedir. Bu rakamlarla Türkiye Avrupa’ nın beşinci en büyük internet kullanıcısıdır [28]. Araştırmalara göre de sosyal medya Türkiye’nin internet kullanımındaki en önemli gerekçesidir. Ayrıca facebook kullanıcıları arasında Türkiye dünyada üçüncü en geniş nüfusa sahip ülkedir [29].

Dünya genelinde olduğu gibi Türkiye’de de internet kullanıcı sayısı yıllara göre ivmeli bir şekilde artış göstermektedir. Abone sayısı itibarı ile 2015 yılsonunda 11,5 milyon sabit abonenin yanında, mobil hizmetlerde abone sayısı 73,6 milyonu aşmış ve penetrasyon oranı %93,5 düzeyinde gerçekleşmiştir. Fiber altyapısı açısından ise, 2014 yılı sonunda 244.847 km olan fiber uzunluğu, 2015 yılı sonunda 268.120 km’ye ulaşarak son bir yılda %10 artmış, son çeyrek itibarı ile de %2,6 oranında artış gerçekleşmiştir [30].

2015 yılı dördüncü üç aylık döneminde Türk Telekom, mobil şebeke işletmecileri ve diğer işletmecilerin net satış gelirleri yaklaşık 10,4 milyar TL olarak gerçekleşmiştir. Toplam yatırım miktarı da yaklaşık 13,6 milyar TL olarak gerçekleşmiştir. Sektördeki bu hızlı büyüme yatırım gerekliliğini de beraberinde getirmektedir. Nitekim sektörde yer alan ilk 5 firmanın; 2015 yılına ait yatırım büyüklükleri ilk çeyrekte 823 milyon TL iken ikinci çeyrekte 1.016 milyon TL, üçüncü çeyrekte 986 milyon TL olarak gerçekleşmiştir [30].

Bu kadar hızla gelişen telekomünikasyon sektöründe teknolojik devrim milyonlarca insanın tabletler, akıllı telefonlar veya bilgisayarlar gibi herhangi bir dijital cihazdan herhangi bir zamanda devasa boyutlarda veri üretmesine olanak sağlamıştır. Cisco’ nun yıllık tahminlerine göre telekomünikasyon veri trafiğindeki büyüme 2014 yılından 2019 yılının sonuna kadar yıllık 3.4 zettabyte büyüme ile 10.4 zettabyte’ a ulaşacaktır [31]. Telekomünikasyon sektörü muazzam boyutlarda veriye sahip olduğundan dolayı veri madenciliği tekniklerinin uygulamasında birçok zorlukla karşı karşıyadır. Ancak bu

büyük boyutlardaki karmaşık veriler içerdiği değerli bilgilerden dolayı fırsatlar da barındırmaktadırlar.

Hem büyük boyutta hem de eş zamanlı olarak veriyi, büyük veri madenciliği yöntemleri ile işlemek gerekliliği ortaya çıkmıştır. Büyük veri madenciliği ise; hacim, hız ve çeşitlilik bakımından büyük veri setlerinden klasik yöntemlerle daha önce yapılamayan değerli bilgi çıkarımını gerçekleştirmektir [33]. Bütün bu bilgiler ışığında birçok telekomünikasyon şirketi müşterilerinden topladıkları büyük boyutlardaki verilerden faydalanarak iş hayatındaki problemlerini çözmede faydalı bilgiye dönüştürebileceklerini fark etmişlerdir. Bu konuda da veri madenciliği tekniklerini etkin bir şekilde uygulayabilmektedirler [34-35].

2.3. Telekomünikasyon Sektöründe Veri Madenciliği Uygulamaları

Telekomünikasyon sektörü dünyadaki en veri yoğun sektörler arasında yer almaktadır ve bu sektör veri madenciliği teknolojilerine en erken adapte olup sayısız veri madenciliği uygulamaları kullanmıştır. Uygulama alanları; pazarlama tekniklerini geliştirmek, müşteri dolandırıcılıklarını tespit etmek ve daha iyi ağ yönetimi olmak üzere üç ana başlıkta incelenebilir [36]. Pazarlama ve müşteri ilişkilerini geliştirmek için yapılan veri madenciliği uygulamaları alanları şöyledir: müşteri değerini ölçmek ve karlı müşterileri elde tutmak, müşterilerin özelliklerini çıkarmak ve bunları pazarlama amaçları için anlamlandırmak, yeni müşteri elde etmek, karlılığı maksimize etmek, müşteri kaybı yönetimini ve tahmin etmeyi içeren müşteri kaybı analizidir. Dolandırıcılık tespiti kapsamında ise dolandırıcılık yapan müşterileri tespit etmek, müşterilerin hesaplarına erişen dolandırıcıları denetlemek ve olağandışı kullanım düzenlerini bulmak için yapılan çalışmalar yer almaktadır. Telekomünikasyon sektörü ağ yönetiminde ise ağ hatalarını belirlemek ve tahmin etmek, sistemin iş yükünü ve kaynak kullanımını yönetmek, kullanıcı gruplarının davranışlarını belirlemek ve veri trafiğini karşılaştırmak veri madenciliğinde uygulama alanı bulmuştur [37].

2.3.1. Telekomünikasyon Sektöründe Pazarlamada Alanında Veri Madenciliği Uygulamaları

Telekomünikasyon sektörü; müşterileri hakkında çok büyük miktarda bilgi edinebilmektedir ve bu nedenle her müşteriden elde edilen karı maksimize etmek, müşteriyi elde tutmak ve müşterisini tanımak için veri madenciliği tekniklerini kullanan öncü sektördür.

Arama detaylarından müşteri profilini çıkarmak ve pazarlama faaliyetleri için bu profilleri veri madenciliği tekniklerini uygulayarak incelemek genelde uygulanan yöntemlerdendir. Bu yaklaşıma basit bir örnek verecek olursak; bir telefon hattının daha çok faks mı yoksa telefon görüşmesi için mi kullanıldığının incelenip daha sonra da kullanılan hattın işyerine mi yoksa eve mi ait olduğunun sınıflandırılması olarak kullanılmıştır [38].

Veri madenciliği teknikleri şirketlere geçmiş verilerden, ayrılmak üzere olan müşteriyi tahmin etmeye olanak sağlamaktadır. Bu teknikler fatura bilgilerinden, arama detay verisinden, abonelik bilgilerinden ve müşteri bilgilerinden yararlanırlar. Bu bilgilere göre müşteri profili oluşturulur. Daha sonra müşteri profilleri pazarlama amacıyla ya da ayrılmak üzere olan müşteriyi daha iyi anlayıp vazgeçirmek için tahminleme modelleri kurmada kullanılabilir. Bu modelleri temel alarak şirketler müşteriler için yeni kampanyalar gibi eylem planları oluşturabilirler.

Telekomünikasyon sektöründe son zamanlarda yapılan pazarlama ve müşterilere yönelik veri madenciliği çalışmalarının başlıcaları:

- Ahn et al. [9] yılında yaptıkları çalışmada çapraz satışta kullanılabilecek bir müşteri sınıflandırma modeli önermişlerdir. Müşterinin daha önce kullanmış olduğu ürünler, demografik özellikleri gibi veriler kullanılmıştır. Çalışmada birçok sınıflandırma tekniği: lojistik regresyon, yapay sinir ağları ve karar ağaçları kullanılmıştır. Bu yöntemlerle birbirinden bağımsız olarak yeni ürün satın alma tahmin edilmiştir. Daha sonra bulunan bütün olasılıkları genetik algoritma ile birleştirerek son karar belirlenmiştir.

- Cheng ve Sun [39] çalışmalarında ilk olarak güncel telekomünikasyon sektör özelliklerine göre değerli müşterileri belirlemek için yeni bir model önermişler ve daha sonra müşterileri segmente etmişlerdir. Tayvan’ daki 3G mobil servis sağlayıcı müşterilerinin arama detay bilgileri kullanılarak kurallar çıkarılmıştır. Bunlara göre de 3G operatörlerine bazı pazarlama önerileri sunulmuştur.
- Lim et al. [40] mobil telekomünikasyon sektörü için Çin’ de yeni bir bölgesel kümeleme yaklaşımı önermiştir.
- Chen et al. [41] sabit hat kullanıcılarının ödeme davranışlarını tahmin etmek için birliktelik kuralları, kümeleme ve karar ağaçlarından yararlanarak bir veri madenciliği stratejisi geliştirmişlerdir.
- Vidya et al. [42] mobil telefon satıcılarının itibar analizi için Twitter verilerini kullanarak duygu analizi yapmışlardır. Duyguları kıyaslamak için 3 ayrı sınıflandırma algoritmasından elde edilen sonuçlar karşılaştırılmıştır.

Geçtiğimiz birkaç yıl içinde, telekomünikasyon sektöründe pazarlama uygulamaları daha çok yeni müşteriler elde etmekten çok mevcut müşterileri korumaya yönelmiştir. Bu yönelimin meydana gelmesinin sebebi ise yeni müşteri elde etmenin maliyeti mevcut müşteriyi elde tutmanın maliyetinden daha yüksek olmasından kaynaklanmaktadır.

Müşterinin kullanmakta olduğu servis sağlayıcıdan ayrılması “müşteri kaybı” olarak ifade edilir ve müşteri kaybı analizi müşteri davranışlarının modellenmesinden oluşur. Müşteri kaybı telekomünikasyon sektörü için çok önemli bir konudur ve şirketler müşterilerine rutin olarak servis sağlayıcılarını değiştirmeleri için teklif vermektedirler. Birçok sistem ve metot müşteri kaybını tahmin etmek için önerilmiştir [43]. Tez konusu olan müşteri kaybı yönetimi ile ilgili çalışmalar ilerleyen bölümlerde daha detaylı incelenmiştir.

2.3.2. Telekomünikasyon Sektöründe Dolandırıcılık Tespitinde Veri Madenciliği Uygulamaları

Telekomünikasyon şirketleri için dolandırıcılık tespiti her yıl milyarlarca dolar kayba sebep olabilen ciddi bir problemdir [36]. Dolandırıcılık tespit uygulamaları standart veri

madenciliği uygulama tekniklerinde modifikasyonlara gereksinim duyar. Daha karmaşık maliyet duyarlı metrik sistemler değerlendirmede kullanılmalıdır. Bu çok önemlidir çünkü sınıfların yanlış sınıflandırma maliyeti birbirinden farklıdır [44]. Bu sebeple dolandırıcılığı belirlemek için bir sınıflayıcıyı modellerken sahte aramaya izin vermenin maliyeti ile meşru bir müşteriden bir çağrıyı engellemenin maliyeti arasındaki göreceli durumu dikkate almak gerekir.

Veri madenciliği teknikleri simetrik olmayan sınıf dağılımlarının olduğu durumlarda nadir sınıfı tahmin etmede zorlanmaktadır. Örneğin sahte aramaların bütün aramalardaki oranı %2' dir. Bu durumla başa çıkabilmek için maliyet duyarlı sınıflandırma tekniklerinden yararlanılabilir. Literatürde bu konuda güncel çalışmalar yer almaktadır.

Farvaresh ve Sepehri 2011 [45] yılındaki çalışmalarında Tahran telekomünikasyon şirketinden alınan gerçek veri ile veri madenciliği tekniklerini kullanarak dolandırıcılık tespiti yapmışlardır. Önişleme, kümeleme ve sınıflandırma aşamalarından oluşan bir hibrid yaklaşım uygulamışlardır. Olszewski [46], iki sınıflı bir problem olan hileli hesapların sınıflandırılması üzerine bir çalışma yapmıştır.

2.3.3. Telekomünikasyon Sektöründe Ağ Yönetiminde Veri Madenciliği Uygulamaları

Telekomünikasyon ağları ziyadesiyle karmaşık yazılım ve donanım yapıları içerir. Bu ağları etkin bir şekilde yönetebilmek için meydana gelmeden önce ağ hataları belirlenmeli ve alarmlar otomatik olarak analiz edilmelidir. Şebekenin güvenilirliğini korumak için önleyici müdahaleler önemlidir. Verinin boyutundan dolayı tek bir hata birçok alarma sebep olabilmektedir. Veri madenciliği teknikleri hataları belirlemede kurallar üretmede önemli rol oynamaktadır. Örneğin; Chao et al. [47] Beijing üniversitesinden elde ettikleri verilerle olağandışı internet kalitesini k-ortalamlar kümeleme algoritması ile analiz etmişlerdir. Ren et al. [48] k-ortalamlar kümeleme yöntemi ile telekomünikasyon ağları için bir kümeleme çalışması yapmıştır. Velmurugan [49] çalışmasında sunucu yerleşimleri için k-Ortalamlar ve Bulanık C-Ortalamlar kümeleme yöntemlerini uygulamıştır.

2.4. Telekomünikasyon Sektöründe Müşteri Kaybı Yönetimi

Telekomünikasyon müşteri kaybı üzerine kurulu bir sektördür. Müşteri kaybının avantajlarından yararlanabilmek için tüketicilere sunulan her ürünün ve hizmetin çok sınırlı bir raf ömrü olduğu gerçeğini kabul etmek gerekir. Telekomünikasyon sağlayıcılarının işi, en iyi telekomünikasyon teknolojisini bulup mümkün olduğunca uzun süre tüketicilerin kullanmasını sağlamak değil, müşterilerin ihtiyacını okumayı öğrenmek ve bunları müşterilere sağlamaktır. Başka bir deyişle, amaç onlara son telekomünikasyon ürününü sunmak değil, bir sonraki telekomünikasyon ürününü sunmaktır.

2.4.1. Müşteri Kaybı

Bir müşteri kullanmakta olduğu servis sağlayıcısını bırakır da başka bir sağlayıcısına geçerse buna müşteri kaybı denir. Müşteri kaybı bankalar, sigorta şirketleri ve tabii ki de telekomünikasyon sektörü için ciddi bir problemdir. Berson et al. [50] müşteri kaybını kablosuz telekomünikasyon servis sektöründe müşterinin bir sağlayıcıdan başka bir sağlayıcıya hareketini ifade eden bir terim olarak kullanmışlardır. Müşteri kaybı firmadaki her bir departman için farklı anlama gelebilmektedir. Finans departmanı için her ne sebeple olursa olsun artık abone olmayan eski müşteriler anlamına gelmektedir. Pazarlama departmanı ise duruma başka bir servis sağlayıcıyı seçen müşteriler olarak bakmaktadır. Ön ödemeli hatlar departmanındaki çalışanlar içinse son üç ayda hiç kontör yüklemeyenler ayrılan müşteridir. Bunların içinde hiçbir resmi görüşme yapmadan ödeme yapmadıkları için telefon görüşmesi yapmayan müşteriler de yer almaktadır.

Müşteri kaybının genel özellikleri şöyledir [51]:

- *Müşteri kaybı yaygındır.* Şüphesiz, müşteri kaybı dünyadaki her servis sağlayıcı için büyük bir sorun haline gelmiştir veya yakında olacaktır.
- *Müşteri kaybı kaçınılmazdır.* Telekomünikasyon sektöründe genel olarak çeşitli nedenlerden dolayı müşteri kaybı kaçınılmazdır. Bunun nedenleri;
 - *Teknoloji:* Müşteri kaybının gerçekleşmesinin birinci sebebi, teknolojinin değişmeye devam etmesidir. İcat edilen daha yeni, daha iyi, daha şık ve

daha ucuz alternatifler olduđu sürece müşterilerin tercih ettikleri hizmet/ürün özelliđi giderek artmaktadır. Telekomünikasyon teknolojisi bugünkü haliyle inanılmaz derecede kısa sürede kendi eskimişliğini üretmektedir.

- o *Müşteriler:* On yıldan az bir sürede kablosuz telefon, milyonerlerin oynacağı olmaktan, dünyanın dört bir yanındaki toplumlarda gençliđin ve modernliđin önde gelen sembolü haline gelmiştir.
- o *Rakipler:* Tabii ki, asıl sebep teknoloji ya da müşteriler deđil, rakiplerdir. Yeni ağlar geliştiren, yeni teklifler sunan, yeni teknolojileri tanıtan ve en önemlisi fiyatları düşüren rakipler. Piyasada yeni iş arayışında olan rakipler varsa, bakacakları ilk yer diđer servis sağlayıcıların müşterileridir.
- *Müşteri kaybı pahalıdır.* Müşteri kaybının birçok sonucu vardır ve çoğunun parasal bedeli vardır. Kaybedilen müşteri, kaybedilen kazanç demektir. Müşteri kaybını önlemeye yönelik çabalara rağmen, kaçınılmaz olarak müşterilerin bir kısmı er ya da geç ayrılacaktır. Bu müşterileri kazanmak için yapılan tekrar kazanma kampanyaları genellikle başarılı ancak elbette ki aynı zamanda maliyetlidir. Daha etkin olan aslında ilk etapta müşterilerin ayrılmasını önlemeye yardımcı olan kampanyalar oluşturmaktadır. Müşteri kaybının önüne geçebilmek için şirketlerin yapacağı ilk şeylerden biri, reklâm giderlerini artırmaktır. Bu savaşın yarattığı giderlerin artmasına ek olarak, fiyatları sürekli olarak düşüren bir rekabet de söz konusudur. Açıkçası, bu aşağı yönlü fiyat spirali giderek daha fazla gelir kaybına neden olmaktadır.
- *Müşteri kaybını yönetmek zordur.* Müşteri kaybı aniden ve tahmin edilemeden ortaya çıkar. Ayrılacak müşteriyi tahmin etmek ve neden ayrılmak istediklerini anlamak çok zordur. Sonuçta, müşterilerin neden ayrıldığı anlaşılabilirse, önlem alınabilir. Ve elbette, ne zaman veya neden olacağı bilinmeyen bir duruma karşı bir planı yapmak mümkün değildir.
- *Müşteri kaybı fırsattır.* Her şeye rağmen aslında müşteri kaybını; pazar pozisyonunda, karlılıkta ve şirketin organizasyonel yapısında köklü değişimler yaparak fırsata çevirmek mümkündür.

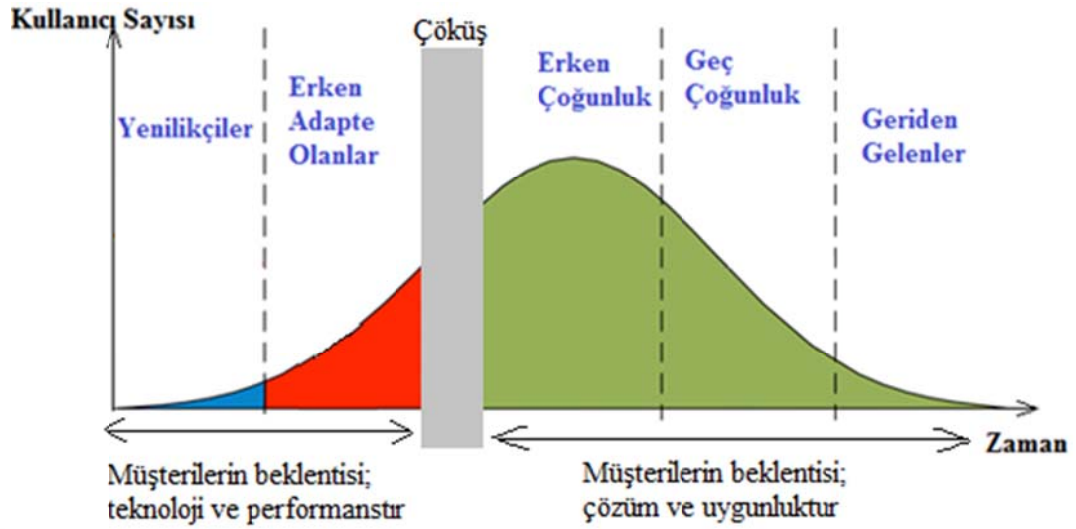
Açıkçası, teknolojinin varlığı ile etkilenmeyen ve değişime uğramayan hiçbir sektör yoktur. Teknoloji seyahat ve turizm sektöründe önemli bir rol oynamaktadır. Mobil cihazlardan, evden ya da bir masaüstü ya da dizüstü bilgisayar kullanarak işyerinden, internet ve teknolojinin kullanımı ile daha günlük yaşam daha kolaylaşmaktadır. Bununla birlikte, küreselleşme ve teknolojinin sınırlarını kaldırması, seyahat ve turizm endüstrisinin gelişmesi için sayısız fırsat sunmuştur.

Teknolojideki gelişmeler müşterilere bağımsız seçim imkânı sağlamıştır. Web siteleri, rezervasyon platformları, uygulamalar ve sosyal medya; gelişmiş bir seyahat deneyimi sağlamak için müşterilere çeşitli düzeylerde müdahale edebilmektedir. Müşteriler artık bilgiye kolayca erişilebilmektedir bu nedenle şirketlerin daha akıllı hale gelmeleri ve rekabet edebilmeleri için de teknolojiye adapte olmaları önemlidir.

Teknoloji, finans sektörünün çalışma biçiminde devrim yaratmaktadır. Hızlı teknolojik gelişmeler; dijital para birimleri, mobil bankacılık, çevrimiçi yatırım ve yeni ödeme sistemleri gibi tamamen yeni iş fırsatları ortaya koymuştur. Müşterilerin kendi tasarruflarını ve yatırımlarını yönetmesine olanak sağlamıştır. Teknoloji, müşteri beklentilerini şekillendiren sosyal medyayı kullanarak finansal kuruluşların pazarlama yeteneğini geliştirmiştir.

Son birkaç on yılda tüketicilerin ve işletmelerin yeni teknolojiyi günlük işlemlerinde nasıl özümstedikleri hakkında birçok araştırma yapılmıştır. Metodolojik açıdan yeni teknolojiye direnç bağımlı bir değişken olduğu kadar bağımsızdır da çünkü teknolojiye karşı kullanıcının tepkisi yeni teknolojiyi daimî bir süreçte etkiler.

Yeni teknolojik ürün geliştirme Ar-Ge yatırımlarının teknolojideki iş tarafını kapsamaktadır. Diğer önemli husus, tüketiciler yeni ürüne adapte olma konusunda farklılaşmaktadır. Bu nedenle, teknolojik ürünün yaşam döngüsü ve müşterilerin bu ürünü nasıl kabul edeceği göz önüne alınmalıdır. Bu modellerden yararlanarak işletmeler ve kurumlar, teknolojil olgunlaştıkça yatırım getirilerini belirlemek için bazı öngörülerde bulunabilirler. Şekil 2.1. yeni teknoloji adaptasyon yaşam döngüsü eğrisini göstermektedir. Eğri, farklı insanların farklı zamanlarda yeni teknolojiye nasıl tepki gösterdikleri hakkında fikir vermektedir. Grafiğin dikey eksen, yeni teknolojiyi kullanan kişilerin sayısını, yatay eksen ise zamanı göstermektedir.



Şekil 2.1. Yeni teknoloji adaptasyon eğrisi [32]

Teknoloji ilk olarak piyasaya sunulduğunda, hemen denemek isteyen az sayıda kullanıcı şekilde yenilikçiler olarak gösterilmiştir. Bu müşteriler, teknolojik gelişmelere son derece ilgi duyan, risk odaklı, öncü fikir sahibi kişilerdir. Yenilikçiler genel tüketici nüfusunun kısıtlı bir parçasıdır. Bu kesim müşteriler önemlidir çünkü yeni ürünü test eden kişilerdir. Ancak erken adapte olanlar yeni teknolojiyi sevdikleri için birçoğu keşfedilecek bir sonraki yeniliği aramaya devam ederler.

Yenilikçilerden sonra daha büyük ama nispeten küçük bir kesim olan erken adapte olan müşteriler genellikle risk odaklıdır ve yeni teknolojiye kolayca adapte olabilirler. Bu müşteriler; yeni ürünleri kucaklayan, yenilikçileri takip eden, genç ve eğitimli müşterilerdir.

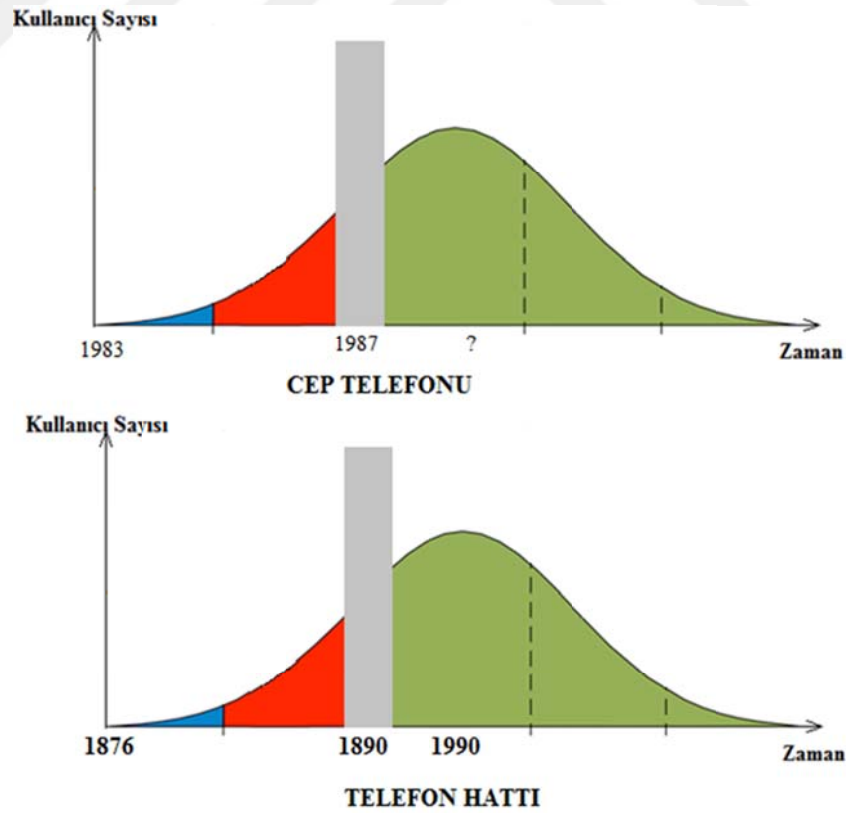
Erken adapte olanlardan sonra çöküş döneminde, zaten nispeten az sayıda olan kullanıcı sayısı düşmektedir. Bu dönem teknolojinin kitlesel pazara hazır hale gelmesinden önceki zaman dilimini göstermektedir ve geçici bir durgunluk dönemidir.

Önceki iki gruba kıyasla daha büyük ve daha dikkatli olan çoğunluğun yer aldığı erken çoğunluğu oluşturan müşteriler yeni fikirlere açıktır, ancak genellikle yatırım yapmadan önce neyle karşılaştıklarını görmek için beklemektedir. Bu müşteriler piyasanın büyüme aşamasını oluşturmaktadır. Büyüme aşamasında, yeni teknolojiyi kullanmaya kaydolun insan sayısının hızla, çoğunlukla katlanarak arttığı görülmektedir. Daha fazla tüketicinin beğenisini çektiği gibi ürünün onaylanması ve arkadaşlarına bu konuda bilgi vermesi, bir ivme oluşturmaktadır ve büyüme daha hızlı ilerler.

Geç çoğunluk; biraz muhafazakâr ve riskten kaçınan, yeni bir şeye yatırım yapmadan önce ikna edilmeye ihtiyaç duyan potansiyel müşterilerin büyük bir grubudur. Bu müşteriler ile birlikte piyasa olgunlaşma aşamasındadır. Bu dönemde pazar doymuş hale geldikçe, talep kesintisi görülmeye başlanır. Bu süre zarfında hızlı büyüme durur.

Son olarak piyasaya geriden gelenler girerler. Bu müşteriler; son derece tutumlu, muhafazakâr, risk sevmeyen, yalnızca yeni fikirlere yatırım yapan genellikle genç olmayan ve eğitimsiz bireylerin oluşturduğu küçük bir müşteri kesimidir. Bu müşterilerle birlikte yeni ürünün düşüş aşaması başlamış olur. Bu dönemin, sonunda, teknolojiye alternatif ürünler icat edilmiştir ve nihayetinde kullanıcı sayısı azalmaktadır.

Telekomünikasyon sektöründe yeni teknolojinin adaptasyon eğrisini iyi analiz edebilmek, müşteri kaybıyla ilgili birçok sorunu anlamaya yardımcı olabilmektedir. Yeni teknolojinin adaptasyon döngüsünü kablosuz telekomünikasyon sektöründeki uygulaması Şekil 2.2.' de gösterilmiştir.

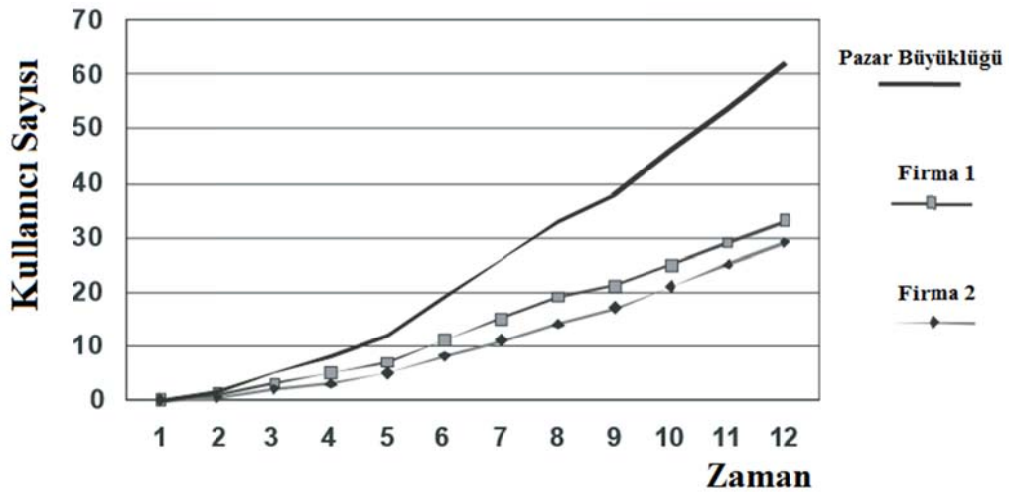


Şekil 2.2. Telekomünikasyon sektörü yeni teknoloji adaptasyon eğrisi [51]

Bu yeni teknoloji adaptasyon eğrisinin şeklini anlamak, mevcut piyasa pozisyonunu eğri üzerinde belirlemek ve eğrinin zaman içinde nasıl yeniden şekilleneceğini tahmin edebilmek; telekomünikasyon şirketlerinin genel pazarlama stratejilerini ve tüketici yönetimini nasıl yapacağına ilişkin kararlarında kritik önem taşımaktadır.

Erken adapte olanlar, yeni teknolojiler sebebiyle ilk günlerde ayrılırlar. Fiyat, uygunluk, sosyal / psikolojik ve kalite nedenlerinden ötürü müşteriler yaşam döngüsünün çok ilerisinde, büyüme ve olgunluk dönemlerinde, ayrılmaya başlarlar.

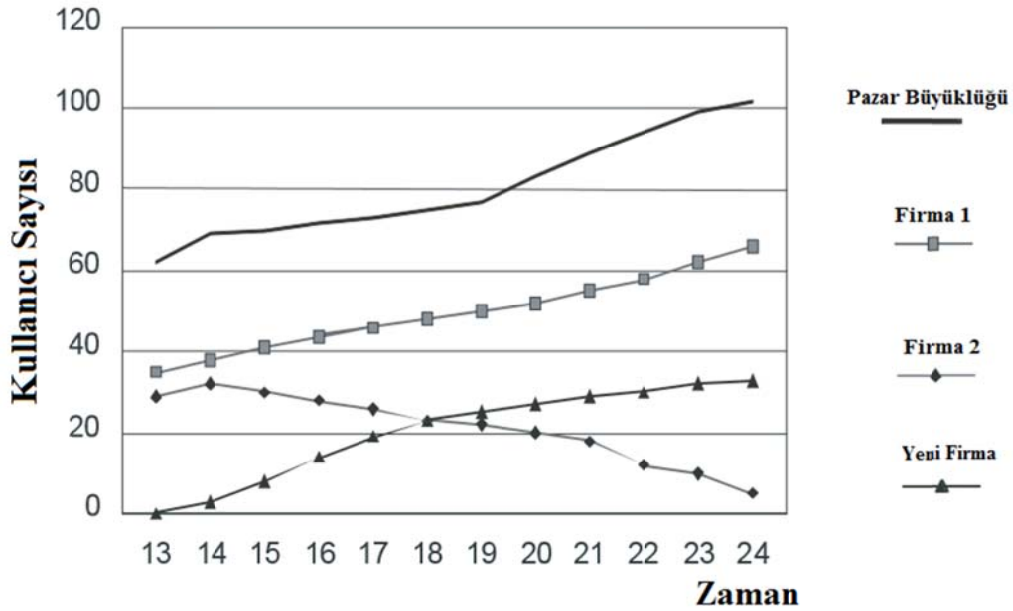
Büyüme döneminde, piyasada rekabet artar ve firmalar mümkün olduğunca çabuk ve çok müşteri çekebilmek için farklı pazarlama yaklaşımları kullanırlar. Bu dönemde müşteriler önceki dönemlere nazaran teknolojik ürün hakkında daha fazla bilgiye sahiptir ve bilinçlenmişlerdir. Ürünlerin fiyatları düşerken, firmaların kar marjları artar. Büyüme döneminde şirketlerin pazardan elde edecekleri pay genellikle pazarlama bütçelerinin büyüklüğüne ve yeteneklerine dayanmaktadır. Şekil 2.3, büyüme dönemindeki rakip firmaların tipik bir pazar payı dağılımını göstermektedir. Üstteki çizgi, genel pazardaki büyümenin zamana göre değişimini gösterir. Noktalı çizgiler ise, her bir firmanın pazar büyümesini göstermektedir.



Şekil 2.3. Büyüme Döneminde Pazar Payı [51]

Büyüme dönemi sona erince, olgunlaşma dönemi başlar ve bu dönemde pazar payı daralmıştır. Teknolojinin firmalarca daha etkin kullanılmasıyla beraber maliyetler de düşmektedir. Satışlar bu dönemde en üst seviyelerini görürken kar marjlarında diğer

dönemlere göre düşüşler yaşanır. Firmalar bu dönemde rakiplerine nazaran farklılıklar yaratarak pazar paylarını korumakla kalmayıp artırma da bilirler. Tüketiciler bu dönemde daha iyi kapsama alanı veya çağrı kalitesi sunan rakiplerin tekliflerine daha açıktırlar. Müşteri nispeten memnun değilse, piyasanın olgunluk evresinde, şirketin müşteri ilişkilerini ne kadar iyi yönettiğine bağlı olarak müşteri aboneliklerinde önemli değişiklikler görülebilmektedir. Şekil 2.4. nispi pazar payının olgunlaşma evresinde nasıl değişebileceğini göstermektedir. Tüketici ihtiyacının farkında olan ve buna göre ayarlamalar yapan şirketler, bu süre zarfında birçok ek müşteri kazanabilir. Müşteri isteklerine cevap veremeyen firmalar ise pazarlarının belirgin bir kısmını kaybedebilirler.

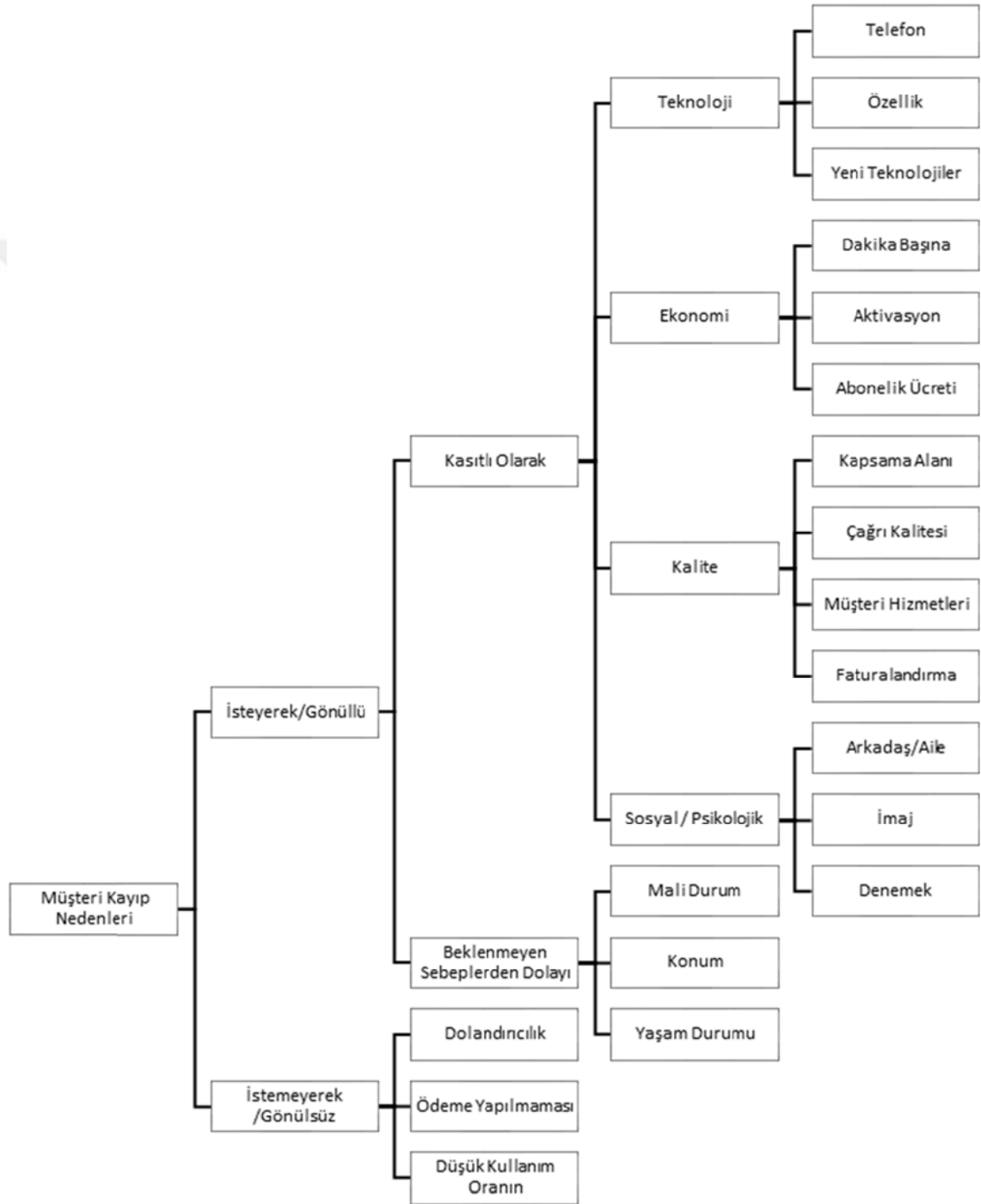


Şekil 2.4. Olgunlaşma Aşamasında Pazar Payı [51]

2.4.2. Müşteri Kaybının Sınıflandırılması

Müşteri yönetiminden bahsetmeden önce farklı türdeki müşteri kayıplarını anlamak çok önemlidir. Müşteri kayıplarını kategorilere ayırmanın birçok yolu vardır. Bunu gerçekleştiği zamana göre (bir ay içinde ayrılan müşteriler, bir yıl sonra ayrılan müşteriler veya sözleşmeleri sona eren müşteriler), müşterilerin ayrılma şekillerine göre (telefonla veya bizzat ayrılanlar ve kayıp kişiler) gruplandırmak mümkündür. Bunlar arasında en yararlı yaklaşım, ayrılma nedenlerine göre kategorize etmektir. Sebeplerine göre kategorize edebilmek için kişinin neden gerçekten ayrıldığının bilinmesi gerekir.

Ancak, müşteriler ayrılmasının asıl sebebini söylemesi pek olası değildir. Gerçek nedenlerden utanıyor olabilir veya şirket çalışanının veya görüşmecinin duygularına zarar vermek istemeyebilirler.



Şekil 2.5. Müşteri kaybının sınıflandırılması [51]

Şekil 2.5.’ te Mattison’ un [51] çalışmasında müşteri kaybının nedenleri sınıflandırılarak gösterilmiştir ve detaylandırılmıştır. Buna göre nedenler iki ana başlıkta incelenmiştir: isteyerek ayrılanlar ve istemeyerek ayrılanlar.

2.4.2.1. İstemeyerek Ayrılan Müşteriler

Bunlar arasında en kolay sınıflandırılanlar istemeyerek ayrılan müşterilerdir. Bu müşteriler firma tarafından abone listelerinden çıkarılanlardır. Bu sınıftaki müşteriler bazı firmalar için çok önemli bir sorunken bazıları içinse bir sınıfı bile oluşturmayabilmektedir. Bu kategorideki müşteriler; dolandırıcılık, ödeme yapmayan ve düşük kullanım oranına sahip kişileri içerir.

- Dolandırıcılık yakın zamana kadar kablosuz iletişim firmaları için önemli sorunlardan biriydi. Suçlular, cep telefonlarının tanımlama kodunu ele geçirip diğer telefonlarda nasıl kullanabileceklerini bulmuşlardı. Klonlama olarak bilinen bu süreç, bir müşterinin telefon numarasının çalınması ve suçluların ücretsiz olarak telefon görüşmesi yapması anlamına geliyordu.
- Ödeme yapılmaması, birçok firma için önemi giderek artan bir sorundur. Bu problem, abonelerin faturalarını genellikle birkaç ay üst üste ödememeleri durumunda ortaya çıkar.
- Bazı telekomünikasyon firmalarının da dikkat etmeye başladığı bir diğer yeni kategori de düşük kullanım oranına sahip müşterilerdir. Müşterilerin sahip oldukları bir telefon veya bir numaradan faydalanmaması, telefon bağlantılarını kesmek için yeterli sebep olabilmektedir. Firma telefonun ne kadar kullanıldığını takip eder ve belli bir süre sonra aboneliklerini iptal eder. (Düşük kullanım yalnızca, ön ödemeli müşteriler veya aylık ücret ödemeyenler için bir sorundur.)

2.4.2.2. İsteyerek Ayrılan Müşteriler

İstemeyerek ayrılan müşteri grubu, müşteri kaybı sorunun büyük bir bölümünü oluştursa da telekomünikasyonda müşteri kaybı düşünüldüğünde genellikle akla gelen isteyerek ayrılan müşteri grubudur. Müşteri, hizmet sözleşmesinin feshini başlattığında isteyerek ayrılma gerçekleşmiş olur.

İsteyerek ayrılma kategorisi, iki ana başlık altında incelenebilir: beklenmeyen sebeplerden dolayı ayrılan müşteriler ve kasıtlı olarak ayrılan müşteriler.

A-Beklenmeyen Sebeplerden Dolayı Ayrılan Müşteriler

Müşteriler planları dışında hayatlarında olan bir değişiklikten dolayı aldıkları hizmetleri sonlandırmaları gerekebilmektedir. Müşteriler;

- Mali durumlarındaki değişiklikten dolayı;
- Konumlarındaki değişiklikten dolayı;
- Yaşam durumundaki önemli değişiklikten dolayı;

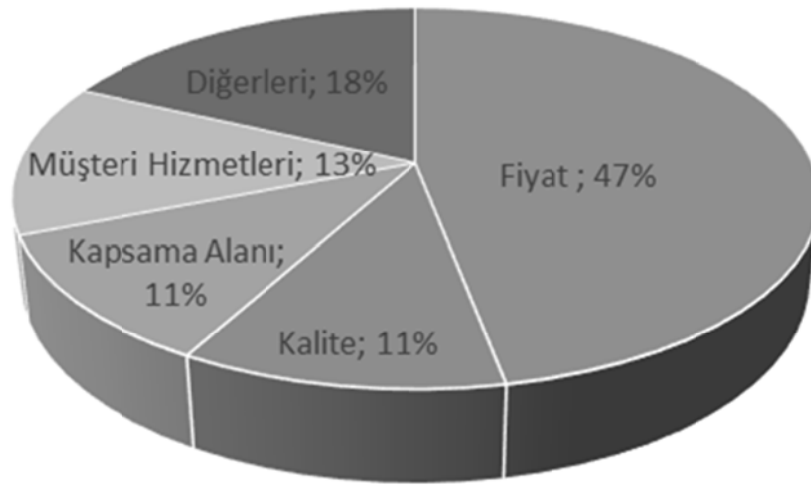
hizmetlerini sonlandırabilmektedir.

B-Kasıtlı Olarak Ayrılan Müşteriler

Çoğunlukla, beklenmeyen sebeplerden dolayı ayrılma çok az sayıda gerçekleşir. Daha da önemlisi, müşterilerin planlayarak ve kasten ayrılmaları durumu söz konusudur. İsteyerek ayrılma konusunda çok fazla pazar araştırması yapılmıştır. Sonuçlar çeşitli pazarlarda, ülkelerde ve kültürlerde farklılıklar göstermektedir; ancak çoğunda, ayrılmanın başlıca nedenleri şunlardır:

1. Fiyat
2. Kalite
3. Kapsama Alanı
4. Müşteri Hizmetleri
5. İmaj

Çok sayıda varyasyon ortaya çıkmasına rağmen, bu beş neden, kaçınılmaz olarak, tüm ayrılma nedenleri listelerinin en üstünde yer almaktadır ve bu tür araştırmaların yapıldığı hemen her ülkede geçerlidir. Şekil 2.6' da ABD'de yapılan kablosuz iletişimle ilgili bir araştırmanın sonuçlarını gösterilmektedir.



Şekil 2.6. Müşteri ayrılma nedenleri

Bu çalışmalar dikkate alınarak kasıtlı olarak ayrılan müşterileri; teknoloji (daha yeni veya daha iyi teknoloji isteyen müşteriler), ekonomi (fiyat hassasiyeti olan müşteriler), kalite, sosyal veya psikolojik nedenlerle ayrılanlar olarak sınıflandırmak mümkündür.

Teknoloji

Teknoloji temelli müşteri ayrılmaları da üç ana kategoriye ayrılabilir: telefon değiştirme, özellik / işlev değiştirme ve ağ değiştirme.

- Telefon değiştirme sebebiyle müşteri ayrılmaları, özellikle Asya'da pek çok pazarda yaygın bir durumdur. Basit olarak eğer müşteri telefonunun yeteri kadar yeni olmadığını düşünüyorsa, yeni bir telefon almanın en kolay yolu yalnızca operatörünü değiştirmektir. Telefonun değiştirilmesi bazı pazarlardaki kablosuz servis sağlayıcıları için son derece önemli bir sorun olabilmektedir. Tüketicilerin telefonunun tam fiyatını veya tam fiyatına yakın ödeme yaptıkları bir pazarda, müşterilerin servis sağlayıcılarını değiştirerek elde edecekleri bir kazanç yoktur. Ancak sübvansiyonlu piyasalarda farklı bir durum söz konusudur. Burada servis sağlayıcısı, müşterinin kullanımlarla geri ödeyeceğini varsayarak telefon maliyetinin çoğunu öder. Burada firmanın müşteriyi servis sağlayıcı hizmeti yüzünden değil, müşterinin kullandığı telefonun teknolojik veya psikolojik olarak yeni olmaması yüzünden kaybetmesi firma açısından üzücü bir durumdur.

- Servis sağlayıcısından özellik veya işlevleri yüzünden ayrılma; telefona benzer şekilde müşteri, bir rakipte yeni özellikler bulunduğunu keşfeder ve daha sonra, bunlara erişmek için operatörünü değiştirir. Sesli posta, çağrı yönlendirmesi, kısa mesaj servisi (SMS) veya başka bir yardımcı servis isteyen müşteriler, onu sunan operatörü ararlar ve bu operatöre geçerler.
- Ağdan dolayı müşteri ayrılmaları, telefon veya özellikten dolayı ayrılmalar gibi başa çıkılabilecek kolay bir ayrılma sebebi değildir. Ağdan dolayı ayrılmalar, alternatif yeni ağ teknolojileri piyasaya girdiğinde ve tüketicilere yeni servise yükseltme fırsatı sunulduğunda gerçekleşir.

Göz önüne alınması gereken başka ayrılma türleri olmasına rağmen, teknoloji temelli ayrılan müşteriler dikkat edilmesi gereken son derece önemli bir gruptur. Teknoloji temelli ayrılmalar, piyasa hakkında önceden uyarı verir. Eğer birçok müşteri teknoloji sebebiyle ayrılıyorsa, ele alınması gereken bazı ciddi uzun vadeli kararlar var demektir.

Ekonomi

Müşterilerin ekonomik sebeplerden ötürü ayrılmaları anlaşılabilir bir durumdur. Hizmet sağlayıcıların şirketlerini benzersiz ve ilginç hale getirmek için gösterdikleri çabalara rağmen tüketicilerin ayrılma sebepleri arasında en can alıcısı her zaman fiyattır. Genişleme aşamasında, rakipler yeni müşterileri davet etmek için fiyatlarını düşürebilir. Ne yazık ki, fiyatın müşteri kazanma aracı olarak kullanılması kararı çok büyük kusurdur. Fiyat indirimlerinin uygulanması kolay olabilir, ancak rakipler için de aynı kolaylıktadır. Bu, müşteri kazanmak için fiyat baskısı yapmaya başlanırsa, rakibin de aynen karşılık verme olasılığı olduğu anlamına gelir. Nihai sonuç açıktır. Ya her firma başladıkları pazardaki göreceli payını koruyacaktır ya da her firma daha önce olduğundan daha düşük bir kâr marjı ile çalışacaktır. Doğrudan fiyat rekabeti ile yaratılan sorunlardan kurtulmak için pek çok şirket, fiyat avantajlarından bahseden programlar, promosyonlar ve kampanyalar oluşturmaktadır. Müşterilerin en çok duyarlı olduğu üç durum söz konusudur.

- Oran planından dolayı ayrılma; bir rakip tarafından müşteriye çağrının süresine ve zamanına dayalı karmaşık değişken oran planları sunulması ve müşterinin başka bir platform üzerinde tercih etmesi durumudur.
- Dakikalık maliyet tüketimi fiyatından dolayı ayrılma; müşterinin, birim başına daha iyi bir ücret almak için servis sağlayıcısını değiştirmesi durumudur.
- Aktivasyon oranından dolayı ayrılma; müşterinin aktivasyon ücretlerini ortadan kaldırdığı veya indirdiği için başka bir servis sağlayıcıyı tercih etmesi durumudur.

Kalite

Fiyat, müşterilerin bir numaralı ayrılma nedeni olsa da ikinci en sık görülen neden kalitedir. Telekomünikasyon sektörü sonuçta bir hizmet sektörüdür ve hizmetin kalitesi müşterilerin satın aldığı şeydir. Tüketici için, kalite ve sosyal / psikolojik değere ilişkin kararlar çok kişiseldir ve son derece zordur. Pazar araştırması, tüketicilerin aldıkları kalite hakkındaki düşünceleri ile gerçekte aldıkları kalite ile ilgili duygularının çok farklı olabileceğini göstermiştir. Çoğu tüketici tarafından kullanılan kalite kriterleri aşağıdaki ana kategorilere girmektedir.

- Kablosuz müşteri için en kolay tanınan kalite kriterlerinden biri kapsama alanıdır. Sonuçta, kablosuz telefon bazı önemli noktalarda çalışmazsa, tüketici memnuniyetsizliği söz konusu olur ve istenilen kapsama alanıyla yeni bir sağlayıcı arayacaktır. Kapsama konuları, fiziksel araziye ve düzenleyici çevreye bağlı olarak karmaşık veya basit olabilir. Diğer düzenleyici ortamlarda, kule yerleşimlerinin ve bant genişliğinin yoğunluğu çok daha kritiktir ve burada tüketiciler, en iyi kapsama alanına sahip taşıyıcılar için net tercihler geliştirecektir.
- Çağrı kalitesi; bir müşterinin istediği kapsama alanının belirlenmesi oldukça kolay olmasına rağmen, bir müşterinin görüşme kalitesi hakkındaki beklentisi tamamen farklı bir konudur. Çağrı kalitesi hakkında nesnel bir fikir edinmek için yapılabilecek en azından birkaç şey vardır. Öncelikle, şebeke içinde kaç çağrı olduğu; ikincisi, müşteri hizmetlerine kalite ile ilgili kaç müşteri şikâyeti geldiği,

üçüncüsü, ağ mühendislerinin olağan teşhis ve testlerini gerçekleştirmesi. Müşterilerin kalite algılamasını belirleyen başka bir ilginç yol da şirketin teknolojik üstünlüğüne ilişkin görüşleridir.

- Telefon servisiyle herhangi bir sorun yaşamış olan müşteri, müşteri hizmetleri personelinin kalitesi veya kalitesizliği ile karşılaşmıştır. Müşteri hizmetleri çalışanları, şirket hakkındaki genel algının kritik bir parçasıdır. Müşteri hizmetleri, bir servis sağlayıcı hakkında müşterilerin şikâyet listesinde çokça yer alır. Tüketicilerin müşteri hizmetleri ile ilgili en büyük şikâyetleri şunlardır: kaba müşteri hizmetleri temsilcileri, şirket veya hizmetleri (tanıtımlar, kampanyalar ve hizmet konuları) hakkında yeterli bilgi sahibi olmayan temsilciler, yanıt vermeyen ya da takip etmeyen temsilciler, müşteriler hakkında bilgi eksikliği olan temsilciler, telekomünikasyon sektörü hakkında genel olarak eğitim eksikliği olan temsilciler. Otomatik çağrı merkezi destek yazılımı, veri tabanları ve eğitim bu sorunlara bir çözüm getirebilir.
- Fatura sistemleri, bir müşterinin şirketinizle olan kalite sorunları listesinin bir parçası değildir. Ancak birkaç faturalandırma sistemi sorunu müşterilerin şikâyet listesinin en üstünde yer alacaktır. Faturalama sistemi sorunları; geçersiz fatura ve faturaları içermektedir. Bunların her ikisi de müşteriyi çok hızlı bir şekilde mutsuz edecek, özellikle tekrarlayan bir sorundur.

Sosyal / Psikolojik

Her ne kadar teknoloji, fiyat ve kalite müşteri kayıplarının en büyük sebeplerinden biri olmasına rağmen, tek başına müşterinin ayrılmasına sebep olmayan ama katkıda bulunan alt faktörler de yer almaktadır. İnsanlar sosyal varlıklardır; arkadaşlar, aile ve diğer toplumsal faktörlerden etkilenirler.

- Arkadaş / Aile; sosyal müşteri kayıp faktörleri arasında en güçlüsü meselesidir. Müşteriler aldıkları kararları bu grupların baskısıyla değiştirebilmektedirler. Aile baskısından daha güçlü olan, özellikle gençler ve genç yetişkinler için akran baskısıdır.
- İmajın arkadaş ve ailenin bir tüketici için oluşturduğu baskıyla büyük bir ilişkisi vardır. Bazı durumlarda, şirket tarafından öngörülen imaj tüketicinin kendi

imajıyla çatışabilir. Bu gibi durumlarda, tüketiciler kendi imkânlarıyla farklı imaja sahip bir şirkete geçebilirler.

- Denemek; yeni şeyler denemek isteyen küçük bir tüketici grubu vardır. Bu tüketiciler, sadece diğer servisin ne olduğunu görmek ve karşılaştırmak için servis sağlayıcılarını değiştirirler. Tüm ayrılma nedenlerinde olduğu gibi, deneme nedeninin gerçekte ne kadar geçerli olduğunu bilmek zordur. Bu nedeni, insanların gerçek sebeplerini söylemek istemediğinde bir bahane olarak söylemesi de mümkündür.

2.4.3. Telekomünikasyon Sektöründe Müşteri Kaybı Yönetimi Alanında Veri Madenciliği Uygulamaları

Küresel pazarda telekomünikasyon sektöründe süregelen bir değişim ve gelişim mevcuttur. Bu gelişmeler ve yeniden yapılanmalar, yeni pazar fırsatları ortaya çıkarmakta ve rekabeti artırmaktadır. Sektörde rekabetin artması beraberinde müşteri kayıplarını, pazar payının korunması ve artırılması gibi konuları gündeme getirmiştir.

Birçok ülkede olduğu gibi Türkiye’de de geçtiğimiz yıllarda yaşanan gelişmelerle telekomünikasyon pazarı doygunluğu ulaşmış ve müşteri kaybının çok yaşandığı ve araştırmaların yoğunlaştığı sektörlerden biri olmuştur. Aynı zamanda, mobil altyapılarındaki ve kamu düzenlemelerindeki standardizasyon bir müşterinin diğer bir servis sağlayıcısına kolayca geçmesine olanak sağlamıştır. Bu sektörde piyasaya yeni kurumların katılımıyla artan rekabetle birlikte müşteri kaybı yönetimi önem kazanmıştır. Mattersion [52]’ in de belirttiği gibi birçok telekomünikasyon yöneticileri için müşteri kaybıyla nasıl başa çıkabileceklerini belirlemek şirketlerinin ayakta kalabilmesi için en önemli konudur. Telekomünikasyon şirketlerinin gelecekteki başarısı veya başarısızlığı, kendi pazarlarını ve müşterilerini yakından tanımlarından geçmektedir.

Müşterileri kazanma, kontrol etme ve elde tutma kabiliyeti kurumun hızına ve etkinliğine bağlıdır. Müşteri kaybı yönetiminde müşteri kazanmaktansa eldeki müşterileri elde tutma yaklaşımı önem kazanmıştır. Dahası literatürde de belirtildiği gibi yeni müşteri kazanmanın maliyeti mevcut müşteriyi elde tutmanın maliyetinin 6 katıdır

[53-56]. Uzun dönem müşteriler ise daha az maliyetli hizmet olarak en yüksek karlılığı getiren müşterilerdir ve de rekabetçi pazar aktivitelerine çok az duyarlıdırlar [57-62]. Ayrıca müşteri kaybının azalan satışlardaki fırsat maliyeti de söz konusudur [63].

Telekomünikasyon sektörü için aylık müşteri kaybı oranı %2,2'dir. Bu da bir şirket yılda müşterilerinin % 27' sini kaybettiği anlamına gelir [43]. Avrupa ve US pazarlarında yıllık müşteri kaybının maliyeti 4 milyar \$ iken bütün dünya genelinde 10 milyar \$'dır [13]. Maliyetlerin yüksek olması ve artan rekabetten dolayı firmalar kaybettikleri müşterilerinin yerine yeni müşteri kazanmak için gerekli finansmanı yapamamaktadırlar. Ayrıca yeni müşteri kazanmanın maliyeti elde tutmaktan daha da yüksek olduğundan müşteri kaybetmemek için daha fazla çaba göstermek etkili olacaktır [3].

Etkili müşteri kaybı yönetimi ile firma hangi müşterilerinin ayrılmaya yakın ve hangilerinin sadık olduğunu belirleyebilmelidir. Firma bu bilgiye sahipse müşteri kaybını önlemek için strateji geliştirebilir ve doğru müşteriye doğru kampanyalarla ulaşabilir. Yerinde karar vermenin yolu da büyük ölçüde doğru bilgiye ulaşmayı sağlayacak yollara sahip olmaya dayanmaktadır [64]. Müşteri kaybı yönetimi, ayrılma potansiyeli olan müşterilerin tahminini yapan ve bunu önlemenin yollarını arayan bir süreç olarak tanımlanabilir. Müşteri kaybı yönetiminin amacı müşteri kaynaklı kayıpların azaltılması ve karlı müşterilerin elde tutulmasıdır [19].

Müşteri kaybı yönetimi modellerinin amacı ayrılma eğiliminde olan müşterileri tahmin etmek, müşteri tutundurma kampanyalarını etkin bir şekilde geliştirmek ve müşteri kaybıyla ilgili maliyetleri azaltmaktır. Maliyeti azaltmak en öncelikli amacı olsa da müşteri ayrılma tahminleme modelleri performans ölçülerini temel alan istatistiklerle genel olarak değerlendirilmiştir [12].

Başarılı bir müşteri kaybı yönetimi; müşteri karlılık ölçüsü, birtakım stratejiler ve bir dizi elde tutma taktikleri ile bir servis sağlayıcıdan diğerine geçme kararı alan müşteriyi azaltabileceğini varsayar [2]. Ayrıca müşteri kaybı tahmin sistemi sadece muhtemel müşterileri belirlemekle kalmamalı aynı zamanda neden ayrılmak isteyebileceklerini de söyleyebilmelidir. Ayrılmak isteyebilecek müşteri belirlendiğinde, pazarlama departmanı o müşteriyi iyi tasarlanmış ve çekici kampanyalarla elde tutmaya çalışır.

Pazarlama departmanının bu yaptığı çabalar netice itibariyle kısıtlı kaynaklar dahilinde yapılması gerektiği için müşteri kaybı tahmin modelleri gerçekten ayrılmayı düşünen müşterileri bulmaya çalışmalıdır.

Literatürde yapılan çalışmalar aşağıdaki şekilde özetlenebilir:

- Gopal ve Meher [65], 100.000 müşteri ve 169 nitelikten oluşan veri seti ile mobil müşteriler kaybını ordinal regresyon ile modellemiştir. Veri seti içerisinde en az 7 aylık eski müşteriler ve 25 aylık veri yer almaktadır. Modelde en iyi 50 nitelik seçilmiş ve veri seti rastgele iki kümeye ayrılmıştır. Ordinal regresyon 86.21% doğruluk verirken çok sınıflı sınıflandırma ile 83.8% doğruluk elde etmiştir.
- C4.5 ve SVM yöntemleri müşteri kaybı tahmini için uygulanmış ve her iki metot da tatmin edici sonuçlar vermiştir [66].
- Random Forest ve PSO tabanlı veri dengeleme araçları müşteri kaybı tahmininde bazı nitelik indirgeme stratejileri ile birlikte uygulanmıştır [67].
- Telekomünikasyon sektörüne etkin bir müşteri kaybı yönetimi sağlamak için anahtar faktörleri çıkarmak için bulanık korelasyon analizlerinin kullanıldığı bir çalışma yapılmıştır [68].
- Tunga ve Karahoca [69], çalışmalarında çok değişkenli sınıflandırma problemi modeli olarak Öklid İndeks HDMR metodunu kullanarak GSM kullanıcılarının ayrılma davranışlarını modellemiştir. Çalışmada kullandıkları veri seti 21 nitelikten oluşmaktadır.

Yukarıda da bahsedildiği üzere farklı veri madenciliği metotları arama detay bilgilerinin farklı niteliklerinden yararlanmışlardır. Nitelik seçimi işlemi göstermiştir ki müşteri kaybı davranışı farklı kullanıcı karakteristiklerine göre değişmektedir. Ayrıca müşteri bağlılık davranışı kayıplarını elimine etmek için promosyonlarla desteklenebilir [70].

Telekomünikasyon sektöründe müşteri kaybı tahmin etme ile ilgili literatürde yapılan daha önceki çalışmalarda daha çok yapay sinir ağları, karar ağaçları ve kümeleme analizleri gibi veri madenciliği teknikleri kullanılmıştır.

- Hung, Yen, ve Wang [6] çalışmalarında Tayvan pazarında müşteri kaybı tahmini için yapay sinir ağlarını kullanmış ve karar ağaçlarına göre daha iyi sonuç elde etmiştir.
- Tayvan telekomünikasyon sektöründe yapılan bir diğer çalışmada abone sözleşme bilgileri ve arama düzenlerindeki değişikliklere göre müşteri kaybı tahminleme modeli geliştirilmiştir [43].
- Nath ve Behara [71], Amerika mobil telekomünikasyon sektöründe 50.000 müşterilik bir veri seti için Naive–Bayes algoritmasını uygulamışlardır.
- Karahoca ve Kara [72], Türkiye’deki müşterileri karlılıklarına göre segmente edip kümeleme teknikleri uygulanmıştır.
- Jahanzeb ve Jabeen [73], Pakistan’daki iki telekomünikasyon şirketinin karşılaştırmasını yapmış ve müşterilerin stratejileri hakkında bir çalışma yapmıştır.
- Huang ve Kechadi [74], müşteri davranışlarını tahmin etmek için denetimli ve de denetimsiz öğrenme yöntemlerini entegre ederek bir hibrid temelli öğrenme modeli önermiştir.
- Huang et al. [67], farklı nitelikler ekleyerek 7 farklı sınıflandırma algoritması ile müşteri kaybı tahmini yapmıştır.
- Keramati et al. [75], İran telekomünikasyon şirketinden aldıkları verilerle karar ağaçları, yapay sinir ağları, en yakın komşu sınıflandırma algoritmaları ile müşteri kaybını tahmin etmiştir.
- Zhang et al. [76], müşteri kayıp tahmininde kişilerarası etkileşimin önemini tahmin etmiştir.
- Tsai ve Chen [77], çalışmalarında multi medya talebi olan müşterilerin kayıp analizini sinir ağları ve karar ağaçları ile yapmıştır.
- Dierkes et al. [4], bir müşterinin diğer bir müşteriye olan etkisini markov lojistik ağları ile incelemiş ve komşu ağların birbirleri üzerinde son derece etkili olduğunu göstermiştir.
- Owczarczuk çalışmasında [78], Polonya’daki mobil telekomünikasyon müşteri kaybının tahmin edilmesi için regresyon ve karar ağacı tekniklerini uygulamış ve etkinliğini incelemiştir.

- Kisioglu ve Topcu [25], Bayes ağlarını kullanarak Türkiye’deki bir telekomünikasyon şirketinden aldıkları verilerle ayrılmaya meyilli müşterilerin davranışlarını belirlemek için bir model oluşturmuşlardır.

Uygun veri tabanı boyutu ve kalitesi ile veri madenciliği teknolojisi, yeni fırsatlar yaratmak için iş zekâsı sağlar. Bütün bu çalışmalar göstermiştir ki müşterileri elde tutmak ve servis sağlayıcı değişimini önlemek için veri madenciliği tekniklerini kullanarak telekomünikasyon sektöründe müşteri kaybını tahmin etmek çok önemlidir.

Tablo 2.1'de, telekomünikasyon sektöründe, müşteri kaybı yönetiminde veri madenciliği uygulamaları incelenmiş; yeni ve ana çalışmaların örnekleri verilmiştir.

Tablo 2.1. Telekomünikasyon sektöründe, müşteri kaybı yönetiminde veri madenciliği uygulamaları.

[illegible]

Müşteri kaybı yönetimi alanında literatürde yer alan nitelik indirgeme çalışmalarının başlıcaları:

- Huang et al. [80] telekomünikasyon sektöründe NSGA-II' yi temel alan çok amaçlı bir nitelik indirgeme algoritması önermişlerdir. Çalışmada farklı büyüklüklerdeki lokal nitelik indirgeme alt kümeleri seçilip modifiye edilmiş ve daha sonra global etkin nitelik alt kümelerin seçilebilmesi için etkin çözümleri aramada kullanılmıştır. Çalışmada dikkate alınan amaçlar:
 - Genel doğruluk oranı
 - Gerçek abone kayıp oranı
 - Gerçek abone kayıp olmadığı durumun doğruluk oranı
- İdris et al. [67] çalışmalarında önerdikleri yaklaşım Chr-PmRF, PSO ile birlikte mRMR nitelik indirgeme algoritması ve Random Forest sınıflandırma algoritmasını temel almaktadır. Chr-PmRF müşteri kayıp analizinde efektif sonuçlar vermiştir.
- Hong ve Zigang [81] nitelik boyutunu etkin bir şekilde indirgeyen iki aşamalı hibrid bir nitelik indirgeme algoritması önermiştir. Algoritma hem Markov örüntüsü keşif tekniğini hem de geleneksel uzman yaklaşımları kullanmıştır. Önerilen yaklaşımın performansı geleneksel yaklaşımdan daha iyi sonuçlar vermiştir.
- Xie v.d. [82], ensemble öğrenme ile transfer öğrenme yaklaşımlarını kombine etmiş aynı zamanda da transfer öğrenmeyi temel alan bir nitelik indirgeme (FSTE) algoritması önermiştir. Deneysel çalışmalar FSTE algoritmasının geleneksel müşteri kayıp tahmin modellerinden ve bazı transfer öğrenme modellerinden daha iyi sonuç verdiğini göstermiştir.
- Xiao v.d. [83] dinamik transfer ensemble yaklaşımını temel alan bir nitelik indirgeme modeli (FSDTE) önermiştir.

2.4.4. Diğer Sektörler ile Telekomünikasyon Sektörünün Müşteri Kaybı Yönetimi Açısından Değerlendirilmesi

Telekomünikasyon sektöründe müşteri sadakatini sağlamanın neden bu kadar zor olduğunu anlayabilmek için diğer sektörlerle telekomünikasyon sektörü arasındaki müşteri ilişkileri açısından farklılıkları incelemek gerekir. Müşteri odaklı birçok sektör vardır örneğin; bankacılık, perakende, havayolları ve otomobil üreticileri. Bu sektörlerin hepsinde de farklı boyutlarda olsa da müşteri kaybıyla karşılaşılır. Telekomünikasyon sektöründeki durum diğer sektörlerden farklıdır ancak diğer sektörlerde hiçbir şey öğrenemeyecek kadar özel bir yapıya sahip değildir. Aslında, servis sağlayıcıların diğer sektörlerden alabileceği önemli dersler ve iyi örnekler vardır. Öte yandan da telekomünikasyon alanının doğası gereği; sağlayıcılar ile tüketicileri arasındaki ilişki çok özeldir ve bu nedenle farklı bir yönetim yaklaşımı da gerektirir [51].

Perakende Sektörü

Telekomünikasyon sektörü dışındaki diğer sektörlerdeki müşteri kaybı yönetimi incelendiğinde; perakende sektöründen daha fazla tüketicinin kararsızlığına aşına sektör yoktur. Perakendeci her an, müşteriyi sadık kalmanın iyi bir yatırım olduğuna yeniden ikna etmek için bir mücadele içindedir. Bu mücadele sırasında karşılaştıkları sorunlar şöyledir [51]:

- Çoğunlukla, sektördeki firmaların hepsi aynı veya benzer tedarikçi havuzuna erişebilir ve ürünlerini buralardan alırlar. Yani müşteriler aynı marka ürünü birçok farklı perakendecide bulabilir ve satın alabilirler. (Tüketiciler aynı telefonları, aynı kapsama alanını ve aynı hizmetleri çeşitli sağlayıcılardan alabilmeleri gibi.)
- Müşteriler mağazalardan sık sık alışveriş yapmaktadırlar. Örneğin; bazı mağazalardan (örneğin bakkaliye) günlük, bazılarında (örneğin marketler) haftalık, bazılarında (örneğin giyim mağazaları) aylık, bazılarında da birkaç yılda bir kez (örneğin yataklar, mobilya ve aletler) alışveriş yaparlar. Müşteriler tarafından bu kadar sık satın alma kararının verildiği pazar şartlarında; perakendeciler müşterilerin dikkatini çekmek için neredeyse günlük olarak rekabet etmek zorundadır. Tüketicilerin satın alma kararlarını yeniden

değerlendirdiği sıklığa satın alma döngüsü denir ve telekomünikasyon müşterisinin satın alma döngüsünün boyutu ve şekli, karşı karşıya bulunduğu büyük zorluklardan biridir.

Bir perakendeci için, müşteri kaybı, olağan bir durumdur ve işletmeler neredeyse bu durumla günlük olarak karşılaşır. Bunun yanı sıra, günlük işlerinin bir parçası olarak müşteri sadakati ile uğraşan perakendecilerin çeşitli avantajları da vardır. Çoğu perakendecinin, işletmeleri müşterilerinin satın alma alışkanlıklarına uygun bir konumda yer almaktadır. Bu durum perakendeci için hem bir avantaj hem de dezavantaj olabilir. Müşterilerin alışkın olduğu bir konumda yer alması erişim kolaylığı sebebiyle müşteriler tarafından tercih sebebi olacağı için avantajlıdır. Müşteriler bir perakendeciden satın alma kararı verirken; işletmeyi, çalışanlarını ve tüm geçmiş satın alma deneyimlerini bir bütün olarak değerlendirir. İşletmenin yakın konumda yer almasının dezavantajı ise, rakiplerin de aynı seçeneğe sahip olmasıdır. Konum avantajından yararlanmak isteyen aynı tür ürünler konusunda uzmanlaşmış birçok perakendecinin şehrin belirli kesimlerinde kümelenmiş olarak bulmak nadir bir durum değildir.

Müşterilerin satın alma döngüsünün sık olması perakendecilerin sahip oldukları diğer bir avantaj olabilmektedir. Tüketicilerin seçim kararını düzenli olarak yeniden gözden geçirmeleri, perakendecilere müşteri sadakatini yeniden sağlamak veya sağlamlaştırmak için fırsat sunar. Müşteriler her alışverişlerinde perakende satış birimlerinde karar vermelerine yardımcı olan çalışanlarla, diğer müşterilerle tanışır ve bütün bunlar kalıcı izlenimler bırakır.

Perakendeciler için, müşteri kaybını yönetmek için belirgin prensipler vardır. Perakendeci için mağazaya yapılan her ziyaret, müşteriyi kuruluşa sadık kalmaya ikna etmek için yeni bir karar noktası ve yeni bir fırsattır. Perakendecilerin müşteri sadakatini artırmak için kullandığı tekniklerden bazıları şunlardır [51]:

- Her konumda iyi hizmet
- Temiz, iyi organize edilen mağazalar ve cazip bir şekilde sergilenen ürünler
- Mağazayı alışveriş yapmak için iyi bir yer olarak tanıtmak için agresif imaj reklamcılığı

- Rakiplerin müşterilerini alma teşebbüslerini önlemek için iyi tasarlanmış tanıtım

Kısacası, perakendeci müşteri kaybı ile müşterilerin alışveriş kararlarını sürekli olarak güçlendiren satış elemanları, mağaza ambiyansı ve yer çekiciliğinden oluşan bir mağaza deneyimi yaratarak başa çıkmaya çalışır. Tüm bunları yaparken aynı zamanda da rakipleri tarafından müşterilerin satın alma kararlarını etkilemek için iyi hazırlananmış tanıtımlarla da mücadele etmeye çalışır.

Perakende sektörü; şirketler ve tüketiciler arasında gelişebilecek ilişkilerin niteliği hakkında geniş bir bilgi birikimi sağlar ancak perakende tekniklerini doğrudan telekomünikasyon deneyimine uygulanması tam olarak mümkün değildir. Telekomünikasyon hizmetleri için alışveriş / satın alma döngüsü, perakende satışlardan çok farklıdır. Tüketiciler telekomünikasyon sektöründe satın alma kararını her gün vermezler. Servis sağlayıcı değiştirmek müşteriler için büyük bir karardır (Bu, ön ödemeli kartların artan popülaritesi ile biraz değişiyor, ancak bu durum hala temelde geçerlidir).

Perakende sektöründe; perakende birimlerine müşteriler tarafından yapılan tekrar ziyaretler, şirket ve müşteri arasında devam eden bir ilişki kurmak için şirkete birçok fırsat verirken telekomünikasyon sektöründe böyle bir durum söz konusu değildir. Birçok müşteri, fiili olarak telekomünikasyon şirketlerine hiç uğramamıştır ve de birçoğunun hiçbir zaman bir müşteri hizmetleri temsilcisi ile konuşmak için bir sebebi olmamıştır. Aslında, müşteriler açısından baktığımızda, satın alma kararı verdikten sonra, servis sağlayıcısını hiç düşünmemeyi tercih ederler. Servis sağlayıcısını düşündükleri tek zaman, fatura geldiği zaman veya hizmetle ilgili sorun yaşadıkları zamandır.

Havacılık Sektörü

Perakende sektörü şüphesiz müşteri sadakati ile uğraşmak zorunda olan endüstrilerin en belirgin örneklerinden biri olmasına rağmen, havacılık sektöründe müşteri ilişkileri telekomünikasyon sektörüne daha fazla benzemektedir. Telekomünikasyon gibi, havacılık sektörünün de çok farklı ihtiyaç ve tercihleri olan geniş müşteri yelpazesi ile uğraşmaları gerekmektedir, ayrıca müşterilerin düzensiz ve öngörülemeyen alışveriş /

satın alma döngüsü ile de ilgilenmektedirler. Havacılık sektörünün farklı yapısı ve zorlu piyasa koşulları nedeniyle, işletmeler müşteri kaybını önlemek için diğer endüstriler tarafından örnek alınan benzersiz müşteri sadakati teknikleri geliştirmişlerdir.

Şimdiye kadar başlatılan müşteri sadakati programlarının en kapsamlı ve en başarılısı havayolu şirketlerine yönelik sık uçuş programlarıdır. Programın arkasındaki temel ilke çok basittir. Tüketicilerin ne sıklıkta uçtukları takip edilir ve yaptıkları uçuş ödüllendirilir. Müşteri ne kadar sık uçuyorsa, o kadar çok kazanç elde ederler.

Müşterilere havayolunun gelecekteki ücretsiz veya indirimli seyahat elde edebilmeleri için mil veya puan biriktirebildikleri bir üyelik numarası verilir. Birikmiş mil sayısı arttıkça, müşteri statüsü de yükseltilir. Yolcu statüsü yükseldikçe, kazançlar da artar. Havalimanı salonuna ücretsiz giriş, check-in sırasında özel muamele; sadık ve sık yapılan seyahatlerin ödüllendirilmesinin bazı yollarıdır.

Sık uçuş programlarının, başarılı olmasını sağlayan sebeplerden ilki, tüketicinin belirli bir havayoluna karar vermesi için düzenli olarak güçlendirilen bir ortam oluşturmalarıdır. Program, yapısı sayesinde, tüketiciyi sevdiği bir havayolunu seçmeye ve onu mümkün olduğunca sık kullanmaya teşvik etmektedir. Programın başarısını sağlayan diğer bir sebep ise, müşteriler arasında hiçbir ayırım yapılmadan fiyattan bağımsız bir sadakat geliştirmelerini sağlamasıdır. Programın başarısını destekleyen ve güçlendiren bir diğer yapısı ise havayolu personelinin daha yüksek puanlı müşterilere daha iyi hizmet etmeye teşvik eden bir ortam hazırlamasıdır. Müşterinin statüsü biletlerine ve kullandığı güzergâhlara yazılır; müşterinin adı her arandığında bilgisayar ekranında görüntülenir. Müşteri statüsünün takip edilmesi ve görünürlüğü; işletmenin tüm bölümleri tarafından en iyi müşterileri en mutlu hale getirmeye odaklamaya yardımcı olur.

Telekomünikasyon perspektifinden havayolu müşteri sadakati uygulamaları incelendiğinde; doğrudan uygulanabilir olmasa da bir servis sağlayıcı havayolu sık uçuş programlarından çok şey öğrenebilir. Havayolunun temel para birimi olan miller, kolayca farklı değer türlerine dönüştürülebilmektedir. Ancak telekomünikasyonun temel para birimi olan telefon kullanımı dakikası, yalnızca daha fazla dakikalara dönüştürülebilir. Bu, daha sonra, her şeyi fiyat boyutuna indirger ve rakiplerin kolayca yarışabileceği bir alandır.

İki sektör arasındaki bir diğerk fark ise, havayollarının perakendecilerin sahip olduđu aynı avantaja sahip olması, yani tüketicileri ile birebir temaslarının olmasıdır. Müşterilerin fiziksel olarak uçağı binmesi gerekliliğı, havayolu personeli ile temasını sağlar. Bu durum havayolu şirketine, seçim kararını güçlendirecek bir fırsat vermektedir. Telekomünikasyon şirketi, ne yazık ki müşteri sadakatini güçlendirecek müşteri ile fiziksel teması olmadığından benzer bir avantaja sahip değildir. Tüm sağlayıcıların yapabileceğı, müşterilerin telefonda harcadıkları dakikaların en ufak bir kesintiye uğramamasını sağlamak ve mümkün olan en iyi kalitede olduğundan emin olmaktır.



3. BÖLÜM

VERİ MADENCİLİĞİ

3.1. Giriş

“Bilgi çağında yaşıyoruz” popüler bir sözdür; ancak aslında “veri çağında yaşıyoruz”. Her gün bilgisayar ağlarımıza, çeşitli veri depolama aygıtlarına bilim ve mühendislik, tıp veya günlük hayatla ilgili çok fazla miktarda veri aktarılmaktadır. Mevcut veri hacmindeki bu ivmelenerek büyüme, bilgisayarlaşmanın ve veri depolama araçlarının hızlı bir şekilde gelişmesinin bir sonucudur. Telekomünikasyon şebekeleri her gün onlarca petabayt veriyi taşımaktadır. Örneğin bilimsel ve mühendislik alanında; uzaktan algılama, süreç ölçümleri, bilimsel deneyler, sistem performansı, mühendislik gözlemleri gibi sürekli boyutta veri üretilir. Tıp ve sağlık sektörü; tıbbi kayıtlar, hasta izlemleri ve tıbbi görüntüler gibi muazzam miktarda veri üretmektedir. Arama motorları tarafından desteklenen milyarlarca web araması her gün onlarca petabayt veriyi işlemektedir. Dijital resimler ve videolar üreten sosyal medya: bloglar, web sayfaları ve çeşitli sosyal ağlar; giderek daha önemli veri kaynakları haline gelmektedir [90].

“Büyük Veri” ifadesi ilk defa 1997 yılında NASA araştırmacıları Cox ve Ellsworth tarafından kullanıldığından bu yana verinin gittikçe artan boyutu yorumlanmasındaki problemleri de beraberinde getirmiştir. Büyük veriyi dört özelliği ile tanımlamak mümkündür: verinin hacmi, verinin hızı, verinin belirsizliği ve verinin çeşitliliği. Günümüzde her gün çok büyük ölçekli veri, üretilmekte ve veri tabanlarına kaydedilmektedir. Bu kadar veri yoğun ortamda, daha çok veya daha iyi bilgiye nasıl ulaşılabilir sorusundan ziyade, mevcut olan bilgiyi daha nasıl etkin kullanılabilir sorusu önem kazanmaktadır.

Büyük miktarda veriyi bilgi haline dönüştürmek için güçlü ve çok yönlü araçlar gereklidir. Bu gereklilik, veri madenciliğinin doğmasına neden olmuştur. Veri madenciliği, büyük veriyi bilgiye dönüştürür. Örneğin; bir arama motoru, her gün yüz milyonlarca sorgu alır. Kullanıcı arama sorgularını tek tek inceleyip paha biçilemez bilgi elde edebilir. Örneğin, herhangi bir hastalıkla ilgili arama yapan kullanıcıların sayısı, bir salgın olup olmadığı ya da hangi bölgede olduğu hakkında bilgi verebilir [84].

Literatürde veri madenciliği ile ilgili birçok tanım söz konusudur. Bu tanımlar arasında ilk olarak; “veri madenciliği kavramı bilginin etkin kullanımı için disiplinler arası izlenen kurallar bütünüdür” tanımı öne çıkmaktadır. Bu kavram hızla gelişen bilgi teknolojilerinin gereği gerek günlük hayata dair gerekse iş ve bilimsel çalışmalardan büyük verinin analiz edilerek içlerinde saklamakta oldukları bilginin ortaya çıkarılması gerekliliği ile gün yüzüne çıkmıştır [84].

Veri madenciliğinde en önemli konu geçmiş verilerin iyi bir şekilde yorumlanması ile geleceği en iyi şekilde tahmin edebilmeyi sağlayacak en uygun girdilerin ortaya çıkarılabilmesidir. Bu anlamda, mevcut verilerden üstü kapalı ya da çok net olmayan ancak faydalı bilginin çıkarılmasıdır [85].

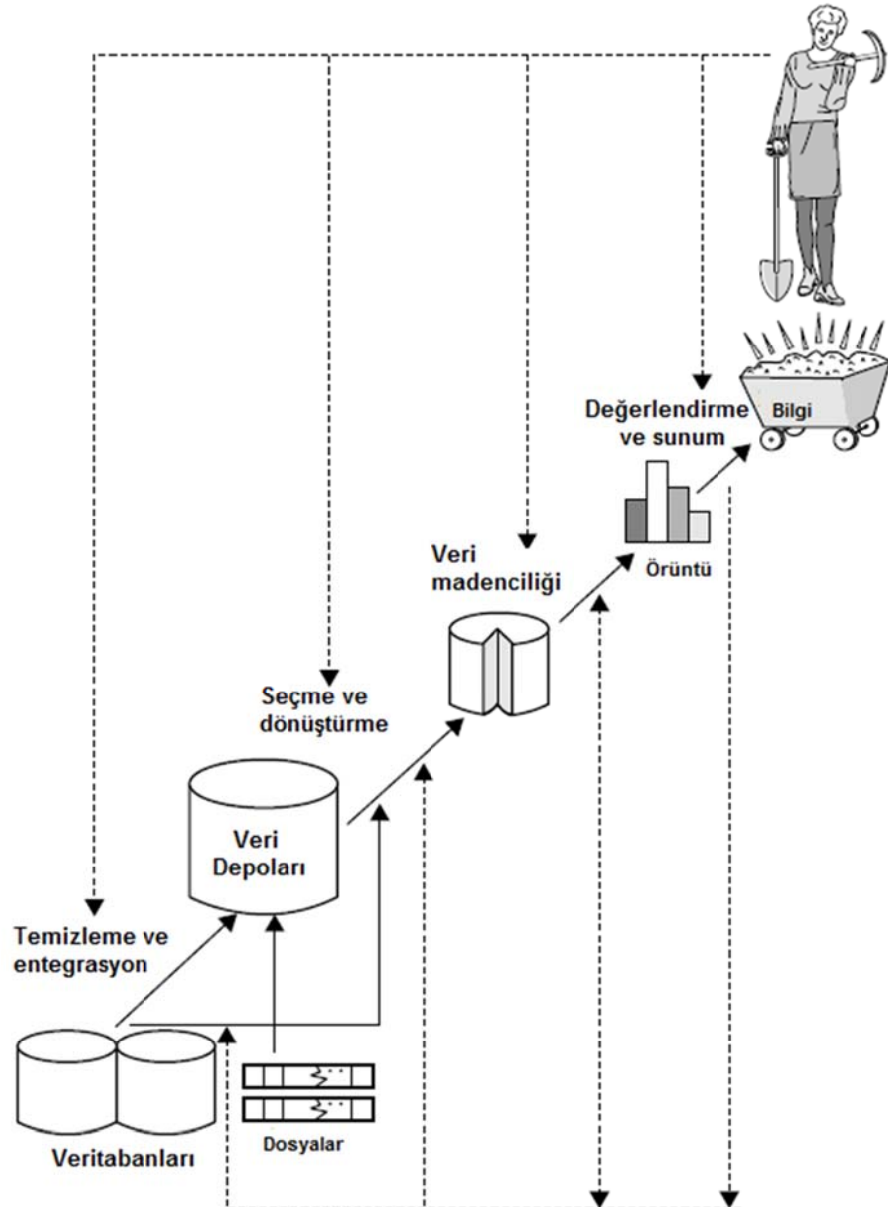
3.2. Veri Madenciliği

Veri madenciliği, bilgi teknolojisinin doğal bir gelişimi ve çeşitli ilgili disiplinlerin ve uygulama alanlarının birleştiği bir alan olarak kabul edilebilir. Veri madenciliğinin birçok boyutu vardır: veri, bilgi, teknolojiler ve uygulamalardır. Teoride veri madenciliği bilgi keşfi sürecinin bir parçası olarak kabul edilirken pratikte veri madenciliği ve bilgi keşfi eş anlamlı olarak kullanılmaktadır [86]. Amaç, daha önceden fark edilmemiş veri desenlerini tespit edebilmektir.

Veri madenciliği, büyük veriden ilginç bilgileri keşfetme sürecinin önemli bir adımıdır. Bilgi bulma süreci aşağıdaki adımların yinelenmeli bir dizisi olarak Şekil 3.1' de gösterilmiştir [84]:

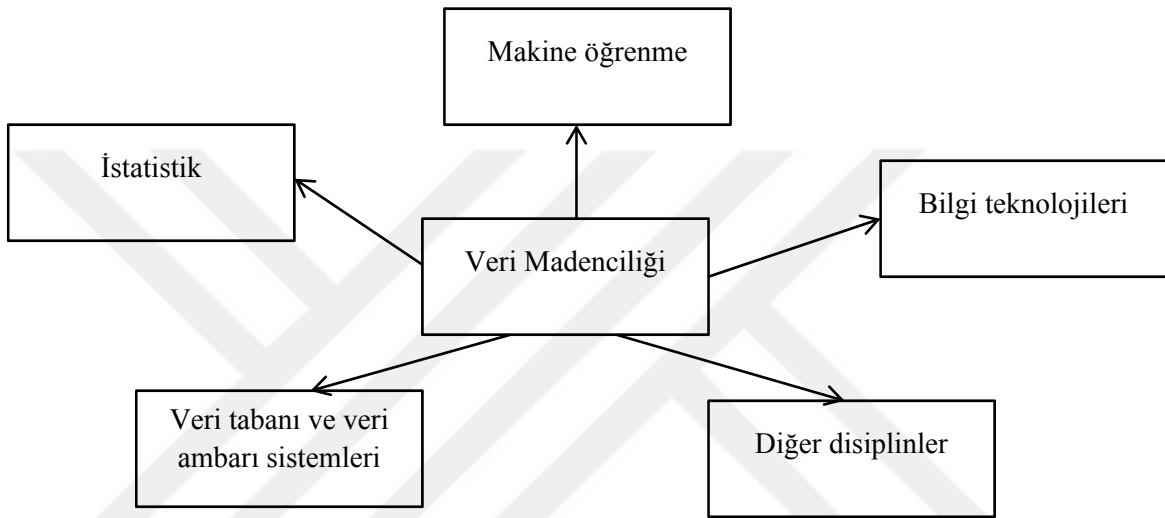
1. Veri temizleme, gürültülü ve tutarsız verilerin elenmesidir.
2. Veri entegrasyonu, birden fazla veri kaynağının birleştirmesidir

3. Veri seçimi, veri tabanından analiz için gerekli verilerin çıkarılmasıdır.
4. Veri dönüşümü, verilerin dönüştürülmesi ve birleştirilmesidir.
5. Veri madenciliği, veri örüntülerini çıkarmak için akıllı yöntemlerin uygulandığı önemli bir süreçtir.
6. Örüntü değerlendirme, bilgiyi temsil edebilecek gerçekten ilginç örüntüleri tanımlamaktır. Eğer bir örüntü, insanlar tarafından kolay anlaşılabilirse ve de test değerleri belirli bir doğruluk düzeyinin üzerindeyse faydalıdır. Faydalı örüntüler bilgiyi temsil eder.
7. Bilgi sunumu, bilgiyi kullanıcılara sunmak için görselleştirme ve bilgi temsil teknikleri kullanılmasıdır.



Şekil 3.1. Bilgi keşfinin adımları [84]

Veri madenciliği, oldukça uygulama odaklı bir alan olarak birçok başka alanın teknolojilerini birleştirmiştir. Bunlar arasında istatistik, makine öğrenme, veri tabanı ve veri ambarı sistemleri ve bilgi teknolojileri bulunmaktadır. Veri madenciliğinin disiplinler arası yapısı, veri madenciliğinin başarısı ve yaygın uygulamalarına katkıda bulunmaktadır. Örneğin; iş zekâsı, Web araması, sağlık bilişimi, finans ve dijital kütüphaneler ve gibi birçok alanda başarılı bir şekilde uygulanmaktadır.



Şekil 3.2. Veri madenciliği sürecine katkı sağlayan alanlar

3.3. Veri Madenciliği Tekniklerinin Sınıflandırılması

3.3.1. Denetimli Öğrenme: Sınıflandırma ve Regresyon

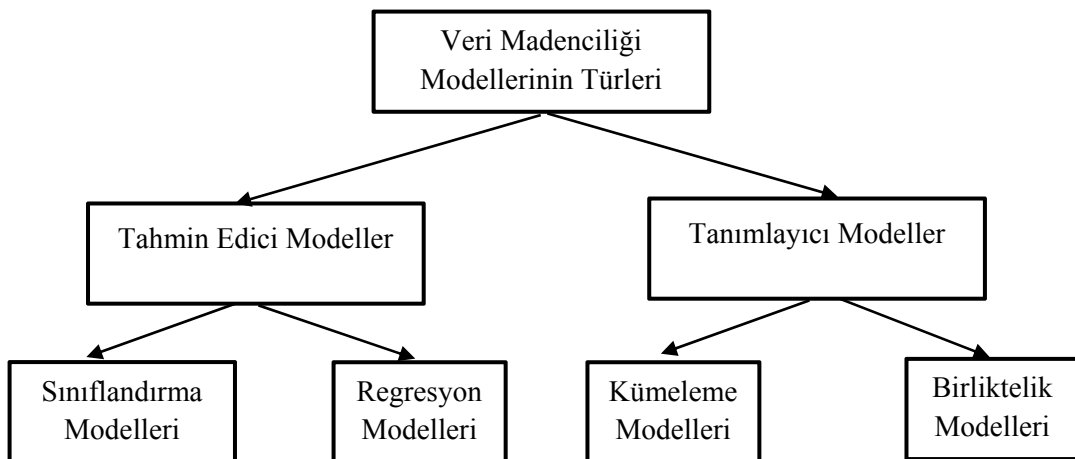
Denetimli öğrenmede sınıf etiketleri dikkate alınır, diğer bir deyişle her bir örneği etiketleyen bir sınıf değişkeni bulunur. Diğer değişkenler öngörücü değişkenlerdir ve sınıf değişkenini tahmin edebilmek için kullanılır. Bazen sınıf değişkeni bağımlı değişken, öngörücü olarak adlandırılan değişkenler bağımsız değişkenler olarak tanımlanabilmektedir. Amaç bağımlı değişkeni bağımsız değişkenlerle açıklayabilmektir. Sınıflandırma teknikleri kategorik bir sınıf değişkenini öngörmekte ve hedef, örnekleri tahmini değişkenlerine göre sınıflandırmaktır. Regresyon teknikleri sayısal bir yanıt değişkenini varsaymaktadır. Amaç, verileri en az hata ile tespit edebilen bir işlevi bulmaktır [87].

3.3.2. Denetimsiz Öğrenme: Kümeleme ve Örüntü Bulma

Denetimsiz öğrenme sınıflandırılmamış veriyi dikkate alır, diğer bir deyişle, değişkenler öngörücü ve sınıf değişkeni olarak ayrılmamıştır. İki tür denetimsiz öğrenme ele alınmıştır: kümeleme ve örüntü bulma. Kümeleme algoritmaları, benzer örneklerin gruplarını bulmak için verileri inceler. Sınıflandırmadan farklı olarak, sınıf değişkeni üzerine yoğunlaşmak yerine tüm örnekleri dikkate alır. Verilerdeki kalıpları keşfetmek için birçok teknik vardır. En iyi bilinen teknik, birliktelik kuralı madenciliğidir. Birliktelik kuralları, denetimsiz öğrenme kullanılarak keşfedilir, diğer bir deyişle, bir sınıf değişkeninin seçilmesine gerek yoktur. Veri madenciliği sonuçları tahmin edici ve tanımlayıcı olabilir. Karar ağaçları, ilişki kuralları, regresyon fonksiyonları model öğrenmek için kullanılan veri seti hakkında bilgi verebilir [8].

3.4. Veri Madenciliğinin Görevleri ve Teknikleri

Veri madenciliği, büyük miktarda veriyi eleyerek, kurumsal bir veri tabanında gizli olan ilginç kalıpları, eğilimleri ve ilişkileri otomatik olarak keşfetme sürecidir. Kuruluşlar için rekabet avantajı yaratmak için artan veri ambarı tabanından yeni bilinmeyen bilgi çıkarma aracı sağlar. Veri madenciliğinin görevleri genel olarak ikiye ayrılır: tahmin edici ve betimleyici. Şekil 3.3.' te veri madenciliği görevleri sınıflandırılmıştır.



Şekil 3.3. Veri madenciliğinin görevleri [88]

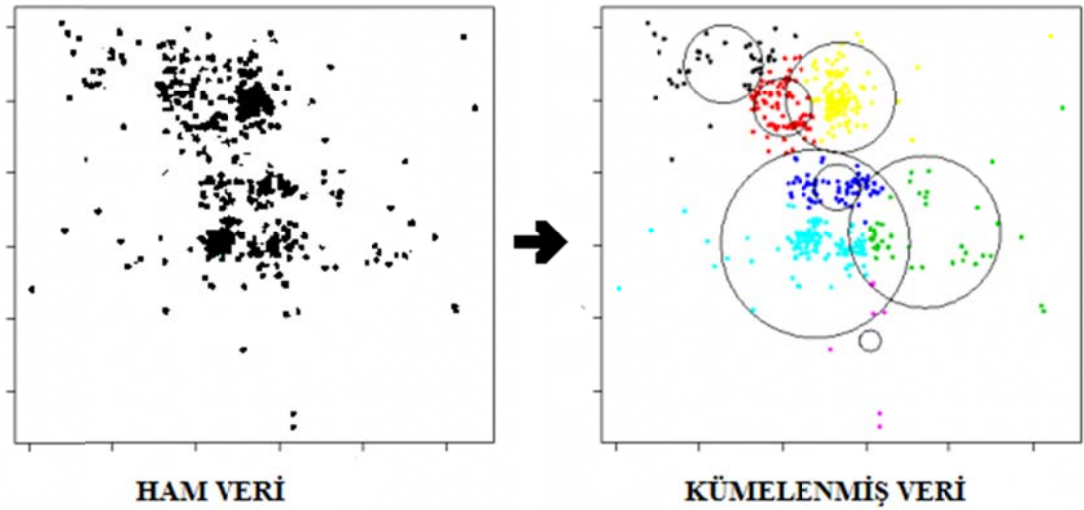
Tahmin edici modeller; veri madenciliğinde gelecekteki sonuçların tahmin edilmesine olanak tanıyan ve verilerdeki eğilimleri, kalıpları ve saklı ilişkileri bulan bir öğrenme sürecidir. Eğer model, sınıf üyeliğini tahmin ederse, sınıflandırma modeli; eğer model olası değerlerden bir sayı tahmin ederse, o zaman regresyon modeli olarak adlandırılır.

Tanımlayıcı modeller; ilginç kalıpları veya ilişkileri bulmak için verileri keşfetmeye ve görselleştirmeye odaklanmış veri madenciliği sürecidir. Bu modeller, mevcut verilerin modellerini tanımlar ve temelde anlamlı alt gruplar oluşturmak için kullanır. Veri karmaşıklığını azaltmak için benzer nesneleri, olayları veya insanları bir araya getirerek kümeler oluşturur. Tanımlayıcı modeller; iki veya daha fazla şeyin ne sıklıkta ilişkilendirildiğini belirleyen modellere birliktelik modelleri denir.

3.4.1. Sınıflandırma ve Tahmin

Sınıflandırma temel olarak, bir veri kümesindeki her bir öğeyi önceden tanımlanmış sınıflar veya gruplardan birine sınıflandırmak için kullanılır. Sınıflandırma yöntemleri; karar ağaçları, doğrusal programlama, sinir ağı ve istatistik gibi matematiksel tekniklerden yararlanmaktadır. Sınıflandırmada, veri öğelerinin sınıflara nasıl sınıflandırılacağını öğrenebilen probleme özgü yazılım geliştirilebilir. Örneğin; şirketten ayrılan çalışanlar hakkında tüm bilgileri verdikten sonra gelecekte kimin ayrılacağını tahmin edilmesini sağlayacak bir sınıflandırma modeli geliştirilebilir. Sınıflandırma ve tahmin yöntemlerini karşılaştırmak için aşağıdaki kriterler önerilmiştir:

- Tahmin doğruluğu, modelin bir veri kümesinin sınıf etiketini doğru olarak tahmin edebilmesidir.
- Hız, modelin hesaplama maliyeti anlamına gelmektedir.
- Sağlamlık, modelin, gürültülü ya da kayıp veri ile başa çıkabilme yeteneğidir.
- Ölçeklendirilebilirlik, modelin büyük veri ile etkin bir şekilde sonuç verebilmesi yeteneğidir.
- Yorumlanabilirlik, modelin çıktılarının anlaşılma seviyesidir.



Şekil 3.4. Kümeleme Analizi

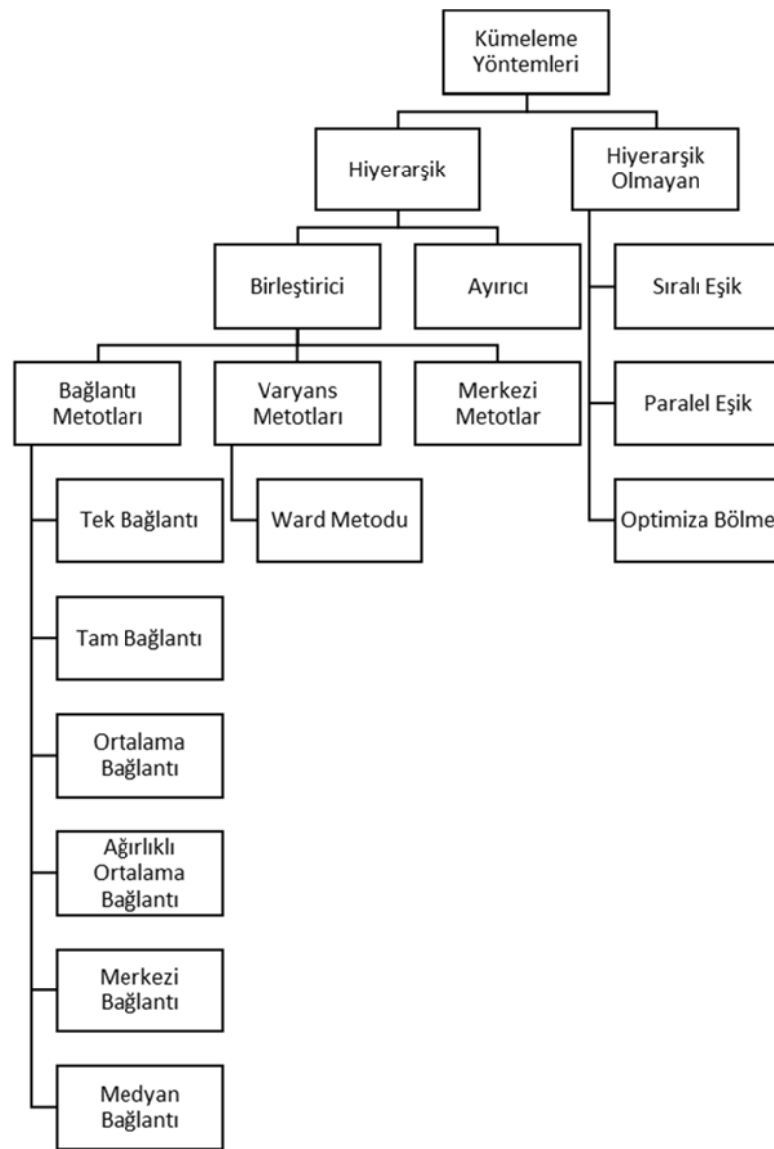
3.4.2. Kümeleme

Kümeleme modelleri benzer özelliklere sahip nesnelerin anlamlı kümelerini oluşturan bir veri madenciliği tekniğidir. Kümeleme tekniğinde, kümeler tanımlanır ve nesnelerin her biri kümelere atanır; sınıflandırma tekniklerinde nesneler önceden tanımlanmış sınıflara atanır. Kümeleme analizinde; ilk olarak nesneler arasındaki benzerliklerin belirlenebilmesi için ölçüler ve değişkenler seçilir ve bu ölçülere göre nesneler kümelendir, daha sonra; belirlenen kümelerin uygunluğu değerlendirilip ve istatistik geçerliliği ortaya konur.

Kümeleme analizinin bazı uygulama alanları,

- Tıp,
- Biyoloji,
- Psikoloji,
- Sosyoloji,
- Arkeoloji.

Kümeleme teknikleri iki ayrı grupta incelenebilir. Bunlar hiyerarşik kümeleme ve hiyerarşik olmayan kümelemedir. En çok kullanılan yöntemler hiyerarşik kümeleme yöntemleridir.



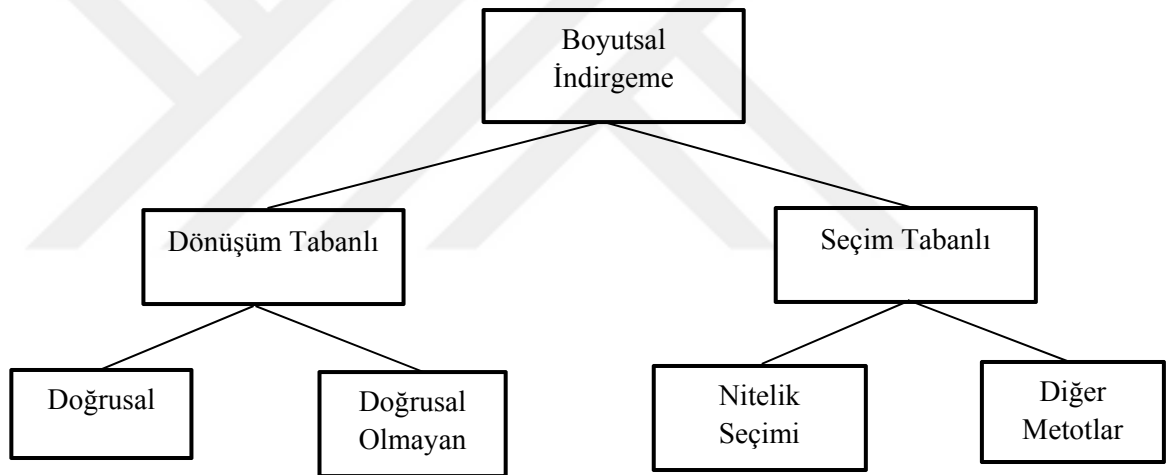
Şekil 3.5. Kümeleme Yöntemleri [84]

3.4.3. Birliktelik Analizi

Birliktelik analizi, en tanınmış veri madenciliği tekniklerinden biridir. Birliktelik analizi, bir veri kümesindeki nesneler arasındaki bağlantıları arayan ve bu bağlantılardan model keşfedilir. Birliktelik analizi, müşterilerin sıklıkla birlikte satın aldıkları bir ürün setini tanımlamak için pazar sepeti analizinde kullanılır. Perakendeciler, müşterinin satın alma alışkanlıklarını araştırmak için birleştirme tekniği kullanmaktadırlar.

3.4.4. Veri Madenciliğinde Nitelik İndirgeme Yaklaşımları

Nitelik indirgeme; veri madenciliği yaklaşımlarından en önemli konulardan biridir. Çok fazla kayıt nitelik arasından çok azının karar kuralı içinde kullanıldığı görülmektedir. Bu durum, sınıflandırma gibi diğer teknikler uygulanmadan önce verinin hazırlanmasının gerekliliğini ortaya koymaktadır. Veri tabanları, gereksiz ve anlamsız niteliklere sahip olabilir. Bu tip gereksiz nitelikler veri madenciliği algoritmasının yavaşlamasına sebep olabilir. Dolayısı ile veri tabanındaki gereksiz verilerin indirgenmesi ve veri kütlesi içerisinde sadece kullanılacak niteliklerin ayırt edilmesi işlemi aslında veri madenciliği algoritmalarının etkinliğinin artırılması için gerekli bir ön işlem olarak yapılmalıdır.



Şekil 3.6. Boyut indirgeme yöntemlerinin sınıflandırılması [85]

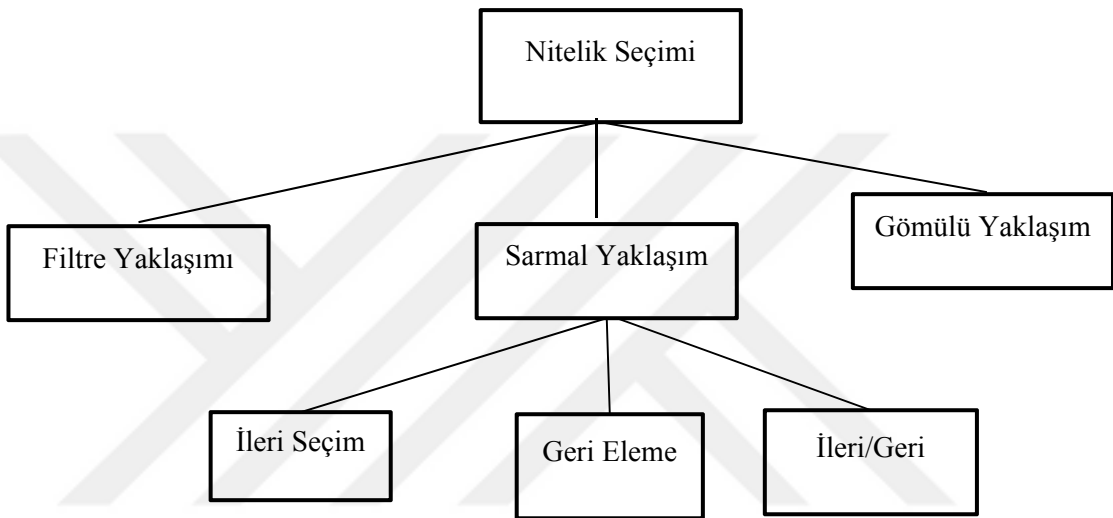
Nitelik indirgeme problemleri üç kategoride değerlendirilir:

- Fazla sayıdaki niteliklerin indirgenmesi için kullanılan metotlar
- Az sayıdaki niteliklerin indirgenmesi için kullanılan metotlar
- Görselleştirme problemleri

Şekil 3.6.'da gösterildiği üzere nitelik indirgeme; verideki anlamsal bütünlüğü bozup bozmamasına göre dönüşüm ve seçim tabanlı olmak üzere iki şekilde incelenebilir. Dönüşüm tabanlı metotlarda indirgeme, orijinal veri kümesinin anlamına ileride

yapılacak işlemlerde ihtiyaç olmadığında kullanılabilir. Temel olarak iki yaklaşımdan bahsedilebilir. Bunlar doğrusal tabanlı ve doğrusal olmayan tabanlı yöntemlerdir.

Doğrusal yöntemler arasında, temel bileşen analizleri, projeksiyon izleme ve çok boyutlu ölçekleme yöntemleri yer almaktadır. Doğrusal yöntemler ile doğrusal olmayan veri ile boyut indirgeme yapılamamaktadır. Doğrusal olmayan ilişkilere sahip veri kümelerinde boyut indirgeme de doğrusal olmayan yöntemlerle uygulanabilmektedir.



Şekil 3.7. Nitelik Seçimi Yaklaşımları [85]

Seçim tabanlı nitelik indirgeme yöntemleri; nitelik seçimi ve diğer metotlar olarak ikiye ayrılır. Nitelik seçimi de Şekil 3.7.' de gösterildiği üzere filtre, sarmal ve gömülü olmak üzere 3 yaklaşıma ayrılır. Filtre yaklaşımında nitelik seçimi herhangi bir öğrenme algoritmasından bağımsız olarak yapılmaktadır. Sarmal yaklaşımda ise değerlendirme işlemi bir sınıflandırma olarak uygulanmaktadır. Gömülü yaklaşımlarda nitelik seçimi ve öğrenme iç içe geçmiş durumdadır [85].

4. BÖLÜM

ÖNERİLEN MÜŞTERİ KAYBI TAHMİN MODELİ

4.1. Giriş

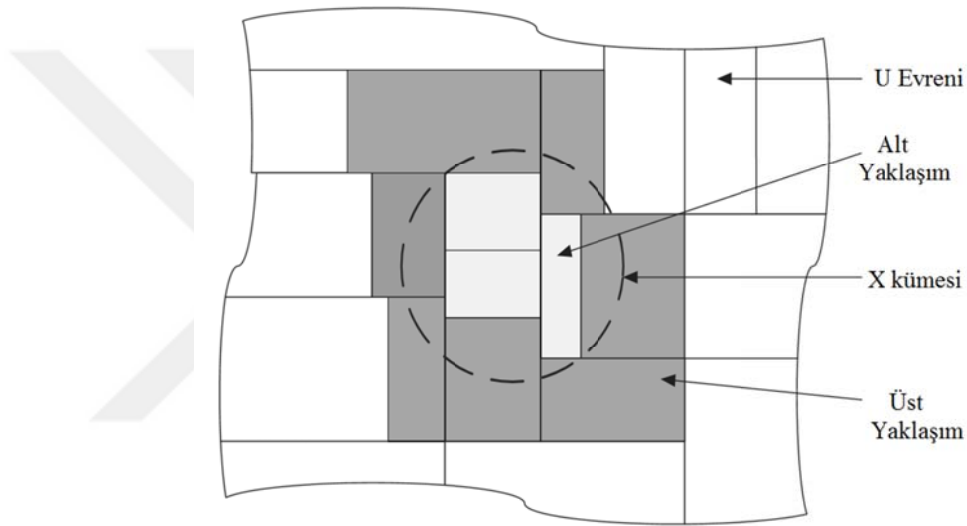
Telekomünikasyon sektöründe piyasa şartları gereği müşteri kaybı analizi giderek önem kazanmaktadır. Bu nedenle hazırlanan tezde telekomünikasyon sektörünün bu problemine yönelik olarak, müşteri verilerinin karakteristiğine uygun entegre müşteri kaybı tahmin modeli önerilmiştir. Telekomünikasyon müşteri verileri dengesiz sınıf dağılımı ve asimetrik yanlış sınıflandırma maliyeti özelliklerine sahiptir. Bu özelliklere sahip verilerle başa çıkabilmek ve yüksek doğrulukla tahmin elde edebilmek için önerilen model maliyet duyarlı olarak kurgulanmıştır. Ayrıca önerilen modelden mümkün olduğunca gerçekçi sonuçlar elde edebilmek için, detaylı çalışmalar ve uzman görüşleriyle, her müşteri için çok sayıda nitelik derlenmiştir. Bu nitelikler veri setine diğer özelliklerine ek olarak büyük boyutluluk özelliğini de eklemiştir. Veri setinin büyük boyutlu karakteristiği ile başa çıkabilmek için de önerilen model, ayrıca nitelik seçimi ve sınıflandırma algoritmalarının entegrasyonu olarak kurgulanmıştır.

Önerilen model iki ana aşamadan oluşmaktadır. Nitelik seçimi aşamasında iki ayrı nitelik seçimi algoritması uygulanmıştır. Bu nitelik seçimi algoritmalarından ilki bulanık kaba küme temelli FRQR nitelik seçimi yaklaşımı diğeri ise çok amaçlı ve maliyet duyarlı olarak uyarlanan NSGA-II temelli nitelik seçimi yaklaşımıdır. Sınıflandırma aşamasında çok amaçlı ve maliyet duyarlı KKO temelli bir sınıflandırma yaklaşımı literatüre kazandırılmış ve ilk aşamada elde edilen nitelik alt kümelerine uygulanmıştır.

Bu bölümde ilk olarak, önerilen model için temel kavramlar incelenmiş daha sonra da önerilen nitelik seçimi ve sınıflandırma yaklaşımları detaylı olarak anlatılmıştır.

4.2. Veri Madenciliğinde Kaba Küme Teorisi Yaklaşımları

Kaba küme analizi, Pawlak tarafından literatüre kazandırılan kaba küme teorisine dayanan bir matematiksel yaklaşımdır. Kaba kümenin amacı, belirsiz veri kaynaklarından bilgiyi keşfetmektir. Kaba küme teorisi, ayırt edilemezlik kavramı ve nesnelerin ayırt edilememesi üzerine kuruludur. Ayrıca hatalı verilerin ele alınmasına yönelik bir yaklaşım olarak kaba küme analizi; olasılık teorisi, kanıt teorisi ve bulanık küme teorisi gibi daha geleneksel teorileri tamamlamaktadır.



Şekil 4.1. Kaba Kümenin Gösterimi [89]

Son yıllarda kaba küme teorisi, gerçekten araştırmacılar için büyük bir ilgi konusu haline gelmiştir ve birçok alanda uygulanmıştır. Bu başarı kısmen teorinin aşağıdaki yönlerine bağlıdır [90].

- Sadece verideki gizlenmiş gerçekler analiz edilir,
- Veriyi analiz edebilmek için eşikler veya uzman görüşü gibi ek bilgiye ihtiyaç duymaz,
- Minimum bilgi temsili elde edilebilir.

Kaba küme teorisi, veri bağımlılıklarını keşfetmek ve yalnızca bir yapısal yöntemle bir veri kümesindeki nitelik sayısını azaltmak için bir teknik olarak kullanılabilmektedir.

4.2.1. Kaba Küme Teorisinde Nitelik Seçimi Yaklaşımı

Kaba küme teorisi; veri analizi için nesnelerin yaklaşık tanımlarını oluşturmada etkili bir tekniktir. Bir kaba küme; alt ve üst yaklaşımlar denilen bir çift kesin kavramlar tarafından tanımlanan belirsiz bir kavramın yaklaşımlarıdır. Bilgi içeriğini koruyarak gereksiz nitelikleri değerli veri kümelerinden çıkarmak için kaba küme kullanılmaktadır. Kaba küme teorisinde nitelik indirgemenin temeli ayırt edilemezlik kavramına dayanmaktadır. Bilgi sistemi $I = (U, A)$ ile ifade edilir ve U boş olmayan örnekler kümesini, A boş olmayan nitelikler kümesini ve $\forall a \in A, V_a$ ise a niteliğinin alabildiği değerler kümesini ifade eder [90].

Karar sisteminde $A = \{C \cup D\}$, C koşullu nitelikleri ve D karar niteliğini göstermektedir. Herhangi bir $P \subseteq A$ ile ilişkili $IND(P)$ şeklinde bir denklik ilişkisi vardır.

$$IND(P) = \{(x, y) \in U^2 \mid \forall a \in P, a(x) = a(y)\} \quad (4.1)$$

$IND(P)$ tarafından üretilen U 'nun bölümü U/P olarak gösterilir ve aşağıdaki gibi hesaplanabilir:

$$U/P = \bigotimes \{a \in P : U/IND(\{a\})\}, \quad (4.2)$$

$$A \otimes B = \{X \cap Y : \forall X \in A, \forall Y \in B, X \cap Y \neq \emptyset\} \quad (4.3)$$

Eğer $(x, y) \in IND(P)$ ise, x ve y , P' 'den gelen nitelikler tarafından ayırt edilemezler. P -ayırt edilemezlik ilişkisinin denklik sınıfı $[x]_P$ olarak gösterilir. $X \subseteq U$ için bir P -alt yaklaşımı şu şekilde tanımlanabilir:

$$\underline{P}X = \{x \mid [x]_P \subseteq X\} \quad (4.4)$$

P ve Q, U üzerinde denklik ilişkileri ise pozitif bölge şu şekilde tanımlanabilir:

$$POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}X \quad (4.5)$$

Sınıflandırma açısından, pozitif bölge, P nitelikleri bilgisini kullanarak U/Q sınıflarına göre sınıflandırılabilen U 'nun tüm nesnelerini içerir. Veri analizinde önemli bir husus

nitelikler arasındaki bağımlılıkları keşfetmektir. Bağımlılık aşağıdaki şekilde tanımlanabilir:

$P, Q \subseteq A$ için; Q, P' 'ye k ($0 \leq k \leq 1$) derecesinde bağlı olup, $P \Rightarrow_k Q$ olarak gösterilir,

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|} \quad (4.6)$$

Eğer $k = 1$ ise, Q' nun tamamen P' 'ye bağlı olduğu; eğer $0 < k < 1$ ise Q' nun kısmen (k derecesinde) P' 'ye bağlı olduğu ve eğer $k = 0$ ise Q' nun P' 'ye bağımlı olmadığı anlamına gelmektedir. Bir nitelik, nitelik kümesinden çıkarıldığında, bağımlılıktaki değişimi hesaplayarak, niteliğin önem değeri elde edilebilir. Bağımlılıktaki değişim ne kadar yüksekse, nitelik o kadar önemli demektir. Eğer önem değeri 0 ise, nitelik gereksiz demektir. P, Q ve bir $x \in P$ niteliği verildiğinde, Q üzerindeki niteliğin önemi aşağıdaki şekilde tanımlanır:

$$\sigma_P(Q, x) = \gamma_P(Q) - \gamma_{P-\{x\}}(Q) \quad (4.7)$$

Niteliklerin azaltılması, nitelik kümeleri tarafından üretilen eşdeğerlik ilişkilerinin karşılaştırılmasıyla elde edilir. Azaltılmış kümede orijinalle aynı kalitede sınıflandırmayı sağlayacak nitelikler çıkarılır. Karar sisteminde, indirgenmiş nitelik alt kümesi, nitelik kümesi C' nin bir alt kümesi R olarak, $\gamma_R(D) = \gamma_C(D)$, tanımlanır. Belirli bir veri kümesi, birçok indirgenmiş nitelik alt kümesine sahip olabilir ve tüm indirgenmiş nitelik alt kümeleri aşağıdaki gibi gösterilir:

$$R = \{X : X \subseteq C, \gamma_X(D) = \gamma_C(D)\} \quad (4.8)$$

R 'deki tüm kümelerin kesişimi, veri kümesinden çıkarılamayan öğeler olan çekirdek olarak adlandırılır. Nitelik indirgemede, en az elemanlı nitelik alt kümesi aranır; diğer bir deyişle, minimum nitelik alt kümesi $R_{min} \subseteq R$ belirlenmeye çalışılır.

$$R_{min} = \{X : X \in R, \forall Y \in R, |X| \leq |Y|\} \quad (4.9)$$

Bunu başarmanın basit bir yolu, C 'nin olası tüm alt kümelerinin bağımlılıklarını hesaplamaktır. Herhangi bir $\gamma_X(D) = 1$ değerine sahip X alt kümesi, nitelik alt kümesidir ancak en az elemanlı alt küme minimal indirgenmiş nitelik alt kümesidir.

4.2.2. FRQR Nitelik Seçimi Algoritması

Bilgi içeriğini koruyarak gereksiz nitelikleri kesikli veri kümelerinden çıkarmak için kaba küme yaklaşımları yaygın olarak literatürde kullanılmaktadır. Temel prensipleri verilen kaba küme temelli nitelik indirgeme süreci, yalnızca kesikli değerler içeren veri kümeleri ile etkin bir şekilde çalışabilmektedir. Ancak çoğu veri kümesi sürekli nitelik içerdiğinden, önceden bir kesiklendirme adımını gerçekleştirmek gerekir. Bulanık kaba küme yaklaşımını kullanarak, veriye kesiklendirme işlemi uygulamadan ve böylelikle bilgi kaybının önüne geçerek sürekli verilere nitelik seçimini uygulayabilmek mümkündür. Telekomünikasyon müşteri veri setinde sürekli nitelikler de yer aldığı için Jensen ve Shen [90] tarafından önerilen bulanık kaba küme temelli FRQR nitelik indirgeme algoritması tercih edilmiştir. FRQR nitelik indirgeme algoritması anlatımından önce bazı temel kavramlar aşağıda verilmiştir.

Bulanık P -alt ve P -üst yaklaşımları:

$$\mu_{\underline{P}X}(x) = \sup_{F \in \frac{U}{P}} \min(\mu_F(x), \inf_{y \in U} \max\{1 - \mu_F(y), \mu_X(y)\}) \quad (4.10)$$

$$\mu_{\overline{P}X}(x) = \sup_{F \in \frac{U}{P}} \min(\mu_F(x), \sup_{y \in U} \min\{\mu_F(y), \mu_X(y)\}) \quad (4.11)$$

Uygulamada, tüm $y \in U'$ ların dikkate alınması gerekmez. Denklik sınıfı F' nin bir bulanık elemanı y için yalnızca $\mu_F(y) \neq 0$ olduğu durumlar dikkate alınır.

Bulanık kaba küme nitelik indirgeme, sürekli nitelikleri içeren veri kümelerinin indirgenmesini sağlamak için bulanık alt yaklaşımı kavramını kullanılır. Geleneksel kaba küme teorisinde pozitif bölge, alt yaklaşımların birleşimi olarak tanımlanmaktadır. Genişletme prensibi ile, $x \in U$ olan bir x nesnesinin, bulanık pozitif bölgeye aitliği aşağıdaki gibi tanımlanır:

$$\mu_{POS_P(Q)}(x) = \sup_{X \in U/Q} \mu_{\underline{P}X}(x) \quad (4.12)$$

Bulanık pozitif bölge tanımını kullanarak bulanık bağımlılık fonksiyonu şu şekilde tanımlanabilir:

$$\gamma'_P(Q) = \frac{|\mu_{\text{POS}_P(Q)}(x)|}{|U|} = \frac{\sum_{x \in U} \mu_{\text{POS}_P(Q)}(x)}{|U|} \quad (4.13)$$

Bulanık kaba indirgeme işleminde birden çok niteliğin, örneğin $P = \{a, b\}$ ile ilgili bağımlılık derecesini belirleyebilmek gereklidir. Bulanık durumda nesneler birçok denklik sınıfına ait olabilir, bu nedenle U/P 'yi belirleyebilmek için $U/IND(\{a\})$ ve $U/IND(\{b\})$ 'nin kartezyen çarpımı dikkate alınmalıdır.

$$U/P = \otimes \{a \in P : U/IND(\{a\})\} \quad (4.14)$$

U/P 'deki her küme denklik sınıfını belirtir. Örneğin, $P = \{a, b\}$, $U/IND(\{a\}) = \{N_a, Z_a\}$ ve $U/IND(\{b\}) = \{N_b, Z_b\}$ için,

$$U/P = \{N_a \cap N_b, N_a \cap Z_b, Z_a \cap N_b, Z_a \cap Z_b\} \quad (4.15)$$

Dolayısıyla, bir nesnenin böyle bir denklik sınıfına ne ölçüde dahil edildiği, $F_i, i = 1, 2, \dots, n$ gibi bulanık denklik sınıflarının birleşimi kullanılarak hesaplanır:

$$\mu_{F_1 \cap \dots \cap F_n}(x) = \min(\mu_{F_1}(x), \mu_{F_2}(x), \dots, \mu_{F_n}(x)) \quad (4.16)$$

Jensen ve Shen [90] tarafından önerilen FRQR: Bulanık Kaba Hızlı Nitelik İndirgeme Algoritmasında, bir indirgenmiş nitelik alt kümesi tüm nitelik kümesi A ile aynı bilgi içeriğine sahip niteliklerin bir altkümesi R olarak tanımlanır. Bağımlılık fonksiyonu açısından, eğer veri kümesi tutarlıysa $\gamma(R)$ ve $\gamma(A)$ değerleri aynı ve 1'e eşittir. Bununla birlikte, bulanık kaba yaklaşımında, nesnelerin birçok bulanık denklik sınıfına ait olduğu zaman karşılaşılan belirsizlik gibi bir durum söz konusu değildir. Bununla mücadele etmenin olası bir yolu, bütün nitelik kümesinin bağımlılık derecesini belirleyip bunu payda olarak kullanmak olacaktır. Bu sorunlar göz önüne alındığında, geliştirilen FRQR Algoritması adımları Şekil 4.2' de verilmiştir.

```

 $R = \emptyset$ 
 $\gamma'_{en\ iyi} = 0$ 
 $\gamma'_{önceki} = 0$ 
Tekrar et
   $T = R$ 
   $\gamma'_{önceki} = \gamma'_{en\ iyi}$ 
   $\forall x \in (C - R)$ 
    Eğer  $\gamma'_{R \cup \{x\}}(D) > \gamma'_T(D)$  ise
       $T = R \cup \{x\}$ 
       $\gamma'_{en\ iyi} = \gamma'_T(D)$ 
   $R = T$ 
 $\gamma'_{en\ iyi} = \gamma'_{önceki}$  olana kadar
R elde edilir

```

Şekil 4.2. FRQR Algoritması [90]

Algoritma mevcut indirgenmiş nitelik alt kümesine hangi niteliklerin ekleneceğini seçmek için γ' bağımlılık fonksiyonunu kullanır. Kalan niteliklerin eklenmesiyle bağımlılık artmadığında algoritma sona erer.

4.3. Maliyet Duyarlı Veri Madenciliği Yaklaşımları

Son zamanlarda, sınıflar arasındaki dengesiz dağılım sorunu makine öğrenme ve veri madenciliği toplulukları açısından ilginç bir konudur. Farklı sınıfların örnek sayılarındaki farklılıklardan dolayı ortaya çıkan, sınıflandırmanın en büyük zorluklarından birisidir. Bu sınıflardan azınlık sınıfı genellikle nadir bulunan bir durumdur ve azınlık sınıfı sınıflandırabilmek daha da çok istenmektedir. Bazen bu azınlık sınıflarının önemi, bu örnekleri elde etme maliyetinden de kaynaklanabilmektedir. Literatürde, verilerin dengesizlik derecesini göstermek için çoğunluk ve azınlık sınıfların örneklerinin sayısının oranı “dengesizlik oranı” kullanılır.

Birçok sınıflandırma algoritması, yüksek tahmin kesinliği ve iyi genelleme yeteneği elde etmeye çalışır. Ancak dengesiz veri kümelerini sınıflandırırken, yalnızca doğruluğu iyileştirmeye çalışmak, azınlık sınıfını görmezden gelmeye neden olabilmektedir. Pek çok standart sınıflandırma algoritması genellikle çoğunluk sınıfını tercih etme eğilimindedir ki bu durum; azınlık sınıfının kötü sınıflandırma performansına yol açmaktadır. Sınıf dengesizliği, birçok veri kümesinin de sahip olduğu bir özellik olduğu gibi ayrılan müşteri veri kümesinin de önemli bir karakteristiğidir. Dengesiz veri

kümelerinden öğrenme problemini çözmek için son birkaç yılda birçok çözüm önerilmiştir. En sık kullanılan çözümler kabaca üç gruba ayrılır: veri düzeyindeki çözümler, algoritma düzeyindeki çözümler ve topluluk çözümleri [91].

Veri düzeyindeki çözümler

Sınıf dağılımını dengelemek ve sınıf dengesizliğinin olumsuz etkilerini azaltmak için orijinal veri kümesine bir ön işleme işlemi uygulayarak yeniden örneklendirir. Veri düzeyinde çözümler; yeniden örnekleme tekniklerinin birçok farklı formundan oluşur [91].

Veri seviyesi çözümlerinin avantajı seçilen sınıflandırma algoritmasından bağımsız kullanıldıkları için daha çok yönlü olmalarıdır. Veri düzeyindeki çözümler üç grupta sınıflanabilir:

- Alt örnekleme metotlarının amacı çoğunluk sınıfından bazı örnekleri elimine ederek azınlık sınıfı ile dengelemektir.
- Üst örnekleme metotlarının amacı azınlık sınıfı örneklerinden yeni örnekler oluşturarak veya bu örnekleri çoğaltarak orijinal veri kümesindeki çoğunluk sınıfı ile dengelemektir.
- Hibrid metotlar ise önceki iki metodun bir kombinasyonu olarak aşırı uyum göstermeyi önlemek için yeniden örneklemeden önce veya sonra örneklerin bazılarını elimine etmektir.

Literatürde sıklıkla kullanılan bazı veri düzeyindeki çözümler şöyledir:

- Sentetik Azınlık Üst Örnekleme Tekniği (SMT) [92]; her bir azınlık sınıf örneği alınarak yeniden örneklenir ve k azınlık sınıfın en yakın komşuları sentetik örnek olarak gösterilir. Üst örnekleme gereken miktara göre k en yakın komşulardan rastgele seçilir.
- SMT+Düzenlenmiş en Yakın Komşu (SMT+ENN) [93]; SMT yöntemini uygularken çoğunluk sınıf bazen azınlık sınıf uzayını ihlal edebildiğinden sınıf kümeleri iyi tanımlanamamaktadır. Bu nedenle, SMT+ENN hibrid

yaklaşımında eğitim setindeki en yakın üç komşusundan yanlış sınıflandırılan herhangi bir örneği çıkararak Wilson's ENN kuralını uygulamıştır.

- Güvenlik Seviyesi SMOTE (SL-SMT) [94]; SMT yönteminde çoğunluk sınıf örnekleri göz ardı edilerek seçilen bir azınlık sınıf örneği ve onun en yakın komşularından rastgele azınlık sınıfı örnekleri sentezlenir. Bunun tersi olarak SL-SMT yönteminde güvenlik seviyesi olarak adlandırılan farklı ağırlık dereceleri ile aynı hat üzerindeki azınlık sınıfı örnekleri dikkatlice örneklenir. Güvenlik seviyesi k en yakın azınlık sınıfı örnekleri kullanılarak hesaplanır. Daha sonra eğer bir örneğin güvenlik seviyesi 0' a çok yakınsa o örnek gürültü olarak düşünülür. Eğer k' ya yakınsa bu örnek güvenli olarak düşünülür. Bu nedenle eğitim setinde sentetik örnekler üretmeyi amaçlayan SMT ön işleme tekniğinin yeni bir varyasyonudur.
- Rastgele Üst Örnekleme (ROS) [91]; sezgisel olmayan bu metot azınlık sınıfındaki örnekleri rastgele çoğaltarak sınıf dağılımını dengelemeye çalışır. Bu yöntemin temel dezavantajı mevcut örneklerin tamı tamına kopyalarının yapılması ve aşırı uyum olasılığını artmasına sebep olmasıdır.
- Rastgele Alt Örnekleme (RUS) [91]; sezgisel olmayan bu metot çoğunluk sınıfındaki örnekleri rastgele elimine ederek sınıf dağılımını dengelemeye çalışır. Bu yöntemin temel eksikliği tanımlama prosesinde önemli olabilecek potansiyel yararlı verilerin kaybedilebilmesidir.
- Komşuluk Temizleme Kuralı (NCL) [93]; iki sınıflı problemler için temizleme algoritması olarak eğitim setindeki her bir örneğin üç yakın komşusu belirlenir. Eğer bu örnek çoğunluk sınıfına dahilse ve bunun üç yakın komşusunun sınıfları örneğin sınıfı ile çelişiyorsa o örnek silinir. Eğer o örnek azınlık sınıfına dahilse ve bunun en yakın komşuları yanlış sınıflanıp çoğunluk sınıfındaysa o örnek silinir.
- Tomek Links (TL) [95] algoritmasında iki örnek e_i ve e_j verilmiş olsun. Bu örnekler arası uzaklık $d(e_i, e_j)$ TL olarak adlandırılır $d(e_i, e_i) < d(e_i, e_j) < d(e_j, e_i)$ olan bir e_i örneği yoksa böyle tanımlanabilen örneklerin dışındakilere gürültülü örnek veya sınır örneği denir. Buna göre çoğunluk sınıfındaki gürültülü ve sınır örnekler elimine edilir.
- Örnek Ağırlıklandırma (Maliyet Duyarlı Öğrenme) [96] yaklaşımında yanlış sınıflandırma maliyetlerine göre her bir sınıfın örneği farklı farklı

ağırlıklandırılır. Böylece sınıflandırma algoritması daha az maliyetlilerle daha az uğraşır, sonuçta genel olarak daha düşük maliyetli sınıflandırma yapılmış olur.

Yukarıdaki yöntemler incelendiğinde rastgele yaklaşımlar basit olmalarına göre literatürdeki çalışmalarda daha iyi sonuçlar vermişlerdir. Eğer eklemeyi mi yoksa çıkarmayı mı tercih etmek gerekirse birçok araştırmacı üst örnekleme alt örnekleme göre birçok avantajı olduğunu göstermiştir [97]. Fernandez et al. [98] çalışmalarında bilinen birçok sınıflandırma algoritmasını maliyet duyarlı hale getirip literatür verileri ile deneysel bir çalışma yapmışlardır. Yapılan çalışmaya göre birçok üst örnekleme yöntemlerinden ROS yöntemi diğer birçok karmaşık yöntemle göre çok daha iyi sonuç vermiştir.

Topluluk çözümleri

Birkaç sınıflandırıcıyı birleştirerek yeni ve daha iyi bir sınıflandırıcı elde etmeyi amaçlamaktadır. Bu yaklaşımlar veri seviyesindeki öğrenme algoritmalarını, çekirdek sınıflandırıcıların öğrenme aşamasından önce ön işleme tabii tutarak modifiye ederler veya toplu öğrenme sürecine maliyet duyarlı bir çerçeve ekleyerek değişiklik yapmaktadırlar [91].

Algoritma düzeyindeki çözümler

Sınıflandırma algoritmalarının öğrenilmesini azınlık sınıfına doğru güçlendirmek için yeni algoritmalar geliştirmeyi veya mevcut algoritmaları değiştirmeyi amaçlamaktadır. Bu gruptaki en iyi bilinen algoritma maliyet duyarlı öğrenme yöntemidir. Maliyet duyarlı yöntemlerde, her bir sınıf için farklı yanlış sınıflandırma maliyetleri belirlenir daha sonra da toplam yanlış sınıflandırma maliyetini en aza indirmeye çalışılır. Veri yapısının doğası azınlık sınıfının maliyetinin yüksek, çoğunluk sınıfının maliyetinin ise düşük olmasını gerektirir [91].

Maliyet duyarlı öğrenme algoritmaları farklı sınıfların yanlış sınıflandırma maliyetlerini dikkate alarak öğrenmeyi sağlarlar. Öğrenme prosesinde özellikle azınlık sınıfın önem kazandığı maliyet fonksiyonunda toplam yanlış sınıflandırma maliyetini minimize etmeye çalışır. Bu nedenle maliyet duyarlı öğrenme farklı hatalar için maliyet matrisine ihtiyaç duyarlar [99].

Literatürde maliyet duyarlı öğrenme ile ilgili yapılan başlıca çalışmalardan aşağıda bahsedilmiştir.

- Zadrozny et al. [100] çalışmasında sınıflandırma algoritması ve teorisini maliyet duyarlı algoritmalara ve metotlara çevirmek için bir grup metot önermiş ve geliştirmiştir. Önerilen metot eğitim datalarını ya sınıflandırma algoritmalarını ağırlıklarla besleyerek ya da örnekleme ile gerçekleştiren maliyet orantılı ağırlıklandırmayı temel alır. Önerilen metotta iyi tahmin performansı ve diğer metotlara nazaran büyük ölçüde hesaplamada azalma görülmüştür.
- Şahin [101] tezinde kredi kartı dolandırıcılığı tespiti problemi için her ana düğüm bölünmesinde yanlış sınıflandırma maliyetlerini en aza indirgeyen yeni bir maliyet duyarlı karar ağacı yaklaşımı geliştirmiştir.
- Domingos [102], MetaCost metodunu farklı yanlış sınıflandırma maliyetlerini dâhil edebilmek için önermiştir.
- Czajkowski et al. [103] çalışmalarında literatürde tahminin altında kalma ile üstünde kalmanın maliyet farklılıkları ile ilgili bir çalışma olmamasını dikkate alarak bu boşluğu kapatmak için bir model önermişlerdir.
- Bahnsen et al. [104] makalesinde yeni bir maliyet duyarlı budama kriteri ve yeni bir maliyet duyarlı kirlilik ölçüsünü içeren örnek tabanlı maliyetlerle birlikte çalışan örneğe bağımlı karar ağacı algoritması önermiştir.
- Jiang et al. [105] makalelerinde birçok Bayes sınıflandırıcıları örnek ağırlıklandırma metodu ile modifiye etmiş ki bu da Bayes sınıflandırıcıları maliyet duyarlı hale getirmiştir.
- Siers ve Islam [106], ensemble ve maliyet duyarlı karar ağaçları CSForest adında tekniği önermiştir. Ayrıca, makalede sınıflandırma maliyetlerini minimize eden ve karar ağaçları kümesinin avantajlarını kullanmak için CSVoting isimli maliyet duyarlı tekniği önermiştir.
- Bolón-Canedo et al. [107] çalışmalarında maliyeti de hesaba katabilmek için filtreleme metotlarının değerlendirme fonksiyonuna yeni bir terim eklemiştir.

- Arar ve Ayan [108] ANN ve ABC algoritmaların kombinasyonu olan farklı bir sınıflandırma yaklaşımı önermişlerdir. Optimal ağırlıkları bulabilmek için ABC algoritması tarafından ANN eğitimi uygulanmıştır. ABC algoritmasının optimizasyon görevi olarak FP ve FN oranları çarpımı parametrik maliyet katsayısı olarak kullanılmıştır.
- Maliyet duyarlı öğrenmede önemli çalışmalardan biri de Elkan' a ait olup yanlış sınıflandırma maliyetlerini minimize etmek için maliyet duyarlı öğrenme metodu önermiştir. Bu metot yazılım hatalarını tespit etmede yazılım mühendisliği alanında çok yaygın bir şekilde uygulanmıştır [109].
- Hu et al. [110] çalışmalarında maliyet duyarlı öğrenme ile ensemble metotlarını yazılım proje risk değerlendirme için uygun bir tahminleme modeli geliştirmek için birleştirmiş ve yazılım proje risk tahmini için ilk olarak maliyet duyarlı ve ensemble temelli hibrit bir model (COSENS) önermiştir. Ayrıca yanlış sınıflandırma maliyetlerini dikkate alarak yazılım risk tahmini modellerini değerlendirmek için yeni titiz bir değerlendirme standardı oluşturmuştur.

4.4. Veri Madenciliğinde Meta Sezgisel Yaklaşımlar

Veri madenciliğinde sezgisel ve meta sezgisel yaklaşımlar etkin bir şekilde uygulanmaktadır. Sezgisel metotlar, problemlere hızlı bir şekilde yüksek kaliteli çözümler sunabilen; uzmanlık veya genel bilgiden elde edilen optimizasyon problemlerinin çözümünde kullanılan yöntemlerdir. Sezgisel metotlar çözüm için gerekli zamandan daha kısa sürede kesin çözüme oldukça yakın bir çözümü garanti ederler. Meta sezgisel; üzerinde veya ötesinde anlamına gelen “meta” ve “sezgisel” terimlerinin bileşimidir. Meta sezgisel metotlar ise bir yaklaşımın farklı optimizasyon problemlerinde başarıyla uygulanabilmesi için yeniden kurgulanıp geliştirildiği yöntemlerdir ve yüksek performanslı sezgisel prosedürler tasarlamak için genel stratejilerdir.

Günümüzde bilinen birçok strateji meta sezgisel olarak sınıflandırılabilir ve giderek artan sayıda bu tür yöntemleri kullanan araştırma makalesi ve uygulama, literatüre kazandırılmaya devam etmektedir. Meta sezgisel yaklaşımı ile birlikte

halihazırda var olan çeşitli optimizasyon yöntemler meta sezgisel olarak yorumlanmıştır. Genetik Algoritmalar, Yapay Sinir Ağları, Yerel Aramalar ve Benzetimli Tavlama klasik meta sezgisellerden bazılarıdır. Birçok modern meta sezgisel yaklaşımlar; endüstri, işletme ve mühendislik ile ilgili optimizasyon problemlerinin çözümü için başarıyla uygulanmıştır. Yapay Sinir Ağları ve Genetik Algoritmaların yanı sıra, Karınca Kolonisi Optimizasyonu ve Parçacık Sürüşü Optimizasyonu gibi doğadan ilham alan yaklaşımlar da en yaygın meta sezgisel yaklaşımlardır.

Birçok araştırmacı önerilen yöntemleri analiz etmek için meta sezgisel yaklaşımları kullanmakta ve bu yaklaşımlardan teorik ve pratik açıdan beklenen özellikler şöyledir:

1. Basitlik: Meta sezgisel yaklaşımlar basit ve açık bir ilkeye dayanmalı, anlaşılması kolay olmalıdır.
2. Kesinlik: Meta sezgisel yaklaşımların adımları veya aşamaları belirsizlik olmaksızın kesin şartlarla belirtilmelidir.
3. Tutarlılık: Meta sezgisel yaklaşımların uyarlandığı problemlerin adımları meta sezgisel algoritmanın temel adımlarından hareket edilerek oluşturulmalıdır.
4. Verimlilik: Problemlerin algoritmaları, orta düzeyde hesaplama süresinde iyi çözümler (optimal veya en yakın optimal) sağlamalıdır.
5. Yararlılık: Algoritmalar; eğer uygunsa, karşılaştırma problemlerinin çoğunu optimal olarak çözebilmelidir.
6. Etkililik: Belirli problemlerin gerçekçi örnekleri için optimal veya optime yakın çözümler elde edebilmelidir.
7. Sağlamlık: Meta sezgisel yaklaşımların birçok çeşitli örneklerde iyi performans göstermesi gerekir.
8. Genellik: Meta sezgisel yaklaşımlar, çok çeşitli sorunlar için iyi sezgisellere öncülük etmelidir.
9. Uyarlanabilirlik: Meta sezgiseller, çeşitli uygulama alanlarına veya farklı türdeki modellere uyum sağlayacak unsurlar içermelidir.
10. Kullanıcı dostu: Meta sezgisellerin kullanımı kolay olmalıdır; parametreler olmadan veya kolayca anlaşılır ve ayarlanabilir olmalıdır.

11. İnovasyon: Meta sezgisellerin ilkeleri ve/veya bunların kullanımı, yeni türdeki uygulamalara öncülük edebilmelidir.
12. İnteraktiflik: Meta sezgiseller, performansını iyileştirmek için, kullanıcının prosedüre dâhil olmasına izin vermelidir.
13. Çokluluk: Yöntemler kullanıcının arasından seçebileceği birkaç optimale yakın çözüm sunabilmelidir.
14. Özerklik: Meta sezgiseller, uygulamalarında parametrelerin olmamasına veya otomatik olarak ayarlanabilmesine izin vermelidir.
15. Uygulanabilirlik: Meta sezgiseller, çeşitli alanlara yaygın olarak uygulanabilir olmalıdır.

Veri madenciliğinin; kümeleme, sınıflandırma ve nitelik seçimi gibi bazı görevleri; optimizasyon problemleri olarak formüle edilmiştir. Bu problemlerin çözümleri, görevi gerçekleştirmek için gerekli parametre değerlerinden oluşur. En yakın komşu kümelemede ve sınıflandırmada; çözümler, kuralın uygulanması için olabilecek muhtemel seçimlerden oluşur. Amaç fonksiyonları ise, ilgili performans ölçütleridir. Nitelik seçiminde çözümler; nitelikler veya durumlar olarak belirlenir ve nitelik kümesinin boyutu veya veri miktarı sabitse, amaç tahmin performansı en yükseğe çıkarmaktır. Bununla birlikte, genel olarak, doğruluk ve basitlik olmak üzere en azından iki amaç vardır ve bu amaçlar da birbiriyle çelişirler. Doğruluk en üst düzeye çıkarılmaya çalışılırken, bilgi miktarı ise en aza indirilmeye çalışılır. Bu gibi durumlarla başa çıkabilmek için çok amaçlı meta sezgiseller geliştirilmiştir [111].

4.4.1. Veri Madenciliğinde Çok Amaçlı Meta Sezgisel Yaklaşımlar

Veri madenciliği problemlerinin çoğunun optimize edilmesi için birden fazla amacı yerine getirmesi gerekir. Örneğin, bir nitelik seçiminde, nitelik alt grubunun boyutunu en aza indirirken, sınıflandırma doğruluğunu en iyi duruma getirmeye çalışabilir. Benzer şekilde, bir sınıflandırma problemi, aynı zamanda doğruluk, anlaşılabilirlik gibi birkaç ölçütü aynı anda optimize etmeye çalışabilir. Dolayısıyla, veri madenciliği problemlerinin çoğunda doğası gereği çok amaç vardır. Bu nedenle, son on yılda, çeşitli araştırmacılar çok amaçlı optimizasyon tekniklerini farklı veri madenciliği problemleri

için uygulamışlardır. Bu bölümde, iki önemli veri madenciliği görevi: nitelik seçimi ve sınıflandırma için çeşitli çok amaçlı meta sezgisel yaklaşımlar incelenmiştir.

Nitelik seçimi problemi için genetik ve diğer meta sezgisel yaklaşımlar yaygın olarak kullanılmaktadır. Çok amaçlı meta sezgisel yaklaşımlar daha çok sarmal nitelik seçimi yaklaşımlarında kullanılır, nitelikler kromozomda kodlanır ve uygunluk fonksiyonu olarak nitelik değerlendirme kriteri kullanılır. Nitelik alt kümeleri, seçilen niteliklerin (denetimli yaklaşımlar için) veya kümenin (denetimsiz yaklaşımlar için) veri kümesini ne kadar iyi sınıflandırıldığını temel alarak değerlendirilir. Bununla birlikte, seçilen niteliklerin tek bir kriterle değerlendirilmesi, tüm veri kümeleri için eşit derecede iyi sonuç vermez. Bu nedenle, birden fazla kriterin eşzamanlı olarak optimize edilmesine olan ihtiyaç ortaya çıkmıştır. Çok amaçlı nitelik seçimi, nitelik seçme yöntemlerinin gücünü geliştirmeye yardımcı olabilmektedir. Yakın geçmişte hem denetimli hem de denetimsiz yaklaşımlarda bir dizi çok amaçlı meta sezgisel yaklaşımları önerilmiştir.

Çok amaçlı meta sezgisel yaklaşımlarını kullanarak bir nitelik seçimi problemini çözmeye yönelik ilk ve en önemli adım, kromozom yapısını belirlemektir. Çok amaçlı meta sezgisel optimizasyon tabanlı nitelik seçim algoritmalarının neredeyse tamamı, bir nitelik alt kümesini kodlamak için ikili kromozom kullanmaktadır. Her kromozomun uzunluğu toplam nitelik sayısı kadardır. Kromozomun her biri 1 veya 0 değer alabilir. Eğer 1 değerine sahipse, o nitelik seçilen nitelik alt grubunun bir parçası olarak kabul edilir. Eğer 0 ise, ilgili nitelik dikkate alınmaz. Örneğin, nitelik sayısı 10 ise, 0010110101 kromozomunda, nitelik alt kümesinde {3, 5, 6, 8, 10} nitelikler yer almaktadır [112].

Veri kümesinin boyutu büyükse, ikili kodlama için kromozom uzunluklarının çok büyük olabileceği eleştirisi yapılmaktadır. İkili kodlamanın kullanılması daha uzun kod çözme sürelerine ve daha düşük yakınsamaya yol açabilir ancak her tür karar değişkenini temsil etmek için kullanılabilen evrensel bir kodlama olma avantajına sahiptir. Tamsayılar gibi alternatif kodlamaların kullanılması, kromozom uzunluğunun azalmasına neden olur, ancak çaprazlama ve mutasyon için özel operatörler ve ek mekanizmalar (örneğin, çoğalmış bireylerin ortadan kaldırılması) gerektirebilir [112].

Denetimli nitelik seçimi yaklaşımlarında sınıf etiketlerinin bilindiği bir eğitim seti söz konusudur. Bu nedenle, seçilen nitelik alt kümelerini değerlendirmek için genellikle bazı sınıflandırma performans metrikleri kullanılmaktadır. Çalışmalarda kullanılan bazı amaç fonksiyonları şunlardır: yanlış sınıflandırma oranı, seçilen nitelik alt kümesi eleman sayısı, doğruluk, FP oranı, FN oranı, kesinlik, F-ölçüsü, nitelikler arası korelasyon, nitelik ve sınıfı arası korelasyon ve sınıflandırma MSE hata vs.

Denetimsiz nitelik seçiminde, çok amaçlı meta sezgisel algoritmalar ile ilgili birkaç çalışma yapılmıştır. Bu yaklaşımlarda sınıf etiketleri göz ardı edilir ve bu nedenle eğitim seti yoktur. Bu durumda genellikle, bu niteliklerin veri kümeleme yapısını ne kadar iyi tanımlayabildiğine bağlı olarak bir nitelik alt kümesini değerlendirmek için bir kümeleme algoritması kullanılır. Bu çalışmalarda kümeleme sayısına, nitelik sayısına, sınıf içi benzerliğe ve sınıflar arası benzerliğe bağlı olmayan farklı kümeleme hedefleri eşzamanlı olarak optimize edilmeye çalışılmaktadır [112].

4.4.2. Nitelik İndirgeme ve Sınıflandırma için Çok Amaçlı Meta Sezgisel Yaklaşımlar

Gerçek hayat problemleri eş zamanlı olarak optimize edilmesi gereken çoklu amaçlar içerir. Birden çok hedefi içeren optimizasyon problemleri söz konusu olduğunda, hedefler birbiriyle çeliştiğinde, tüm hedefleri optimize edecek bir çözüm bulunamaz. Bu çözümler pareto optimal çözümler olarak adlandırılır. Son yıllarda, çeşitli çok amaçlı sezgisel algoritmalar başarıyla çok amaçlı optimizasyon problemlerine uygulanmıştır. Yakın zamanda, işlem gücü ve potansiyelindeki artış; çok amaçlı tekniklerin veri madenciliği alanında etkili bir şekilde uygulanmasına olanak sağlamıştır. Nitelik seçimi, kural çıkarımı ve sınıflandırma problemleri, çelişen birçok amacı aynı anda optimize edebilmek için çok amaçlı optimizasyon problemi olarak formüle edilebilmektedir [113].

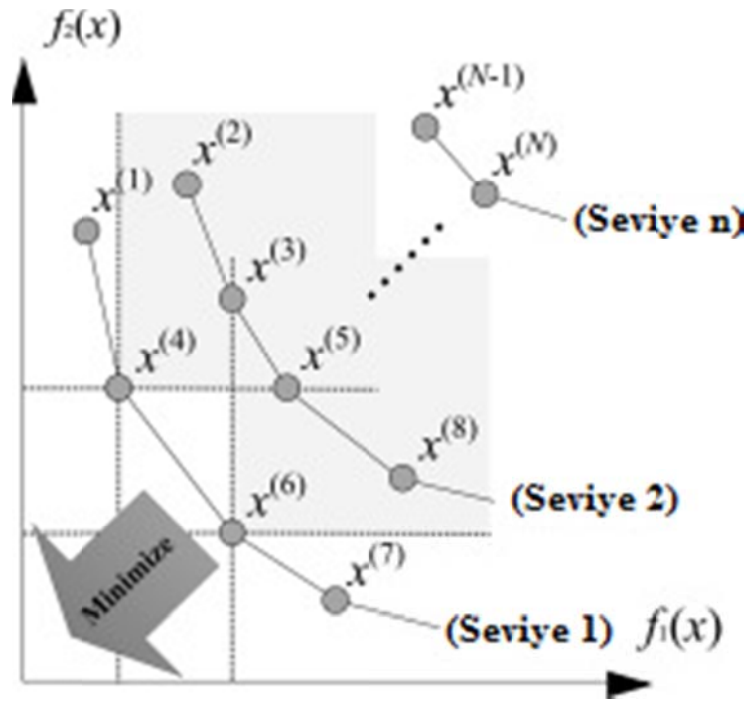
Bu bölümün alt başlıklarında ilk olarak NSGA-II ve KKO algoritmalarının temel prensipleri incelenmiştir. Önerilen müşteri kaybı tahmin modeli alt başlığında ise, müşteri kaybı tahmin modeli için geliştirilen çok amaçlı ve maliyet duyarlı NSGA-II nitelik seçimi yaklaşımı ve yine bu tezde geliştirilen çok amaçlı ve maliyet duyarlı KKO sınıflandırma yaklaşımları detaylı açıklanmıştır.

Çok amaçlı optimizasyon yaklaşımları sınıflandırma için yaygın olarak kullanılmaktadır. Temel olarak üç farklı yaklaşım vardır: sınıflandırma kuralları geliştirmek, eğitim verisinde sınıf sınırlarını (aşırı düzlemler) tanımlamak, standart sınıflandırıcıların modellenmesi.

Bir sınıflandırma kuralı If-then kuralı If <koşul> Then <sınıf> biçiminde temsil edilebilir. <koşul>, kuralın öncülünü temsil eder ve genellikle bir *ve* operatörü ile birleştirilip nitelik-değer çifti grubunu gösterir. Çok amaçlı optimizasyon yaklaşımlarını kullanarak sınıflandırma kuralı geliştirmek için önce kromozom yapısı tanımlanmalıdır. Bu konuda başlıca iki yaklaşım vardır. İlki, tek bir kromozomda bir dizi kuralın kodlandığı Pittsburgh yaklaşımıdır. İkincisi, her kromozomun bir kuralı kodlayan Michigan yaklaşımıdır. Pittsburgh kodlama stratejisi iki kategoriye ayrılabilir. Biri kural seçimi ve diğeri de kural öğrenme yöntemidir. İlk stratejide amaç, önceden tanımlanmış kuralların bir alt kümesini seçmektir ve ikincisinde amaç, meta sezgisel algoritmalar yoluyla kuralları öğrenmektir. Öte yandan, Michigan yaklaşımı genellikle çok daha kolay ve daha az karmaşıktır; çünkü her kromozom tek bir kural kodlamaktadır. Bu nedenle, bu yaklaşımda nihai çözüm, tahsis edilmemiş sınıflandırma kurallarının tamamından oluşur. Çok amaçlı meta sezgisel sınıflandırma yaklaşımlarında birçok amaç fonksiyonu kullanılmıştır: sınıflandırma doğruluğu, AUC, kural sayısı, öncül koşulların sayısı, aday kural kümesinin karmaşıklığı, yanlış sınıflandırma hataları, anlaşılabilirlik, duyarlılık [112].

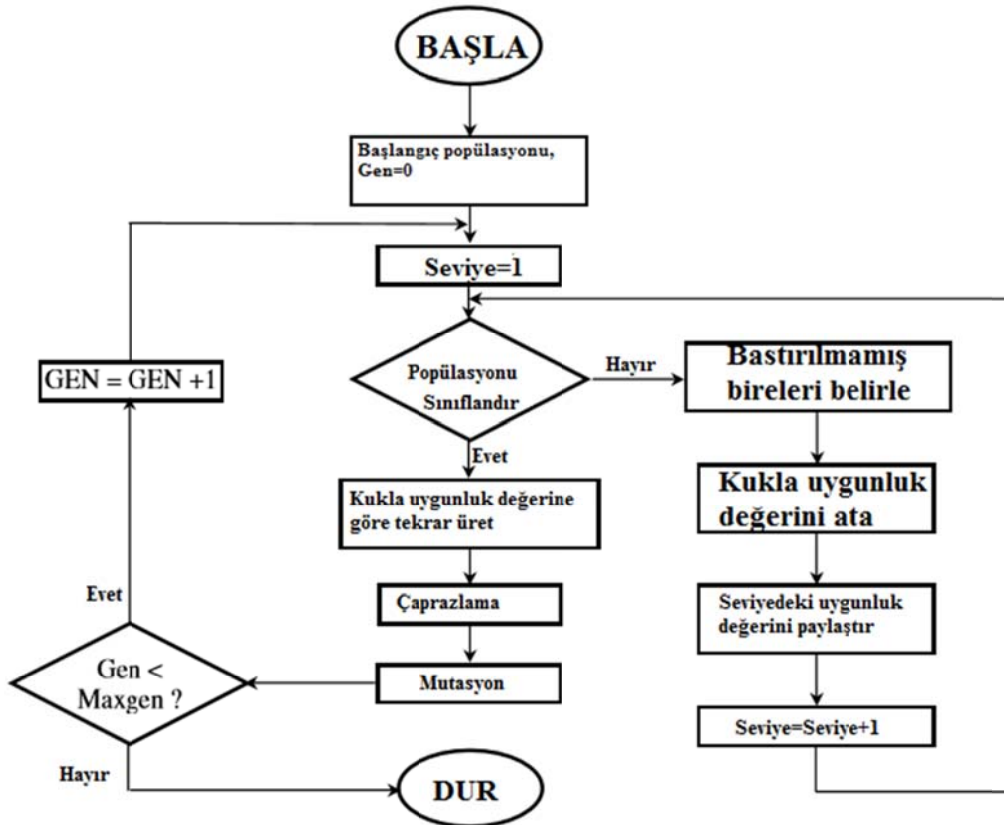
4.5. NSGA-II

NSGA iyi bilinen genetik algoritma tabanlı çok amaçlı bir meta sezgisel algoritmadır. Bastırılmamış çözümleri sıralamasına dayanmaktadır. NSGA' da bu sıralama kavramının yanı sıra "kukla uygunluk değeri" da tanımlanmıştır. NSGA' da ilk popülasyon rasgele oluşturulur ve bu popülasyonun bastırılmamış çözümleri 1. seviyeye atanır. Bu adımdan sonra birinci dereceden bireyler geçici olarak çıkarılır ve ikinci seviyeye atanan bastırılmamış çözümler belirlenir. Şekil 4.3.' te gösterilen bu sıralama mekanizması nüfustaki tüm bireyler sıralanana kadar devam eder. Tüm bireylere seviyelerine göre, kukla uygunluk değeri atanır.



Şekil 4.3. İki amaçlı bir problem için NSGA sıralama mekanizması [114]

NSGA' nın nasıl çalıştığını açıklayan ayrıntılı akış şeması Şekil 4.4.'te verilmiştir.



Şekil 4.4. NSGA akış şeması [114]

NSGA, bastırılmamış çözüm kümesini bulmak için elitist strateji kullanmaz. NSGA' nın aksine, elitist bastırılmamış sıralama algoritması (ENGA- NSGA' nın geliştirilmiş hali) elitist stratejiyi kullanmaktadır. NSGA gibi, ENGA da yeniden üretme aşamasında çaprazlama ve mutasyon uygular ancak farklı bastırılmamış çözüm sıralama tekniği kullanır. İlk önce, ebeveynler ve yeni nesillerden oluşan havuza ek bir bastırılmamış çözüm sıralaması uygulayarak bir sonraki popülasyonun adayları belirlenir. Bu seçim prosedürü iyi ebeveynlerin bir sonraki nesilde yaşamasına ve bu da algoritmayı elitist yapmasına izin verir.

NSGA, iyi sonuçlara ulaşabilmesine rağmen, hesaplamanın karmaşıklığı ve elitizm eksikliği nedeniyle eleştirilmiştir. Yenilenmiş bir versiyonu NSGA-II, elitist koruma yaklaşımı olarak hızlı bastırılmamış sıralama prosedürü ve yakınlık mesafesi yaklaşımları ile geliştirilmiştir. NSGA-II algoritmasında, algoritma sonucunda elde edilen Pareto etkin küme üzerindeki bastırılmayan noktalar, karar vericilere çözüm alternatifleri olarak sunulmaktadır.

Hızlı Bastırılmamış Sıralama Yaklaşımı:

Başlangıçta ebeveyn popülasyon rastgele oluşturulur ve daha sonra popülasyon pareto üstünlükler temel alınarak sıralanır. Bireyler, baskın olma derecelerine göre, farklı baskınlık kümelerinde sınıflandırılır. Popülasyondaki bireyler her amaç için kıyaslanarak, bireyin baskın olduğu bireylerin sayısı ve bireye baskın olan bireyler hafızada tutulur. Eğer bireye baskın olan bireylerin sayısı 0 ise, o birey 1. baskılanamayan bireyler pareto yüzeyine eklenir. $n_p = 0$ olan her p bireyine ait S_p kümesindeki her q bireyi ziyaret edilir ve n_p değeri 1 azaltılır. Bu işlemler sırasında, eğer herhangi bir elemanın q bireyinin baskınlık sayısı 0 ise ayrı bir liste Q 'ya yerleştirilir. Q kümesindeki bireyler, 2. baskılanamayan bireyler pareto yüzeyini oluşturur. Yukarıda anlatılan süreç, Q kümesinin tüm elemanları için uygulanarak, bir sonraki baskılanamayan bireyler pareto yüzeyi oluşturulur. Her pareto yüzey birbirlerine baskın olmayan bireylerden oluşmaktadır. Bu işlemler tüm bireyler pareto yüzeylere atanıp, tüm pareto yüzeyler belirleninceye kadar devam ettirilir [115].

N : popülasyon büyüklüğü

n_p : baskınlık sayısı

S_p : p bireyinin baskın olduğu bireylerin kümesi olmak üzere;

her $p \in P$ için

$S_p = \emptyset$

$n_p = 0$

her $q \in P$ için

eğer $(p < q)$ ise

$S_p = S_p \cup \{q\}$

eğer $(q < p)$ ise

$n_p = n_p + 1$

eğer $n_p = 0$ ise

$p_{derece} = 1$

$F_1 = F_1 \cup \{p\}$

$i = 1$

$F_i \neq \emptyset$ iken

$Q = \emptyset$

her $p \in F_i$ için

her $q \in S_p$ için

$n_q = n_q - 1$

eğer $n_q = 0$ ise

$q_{derece} = i + 1$

$Q = Q \cup \{q\}$

$i = i + 1$

$F_i = Q$

Eğer p, q 'yu domine ediyorsa

q 'yu p tarafından domine edilenler kümesine ekle

q ' nun domine ettikleri sayısını bir artır

Eğer p ilk pareto yüzeyine aitse

Yüzey sayıcıyı başlat

Bir sonraki yüzeyin elemanlarını sakla

q bir sonraki pareto yüzeyine aittir

Şekil 4.5. Hızlı bastırılmamış sıralama prosedürü

Daha net bir şekilde açıklayabilmek için, bir popülasyonu farklı bastırılmamışlık düzeylerine göre sıralama prosedürü Şekil 4.5' te tanımlanmıştır.

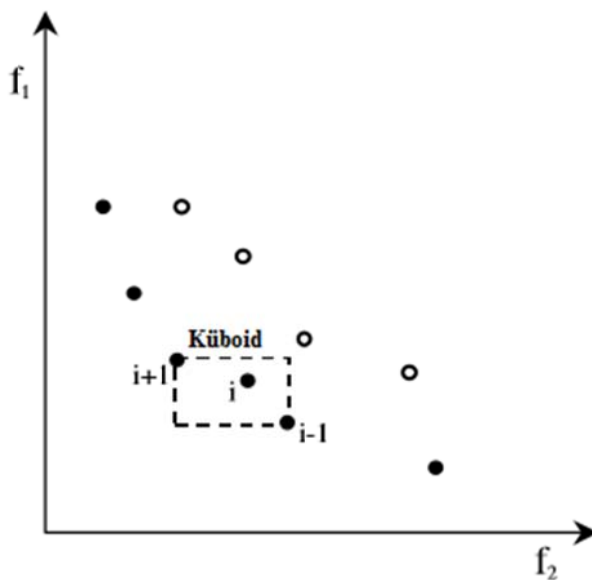
Bütün bireyler baskın oldukları bireylerin sayısına göre gruplandırılmıştır. İlk yüzeyde, en iyi bireyler yer almaktadır. Yeni ebeveyn popülasyonu ilk yüzeydeki elemanların eklenmesiyle oluşturulur ve elitist stratejiyi güvence altına almak için popülasyon büyüklüğünü aşana kadar diğer yüzeylerden sırasıyla eklenir. Bir yüzeydeki bütün bireyler, nüfus büyüklüğünü aşması nedeniyle yeni ebeveyn popülasyona eklenemezlerse, NSGA-II, "yakınlık mesafesine" dayalı bir çeşitlilik mekanizmasını dikkate alarak o yüzeydeki bireyler arasından seçim yapar.

Yakınlık mesafesi:

Her baskılanamayan küme, kendi içinde yakınlık mesafesine göre sıralanır. Bu mesafe bir bireyin komşularına ne kadar yakın olduğunu göstermektedir. Düşük ortalama yakınlık mesafesi "iyi çevrilmiş birey" in göstergesidir. Yüksek ortalama yakınlık mesafesi, popülasyonda daha iyi çeşitliliğin olduğunu gösterir. Her amaç fonksiyonu için, sınır çözümlere (en küçük ve en büyük değere sahip çözümlere) sonsuz uzaklık değeri atanır. Böylece, uç değerlerin seçilmesi garantilenmiş olur. Arada kalan tüm diğer çözümler için, tüm amaç fonksiyonları için bir uzaklık değeri hesaplanır. Toplam yığılma uzaklığı değeri, her amaç için hesaplanan bireysel uzaklıkların toplamıdır.

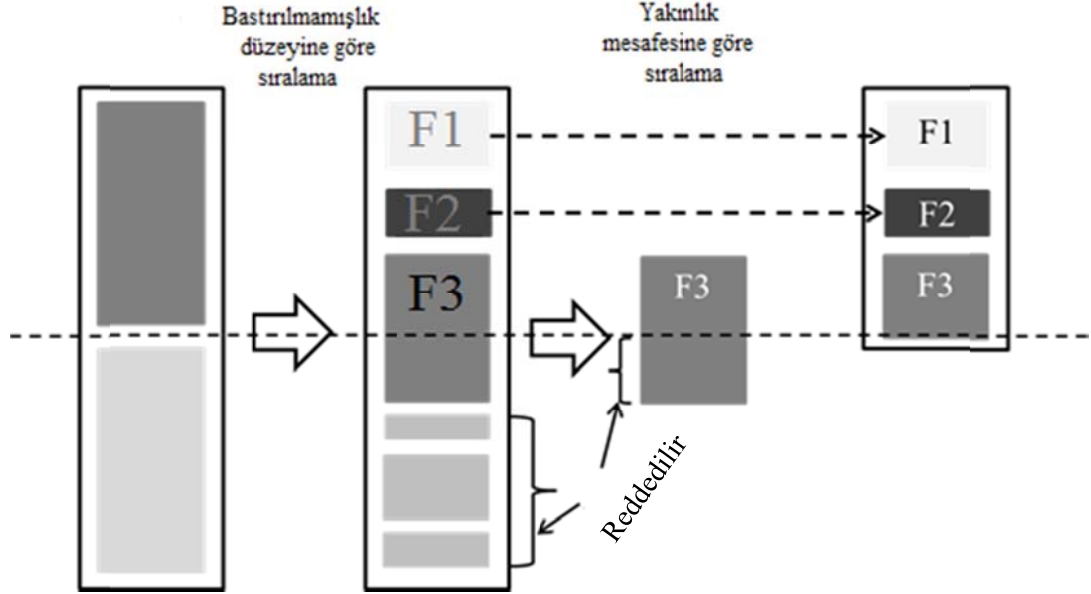
Popülasyondaki iki birey farklı bastırılmamış pareto yüzeylerde ise iki çözümden, daha düşük pareto yüzeyde (daha iyi) olanı tercih edilir. Eğer iki birey de aynı bastırılmamış pareto yüzeyde ise, yakınlık mesafesi daha yüksek olan, başka bir deyişle çevresindeki çözüm kalabalığı daha düşük olan birey tercih edilir ve yeni popülasyona eklenir.

Şekil 4.6.'da, i çözümü ile aynı yüzeyde olan çözümlere (dolu dairelerle işaretlenmiştir) olan yakınlık mesafesi, küboidin ortalama yan uzunluğudur (kesikli bir kutuda gösterilmiştir). Şekilde, dolu dairelerde işaretlenmiş olan noktalar aynı yüzeydeki çözümlerdir.



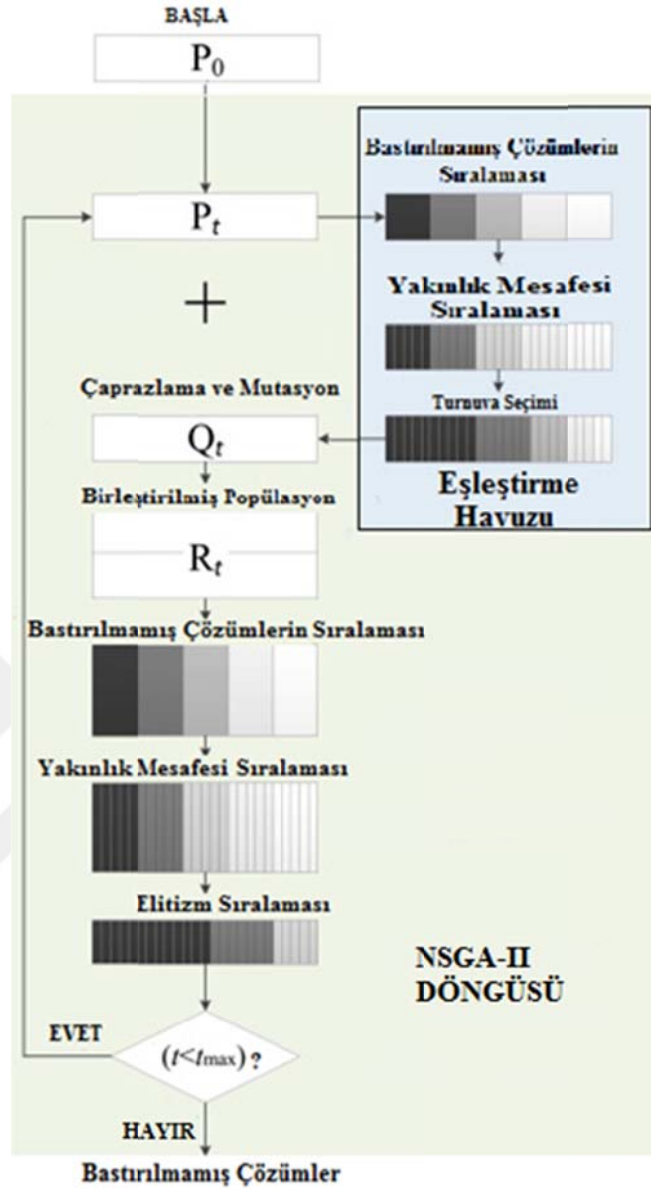
Şekil 4.6. Yakınlık mesafesi hesaplama [116]

NSGA-II 'nin ana adımları seçme, çaprazlama, mutasyon ve değerlendirilmedir [116]. Şekil 4.7' de NSGA-II' nin bastırılmamış çözümleri sıralama algoritması gösterilmiştir.



Şekil 4.7. NSGA-II sıralama akış şeması [116]

- Elde edilen popülasyona, yeniden üretim operatörü, çaprazlama ve mutasyon operatörleri kullanılarak, N büyüklüğünde (Q_0) yeni popülasyon üretilir.
- Mevcut ebeveyn ve yeni birey popülasyonları ile oluşturulan $2*N$ elemanlı R_t popülasyonu tekrar sıralanır ve en iyi bireyler popülasyon büyüklüğü N kadar seçilir. Seçim süreci baskılanamama derecesine ve yakınlık mesafesine dayanmaktadır.



Şekil 4.8. NSGA-II akış şeması [114]

- FI , en iyi bastırılmamış pareto yüzeyi ifade etmektedir. İlk olarak FI ’de yer alan bireyler bir sonraki ebeveyn popülasyonuna (P_{t+1}) aktarılır.
- P_{t+1} ’in eleman sayısı tamamlana kadar sırasıyla bastırılmamış pareto yüzeylerdeki bireyler aktarılır.
- Son bastırılmamış pareto yüzeyin bireyleri aktarılırken bireyler yakınlık mesafesine göre sıralanır ve ilk sıradan başlanarak P_{t+1} ’e aktarılır.
- Yeni oluşturulan N büyüklüğündeki P_{t+1} popülasyonunda seçim, çaprazlama ve mutasyon işlemleri uygulanarak, N büyüklüğünde bir Q_{t+1} popülasyonu oluşturulur. Daha sonra P_{t+1} ebeveyn popülasyonu ile birleştirilerek $2*N$ büyüklüğünde bir R_{t+1} popülasyonu oluşturulur ve yukarıda anlatılan işlemler uygulanır.

- Yeni oluşturulan N büyüklüğündeki P_{t+1} popülasyonunda seçim, çaprazlama ve mutasyon işlemleri uygulanarak, N büyüklüğünde bir Q_{t+1} popülasyonu oluşturulur. Daha sonra P_{t+1} ebeveyn popülasyonu ile birleştirilerek $2*N$ büyüklüğünde bir R_{t+1} popülasyonu oluşturulur ve yukarıda anlatılan işlemler uygulanır.

Tablo 4.1. Çok amaçlı meta sezgisel yaklaşımlarda nitelik seçimi alanında çalışmalar

Algoritma	Amaç Fonksiyonları
Oliveira et al. [117]	Yanlış sınıflandırma oranı, seçilen nitelik sayısı
Wang ve Huang [118]	Nitelikler arası korelasyon, nitelik ve sınıfı arası korelasyon
Venkatadri ve Rao [119]	Farklı desen çiftlerinin sayısının farklı kombinasyonları
Morita et al. [120]	DB index, seçilen nitelik sayısı
Mierswa ve Wurst [121]	DB index, seçilen nitelik sayısı
Shi et al. [122]	Çapraz doğrulama sınıflandırma doğruluğu, test sınıflandırma doğruluğu ve seçilen nitelik sayısı
Ekbal et al. [123]	Arama prosesi sırasında sınıflayıcının kesinliği ve duyarlılığı amaç fonksiyonu olarak kullanılmıştır ve F-ölçütü ise sonuç pareto kümesinden nitelik alt kümelerinden birini seçmek için kullanılmıştır.
Li et al. [124]	Hata belirleme oranı ve seçilen nitelik sayısı
Castro ve Von Zuben [125]	Sınıflandırma hatası ve seçilen nitelik sayısı
Zhao et al. [126]	FPR, FNR ve sınıflandırma karmaşıklık oranı
Kimovski et al. [127]	Sınıflandırma doğruluğu ve seçilen nitelik sayısı
Hamdani et al. [128]	Sınıflandırma doğruluğu ve seçilen nitelik sayısı
Huang et al. [80]	Sınıflandırma doğruluğu, TPR ve TNR
Radtko et al. [129]	Sınıflandırma doğruluğu ve seçilen nitelik sayısı
Spolaor et al. [130]	Veri kümesindeki tutarsız örnek çiftlerinin oranı, nitelik-sınıf arasındaki korelasyon, veri kümesindeki örneklerin Laplace skoru ve niteliklerin entropi değeri
Vatolkin et al. [131]	Seçilen nitelik sayısı ve sınıflandırma MSE hata

4.5.1. Literatürde NSGA-II ile Nitelik Seçim Yaklaşımları

Nitelik seçimi, veri setinden uygun olmayan niteliklerin elimine edilerek verilerin daha kolay anlaşılmasını sağlayan veri madenciliği yaklaşımıdır. Çok yüksek sayıda nitelikler olduğunda bu yaklaşım daha da önem kazanmaktadır. Ayrıca, nitelik seçimi ile veri boyutu da daha küçüldüğü için depolama gereksinimlerini ve hesaplama süresini azaltır. Böylece, öğrenme daha kolay bir süreç haline gelir. Nitelik seçimi,

sınıflandırmanın doğruluğunu ve hesaplama karmaşıklığını olumlu yönde etkileyebilecek önemli bir rol oynamaktadır. Bu nedenle, literatürde birçok çok amaçlı meta sezgisel nitelik seçimi yaklaşımları geliştirilmiştir. Bu yaklaşımlar ve yaklaşımlarda uygulanan amaç fonksiyonları Tablo 4.1'de özetlenmiştir.

4.5.2. Önerilen Çok Amaçlı ve Maliyet Duyarlı NSGA-II Nitelik Seçim Yaklaşımı

Nitelik seçimi modelleri, denetimli, denetimsiz ve yarı denetimli olarak sınıflandırılır; bu eğitim setinin etiketlenip etiketlenmediğine bağlıdır. Ayrıca, nitelik seçimi için üç temel yaklaşım vardır: filtreleme, sarmal ve gömülü modeller. Sarmal türü algoritmalar, bir niteliği gizlemek veya silmek yerine, belirli bir nitelik kümesinin optimumluğunu değerlendirmek için bir öğrenme modeli kullanır. Bu yapılar, sürekli bir arama süreci içeren ve bu aramayı gerçekleştiren bir sınıflandırma algoritmasına bağlı oldukları için sarmal yöntemler olarak adlandırılır. Sarmal yöntemler genel olarak sınıflandırma performansı bakımından diğer iki yöntemden daha iyi sonuçlar verirken bu yöntem hesaplama karmaşıklığı açısından dezavantajlıdır [132]. Önerilen modelde müşteri kaybı tahmininde nitelik alt kümesini belirlemek için sarmal ve denetimli NSGA-II yaklaşımı önerilmiştir.

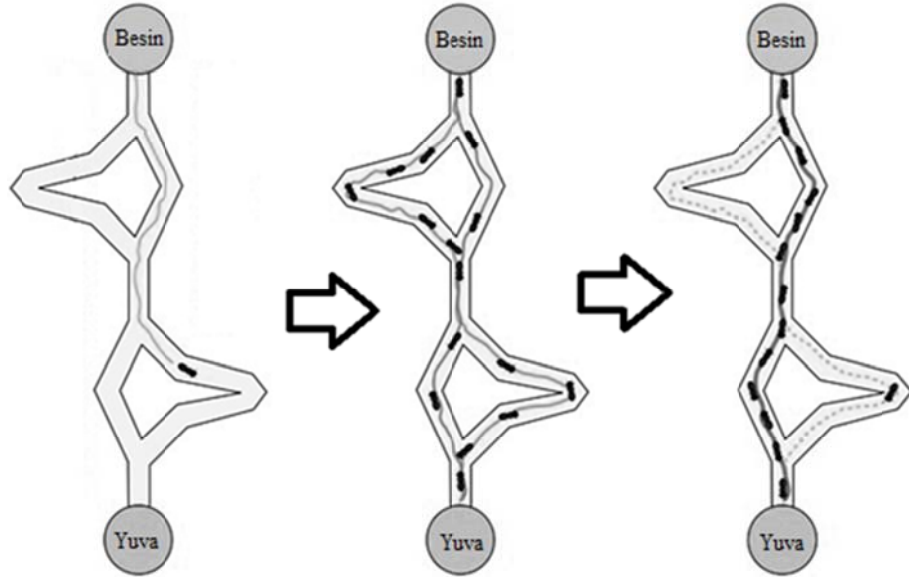
Genetik algoritmalarda bireylerin kalitesi, uygunluk değerlendirme fonksiyonlarıyla belirlenmekte ve uygunluk değerlendirme fonksiyonları, genellikle amaç fonksiyonu olarak seçilmektedir. Bu önerilen nitelik seçimi yaklaşımında yanlış sınıflandırma maliyeti ve nitelik alt küme eleman sayısı olmak üzere iki amaç fonksiyonu tercih edilmiştir. Az sayıda nitelik ile en düşük yanlış sınıflandırma maliyetini elde etmeye çalışmak, çelişen hedeflerdir. Çok amaçlı meta sezgisel yaklaşımlar arasında NSGA-II, telekomünikasyon sektöründe bazı müşteri kaybı yönetimi çalışmalarında başarıyla uygulanmış olması nedeniyle tercih edilmiştir [80]. Şu anki bilgimize göre literatürdeki uygulamalarda yanlış sınıflandırma maliyeti NSGA-II nitelik seçimi yaklaşımlarında uygunluk değerlendirme fonksiyonu olarak uygulanmamıştır. Böylelikle önerilen modelle literatüre farklı bir nitelik seçimi yaklaşımı kazandırılmıştır.

Amaç fonksiyonlarından; yanlış sınıflandırma maliyeti, nitelik seçimi yaklaşımını etkin bir şekilde maliyet duyarlı hale getirmiştir ve yanlış sınıflandırma maliyetini elde edebilmek için, düşük hesaplama karmaşıklığına sahip klasik ve popüler bir

sınıflandırma aracı olan karar ağacı, elde edilen nitelik alt kümelerini değerlendirmek üzere seçilmiştir.

4.6. KKO

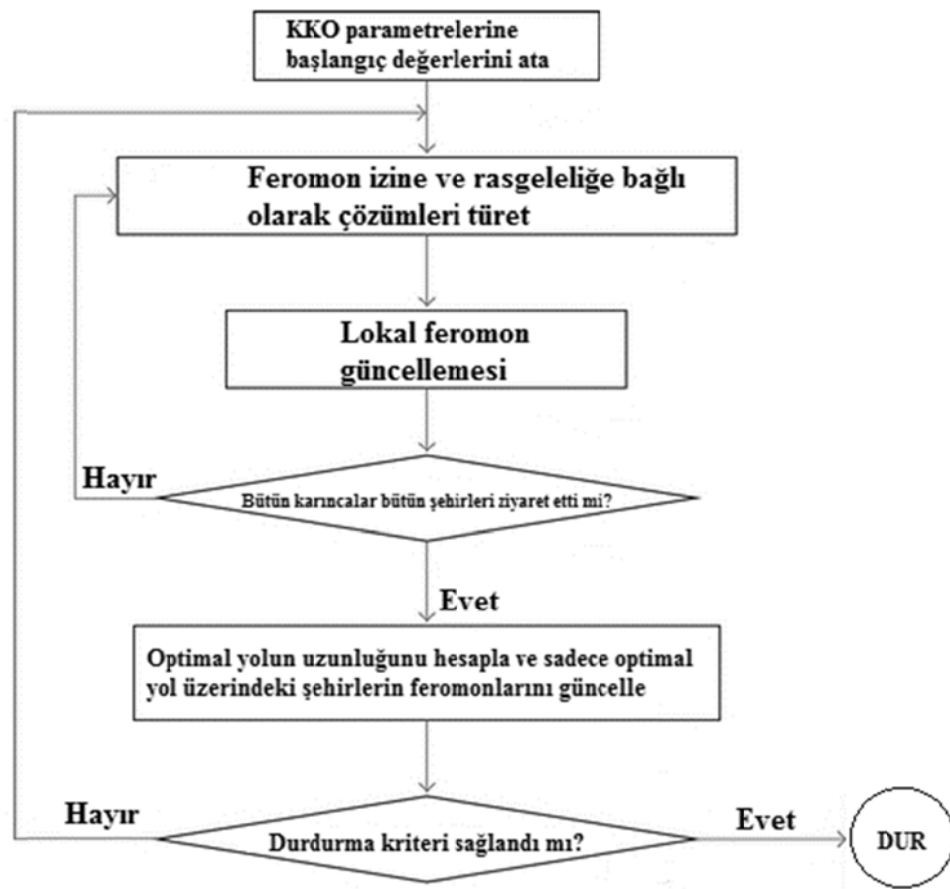
KKO NP-zor optimizasyon problemlerine başarıyla uygulanabilen popülasyona dayalı bir meta sezgisel yaklaşımdır. KKO, karıncaların avlanma davranışını araştıran ve gerçek karınca kolonilerinin yuvaları ile yiyecek noktaları arasındaki en kısa yolu bulma kabiliyetlerinden esinlenilerek geliştirilmiştir [133]. Böcek kolonilerinde, her böcek kendiliğinden hareket ediyor gibi görünse de oldukça organize hareket etmektedirler. Açıkçası, tüm bireysel faaliyetler herhangi bir denetim gerektirmeden sorunsuz bir şekilde entegredir. Bilim adamları, böceklerin davranışlarından, büyük oranda kendi kendine düzenlenen bir iş birlikleri olduğunu görmüşlerdir. Bireyler arasındaki koordinasyon, aralarındaki etkileşim sayesinde sağlanmaktadır. Aralarındaki etkileşim basit olsa da zor sorunları çözebilmektedirler (sayısız yollardan bir gıda kaynağına en kısa yoldan gitmek gibi) [134].



Şekil 4.9. Gerçek bir karıncanın davranışına örnek

Şekil 4.9.' da gösterildiği gibi, karıncalar alternatif yollar söz konusu olduğunda ilk olarak bu yollara eşit olasılıkla dağılırlar. Besin kaynağı ve yuva arasında yürürken geçtikleri yere feromon adı verilen bir madde bırakırlar. Belli bir süre sonra en kısa olan yolda feromon miktarı yoğunlaşır ve yollardaki feromon miktarı ile doğru orantılı olarak zamanla tüm karıncalar en kısa yolu tercih ederler.

KKO yaklaşımının temelini oluşturan unsurlarından biri haberleşme aracı olarak kullanılan ve problemlerde çözümün kalitesini gösteren feromon miktarıdır. Feromon, karıncaların bir haberleşme ve yön bulma aracı olarak kullandıkları, vücutlarından salgıladıkları kimyasaldır. Karıncalar alternatifler arasından geçecekleri yolu seçerken güçlü feromon konsantrasyonuna sahip olan yolu tercih ederler. Feromon izi, karıncaların yiyecek kaynağına (veya yuvaya) doğru en kısa yolu bulmalarını sağlar.



Şekil 4.10. KKO Akış Şeması (Gezgin Satıcı Problemi için) [133]

KKO sayesinde hızlı bir şekilde en uygun çözümler keşfedebilir, bu nedenle klasik gezgin satıcı problemlerinden zor veri sınıflandırma problemlerine kadar birçok sorunu çözmek için kullanılmaktadır. Şekil 4.10.' da KKO yaklaşımının gezgin satıcı probleminde akış şeması verilmiştir. Costa et al. [135] büyük veri tabanlarındaki değişkenler arasındaki ilginç ilişkileri keşfetmek için KKO sınıflandırma algoritmasını önermiştir.

4.6.1. Çok Amaçlı KKO Algoritması

Son zamanlarda, bir dizi çok amaçlı optimizasyon KKO uygulamaları literatürde yer almaktadır. Bu algoritmalar ağırlıklı olarak aşağıdaki üç özelliğe göre farklılık göstermektedir.

Feromon güncellemesi: Feromon izleri güncellerken, çözümlerin hangisinde feromon miktarının güncelleneceğine karar vermek gerekir. Feromon izlerini güncellemek için genellikle iki strateji uygulanmaktadır. İlk strateji, her bir amaca için ya her iterasyonda elde edilen en iyi çözümün ya da şimdiye kadar elde edilen en iyi çözümün feromon izini güncellemektir. İkinci bir strateji ise, bastırılmamış çözümleri ayrı bir kümede depolamak ve feromon güncellemesi için yalnızca bastırılmamış kümedeki çözümlere izin verilmesidir.

Feromon tanımı ve sezgisel bilgiler: Bir çözüm inşa etmenin her adımında, bir aday, iki faktöre bağlı bir geçiş olasılığına göre seçilir: feromon faktörü ve sezgisel faktör. Feromon / sezgisel bilgiyi tanımlamak için iki yaklaşım vardır: bir veya birden çok matris kullanmak. Yalnızca bir matris kullanıldığında, her bir amaçla ilişkili olan feromon bilgileri birleştirilerek birden fazla hedef tek bir hedefe indirgenir. Eğer birden fazla matris kullanılırsa, genellikle her matris bir hedefe karşılık gelir. Feromon bilgileri ile ilgili olarak, her matris uygulanan uygulama stratejisine bağlı olarak farklı değerler içerebilir. Aynı şey sezgisel bilgiler için de geçerlidir.

Feromon ve sezgisel birleştirme: Çoklu matrisler kullanıldığında, feromon / sezgisel matrisleri birleştirmek için bazı işlemler kullanılmalıdır. Bunun için üç yaygın strateji vardır: (1) ağırlıklı toplam, burada matrisler ağırlıklı bir toplamla toplanır; (2) matrislerin ağırlıklı bir ürünle toplandığı ağırlıklı ürün ve (3) rastgele, burada her adımda rastgele bir hedef optimize edilecek şekilde seçilir. Ağırlıklar birden çok matrisin toplanması için kullanıldığında, algoritmanın her iterasyonunda kullanılan ağırlıkların ayarlanması için iki strateji uygulanabilir: (a) dinamik olarak, her bir karınca, her iterasyonda diğer karıncalara göre farklı bir ağırlık verilebilir; (b) sabit olarak, burada tüm karıncalara aynı ağırlık verilebilir ve her bir amaç, tüm algoritma çalıştırması sırasında aynı önemdedir.

4.6.2. Literatürde KKO Sınıflandırma Algoritması

KKO, veri tabanından sınıflandırma kuralları çıkarmak ve gelecek veriler için sınıflandırmayı tahmin etmek için kullanılabilmektedir. KKO sınıflandırma algoritması, verilerin dağılımı hakkında herhangi bir bilgiye ihtiyaç duymadığı için geleneksel istatistiksel tekniklerden daha fazla avantaja sahiptir [134]. Ayrıca, güçlü uyarlanabilirlik ve olumlu geribildirim mekanizması olan KKO sınıflandırma algoritması, araştırma topluluklarının dikkatini çekmiştir ve birçok probleme uygulanmıştır. Proteinlerin hiyerarşik sınıflandırması [135], hücrel otomata geçiş kuralları [136], akustik emisyon sinyallerinin sınıflandırması [134], uzaktan algılama imgelerinin sınıflandırılması [137] ve taşkın riski imar planı [138] alanlarında uygulanmıştır. Bu uygulamalar, KKO sınıflandırma algoritmasının, sınıflandırma problemlerini çözmede büyük avantajları olduğunu göstermektedir. KKO sınıflandırma algoritmasının kısa bir açıklaması Şekil 4.11.' de verilmiştir.

KKO Sınıflandırma Algoritması

Eğitim seti = {tüm eğitim örnekleri}

Kural Listesi = {}

Başlangıçta kural listesi boş bir listeyle başlatılır

(Eğitim setindeki kapsanmayan örnek sayısı > Kapsanan Maks Örnek sayısı) iken TEKRAR ET

$t=1$;

karınca indeksi

$j=1$;

yakınsama test indeksi

Başlangıçta tüm yollara aynı miktarda feromon atanır;

TEKRAR ET

$Karınca_t$ boş bir kuralla başlar ve her seferinde kural R_t 'ye yeni bir terim ekleyerek sınıflandırma kuralını aşamalı olarak oluşturur;

Elde edilen kural budanır;

Karıncanın geçtiği yollardaki feromon miktarı (R_t 'nin kalitesiyle orantılı olarak) artırılır ve ardından diğer yollardaki feromon miktarı da (feromon buharlaşmasını taklit edilerek) azaltılır;

Eğer ($R_t = R_{t-1}$) ise

yakınsama testi güncellenir

$j=j+1$;

Eğer $j=1$ ise;

$t=t+1$;

karınca indeksi artırılır

($t > \text{karınca sayısı}$) veya ($j > \text{yakınsayan kural sayısı}$) ise DUR;

Karıncalar tarafından oluşturulan tüm kurallar R_t , arasından en iyi $R_{en\ iyi}$ kuralı seçilir;

Keşfedilen kural listesine kural $R_{en\ iyi}$ eklenir;

Eğitim Seti = Eğitim Seti- { $R_{en\ iyi}$ tarafından doğru olarak kapsanan eğitim örnekleri};

DUR

Şekil 4.11. KKO sınıflandırma algoritması [134]

Sınıflandırma algoritmalarının çoğu ya yanlış sınıflandırma hatalarını görmezden gelmekte ya da tüm yanlış sınıflandırma hatalarının maliyetlerinin eşit olduğunu varsaymaktadırlar. Fakat gerçek hayatta pek çok sınıflandırma problemi farklı yanlış sınıflandırma maliyetlerine sahiptir ve bu farklılık göz ardı edilemez. Bu nedenle son yıllarda maliyet duyarlı sınıflandırma yaklaşımları ile ilgili çalışmalar önem kazanmıştır. Maliyet duyarlı öğrenme ile maliyet duyarsız öğrenme arasındaki en önemli fark, yanlış sınıflandırma maliyetlerini dikkate almaları ve beklenen yanlış sınıflandırma sayısını asgariye indirmek yerine, beklenen yanlış sınıflandırma maliyetini en aza indirmeye çalışmasıdır.

4.6.3. Değerlendirme Ölçütleri

Değerlendirme ölçütleri, sınıflandırıcının verimliliğini karşılaştırma ve anlamak için temel bir unsurdur. Sınıflandırma konusundaki daha önceki araştırmalarda, birincil değerlendirme ölçütü olarak sınıflandırma doğruluğu (Denklem (4.17)) kullanılır.

$$\text{Sınıflama Doğruluğu} = \frac{TP+TN}{TP+FN+FP+TN} \quad (4.17)$$

Her iki sınıf için doğru ve yanlış sınıflandırma sonuçları karışıklık matrisi Tablo 4.2 'de verilmektedir. Tablo 4.2' de, TP ve TN yapılan doğru tahminlerini temsil etmektedir: TP, aslında hem pozitif hem de pozitif olarak tahmin edilen örneklerin sayısıdır ve TN, aslında hem negatif hem de negatif tahmin edilen örneklerin sayısıdır. Ayrıca, FP ve FN sınıflandırma hatalarının sayısını temsil eder.

Tablo 4.2. İkili sınıflandırma problemi için karışıklık matrisi

	Tahmin edilen	
	Pozitif sınıf	Negatif sınıf
Pozitif sınıf	TP	FN
Negatif sınıf	FP	TN

ROC eğrisi, fayda ve maliyet arasındaki göreceli dengeyi vurgulayan, dengesiz veriler için en uygun değerlendirme kriteri olarak literatürde yer almaktadır. ROC eğrisi, X-ekseni üzerinde $FP / TN + FP$ (FPR) ve Y ekseninde $TP / TP + FN$ (TPR) çizimidir; burada (FPR, TPR) sınıflandırıcının tüm muhtemel eşik değerlerine göre elde

edilir. Sınıflandırıcının eşiği, bir örneğin bir sınıfa üyelik derecesini gösterir. Dengesiz veri setlerinde öğrenmede, ROC Eğrisi Altındaki Alan (AUC) genellikle ROC eğrisi yerine sayısal bir ölçüt olarak kullanılır, daha büyük AUC değeri (Denklem (4.18)) daha iyidir.

$$AUC = \frac{1+TPR-FPR}{2} \quad (4.18)$$

Maliyet duyarlı öğrenme, farklı sınıflara göre değişen yanlış sınıflandırma maliyetlerini göz önünde bulundurur. Bu durumda, bir maliyet matrisindeki $C(i, j)$ maliyeti sınıf i ' ye ait olan bir örneği sınıf j olarak sınıflandırmanın maliyetini temsil eder. Bu yanlış sınıflandırma maliyeti değerleri uzmanlardan veya diğer yaklaşımlardan elde edilebilir. Özellikle veri setinin dengesiz olduğu problemlerle uğraşırken, pozitif örnekleri sınıflayabilmek negatif örnekleri sınıflayabilmekten daha önemlidir [91]. Bu nedenle, pozitif bir örneğin yanlış sınıflandırılma maliyeti negatif bir örneğin yanlış sınıflandırılma maliyetinden daha yüksek olması makul olanıdır ($C(+, -) > C(-, +)$). Bir sınıflandırma algoritması tarafından bir x örneğini sınıf i olarak sınıflandırmanın toplam yanlış sınıflandırma maliyeti $R(i|x)$ aşağıdaki gibi ifade edilir:

$$R(i|x) = \sum_j P(j/x) \cdot C(i, j) \quad (4.19)$$

Sınıflandırma doğruluğu ile değerlendirme esnasında, hata maliyetleri eşit kabul edilir. Aslında, bir sınıflandırma sistemindeki hata maliyetlerindeki eşitsizliklere karşı kayıtsız kalmak çok zordur. Bu gibi durumlarda doğruluk maksimizasyonu tek başına yetersiz olabilir. Bu nedenle tezde AUC ve yanlış sınıflandırma maliyeti ile sınıflandırma sonuçları değerlendirilmiştir.

4.6.4. Önerilen Çok Amaçlı ve Maliyet Duyarlı KKO Sınıflandırma Yaklaşımı

Tezde, özgün kural yapısı ile eğitim verilerinden kurallar çıkarmak için çok amaçlı ve maliyet duyarlı KKO sınıflandırma algoritması geliştirilmiş ve uygulanmıştır. Nitelik seçimi aşamasında elde edilen bütün nitelik alt kümeleri için ayrı ayrı KKO sınıflandırma algoritması uygulanır ve amaç fonksiyonu değerleri her bir nitelik alt kümesi için ayrı ayrı elde edilir. Önerilen algoritmada kuralların kalitesi değerlendirilirken amaç fonksiyonu olarak yanlış sınıflandırma maliyeti ve AUC değeri uygulanmıştır.

Önerilen KKO sınıflandırma algoritmasında, kural yapısı IF-THEN kuralları biçiminde tanımlanmış ve her koşul üç alt öğeden oluşmaktadır: nitelik, ilişkisel operatör, niteliğin değeri. Bu tanımlanan kural yapısı KKO sınıflandırma algoritması ile bu tezde uygulanmıştır ve kural yapısının sınıflandırma algoritmasına olan katkısı sayesinde etkin sınıflandırma sonuçları elde edilebilmiştir.

Müşteri veri kümesi eğitime tabi tutulmadan önce, eğitim ve test olarak 10 katmanlı çapraz doğrulama yaklaşımı ile ayrılmıştır. Eğitim aşamasının başlangıcında, terimler veri kümesi rastgele oluşturulur ve her bir terime aynı miktarda feromon atanır. Terimler IF-THEN kural yapısındaki bir koşulu ifade etmektedir. Maximum iterasyon sayısı kadar her bir karınca bu terimlerle kurallar oluşturur.

Karıncaların oluşturdukları kurallarda yer alan gereksiz koşulların çıkarılmasıyla kurallar budanır. Elde edilen kurallarda yer alan terimlerin feromon miktarları kuralların sınıflama kalitesi ile doğru orantılı olarak güncellenir. Sınıflandırma kalitesi; uygunluk fonksiyonları: yanlış sınıflandırma maliyeti ve AUC değeri ile belirlenir. Tüm iterasyonların sonunda bu uygunluk fonksiyonlarına göre tüm kurallar arasından pareto çözümler seçilir. Bu kümedeki kurallara global feromon güncellemesi uygulanır. Eğitim aşamasında elde edilen pareto kümesindeki kurallar test veri setine uygulanarak algoritmanın sınıflandırma gücü elde edilmiş olur. KKO sınıflandırma algoritmasında seçilen uygunluk fonksiyonları, KKO sınıflandırma algoritmasını başarılı bir şekilde maliyet duyarlı hale getirmiştir.

5. BÖLÜM

ÖNERİLEN MÜŞTERİ KAYBI TAHMİN MODELİNİN UYGULAMASI

5.1. Giriş

Literatürde, müşteri kaybı yönetimi için çoğunlukla müşteri kaybı olasılıklarına göre ayrılma ihtimali olan müşteri tahmin edilmeye yönelik çalışmalar yer almaktadır. Ancak müşteri kaybı analizi uygulamalarında, veri setinin boyutları ve özellikleri dikkate alındığında arzulanan amaçlara ulaşabilmek için kompleks ve güçlü yöntemlerin uygulanması gereklidir. Hazırlanan tezde, Türkiye'de müşteri kaybı ile ilgili sorun yaşayan gerçek bir telekomünikasyon şirketi için, müşteri kaybı yönetim modeli oluşturulup çözümlenmiştir. Bu tezde müşteri kaybına neden olan 66 adet nitelik ayrıntılı olarak belirlenmiştir. Bu niteliklerle derlenen müşteri veri kümesine, önerilen entegre model ile müşteri kaybı tahmini yapılmıştır.

Bu bölümde ilk olarak uygulamada kullanılan müşteri veri setinin derlenmesi ve hazırlanması yer almaktadır. Devamında ise müşteri veri seti için belirlenen nitelikler detaylı olarak açıklanmıştır. Son olarak da önerilen müşteri kaybı modeli aşamaları detaylı olarak incelenmiş ve müşteri kaybı tahmini yapılmıştır. Önerilen model ile elde edilen sonuçlar iyi bilinen yaklaşımlarla karşılaştırılmıştır.

5.2. Veri Hazırlama ve Örneklem

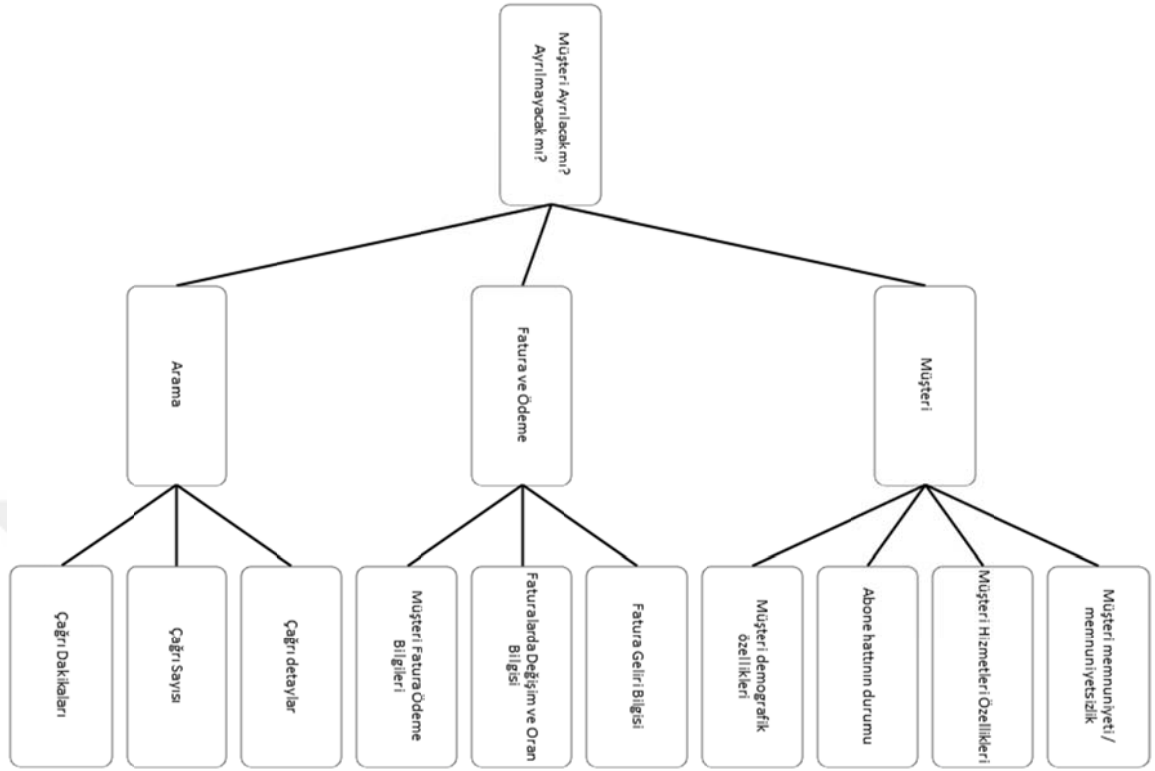
Müşteri kaybı tahminini mümkün olduğunca yüksek performanslı ve gerçeğe yakın gerçekleştirebilmek için müşteri veri setinde yer alan niteliklerin eksiksiz ve kapsayıcı olarak belirlenmesi çok önemlidir. Müşteri veri setinde yer alan niteliklerin belirlenmesi için detaylı bir çalışma yürütülmüştür. İlk olarak literatürde yer alan telekomünikasyon

ve diğer bütün sektörlerdeki müşteri kaybı analizi çalışmaları detaylı olarak incelenmiş ve bu çalışmalarda yer alan nitelikler listelenmiştir. Listelenen nitelikler üzerinde 18 BT sektöründe uzman ve 5 akademisyenle birlikte tartışılmış ve gerekli görülen nitelikler belirlenmiştir. Belirlenen niteliklere ek olarak uzman ekibin uygun gördüğü nitelikler de eklenerek 66 nitelikte müşteri veri seti daha da zenginleştirilmiştir. Literatürde, müşteri veri seti nitelikleri açısından bu kadar ayrıntılı müşteri kaybı yönetimi çalışması şu anki bilgimiz dâhilinde mevcut değildir.

Niteliklerin belirlenmesi kadar, doğru ve yeterli verinin elde edilmesi de tahmin modellerinin gerçek beklentiyi yansıtmasında önem arz etmektedir. Bunu sağlamak için de toplanan örneklem, abone portföyünün bütünsel resmini temsil etmeli ve model oluşturmak için yeterli sayıda müşteri sayısına sahip olmalıdır. Tahmin modeli verisi gözlem ve tahmin periyotları olmak üzere iki ana periyotta derlenmiştir. İlk aşamada, müşterilerin ayrılıp ayrılmayacağı gözlemlenmiş ve bu gözlem verisi ile kurallar elde edilmiştir. Tahmin periyodunda ise elde edilen kuralların doğruluğu test edilmiştir. Gözlem ve tahmin periyodunda tahmin modeli geliştirmek için istatistiksel olarak yeterli sayıda müşteri verisi derlenmiştir. Uygulamada kullanılan veri seti, 6496 müşteri verisinden oluşmaktadır.

Müşteri kaybı analizi için gerekli veriler ETL (Extraction-Transformation-Loading) araçlarını kullanarak şirket veri ambarlarından çekilmiştir. Modelin ETL işlemleri şunlardır: (a) ilgili müşteri bilgilerini kaynaktan tanımlama; (b) bu bilgilerin çıkarılması; (c) çeşitli kaynaklardan gelen verilerin entegrasyonu; (d) veri kümesinin temizlenmesi ve (e) verilerin veri ambarına yüklenmesi.

Şirket CDR (arama kaydı) verileri, kaynak veri tabanlarında mevcuttur. Bu veriler, milyonlarca arama ve kayıt şeklinde olup, operasyonel amaçlar için kullanılmaktadır. Burada; veri aktarımı, özetleme, denormalizasyon gibi ETL araçları, verileri tek tek tablolarda aylık özetler haline getirmek için kullanılmıştır.



Şekil 5.1. Bir müşteri ayrılma eğiliminde mi değil mi?

5.3. Müşteri Veri Seti Nitelikleri

Uygulamada Türkiye'nin ilk 100 BT şirketinden biri olan firmanın müşteri verileri kullanılmıştır. Veri kümesi 2008-2016 yılları arasındaki 6496 müşteri verisini ve 66 nitelik değerini içermektedir. Şekil 5.1' de gösterildiği üzere müşteri veri seti; müşteri, fatura ve ödeme, arama olmak üzere 3 ana nitelik başlığı altında 66 nitelik olarak derlenmiştir.

Tablo 5.1-a' da "müşteri" ana başlığı altında; "müşteri memnuniyeti", "hizmetlerin özellikleri", "abone hattının özellikleri" ve "müşterinin demografik özellikleri" olarak 4 alt başlık; 22 nitelik ve tanımları ile birlikte yer almaktadır.

Tablo 5.1-b' de "fatura ve ödeme" ana başlığı altında; "fatura geliri bilgisi", "faturalarda değişim ve oran bilgisi", "müşteri fatura ödeme bilgileri" olarak 3 alt başlık 12 nitelik ve tanımları ile birlikte yer almaktadır.

Tablo 5.1-c' de "arama" ana başlığı altında; "çağrı detayları", "çağrı sayısı", "çağrı dakikaları" olarak 3 alt başlık 32 nitelik ve tanımları ile birlikte yer almaktadır.

Tablo 5.1.-a Müşteri Niteliklerinin Tanımları

NİTELİKLER		TANIMI
MÜŞTERİ		
A- Müşteri memnuniyeti / memnuniyetsizlik		
1-	Hizmet tatmin faktörü [9-10]	Şikâyet sayısı, hizmet kalitesi skoru ve sadakat skoru gibi hizmet memnuniyeti faktörleri. Müşteri şikâyetleri, müşteri dönüşüm olasılığı ile pozitif ilişkilidir.
2-	Müşteriye yeni ürün pazarlanabilme potansiyeli	Müşteri doğrudan pazarlama eylemlerine yanıt veriyor ve ürün tanıtımını kabul ediyorsa pazarlanabilme potansiyelini gösterir.
B- Müşteri Hizmetleri Özellikleri		
1-	Müşterinin sahip olduğu aktif telefonların toplam sayısı [10,139]	Bireysel kullanıcılar için, çoğunlukla 1'dir.
2-	Hizmet Kullanım Süresi [140]	Bu nitelik, hizmet kullanımının tahmini başlangıç tarihini araştırma tarihinden çıkararak elde edilmiştir. Nispeten daha kısa aboneliklerin ayrılması daha olasıdır.
3-	Müşterinin taahhütlü prog. üye olup olmaması [9-10]	Taahhütlü programlara üye olmak müşterinin ayrılmasını olumsuz yönde etkilemektedir.
4-	Tarife grubu	Müşterinin kullanmakta olduğu tarife grubu sadık ve sadık olmayan müşterilere göre değişkenlik göstermektedir.
5-	Veri indirme hızı [77,141]	Veri yükleme ve indirme hızı
6-	Kampanya [6,10,11,142]	Müşterinin kampanyalara dâhil olup olmaması
7-	Kullanıcının aktif olduğu ay sayısı	Müşteri, hizmet kullanmaya başladıktan sonra geçen süre (ay cinsinden)
8-	İlk telefon görüşmesi	Müşterinin ilk telefon görüşmesi yaptığı tarih
C- Abone hattının durumu		
1-	Hizmet süresinin uzunluğu [6,10]	Müşterinin sözleşmeye başlamasından bu yana geçen süre (ay cinsinden)
2-	Müşteri segmenti [10,11]	
3-	Fesih için bir cezası var mı?	Müşterinin kullanmakta olduğu tarifeden dolayı ayrılması durumunda fesih cezası tahakkuk eder mi?
4-	Broadband service (ADSL veya FTTB) evet/hayır [77]	ADSL veya FTTB broadband servisini kullanıyor mu?
5-	Sözleşme iptal nedeni [10]	Müşterinin sözleşme iptal nedeni
6-	Sözleşme başlangıç tarihi [10]	Müşterinin sözleşme başlangıç tarihi
7-	Sözleşme iptal tarihi [10]	Müşterinin sözleşme iptal tarihi
8-	Sözleşme tipi [10,143]	Müşterinin sözleşme tipi
9-	Sözleşmenin son yenileme tarihi	Sözleşmesini yakın zamanda yenileyen müşterilerin ayrılma olasılığı daha düşüktür
10-	Uluslararası plan evet / hayır [21]	Uluslararası aramalar için telefon ayarlarının açık olup olmaması
D- Müşteri demografik özellikleri [6,80, 143,144,145]		
1-	Müşterinin şehri [11]	Müşterilerin kırsal veya kentsel alanda yaşayıp yaşamadığı bilgisi müşteri kaybı tahmini için gereklidir. Çünkü kentsel alanlarda yaşayan müşteriler kırsal kesimden daha fazla ayrılma eğilimlidir.
2-	Müşterinin doğum günü [6,10,11,77,80,139,143]	Müşterinin doğum günü

Tablo 5.1.-b Fatura ve Ödeme Niteliklerinin Tanımları

NİTELİKLER	TANIMI
FATURA VE ÖDEME [80]	Müşteri fatura tarihleri, her fatura maliyet, fiyatlar, ödenen ücretler vb. bilgi içerir.
A- Fatura Gelir Bilgisi	
1- Ortalama aylık fatura [9,139, 142]	Aylık faturalandırılmış tutarlar, müşterinin ayrılma ihtimaliyle ilişkilidir
2- Toplam fatura tutarı	Müşteriye faturalanan tüm faturaların toplamıdır. Bu tutar ne kadar fazla ise müşterinin ayrılma ihtimali o kadar düşüktür.
3- Önceki ay ödenen fatura tutarı	Önceki ay ödenen fatura tutarı eğer düşükse müşteri başka servis sağlayıcısı arayışında olduğunu gösterir
4- İki ay önceki dakika başına ödenen miktar [10]	Bir önceki aya ait toplam faturalandırma / çağrı dakika sayısı
5- Dakikada en son ödenen ücret [10]	Bir önceki aya ait toplam fatura / çağrı dakika sayısı
6- Günlük aramalarının toplam ücreti [21]	Günlük aramalarının toplam ücreti
7- Ortalama paket dışı gelir	Paket dışındaki sesli aramalardan elde edilen ortalama gelir
B- Faturalarda Değişim ve Oran Bilgisi	
1- Ödenmemiş aylık fatura sayısı [9]	Ödenmemiş aylık faturaların sayısı
2- Gecikme ücreti toplamı	Müşterinin gecikmeden dolayı yaptığı ödeme
3- Gecikmiş ödeme sayısı	Müşterinin gecikmiş ödeme sayısı eğer fazlaysa müşteri istemeyerek de olsa ayrılmak zorunda kalacaktır
4- Ortalama fatura tutarı [9]	Müşterilerin şimdiye kadar ödediği faturalarının ortalama tutarı
C- Müşteri Fatura Ödeme Bilgileri [74]	
1- Ödeme yöntemi [9,77, 80, 140,143]	Müşteri hesaplarıyla ilgili hesap bilgileri, örneğin: Başlangıç zamanı, hesap numarası / türü, ödeme türü, hesap bakiyesi, çağrı bilgisi vb.
1- Ödeme yöntemi [9,77, 80, 140,143]	Bir banka hesabından veya bir kredi kartından doğrudan ödeme

Tablo 5.1.-c Arama Niteliklerinin Tanımları

NİTELİKLER	TANIMI
ARAMA	
A- Çağrı detayları [80]	
1- Son çağrı [146]	Telekomünikasyon şirketlerinin çoğu çağrı detaylarını saklarlar. Çağrı detayları, müşteri hizmetleri kullanımının sıklığını gösterebilir.
2- Ortalama arama uzunluğu (saniye olarak) [11]	Müşterinin belirli bir süre içinde yaptığı son arama tarihi
3- Toplam uluslararası arama ücreti [21]	Geçen ayda ortalama arama uzunluğu (saniye olarak)
4- Akşam arama ücretleri toplamı [21]	Toplam uluslararası arama ücreti
B- Çağrı Sayısı [4]	
1- Toplam çağrı yaptığı gün sayısı	Akşam arama ücretleri toplamı. Önceki ayın akşamları 19:00 ile 24:00 arasında yapılan aramalardan elde edilen gelir.
2- Aylık bazda çağrı adedin ortalaması	Müşterinin firmayı her gün kullanıyor olması ayrılma olasılığını azaltır
3- Geçen 6 ay boyunca aylık	Çağrı adedi ile müşteri kaybı ters orantılıdır
	Geçen 6 ay boyunca aylık ortalama arama sayısı

	ortalama arama sayısı	
4-	Son 6 aydaki toplam görüşme sayısı	Son 6 ayda müşteri için yapılan toplam arama sayısı
5-	Toplam uluslararası çağrı sayısı	Toplam uluslararası çağrı sayısı
6-	Toplam gece görüşme sayısı [21]	00:00 ile 09:00 saatleri arasında yapılan toplam görüşme sayısı
7-	Toplam arama yaptığı gün sayısı [21]	Bir önceki ay için 09.00 ile 19.00 arasında sabah yapılan aramaların toplam sayısı
8-	Toplam akşam görüşme sayısı [21]	Geçen ay 19.00 ile 24.00 arasında gece boyunca yapılan toplam görüşme sayısı
9-	Müşteri hizmetleri ortalama arama sayısı [21, 139]	Müşteri hizmetleri ortalama arama sayısı
10-	Müşteri hizmetleri arama sayısı [139]	Müşterinin ömrü boyunca müşteri hizmetlerine yapmış olduğu toplam çağrı sayısı
11-	Toplam mesaj sayısı [10]	Müşterinin geçen ay aldığı toplam mesaj sayısı
C- Çağrı Dakikaları		
1-	Yapılan toplam çağrılarının aylık ortalaması [139]	Müşterinin ömrü boyunca aylık ortalama dakikalarının ortalaması (saniye cinsinden)
2-	Son ay arama süresi	Son ayda yapılan toplam görüşme süresi (saniye olarak)
3-	Müşterinin aylık arama süreleri toplamı	Müşterinin ömrü boyunca toplam konuşma süresi (dakika cinsinden)
4-	Faturalandırılmış toplam dakika [139]	Müşterinin ömrü boyunca fatura düzenlenmiş toplam konuşma süresi (dakika cinsinden)
5-	Paket dışı yapılan ortalama çağrı süresi	Müşterinin paket dışında yapmış olduğu çağrılarının ortalama süresi (dakika cinsinden)
6-	Paket dışı yapılan toplam çağrı süresi	Müşterinin paket dışında yapmış olduğu bütün çağrılarının toplam süresi (dakika cinsinden)
7-	Ortalama arama süresi [142]	Müşterinin aylık ortalama arama süresi (sn cinsinden)
8-	Operatör içi toplam arama süresi	Operatör içi yapılan toplam arama süresi (sn cinsinden)
9-	6 aylık ortalama arama süresi [143]	Müşteriye ait önceki 6 aydaki ortalama arama süresi (sn cinsinden)
10-	3 aylık ortalama arama süresi [11]	Müşteriye ait önceki 3 aydaki ortalama arama süresi (sn cinsinden)
11-	Önceki 3 aylık dönemde kullanılan aylık ortalama dakika sayısı [11]	Önceki 3 aylık dönemde kullanılan aylık ortalama dakika sayısı
12-	Uluslararası toplam çağrı dakikaları [21]	Uluslararası toplam çağrı dakikaları
13-	Gece arama süreleri toplamı [21]	Bir önceki aya ait 00: 00-09: 00 saatleri arasında arama süresi (dakika cinsinden)
14-	Gündüz arama süreleri toplamı [21]	Bir önceki aya ait sabah 9'dan akşam 9'a kadar arama süresi (dakika cinsinden)
15-	Akşam arama süreleri toplamı [21]	Önceki ay için gece saat 19.00 ile 24.00 arası görüşmelerin toplam süresi (dakika cinsinden)
16-	Müşteri hizmetleri çağrılarında kullanılan arama süresi toplamı [139]	Müşteri hizmetleri çağrılarında kullanılan arama süreleri toplamı (dakika cinsinden)
17-	Müşteri hizmetleri çağrılarında kullanılan ortalama arama süresi [139]	Müşteri hizmetleri çağrılarında kullanılan arama süreleri ortalaması (dakika cinsinden)

5.4. Önerilen Müşteri Kaybı Tahmin Modeli

Bu tezde, yüksek boyutlu ve dengesiz dağılıma sahip telekomünikasyon müşteri verilerini işlemek üzere, nitelik seçimi ve sınıflandırılmasına yönelik entegre bir yaklaşım önerilmiştir. Önerilen model temel olarak Şekil 5.2.'de gösterilmiştir. Nitelik seçimi aşamasında, NSGA-II ve bulanık kaba küme tabanlı yaklaşımlarla nitelik alt kümeleri elde edilmiştir. Sınıflandırma aşamasında ise, çok amaçlı ve maliyet duyarlı KKO sınıflandırma algoritması ile müşteri kaybı tahmini için kurallar elde edilmiştir.

Önerilen model, veri toplama, ön işleme, nitelik seçimi ve sınıflandırma olmak üzere 4 ana aşamadan oluşur. Önerilen modelin temel adımları aşağıda verilmiştir:

ÖNERİLEN MÜŞTERİ KAYBI TAHMİN MODELİ

Aşama 1: Veri Toplama

Müşteri kaybı tahmini için uygun niteliklerin belirlenmesi

Veri setinin, ETL araçları kullanılarak şirket veri ambarlarından çekilmesi

Aşama 2: Veri Temizleme ve Ön işleme

Eksik değerlerin doldurulması ve hatalı verilerin düzeltilmesi

Aşama 3-a: Önerilen NSGA-II nitelik seçimi prosedürü

Nitelik alt kümelerinden oluşan, N boyutlu P_0 başlangıç popülasyonu rastlantısal olarak üretilir

Çaprazlama yapılacak bireylerin seçimi için ikili turnuva seçimi uygulanması

Çaprazlama ve mutasyon işlemcilerinden sonra N boyutlu Q_0 yeni popülasyon üretilir

$R_0 = P_0 \cup Q_0$

Durdurma kriteri (max iterasyon sayısı) sağlanana kadar TEKRAR ET

t . nesil için

$R_t = P_t \cup Q_t$

F =Hızlı bastırılmamış sıralama prosedürü (R_t)

$P_{t+1} = \{ \}$ ve $i=1$

$|P_{t+1}| + |F_i| \leq N$ olana kadar TEKRAR ET

$P_{t+1} = P_{t+1} \cup F_i$

$i=i+1$

DUR

Sırala (F_i, a_n)

$P_{t+1} = P_{t+1} \cup F_i[1:(N-|P_{t+1}|)]$

P_{t+1} popülasyonuna çaprazlama ve mutasyon işlemcileri uygulanarak Q_{t+1} popülasyonu üretilir

$t=t+1$

DUR

Amaç fonksiyonlarına göre bastırılmamış tüm nitelik alt kümeleri elde edilir

Aşama 3-b: FRQR algoritması ile nitelik seçimi prosedürü

Başlangıçta nitelik alt kümesi R boş küme olarak tanımlanır

En iyi bağımlılık değeri ($\gamma'_{en iyi}$) ve önceki bağımlılık değeri ($\gamma'_{önceki}$) 0'a eşitlenir

$\gamma'_{en iyi} = \gamma'_{önceki}$ olana kadar TEKRAR ET

Nitelik alt kümesi T = Nitelik alt kümesi R

Önceki bağımlılık değeri ($\gamma'_{\text{önceki}}$) = en iyi bağımlılık değeri ($\gamma'_{\text{en iyi}}$)

R nitelik alt kümesinde yer almayan her x niteliği için

Eğer $\gamma'_{R \cup \{x\}}(D) > \gamma'_T(D)$ ise

$T = R \cup \{x\}$

$\gamma'_{\text{en iyi}} = \gamma'_T(D)$

$R = T$

DUR

R nitelik alt kümesi elde edilir

Aşama 4: Çok amaçlı ve maliyet duyarlı KKO Sınıflandırma Algoritması

Her bir sınıf için TEKRAR ET

Her bir nitelik alt kümesi için TEKRAR ET

Terim veri kümesi oluşturulur

Tüm terim veri kümesi elemanlarına aynı miktarda feromon atanır

Pareto kural listesi boş bir listeye başlatılır

Durdurma kriteri (max iterasyon sayısı) sağlanana kadar TEKRAR ET

TEKRAR ET

$Karınca_t$ sınıflandırma kuralını oluşturur;

Elde edilen kural budanır;

Bütün yollardaki feromon miktarı güncellenir;

Eğer ($R_t = R_{t-1}$) ise

yakınsama testi güncellenir

$t=t+1$;

karınca indeksi artırılır

($t \geq \text{karınca sayısı}$) veya ($j \geq \text{yakınsayan kural sayısı}$) ise DUR;

Karıncalar tarafından oluşturulan tüm kurallar arasından pareto çözümler seçilir;

Şimdiye kadar elde edilen bütün pareto çözüm kümesi güncellenir;

Pareto çözüm kümesindeki kurallara global feromon güncellenmesi uygulanır

Eğitim veri setinden pareto kümesindeki kuralların içerdiği terimler çıkartılır

DUR

DU

R

DU

R

Tüm sınıflar için elde edilmiş kurallar problem test verisi kümesine uygulanır

AUC ve yanlış sınıflandırma maliyet değerleri hesaplanır

Aşama 1: Veri Toplama

Müşteri verileri şirket veri ambarlarından çekilmeden önce müşteri kaybını etkileyebilecek niteliklerin belirlenmesi için detaylı araştırma yapılmıştır. Öncelikle literatürde yer alan makaleler incelenmiş ve müşteri kaybı yönetiminde kullanılan nitelikler listelenmiştir. BT sektöründen 18 uzman ve 5 akademisyenle toplantılar yapılarak listelenen nitelikler değerlendirilmiş, aralarından uygun olanlar seçilmiştir. Bu niteliklere ek olarak da yine uzman ekibin gerekli gördüğü diğer nitelikler de müşteri veri setine dahil edilmiştir. Nitelikler belirlendikten sonra, veri kümesi, ETL araçları kullanılarak şirket veri ambarlarından çekilmiştir. ETL süreci ve niteliklerin belirlenmesi “Veri Hazırlama ve Örneklem” bölümünde ayrıntılı olarak açıklanmıştır.

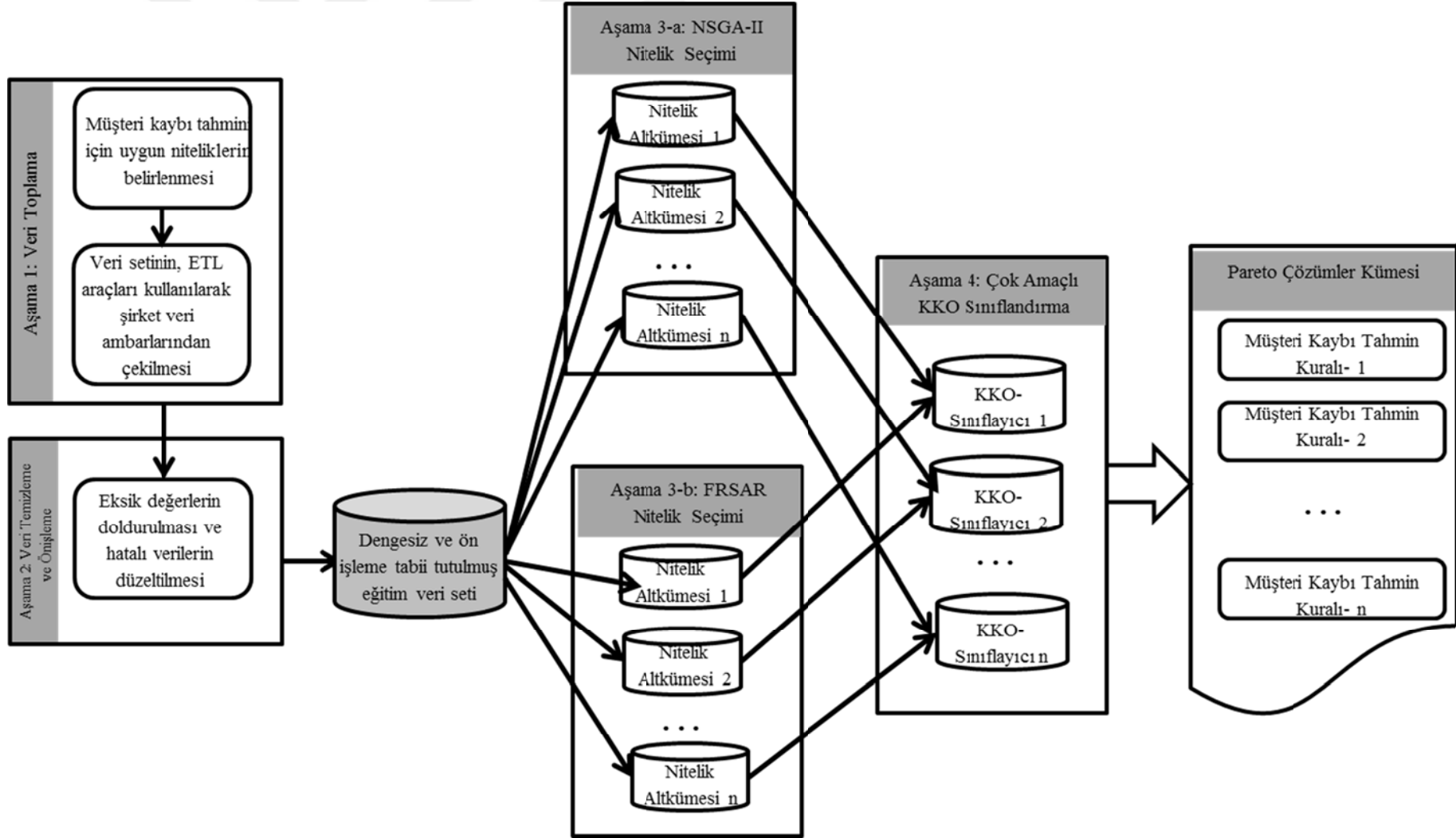
Aşama 2: Veri Temizleme ve Önleme

Veri madenciliğinde, iyi sonuç elde edebilmek için ham veriyi ön işleme tabi tutarak kaliteli veri hazırlanması elzem bir aşamadır. Genellikle, veri temizleme ve hazırlama, toplam veri mühendisliği çabasının yaklaşık %80' ini kadardır. Müşteri kaybı tahmini için kullanılan müşteri veri setinde de eksik, gürültülü ve yanlış değerler bulunmaktadır.

Veri setlerinde eksik veri olması sık görülen bir sorundur. Eksik veri oranının % 1'den az olması genellikle önemsiz olarak kabul edilir, %1-5'lik eksik veriye sahip veri setleri yönetilebilir. Bununla birlikte, %5-15'lik eksik veriye sahip veri setlerinin üstesinden gelinmesi için gelişmiş yöntemler gerekmede ve %15'den fazlası ciddi şekilde sonuçları etkileyebilmektedir [147]. Bu çalışmada uygulanan müşteri veri setinde eksik veri oranı %8.9' dur ve %5-15'lik aralığında kaldığı için de gelişmiş yöntemler ile bu durum yönetilmiştir.

Önerilen modelde kullanılan veri setinin eksik veri oranı da dikkate alındığında, iyi sonuç elde edebilmek için veri temizleme ve ön işleme aşamasına ayrıca önem verilmiştir. Veri setindeki kayıp verilerle başa çıkabilmek için; veri setimizden silmek yerine, eksik veriler tamamlanmıştır, böylece bilgi kaybının önüne geçilmiştir. Eksik değer; nitelik süreklilyse ortalama değerlerle, nitelik kategorik veya ikili ise en sık kullanılan değer ile tamamlanmıştır. Veri kümesindeki gürültülü ve yanlış değerlerle başa çıkabilmek için, hatalı değerler düzeltilmiştir ve aykırı değerler çıkarılmıştır. Aykırı değerler; mekanik arızalar, sistem davranışındaki değişiklikler, dolandırıcılık davranışı, algılayıcı gürültüsü, proses bozulmaları, cihaz bozulması ve/veya insanla ilgili hatalar nedeniyle ortaya çıkabilen, gözlemlerin çoğundan belirgin olarak sapma gösteren değerlerdir.

Aşama 2'de, eksik değerler doldurulmuş ve hatalı veriler düzeltilmiştir. Bu aşamanın modele eklenmesiyle önerilen modele aynı zamanda eksik, gürültülü ve yanlış verilerin de üstesinden gelebilme yetisi kazandırılmıştır. Veri kümesi düzenlendikten sonra, müşteri kaybı analizinde kullanılmak üzere veri tabanına yüklenmiştir.



Şekil 5.2. Önerilen Müşteri Kaybı Tahmin Modeli Yapısı

Aşama 3: Nitelik Seçimi

Aşama 3-a: Önerilen NSGA-II nitelik seçimi prosedürü

Önerilen NSGA-II nitelik seçim yaklaşımı, en düşük yanlış sınıflandırma maliyetine ve nitelik alt küme eleman sayısına sahip global bastırılmamış çözümleri araştırarak nitelik alt kümelerini bulur ve bireylerin birbirlerine olan baskınlıkları bu amaç fonksiyonlarına göre değerlendirir. Nitelik alt kümelerini elde edebilmek için kullanılan kromozom gösterimi aşağıda kodlanmıştır. Her kromozom 2 sıradan oluşmaktadır. İlk satır nitelikler, diğeri ise bu niteliğin alt kümede yer alıp almadığını gösterir. Nitelik alt kümesinde o nitelik yer alıyorsa, aktif (1) veya pasif (0) olarak etiketlenir.

Kromozom gösterimi									
Nitelikler	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
aktif/pasif	1	0	1	1	0	0	0	1	0

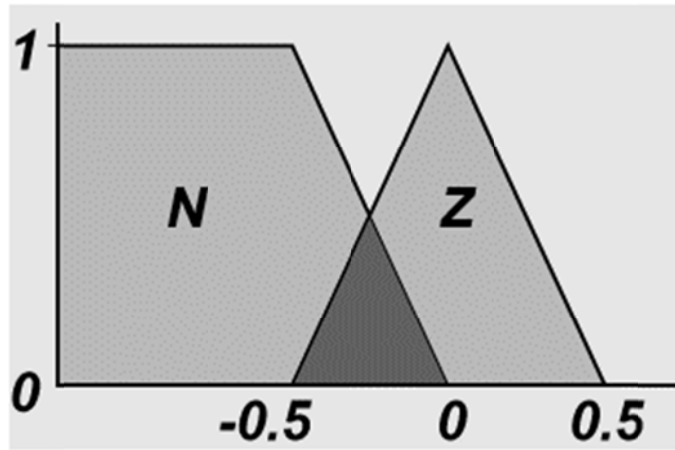
Önerilen NSGA-II nitelik seçiminde uygulanan genetik operatörler:

- NSGA-II nitelik seçimi algoritmasında eşleşme havuzuna alınacak bireylerin seçim işlemi için ilk iterasyonda ikili turnuva seçimi yöntemi tercih edilmiştir. İkili turnuva seçiminde, popülasyondan rastgele olarak seçilen iki birey, nitelik alt küme eleman sayısı ve yanlış sınıflandırma maliyeti açısından karşılaştırılmakta ve daha iyi olan bireyler belirlenmektedir. Belirlenen bireyler arasından eşit olasılıkla bir amaç fonksiyonunun rastgele seçilmesi sağlanmıştır. Eşleşme havuzunda popülasyon büyüklüğü sayısı kadar birey olana kadar bu işlem tekrarlanır.
- Sonraki iterasyonlarda seçim işlemi, NSGA-II algoritmasının kendi mekanizmalarıyla gerçekleştirilmektedir.
- NSGA-II algoritmasının seçim mekanizmasıyla oluşturulan eşleşme havuzuna, belirli oranda çaprazlama işlemi uygulanmıştır. Bu oranda yapılan mutasyon operatörü, popülasyondan seçilen bireylerde, küçük değişiklikler yapılmasını sağlar ve işlem yalnızca seçilen bireyi etkiler.
- Popülasyondaki her birey, denemeler sonucunda belirlenen 0,05 oranında mutasyon işlemine tabi tutulur. Bu çalışmada, bireyde rastgele seçilen bir

konumdaki deęer 0 ise 1 olarak, eęer 1 ise 0 olarak deęiřtirilerek mutasyon iřlemi gerekleřtirilir.

Ařama 3-b: FRQR Nitelik Seim prosedürü

Kaba küme teorisi; veri analizi için nesnelerin yaklařık tanımlarını oluřturmada iyi bir tekniktir ve literatürde nitelik indirgeme alanında etkili bir řekilde uygulanmıřtır. Klasik kaba küme temelli nitelik indirgeme algoritmaları ancak kesikli veri kümeleri ile etkin bir řekilde alıřabilmektedir. Ancak telekomünikasyon müřteri veri setinde sürekli nitelikler de yer aldıęı için Jensen ve Shen [90] tarafından önerilen FRQR nitelik indirgeme algoritması tercih edilmiřtir. Kaba küme teorisinde hibir ilave bilgiye ihtiya duymadıęı için girdi verilerinin bulanıklařtırılması için řekil 5.3.' deki üyelik fonksiyonu kullanılmıřtır.



řekil 5.3. Üyelik fonksiyonu

FRQR prosedüründe ilk olarak baęımlılık deęerlerine 0 bařlangı deęeri verilir. Nitelik alt kümesi R de bařlangıta boş küme olarak belirlenir. Her bir adımda nitelik alt kümesine bir nitelik daha eklenerek baęımlılık deęeri hesaplanır. Eęer elde edilen yeni nitelik alt kümesinin γ' deęeri nitelik eklenmeden önceki γ' deęerinden daha büyükse nitelik alt kümesi güncellenir. Yeni nitelik alt kümesi baęımlılık deęeri, en iyi baęımlılık deęerine eřit olana kadar algoritmanın adımları tekrar edilir.

Aşama 4: Önerilen Çok amaçlı ve maliyet duyarlı KKO Sınıflandırma Prosedürü

Bu tezde, çok amaçlı ve maliyet duyarlı KKO sınıflandırma algoritması geliştirilmiş ve uygulanmıştır. Bir önceki aşamada elde edilen bütün nitelik alt kümeleri için ayrı ayrı KKO sınıflandırma algoritması uygulanır ve amaç fonksiyonu değerleri her bir nitelik alt kümesi için ayrı ayrı elde edilir.

Önerilen KKO sınıflandırma algoritmasında uygulanan kural yapısı:

Önerilen KKO sınıflandırma algoritmasında, kural yapısı IF-THEN kuralları biçiminde tanımlanmıştır ve aşağıdaki şekilde ifade edilir:

IF <koşullar> THEN <sınıf>

Kural öncülü (IF bölümü), mantıksal birleştirme operatörleri (VE) / (VEYA) ile bağlı bir dizi koşul içerir.

IF koşul 1 VE/VEYA koşul 2 VE/VEYA. . .

Bir kuralın her ögesi üç alt öğeden oluşur: nitelik, ilişkisel operatör, niteliğin değeri. İlişkisel operatörlere: küçük veya eşit ($\leq$), büyük ($>$) operatörlerinden atanabilir. THEN kısmı, kuralın öncülünde belirtilen tüm şartları sağlayan durumlar için tahmin edilen sınıfı belirtir. Alt öğelerden ve mantıksal birleştirme operatörlerinden oluşan kural yapısı literatürde KKO sınıflandırma algoritmasına bu tezde uygulanmıştır.

Önerilen çok amaçlı KKO sınıflandırma algoritması eğitim aşaması adımları:

“ÖNERİLEN MÜŞTERİ KAYBI TAHMIN MODELİ” nde yapay kodu verilen KKO sınıflandırma algoritması her nitelik alt kümesi ve her sınıf için ayrı ayrı çalıştırılır.

Başlangıçta, terimler veri kümesi rastgele oluşturulur. Terimler veri kümesindeki her bir terim 3 alt öge ve bir nitelikten oluşur.

$Terim_{ijkl} = Nitelik_i, Değer_j, operatör_k \text{ ve işlemci}_j$

Nitelikler, alt kümede yer alan nitelikler arasından seçilir. Operatörlere ise rastgele küçük veya eşit ($\leq$) veya büyük ($>$) atanır. Nitelik değerleri, ilgili niteliğin sahip

olabileceği minimum ve maksimum değer aralığından rastgele atanır. Ayrıca, aktif niteliğe atanan bu operatörler ve değerler, sınıflandırma kurallarını elde etmek için kullanılır. Önerilen algoritma kural yapısında iki farklı işlemci grubu vardır: 've' ve 'veya'. Aktif niteliklerin işlemcileri de rastgele atanır, bu sayede farklı kurallar ortaya çıkar ve çeşitlilik sağlanır.

Başlangıç terimler kümesindeki her bir terime aynı miktarda feromon atanır. Pareto kural listesi boş küme olarak algoritmaya başlanır. Maximum iterasyon sayısı kadar her bir karıncanın kural oluşturmaları için aşağıdaki adımlar tekrarlanır.

Karıncanın sayısı kadar veya belirlenen yakınsayan kural sayısı alt limit aşılanaya kadar her bir karınca bir sınıflandırma kuralı türetir. Karınca ilk olarak boş bir kuralla başlar ve her seferinde kural R_t 'ye terim veri setinden $\eta_{ij} \times \tau_{ij}(t)$ ile orantılı olarak yeni bir $terim_{ijkl}$ seçilir ve sınıflandırma kuralına eklenir böylece kural aşamalı olarak oluşturulur. η_{ij} probleme özgü sezgisel fonksiyon ve $\tau_{ij}(t)$ ise o iterasyondaki $terim_{ijkl}$ ile ilgili feromon miktarıdır. Yüksek η_{ij} değeri $terim_{ij}$ 'nin sınıflandırma için daha uygun olduğu anlamına gelmektedir ve seçilme olasılığı daha yüksektir. Bir karınca tarafından oluşturulan kuralın kalitesi ne kadar iyi olursa, karınca tarafından bırakılan feromon miktarı da artar. Bu nedenle, zaman geçtikçe, bir kurala eklenecek en iyi terimlerin, seçilme olasılığı artarak, daha büyük miktarda feromona sahip olacaktır.

Elde edilen kural budanır yani uygun olmayan terimler çıkartılır. Bu, daha önce tanımlandığı üzere, yolların feromonlarını güncellemek için kullanılan aynı kural kalitesini kullanarak yapılır. Özünde, bir terim kuralın kalitesini düşürmezse, o terim kuraldan çıkarılır.

Her bir karınca bir kural oluşturduktan sonra yerel feromon güncellemesi yapılır. Bütün yollardaki feromon miktarı (R_t 'nin sınıflandırma kalitesiyle orantılı olarak) artırılır ve ardından diğer yollardaki feromon miktarı da (feromon buharlaşmasını taklit edilerek) azaltılır. Başka bir deyişle, kuralın kalitesi ne kadar yüksek olursa, kuralın öncülünde yer alan terimlerin feromonundaki artış o kadar yüksek olur. Önerilen algoritmada sınıflandırma kalitesi yanlış sınıflandırma maliyeti ve AUC değeri olarak belirlenmiştir.

Karıncalar tarafından oluşturulan kuralların amaç fonksiyon değerlerine göre tüm kurallar arasında pareto çözümler seçilir. Şimdiye kadar elde edilen bütün pareto

çözüm kümesi güncellenir. Bu kümedeki kurallara global feromon güncellenmesi uygulanır. Bir sonraki iterasyona geçmeden önce de terim veri setinden pareto kümesindeki kurallardaki terimler çıkartılır. Böylelikle bir sonraki iterasyon daha küçük bir terim seti ile iterasyon daha etkin bir şekilde ilerletilir.

Eğitim sürecinde aday kurallar oluşturulduktan sonra, sınıflandırma prosedürü eğitim veri setine uygulanır. Uygunluk kontrolü olarak kullanılan değerlendirme süreci, her bir sınıf için ayrı ayrı uygulanır. Bu aşamada, veri kümesini eğitim ve test olarak ayırmak için 10 katmanlı çapraz doğrulama yaklaşımı kullanılmıştır. Kuralların kalitesi, yanlış sınıflandırma maliyeti ve AUC ile birlikte değerlendirilmiştir. Algoritma olası tüm sınıf değerleri için ayrı olarak çalıştırılır ve pareto optimal küme değerleri için kuralların kalitesi elde edilir. Elde edilen kural yapısı ve pareto optimal küme saklanır, algoritmanın sonunda en iyi sınıflandırma çözüm kuralı olarak sunulur. Şimdiye kadar elde edilen pareto optimal küme güncellenir. Bu kümede birbirine bastıramayan kurallar yer almaktadır

Önerilen çok amaçlı KKO sınıflandırma algoritması test aşaması adımları:

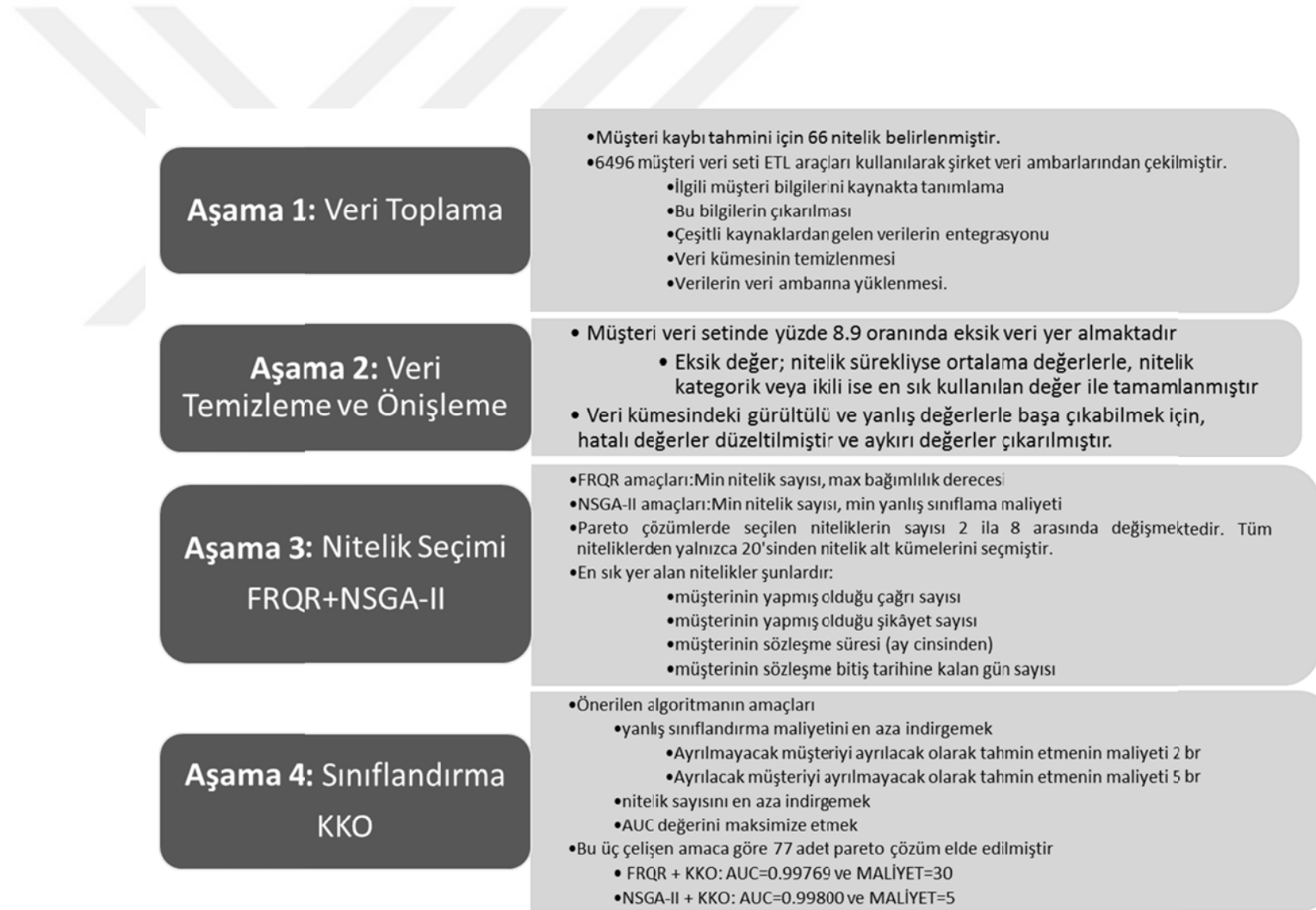
Eğitim aşamasında elde edilen pareto kümesindeki kurallar test veri setine uygulanır. KKO sınıflandırma algoritması, eğitim aşamasında her sınıf için pareto kurallar üretir. Bir sınıf için birden fazla kural pareto kümesinde yer alabilir. Muhtemel her kural çifti için aşağıdaki temel kural yapısı uygulanır.

```

IF öncül1 THEN sınıf1
ELSE IF öncül2 THEN sınıf2
....
....
ELSE sınıfvarsayılan

```

Örnekler, ilgili kuralla tam olarak eşleşene kadar, sırayla listedeki kurallar denenir. Bir örnek listedeki kurallardan hiçbirini sağlamazsa, varsayılan sınıf olarak sınıflandırılır. KKO sınıflandırma algoritmasında varsayılan sınıf, veri kümesindeki en büyük sınıf olarak belirlenmiştir.

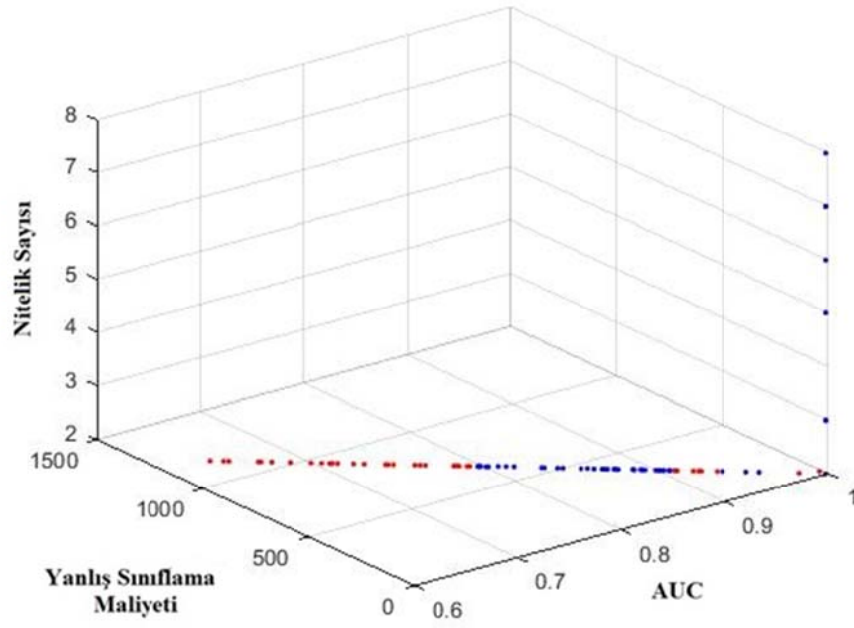


Şekil 5.4. Önerilen müşteri kaybı tahmin modelinin gerçekleşme adımları

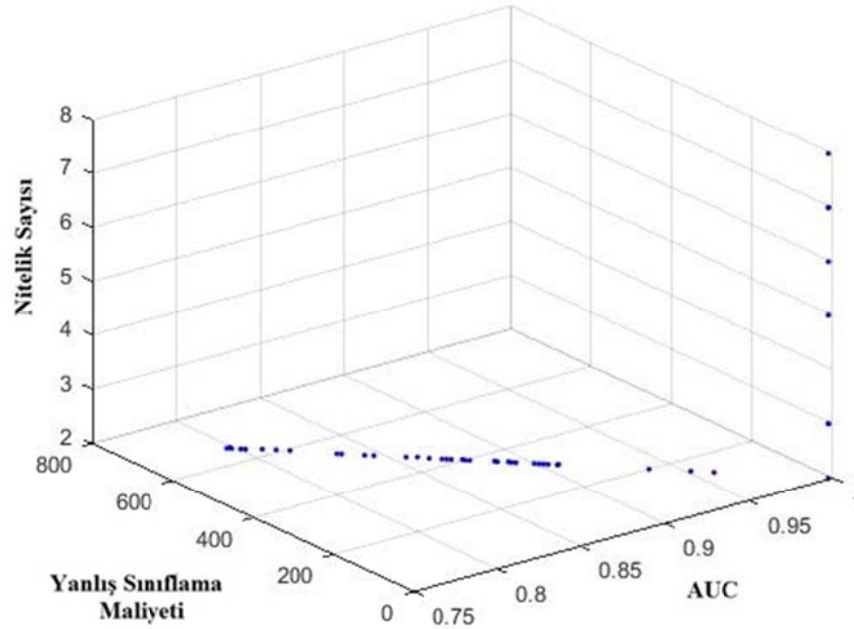
5.5. Önerilen Modelin Müşteri Kaybı Tahminine Uygulaması

Yukarıda detayı verilen müşteri veri setine önerilen model uygulanarak müşteri kaybı tahmini gerçekleştirilmiştir. Uygulamada önerilen modelin gerçekleşme adımları Şekil 5.4.'te verilmiştir.

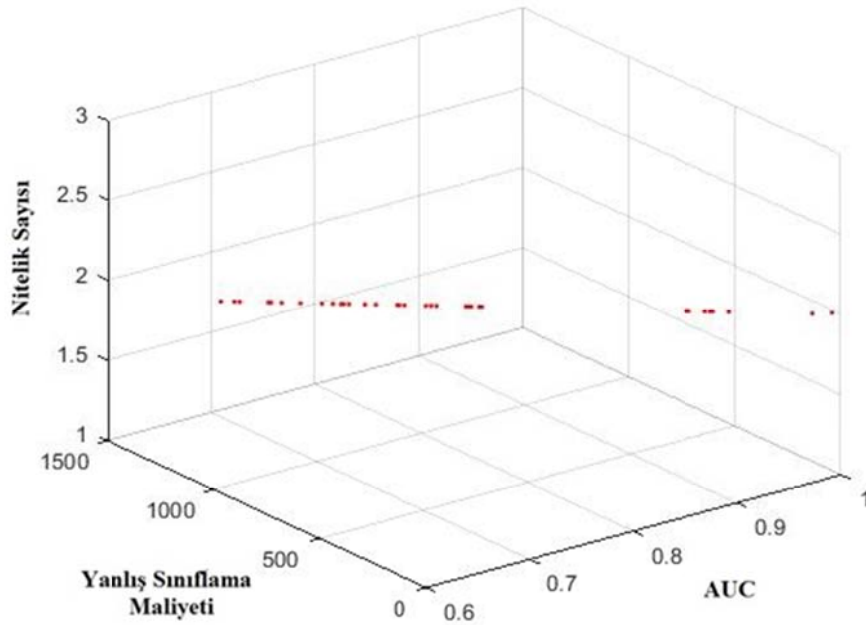
Uygulamada, 66 nitelik ve sınıftan oluşan 6496 müşterinin verileri kullanılmıştır. FRQR ile en az nitelik sayısı ile bağımlılık derecesi en yüksek olan nitelik alt kümesi elde edilmeye çalışılırken; NSGA-II tabanlı nitelik seçimi ile de yanlış sınıflandırma maliyeti ve nitelik sayısı minimize edilmeye çalışılmaktadır. Her iki nitelik seçimi yaklaşımlarından elde edilen nitelik alt kümeleri KKO sınıflandırma algoritmasına uygulanır. KKO sınıflandırma algoritması ise AUC değerini maksimize ederken aynı zamanda da yanlış sınıflandırma maliyetinin minimize etmeyi amaçlamaktadır. Önerilen algoritma aynı zamanda yanlış sınıflandırma maliyetini minimize etmeye çalışarak, nitelik sayısını minimize etmeye çalışarak ve AUC değerini maksimize etmeye çalışarak üç çelişen hedefi göz önüne alarak 77 adet pareto çözüm elde etmiştir. Bu modelin pareto optimal çözümleri, AUC, nitelik sayısı ve yanlış sınıflandırma maliyeti düşünüldüğünde geri kalan arama uzayı çözümlerinin üstünlüğü olarak ifade edilebilir. Her iki nitelik indirgeme yaklaşımından elde edilen nitelik alt kümelerinin KKO ile sınıflandırılması sonucu bulunan bütün pareto çözümler Şekil 5.5' te gösterilmiştir. Mavi noktalar NSGA-II nitelik indirgeme algoritması ile bulunan nitelik alt kümeleri ile elde edilen pareto çözümleri gösterirken, kırmızı noktalar ise FRQR nitelik indirgeme algoritması ile bulunan nitelik alt kümeleri ile elde edilen pareto çözümleri göstermektedir. Şekil 5.6 ve Şekil 5.7.' de ise pareto çözümler nitelik indirgeme yaklaşımlarına göre ayrı ayrı gösterilmiştir.



Şekil 5.5. NSGA-II ve FRQR + KKO ile elde edilen Pareto Çözümler



Şekil 5.6. NSGA-II + KKO ile elde edilen Pareto Çözümler



Şekil 5.7. FRQR + KKO ile elde edilen Pareto Çözümler

Önerilen algorithmadan elde edilen, pareto çözümlerde seçilen niteliklerin sayısı 2 ila 8 arasında değişmektedir. NSGA-II ve FRQR, tüm niteliklerden yalnızca 20'sinden nitelik alt kümelerini seçmiştir. Nitelik alt kümelerinde yer alan niteliklerin önemli bir kısmı da bu tezde önerilen niteliklerden oluşmaktadır. Bu durum önerilen niteliklerin anlamlı ve sınıflandırma kalitesine olumlu katkısı olduğunu göstermektedir. Ayrıca önerilen niteliklerin de yer aldığı alt kümelerde en sık yer alan nitelikler “müşterinin yapmış olduğu çağrı sayısı”, “tarife grubu”, “müşterinin yapmış olduğu şikâyet sayısı”, “müşterinin sözleşme süresi (ay cinsinden)”, “sözleşmenin son yenileme tarihi” ve “paket dışı yapılan toplam çağrı süresi” dir.

Tablo 5.2. Müşteri Kaybı Tahmininde Uygulanan Maliyet Matrisi

	Tahmin Edilen	
	Ayrılan Müşteri	Ayrılmayan Müşteri
Ayrılan Müşteri	0	5
Ayrılmayan Müşteri	2	0

Ayrılan müşteriye sınıflandırmak ile ayrılmayan müşteriye sınıflandırmak doğal olarak eşit maliyete sahip değildir. Ayrılmayacak müşteri ayrılacak olarak tahmin edildiğinden, müşteriye kaybetmemek için gereksiz çabalar gösterilecek ancak ayrılmayı düşünen bir müşteri ayrılmayacak olarak değerlendirildiğinde müşteriye elde tutmak için çaba harcanamayacak ve müşteri kaybedilecektir. Bu durum, diğerinden daha önemli ve sonuçları daha da pahalıya mal olacaktır. Uzmanların ve şirket yöneticilerinin görüşleri alınarak deneyler esnasında Tablo 5.2.'deki maliyet matrisi kullanılmıştır. Maliyet matrisinde verilen değerler, birim değerlerdir ve göreceli önemi göstermektedir. Bu maliyet matrisi belirlenirken müşteriye kaybetmenin ve elde tutmanın maliyet kalemleri tek tek hesaplanmış neticede göreceli önem değerleri elde edilmiştir. Müşteri kaybını önlemek için yapılan tekrar kazanma kampanyaları genellikle başarılı ancak elbette ki maliyetli faaliyetlerdir. Müşterilerin ilk etapta gitmesini önlemeye yardımcı olan sadakat kampanyaları, şirketin mevcut sağlayıcıdan ayrılmamaya karar vermesine yardımcı olan müşteri ilişkileri deneyimleri veya teklifler vermesine izin verir. Bunu sağlayabilmek için de reklam kampanyaları artırılabilir. Rakiplerle yarışabilmek için fiyatlar düşürülebilir. Bunların hepsi birer maliyet kalemidir.

Zhu et al. [148] müşteri kaybı tahmininde farklı çözümleri karşılaştırmak için bazı öncü çalışmalar yapmışlardır. Bu çalışmayı dikkate alarak deneyler sırasında önerilen algoritmayı performansını sınamak için veri seviyesi, algoritma seviyesi ve topluluk çözümlerinden 22 yöntem seçilmiştir. Deneyler sırasında veri kümesini eğitim ve test olarak ayırabilmek için 10 katmanlı çapraz doğrulama yaklaşımı uygulanmıştır. FRQR ve NSGA-II elde edilen bütün nitelik alt kümeleri için bu 22 yöntem uygulanmış ve böylece önerilen yöntemle karşılaştırması daha anlamlı hale gelmiştir. Seçilen tüm yöntemler, kısa açıklamaları ile birlikte Tablo 5.3. 'te listelenmiştir ve modellerin tanımları (<http://www.keel.es>)' da mevcuttur. Bu seçilen yöntemlerle sınıflandırma işleminde yapılan deneyler KEEL yazılım aracı ile yürütülmüştür. Örnekleme yöntemlerinin etkisini araştırmak için, iki sınıflandırma algoritması temel olarak kullanılmıştır: C4.5 karar ağacı ve Destek vektör makinesi (SVM).

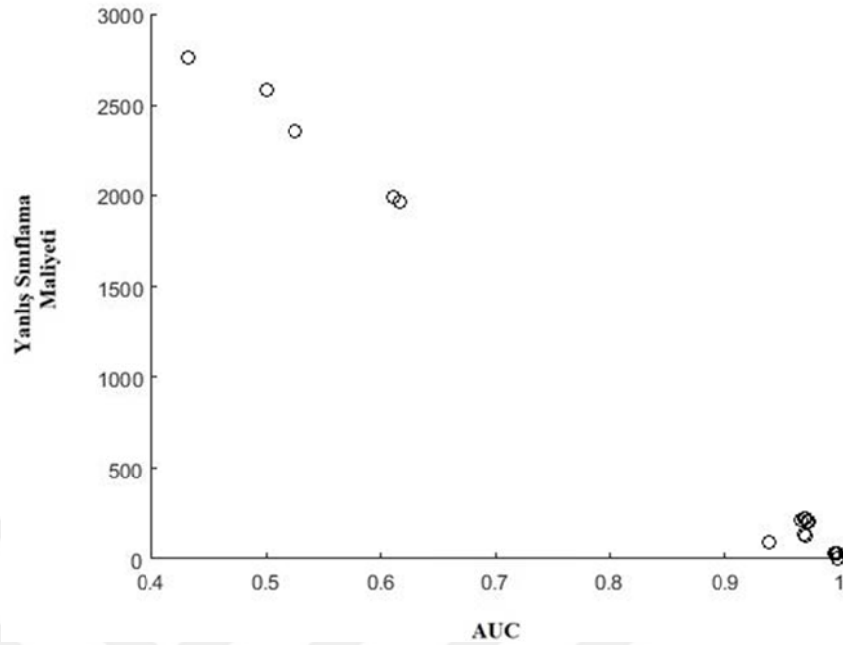
Tablo 5.3. Deneylerde Kullanılan Yöntemler

<i>(a) Örneklemeye yaklaşımları</i>	<i>Tanımı</i>
ROS	Rastgele üst örneklemeye
RUS	Rastgele alt örneklemeye
SMOTE	Synthetic azınlık üst örneklemeye
SMOTE-ENN	ENN eleme ile SMOTE örneklemeye
SMOTE-Tomek	Tomek-link eleme ile SMOTE örneklemeye
<i>(b) Maliyet duyarlı çözümler</i>	<i>Tanımı</i>
CS-Tree	Örnek ağırlıklandırmaya dayalı maliyet duyarlı karar ağaçları
CS-SVM	Maliyet-duyarlı destek vektör makinesi
<i>(c) Hybrid Ensemble çözümler</i>	<i>Tanımı</i>
EasyEnsemble	EasyEnsemble
BalanceCascade	BalanceCascade
<i>(d) Ensemble çözümler</i>	<i>Tanımı</i>
Bagging	Bagging
OverBagging	OverBagging
SMOTEBagging	SMOTEBagging
UnderBagging	UnderBagging
AdaBoost	AdaBoost
AdaC2	AdaC2
RUSBoost	RUSBoost
SMOTEBoost	SMOTEBoost

Önerilen algoritmada iki ayrı nitelik seçimi yaklaşımı uygulanmıştır. Bu yaklaşımlardan elde edilen nitelik alt kümelerinin sonuçları Tablo 5.4. 'te ayrı olarak gösterilmiştir. FRQR + KKO Sınıflandırma Algoritması ile elde edilen pareto çözümün amaç fonksiyonu değerleri: AUC: % 0.99769; yanlış sınıflandırma maliyeti:30 ve nitelik alt küme elaman sayısı:2 olarak elde edilmiştir. NSGA-II + KKO Sınıflandırma Algoritması ile elde edilen pareto çözümün amaç fonksiyonu değerleri: AUC: % 0.99800; yanlış sınıflandırma maliyeti: 5 ve nitelik alt küme elaman sayısı: 3 olarak elde edilmiştir. İki ayrı nitelik seçimi yaklaşımından elde edilen sonuçlar incelendiğinde NSGA-II nitelik seçimi yaklaşımının maliyet duyarlı kurgulanması sayesinde elde edilen sonuçlarda yanlış sınıflandırma maliyetinin daha düşük olduğu; FRQR yaklaşımının ise nitelikler arası ilişki gücünü daha iyi analiz edebilmesi sayesinde de daha küçük nitelik alt kümeleri elde edilebildiği görülmektedir. Bu nedenle her iki yaklaşımın avantajlarından yararlanabilmek için sınıflandırma aşamasında her iki yaklaşımdan elde edilen nitelik alt kümeleri kullanılmıştır.

Tablo 5.4. Telekomünikasyon Müşteri Veri Setinin Deney Sonuçları

			AUC	MALİYET
FRQR + KKO Sınıflandırma Algoritması			0.99769	30
NSGA-II + KKO Sınıflandırma Algoritması			0.99800	5
ÖRNEKLEME YAKLAŞIMLARI	ROS	C45	0.97072	220
		SVM	0.61145	1988
	RUS	C45	0.97072	220
		SVM	0.61100	1988
	SMOTE	C45	0.97072	220
		SVM	0.61687	1962
	SMOTE-ENN	C45	0.50000	2584
		SVM	0.43273	2757
MALİYET DUYARLI YAKLAŞIMLAR	CS-TREE		0.96843	226
	CS-SVM		0.93911	96
ENSEMBLE ÇÖZÜMLER	BAGGING	Bagging	0.97149	209
		Over Bagging	0.97302	207
		SMOTE Bagging	0.52476	2356
		Under Bagging	0.96611	216
	BOOSTING	AdaBoost	0.97072	129
		AdaC2	0.97072	129
		RUS Boost	0.99512	30
		SMOTE Boost	0.99601	31
HYBRID ENSEMBLE ÇÖZÜMLER	EasyEnsemble		0.96919	137
	BalanceCascade		0.97013	220



Şekil 5.8. Telekomünikasyon Müşteri Veri Setinin Deney Sonuçları Grafik Gösterimi

Tablo 5.4.' te karşılaştırma yapılan 22 algoritma ile birlikte 24 test sonucunun AUC ve toplam yanlış sınıflandırma maliyeti değerleri yer almakta ve bu değerler Şekil 5.8.' de grafik üzerinde verilmiştir.

Tablodaki verilerden, önerilen algoritmanın sonuçlarının diğer 22 algoritmaya göre daha iyi AUC ve toplam yanlış sınıflandırma maliyeti değerlerine sahip olduğu görülmektedir. Elde edilen bu sonuçlarla müşteri kaybı yönetimi için literatüre etkin bir tahmin modeli kazandırılması sağlanmıştır.

6. BÖLÜM

SONUÇ

Müşteri ilişkileri yönetiminde müşteri kaybı tahmini, tüm işletmeler için vazgeçilmez yönetsel bir araçtır. Sınıf dengesizliği, müşteri kaybını tahmin etmede karşılaşılan sorunlardan birisidir. Bu konuyla ilgilenmek ve azınlık sınıf olan ayrılan müşterileri tahmin edebilmek, başarılı müşteri kaybı tahmini modellemenin özünü oluşturmaktadır.

Bu çalışmada, telekomünikasyon sektörü müşteri kaybı tahmini problemine özgü entegre bir model önerilmiş ve önerilen model Türkiye'de bir telekomünikasyon şirketinin müşteri verilerine uygulanmıştır. İlk olarak, şirketin müşteri kaybını tahmin etmesine katkı sağlayacak en etkili nitelikler belirlenmiştir.

Önerilen tahmin modeli nitelik indirgeme ve sınıflandırma olmak üzere iki ana aşamadan oluşmaktadır ayrıca yüksek boyutlu ve dengesiz verileri analiz edebilme özelliğine sahiptir. Nitelik indirgeme aşamasında iki ayrı yaklaşımdan yararlanılmıştır. Bunlardan ilki bulanık kaba küme temelli FRQR yaklaşımı diğeri ise maliyet duyarlı NSGA-II tabanlı yaklaşımdır. Bu yaklaşımlarla elde edilen indirgenmiş nitelik kümelerine maliyet duyarlı ve çok amaçlı olarak geliştirilen KKO sınıflandırma algoritması uygulanmıştır. Uygulama sonuçları, önerilen modelin müşteri kaybı veri kümesini doğru ve düşük maliyetle sınıflandırmak için basit kurallar üretebildiğini göstermektedir.

İleriki çalışmalarda, müşteri kaybı yönetimi stratejileri ortaya konabilir. Gerçekleştirilen uygulama müşteriyi daha iyi tanımak ve ihtiyaçlarını karşılamak için müşteri segmentasyonu ile iyileştirilebilir. Çalışmamızda çok amaçlı ve maliyet duyarlı yaklaşımlar uygulanmıştır. Farklı sezgisel algoritmalar maliyet duyarlı bir şekilde uygulanabilir ve bu çalışmada elde edilen sonuçlar ile karşılaştırılabilir. Ayrıca, diğer nitelik indirgeme algoritmaları, maliyet duyarlılığı ve verimliliği açısından bu çalışmada elde edilen sonuçlarla karşılaştırılabilir.

KAYNAKLAR

1. Diskaya, F., Emir, S., & Orhan, N., 2011. Measuring the technical efficiency of telecommunication sector within global crisis: comparison of G8 countries and Turkey. **Procedia-Social and Behavioral Sciences**, **24**, 206-218.
2. Larivière, B., & Van den Poel, D., 2004. Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services. **Expert Systems with Applications**, **27**(2), 277-285.
3. Richeldi, M., & Perrucci, A., 2002. Churn analysis case study. Deliverable D17, 2.
4. Dierkes, T., Bichler, M., & Krishnan, R., 2011. Estimating the effect of word of mouth on churn and cross-buying in the mobile phone market with Markov logic networks. **Decision Support Systems**, **51**(3), 361-371.
5. Piotr, S., 2008. Global Perspectives Mobile Operator Customer Classification in Churn Analysis. **In Technical University of Szczecin, Poland SAS Global Forum** (pp. 1-5).
6. Hung, S. Y., Yen, D. C., & Wang, H. Y., 2006. Applying data mining to telecom churn management. **Expert Systems with Applications**, **31**(3), 515-524.
7. Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Nanavati, A. A., & Joshi, A., 2008, March. Social ties and their relevance to churn in mobile telecom networks. **In Proceedings of the 11th international conference on Extending database technology: Advances in database technology** (pp. 668-677). ACM.
8. Kim, H. S., & Yoon, C. H., 2004. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. **Telecommunications policy**, **28**(9), 751-765.
9. Ahn, J. H., Han, S. P., & Lee, Y. S., 2006. Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. **Telecommunications policy**, **30**(10), 552-568.

10. Özmen, M., 2006. Churn Modeling in Telecommunications Sector (Doctoral dissertation).
11. Karahoca, A., & Karahoca, D., 2011. GSM churn management by using fuzzy c-means clustering and adaptive neuro fuzzy inference system. **Expert Systems with Applications**, **38**(3), 1814-1822.
12. Verbeke, W., Martens, D., Mues, C., & Baesens, B., 2011. Building comprehensible customer churn prediction models with advanced rule induction techniques. **Expert Systems with Applications**, **38**(3), 2354-2364.
13. SAS International, 2001. Predicting Churn, SAS Institute, Cary, NC.
14. Sharma, A., Panigrahi, D., & Kumar, P., 2013. A neural network based approach for predicting customer churn in cellular network services. arXiv preprint arXiv:1309.3945.
15. Kirui, C., Hong, L., Cheruiyot, W., & Kirui, H., 2013. Predicting customer churn in mobile telephony industry using probabilistic classifiers in data mining. **IJCSI International Journal of Computer Science Issues**, **10**(2), 1694-0814.
16. Abbasimehr, H., Setak, M., & Tarokh, M. J., 2011. A neuro-fuzzy classifier for customer churn prediction. **Int J Comput Appl**, **19**(8), 35-41.
17. Au, W. H., Chan, K. C., & Yao, X., 2003. A novel evolutionary data mining algorithm with applications to churn prediction. **IEEE transactions on evolutionary computation**, **7**(6), 532-545.
18. Shaaban, E., Helmy, Y., Khedr, A., & Nasr, M., 2012. A proposed churn prediction model. **International Journal of Engineering Research and Applications**, **2**(4), 693-697.
19. Lazarov, V., & Capota, M., 2007. Churn prediction. Bus. Anal. Course. TUM Comput. Sci.
20. Xia, G. E., & Jin, W. D., 2008. Model of customer churn prediction on support vector machine. **Systems Engineering-Theory & Practice**, **28**(1), 71-77.

21. Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. C., 2015. A comparison of machine learning techniques for customer churn prediction. **Simulation Modelling Practice and Theory**, **55**, 1-9.
22. Cao, K., & Shao, P. J., 2008, December. Customer churn prediction based on svm-rfe. In Business and Information Management, 2008. ISBIM'08. International Seminar on (Vol. 1, pp. 306-309). IEEE.
23. Burez, J., & Van den Poel, D., 2009. Handling class imbalance in customer churn prediction. **Expert Systems with Applications**, **36**(3), 4626-4636.
24. Gürsoy, U. T. Ş., 2010. Customer churn analysis in telecommunication sector. **Istanbul University Journal of the School of Business**, **39**(1), 35-49.
25. Kisioglu, P., & Topcu, Y. I., 2011. Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey. **Expert Systems with Applications**, **38**(6), 7151-7157.
26. Kulali, İ., & Bilir, H., 2010. Bilgi ve İletişim (Telekomünikasyon) sektöründeki Gelişmeler ve Eğilimler. Global Finansal Kriz Sürecinde Yeniden Yapılanma ve Çözüm Arayışları, TOBB.
27. Sumathi, S., & Sivanandam, S. N., 2006. Introduction to data mining and its applications (Vol. 29). Springer.
28. Internet World Stats., 2012. Top 20 Countries with Highest number of Internet Users - June 30, 2012. Retrieved 18 March 2013, from <http://www.internetworldstats.com/top20.htm>
29. Alikilic, O., & Atabek, U., 2012. Social media adoption among Turkish public relations professionals: A survey of practitioners. **Public Relations Review**, **38**(1), 56-63.
30. BTK., 2015), www.btk.gov.tr/File/?path=ROOT%2F1%2Fdocuments%2FSayfalar%2FPazar_Verileri%2F2015-Q3_v1.pdf , 19. 02. 2016
31. Networking, C. V. 2015. Cisco Global Cloud Index: Forecast and Methodology, 2014-2019, (White Paper).

32. Norman, D. A., 1998. The invisible computer: why good products can fail, the personal computer is so complex, and information appliances are the solution. MIT press.
33. Fan, W., & Bifet, A., 2013. Mining big data: current status, and forecast to the future. **ACM SIGKDD Explorations Newsletter**, **14**(2), 1-5.
34. Günay, Z., & Ünal, Ö. F., 2016. AHP-TOPSIS Yöntemi İle Tedarikçi Seçimi (Bir Telekomünikasyon Şirketi Örneği). **PESA Uluslararası Sosyal Araştırmalar Dergisi**, **2**(1).
35. Kurt, A., 2007. Türk Telekomünikasyon Sektörü İle Ülke Ekonomisindeki Gelişmeler Arasındaki İlişkinin Varlığının Ekonometrik Analizi. I. Haberleşme Teknolojileri ve Uygulamaları Sempozyumu (HABTEKUS'07), 96-106.
36. El Emary, I. M. M., 2010. Role of Data Mining and Knowledge Discovery in Managing Telecommunication Systems. *Wireless Technologies for Ambient Assisted Living and Healthcare: Systems and Applications: Systems and Applications*, 14.
37. Joseph, M. V., 2013. Data mining and business intelligence applications in telecommunication industry. **International Journal of Engineering and Advanced Technology (IJEAT)** ISSN, 2249-8958.
38. Kaplan, H., Strauss, M., & Szegedy, M., 1999, January. Just the Fax-Differentiating Voice and Fax Phone Lines Using Call Billing Data. In SODA (pp. 935-936).
39. Cheng, L. C., & Sun, L. M., 2012. Exploring consumer adoption of new services by analyzing the behavior of 3G subscribers: An empirical case study. **Electronic Commerce Research and Applications**, **11**(2), 89-100.
40. Lim, J., Nam, C., Kim, S., Lee, E., & Lee, H., 2015. A new regional clustering approach for mobile telecommunications policy in China. **Telecommunications Policy**, **39**(3), 296-304.

41. Chen, C. H., Chiang, R. D., Wu, T. F., & Chu, H. C., 2013. A combined mining-based framework for predicting telecommunications customer payment behaviors. **Expert Systems with Applications**, **40**(16), 6561-6569.
42. Vidya, N. A., Fanany, M. I., & Budi, I., 2015. Twitter Sentiment to Analyze Net Brand Reputation of Mobile Phone Providers. **Procedia Computer Science**, **72**, 519-526.
43. Wei, C. P., & Chiu, I. T., 2002. Turning telecommunications call details to churn prediction: a data mining approach. **Expert systems with applications**, **23**(2), 103-112.
44. Weiss, G. M., & Hirsh, H., 1998, August. Learning to Predict Rare Events in Event Sequences. In KDD (pp. 359-363).
45. Farvaresh, H., & Sepehri, M. M., 2011. A data mining framework for detecting subscription fraud in telecommunication. **Engineering Applications of Artificial Intelligence**, **24**(1), 182-194.
46. Olszewski, D., 2012. A probabilistic approach to fraud detection in telecommunications. **Knowledge-Based Systems**, **26**, 246-258.
47. Chao, D. O. N. G., LEI, Z. M., & Feng, L. I. U., 2011. Internet quality abnormal analysis with k-means clustering. **The Journal of China Universities of Posts and Telecommunications**, **18**, 94-100.
48. Ren, D. Q., Zheng, D., Huang, G., Zhang, S., & Wei, Z., 2013, November. Parallel Set Determination and K-Means Clustering for Data Mining on Telecommunication Networks. In High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCC_EUC), 2013 IEEE 10th International Conference on (pp. 1553-1557). IEEE.
49. Velmurugan, T., 2014. Performance based analysis between k-Means and Fuzzy C-Means clustering algorithms for connection oriented telecommunication data. **Applied Soft Computing**, **19**, 134-146.
50. Berson, A., & Smith, S. J., 2002. Building data mining applications for CRM. McGraw-Hill, Inc..

51. Mattison, R., 2006. The telco churn management handbook. Lulu. com.
52. Mattering, R., 2001. Telecom churn management. Fuquay-Varina, NC: APDG Publishing.
53. Bhattacharya, C. B., 1998. When customers are members: Customer retention in paid membership contexts. **Journal of the academy of marketing science**, **26**(1), 31-44.
54. Rasmusson, E., 1999. Complaints Can Build Relationships. **Sales & Marketing Management**, **151**(9), 89-89.
55. Colgate, M., Stewart, K., & Kinsella, R., 1996. Customer defection: a study of the student market in Ireland. **International Journal of Bank Marketing**, **14**(3), 23-29.
56. Athanassopoulos, A. D., 2000. Customer satisfaction cues to support market segmentation and explain switching behavior. **Journal of business research**, **47**(3), 191-207.
57. Mizerski, R. W., 1982. An attribution explanation of the disproportionate influence of unfavorable information. **Journal of Consumer Research**, **9**(3), 301-310.
58. Stum, D. L., & Thiry, A., 1991. Building customer loyalty. **Training and development Journal**, **45**(4), 34-36.
59. Reichheld, F. F., 1996. Learning from customer defections.
60. Zeithaml, V. A., Berry, L. L., & Parasuraman, A., 1996. The behavioral consequences of service quality. **the Journal of Marketing**, 31-46.
61. Paulin, M., Perrien, J., Ferguson, R. J., Alvarez Salazar, A. M., & Michel Seruya, L., 1998. Relational norms and client retention: external effectiveness of commercial banking in Canada and Mexico. **International Journal of Bank Marketing**, **16**(1), 24-31.
62. Ganesh, J., Arnold, M. J., & Reynolds, K. E., 2000. Understanding the customer base of service providers: an examination of the differences between switchers and stayers. **Journal of marketing**, **64**(3), 65-87.

63. Rust, R. T., & Zahorik, A. J., 1993. Customer satisfaction, customer retention, and market share. **Journal of retailing**, **69**(2), 193-215.
64. Thearling, K., 1999. An introduction to data mining. **Direct Marketing Magazine**, 28-31.
65. Gopal, R. K., & Meher, S. K., 2008, May. Customer churn time prediction in mobile telecommunication industry using ordinal regression. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 884-889). Springer, Berlin, Heidelberg.
66. Huang, B., Kechadi, M. T., & Buckley, B., 2012. Customer churn prediction in telecommunications. **Expert Systems with Applications**, **39**(1), 1414-1425.
67. Idris, A., Rizwan, M., & Khan, A., 2012. Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. **Computers & Electrical Engineering**, **38**(6), 1808-1819.
68. Chueh, H. E., 2011. Analysis of marketing data to extract key factors of telecom churn management. **African Journal of Business Management**, **5**(20), 8242.
69. Tunga, M. A., & Karahoca, A., 2015. Detecting GSM churners by using Euclidean Indexing HDMR. **Applied Soft Computing**, **27**, 38-46.
70. Van Doorn, J., Lemon, K. N., Mittal, V., Nass, S., Pick, D., Pirner, P., & Verhoef, P. C., 2010. Customer engagement behavior: Theoretical foundations and research directions. **Journal of service research**, **13**(3), 253-266.
71. Nath, S. V., & Behara, R. S., 2003, November. Customer churn analysis in the wireless industry: A data mining approach. In **Proceedings-annual meeting of the decision sciences institute** (pp. 505-510).
72. Karahoca, A., & Kara, A., 2006, May. Comparing clustering techniques for telecom churn management. In **Proceedings of the 5th WSEAS International Conference on Telecommunications and Informatics**, (pp. 27-29).

73. Jahanzeb, S., & Jabeen, S., 2007. Churn management in the telecom industry of Pakistan: A comparative study of Ufone and Telenor. **Journal of Database Marketing & Customer Strategy Management**, **14**(2), 120-129.
74. Huang, Y., & Kechadi, T., 2013. An effective hybrid learning system for telecommunication churn prediction. **Expert Systems with Applications**, **40**(14), 5635-5647.
75. Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., & Abbasi, U., 2014. Improved churn prediction in telecommunication industry using data mining techniques. **Applied Soft Computing**, **24**, 994-1012.
76. Zhang, X., Zhu, J., Xu, S., & Wan, Y., 2012. Predicting customer churn through interpersonal influence. **Knowledge-Based Systems**, **28**, 97-104.
77. Tsai, C. F., & Chen, M. Y., 2010. Variable selection by association rules for customer churn prediction of multimedia on demand. **Expert Systems with Applications**, **37**(3), 2006-2015.
78. Owczarczuk, M., 2010. Churn models for prepaid customers in the cellular telecommunication industry using large data marts. **Expert Systems with Applications**, **37**(6), 4710-4712.
79. Amin, A., Khan, C., Ali, I., & Anwar, S., 2014, November. Customer churn prediction in telecommunication industry: With and without counter-example. In Mexican International Conference on Artificial Intelligence (pp. 206-218). Springer, Cham.
80. Huang, B., Buckley, B., & Kechadi, T. M., 2010. Multi-objective feature selection by using NSGA-II for customer churn prediction in telecommunications. **Expert Systems with Applications**, **37**(5), 3638-3646.
81. Xu, H., Zhang, Z., & Zhang, Y., 2009, November. Churn prediction in telecom using a hybrid two-phase feature selection method. In Intelligent Information Technology Application, 2009. IITA 2009. Third International Symposium on (Vol. 3, pp. 576-579). IEEE.

82. Xie, L., Li, D., & Xiao, J., 2011, October. Feature selection based transfer ensemble model for customer churn prediction. In System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2011 International Conference on (Vol. 2, pp. 134-137. IEEE.
83. Xiao, J., Xiao, Y., Huang, A., Liu, D., & Wang, S., 2015. Feature-selection-based dynamic transfer ensemble model for customer churn prediction. **Knowledge and information systems**, **43**(1), 29-51.
84. Malhotra, N. K., Birks, D. F., Palmer, A., & Koenig-Lewis, N., 2003. Market research: an applied approach. **Journal of marketing management**, **27**, 1208-1213.
85. Kızılkaya Aydoğan, E., 2008. Veri Madenciliğinde Sınıflandırma Problemleri için Evrimsel Algoritma Tabanlı Yeni Bir Yaklaşım: Rough-Mep Algoritması, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, Ankara, 149 s.
86. Lee, W., & Stolfo, S. J., 1998, January. Data mining approaches for intrusion detection. In USENIX Security Symposium (pp. 79-93).
87. Van der Aalst, W. M., 2011. Data Mining. In Process Mining (pp. 59-91). Springer Berlin Heidelberg.
88. Patnaik, D. A. K., & Nayak, B., 2011. Data mining and its current research directions. **Information Sciences**, **181**(7), 1264-1284.
89. Olson, D. L., & Delen, D., 2008. Advanced data mining techniques. Springer Science & Business Media.
90. Jensen, R., & Shen, Q., 2004. Fuzzy-rough attribute reduction with application to web categorization. **Fuzzy sets and systems**, **141**(3), 469-485.
91. López, V., Fernández, A., García, S., Palade, V., & Herrera, F., 2013. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. **Information Sciences**, **250**, 113-141.
92. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P., 2002. SMOTE: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, **16**, 321-357.

93. Wilson, D. L., 1972. Asymptotic properties of nearest neighbor rules using edited data. **IEEE Transactions on Systems, Man, and Cybernetics**, 2(3), 408-421.
94. Bunkhumpornpat, C., Sinapiromsaran, K., & Lursinsap, C., 2009. Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem. **Advances in knowledge discovery and data mining**, 475-482.
95. Tomek, I., 1976. Two modifications of CNN. **IEEE Trans. Systems, Man and Cybernetics**, 6, 769-772.
96. Zhao, H., 2008. Instance weighting versus threshold adjusting for cost-sensitive classification. **Knowledge and Information Systems**, 15(3), 321-334.
97. Fernández, A., García, S., del Jesus, M. J., & Herrera, F., 2008. A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. **Fuzzy Sets and Systems**, 159(18), 2378-2398.
98. FernáNdez, A., LóPez, V., Galar, M., Del Jesus, M. J., & Herrera, F., 2013. Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches. **Knowledge-based systems**, 42, 97-110.
99. Sun, Y., Wong, A. K., & Kamel, M. S., 2009. Classification of imbalanced data: A review. **International Journal of Pattern Recognition and Artificial Intelligence**, 23(04), 687-719.
100. Zadrozny, B., Langford, J., & Abe, N., 2003, November. Cost-sensitive learning by cost-proportionate example weighting. In Data Mining, 2003. ICDM 2003. Third IEEE International Conference on(pp. 435-442). IEEE.
101. Şahin, Y., 2013. Intelligent Ways of Detecting Fraud. Marmara Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, İstanbul, 94 s.
102. Domingos, P., 1999, August. Metacost: A general method for making classifiers cost-sensitive. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, (pp. 155-164). ACM.

103. Czajkowski, M., Czerwinka, M., & Kretowski, M., 2015. Cost-sensitive global model trees applied to loan charge-off forecasting. **Decision Support Systems**, **74**, 57-66.
104. Bahnsen, A. C., Aouada, D., & Ottersten, B., 2015. Example-dependent cost-sensitive decision trees. **Expert Systems with Applications**, **42**(19), 6609-6619.
105. Jiang, L., Li, C., & Wang, S., 2014. Cost-sensitive Bayesian network classifiers. **Pattern Recognition Letters**, **45**, 211-216.
106. Siers, M. J., & Islam, M. Z., 2015. Software defect prediction using a cost sensitive decision forest and voting, and a potential solution to the class imbalance problem. **Information Systems**, **51**, 62-71.
107. Bolón-Canedo, V., Porto-Díaz, I., Sánchez-Marño, N., & Alonso-Betanzos, A., 2014. A framework for cost-based feature selection. **Pattern Recognition**, **47**(7), 2481-2489.
108. Arar, Ö. F., & Ayan, K., 2015. Software defect prediction using cost-sensitive neural network. **Applied Soft Computing**, **33**, 263-277.
109. Elkan, C., 2001, August. The foundations of cost-sensitive learning. In International joint conference on artificial intelligence (Vol. 17, No. 1, pp. 973-978). Lawrence Erlbaum Associates Ltd.
110. Hu, Y., Feng, B., Mo, X., Zhang, X., Ngai, E. W. T., Fan, M., & Liu, M., 2015. Cost-sensitive and ensemble-based prediction model for outsourced software project risk prediction. **Decision Support Systems**, **72**, 11-23.
111. Torres, M. G., 2009. Metaheuristics in Data Mining. In Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1200-1206). IGI Global.
112. Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S., & Coello, C. A. C., 2014. A survey of multiobjective evolutionary algorithms for data mining: Part I. **IEEE Transactions on Evolutionary Computation**, **18**(1), 4-19.
113. Bandaru, S., Ng, A. H., & Deb, K., 2017. Data mining methods for knowledge discovery in multi-objective optimization: Part A-Survey. **Expert Systems with Applications**, **70**, 139-159.

114. Honda, S., Igarashi, T., & Narita, Y., 2013. Multi-objective optimization of curvilinear fiber shapes for laminated composite plates by using NSGA-II. **Composites Part B: Engineering**, **45**(1), 1071-1078.
115. Durmaz, E. D., & Şahin, R., 2017. ÇOK AMAÇLI TEK SIRA TESİS DÜZENLEME PROBLEMİNİN ÇÖZÜMÜ İÇİN NSGA-II VE HEDEF PROGRAMLAMA YAKLAŞIMI. **Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi**, **32**(3).
116. Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. **IEEE transactions on evolutionary computation**, **6**(2), 182-197.
117. Oliveira, L. S., Sabourin, R., Bortolozzi, F., & Suen, C. Y., 2003, August. Feature selection for ensembles: A hierarchical multi-objective genetic algorithm approach. In **Proceedings of the Seventh International Conference on Document Analysis and Recognition-Volume 2** (p. 676). IEEE Computer Society.
118. Wang, C. M., & Huang, Y. F., 2009. Evolutionary-based feature selection approaches with new criteria for data mining: A case study of credit approval data. **Expert Systems with Applications**, **36**(3), 5900-5908.
119. Venkatadri, M., & Rao, K. S., 2010. A multiobjective genetic algorithm for feature selection in data mining. **International Journal of Computer Science and Information Technologies**, **1**(5), 443-448.
120. Morita, M., Sabourin, R., Bortolozzi, F., & Suen, C. Y., 2003, August. Unsupervised feature selection using multi-objective genetic algorithms for handwritten word recognition. In **Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on** (pp. 666-670). IEEE.
121. Mierswa, I., & Wurst, M., 2006, July. Information preserving multi-objective feature selection for unsupervised learning. In **Proceedings of the 8th annual conference on Genetic and evolutionary computation** (pp. 1545-1552). ACM.

122. Shi, S. Y., Suganthan, P. N., & Deb, K., 2004, October. Multiclass protein fold recognition using multiobjective evolutionary algorithms. In **Computational Intelligence in Bioinformatics and Computational Biology, 2004. CIBCB'04. Proceedings of the 2004 IEEE Symposium on** (pp. 61-66). IEEE.
123. Ekbal, A., Saha, S., & Garbe, C. S., 2010, August. Feature selection using multiobjective optimization for named entity recognition. In **Pattern Recognition (ICPR), 2010 20th International Conference on** (pp. 1937-1940). IEEE.
124. Li, L., Li, M., Lu, Y., & Zhang, Y., 2010. A New Multi-Objective Genetic Algorithm for Feature Subset Selection in Fatigue Fracture Image Identification. **JCP**, **5**(7), 1105-1111.
125. Castro, P. A., & Von Zuben, F. J., 2010. Multi-objective feature selection using a Bayesian artificial immune system. **International Journal of Intelligent Computing and Cybernetics**, **3**(2), 235-256.
126. Zhao, J., Fernandes, V. B., Jiao, L., Yevseyeva, I., Maulana, A., Li, R., ... & Emmerich, M. T., 2016. Multiobjective optimization of classifiers by means of 3D convex-hull-based evolutionary algorithms. **Information Sciences**, **367**, 80-104.
127. Kimovski, D., Ortega, J., Ortiz, A., & Baños, R., 2015. Parallel alternatives for evolutionary multi-objective optimization in unsupervised feature selection. **Expert Systems with Applications**, **42**(9), 4239-4252.
128. Hamdani, T. M., Won, J. M., Alimi, A. M., & Karray, F., 2007, April. Multi-objective feature selection with NSGA II. In **International Conference on Adaptive and Natural Computing Algorithms** (pp. 240-247). Springer, Berlin, Heidelberg.
129. Radtke, P. V., Wong, T., & Sabourin, R., 2009. Solution over-fit control in evolutionary multiobjective optimization of pattern classification systems. **International Journal of Pattern Recognition and Artificial Intelligence**, **23**(06), 1107-1127.

130. Spolaôr, N., Lorena, A. C., & Lee, H. D., 2011, April. Multi-objective genetic algorithm evaluation in feature selection. In *International Conference on Evolutionary Multi-Criterion Optimization* (pp. 462-476). Springer Berlin Heidelberg.
131. Vatulkin, I., Preuß, M., Rudolph, G., Eichhoff, M., & Weihs, C., 2012. Multi-objective evolutionary feature selection for instrument recognition in polyphonic audio mixtures. **Soft Computing**, **16**(12), 2027-2047.
132. Saeys, Y., Inza, I., & Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. **bioinformatics**, **23**(19), 2507-2517.
133. Dorigo, M., & Blum, C., 2005. Ant colony optimization theory: A survey. **Theoretical computer science**, **344**(2-3), 243-278.
134. Omkar, S. N., & Karanth, R., 2008. Rule extraction for classification of acoustic emission signals using ant colony optimisation. *Engineering Applications of Artificial Intelligence*, **21**(8), 1381-1388.
135. Costa, E., Lorena, A., Carvalho, A., Freitas, A., & Holden, N., 2007. Comparing several approaches for hierarchical classification of proteins with decision trees. *Advances in Bioinformatics and Computational Biology*, 126-137.
136. Liu, X., Li, X., Yeh, A. G. O., He, J., & Tao, J., 2007. Discovery of transition rules for geographical cellular automata by using ant colony optimization. **Science China. Earth Sciences**, **50**(10), 1578.
137. Liu, X., Li, X., Liu, L., He, J., & Ai, B., 2008. An innovative method to classify remote-sensing images using ant colony optimization. **IEEE transactions on geoscience and remote sensing**, **46**(12), 4198-4208.
138. Lai, C., Shao, Q., Chen, X., Wang, Z., Zhou, X., Yang, B., & Zhang, L., 2016. Flood risk zoning using a rule mining based on ant colony algorithm. **Journal of Hydrology**, **542**, 268-280.
139. Lee, H., Lee, Y., Cho, H., Im, K., & Kim, Y. S., 2011. Mining churning behaviors and developing retention strategies based on a partial least squares (PLS) model. **Decision Support Systems**, **52**(1), 207-216.

140. Zhang, Y., Qi, J., Shu, H., & Li, Y., 2006, October. Case study on crm: Detecting likely churners with limited information of fixed-line subscriber. In *Service Systems and Service Management, 2006 International Conference on* (Vol. 2, pp. 1495-1500). IEEE.
141. Gok M., 2014. İnternet Servis Sağlayıcısı için İptal Analizi Modeli, TOBB, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul, 71s.
142. Kişioğlu, P., 2009. Telekomünikasyon Sektöründe İptal Analizi, İTÜ, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul, 109s.
143. Kim, N., Jung, K. H., Kim, Y. S., & Lee, J., 2012. Uniformly subsampled ensemble (USE) for churn management: Theory and implementation. **Expert Systems with Applications**, **39**(15), 11839-11845.
144. Keramati, A., & Ardabili, S. M., 2011. Churn analysis for an Iranian mobile operator. **Telecommunications Policy**, **35**(4), 344-356.
145. Kumar, A., Kabra, G., Mussada, E. K., Dash, M. K., & Rana, P. S., 2017. Combined artificial bee colony algorithm and machine learning techniques for prediction of online consumer repurchase intention. **Neural Computing and Applications**, 1-14.
146. Tamaddoni Jahromi, A., 2009. Predicting Customer Churn in Telecommunications Service Providers, Yüksek Lisans Tezi, İsveç, 88s.
147. Acuna, E., & Rodriguez, C., 2004. The treatment of missing values and its effect on classifier accuracy. **Classification, clustering, and data mining applications**, 639-647.
148. Zhu, B., Baesens, B., & vanden Broucke, S. K., 2017. An empirical comparison of techniques for the class imbalance problem in churn prediction. **Information Sciences**, 408, 84-99.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı, Soyadı: Mihrimah ÖZMEN

Uyruğu: Türkiye (TC)

Doğum Tarihi ve Yeri: 19 Ağustos 1983, Kayseri

Medeni Durumu: Evli

Tel: +90 352 207 66 66/32463

email: mihrimah@erciyes.edu.tr

Yazışma Adresi: Erciyes Üniversitesi Mühendislik Fakültesi 38039 Talas/KAYSERİ

EĞİTİM

Derece	Kurum	Mezuniyet Tarihi
Yüksek Lisans	EÜ Fen Bilimler Enstitüsü	2013
Lisans	EÜ Endüstri Müh. Böl.	2004
Lise	N.M.Küçükçalık A.L.	2000

İŞ DENEYİMLERİ

Yıl	Kurum	Görev
2013- Halen	Erciyes Üniversitesi End. Müh. Böl.	Araş.Gör.
2006–2011	Türkiye İş Bankası	Müşteri Temsilcisi
2004-2006	Yataş	Mühendis

PROJELER

1. TurkNet Kaba Churn Tabanlı Karar Destek Bilişim Sistemi, SAN-TEZ, 0777.STZ.2014, Bursiyer, 2016.
2. Kayseri İli Lojistik Merkez Ön Fizibilite Raporu, Kalkınma Bakanlığı (DPT), Araştırmacı, 2015.
3. Alüminyum Mamül Üreten İşletmelerde Öncelikli Atıkların Belirlenmesi ve Entegre Atık Yönetimi", TÜBİTAK Projesi, 111R003, Bursiyer, 2016.

YABANCI DİL

İngilizce

YAYINLAR

SCI, SSCI, AHCI INDEXLERİNE GİREN DERGİLERDE YAYIMLANAN MAKALELER

1. Kizilkaya Aydoğan E., Özmen M., "The stochastic Vikor method and its use in reverse logistic option selection problem", **RAIRO-OPERATIONS RESEARCH**, vol.1, no.1, pp.1-10, 2016
2. Özmen M., Kizilkaya Aydoğan E., Ateş N., Uzal N., "Developing a Decision-Support System for Waste Management in Aluminum Production", **ENVIRONMENTAL MODELING & ASSESSMENT**, vol.21, pp.803-817, 2016
3. Kizilkaya Aydoğan E., Özmen M., "TWO NEW METHOD FOR MULTI CRITERIA STOCHASTIC DECISION MAKING: SMAA-GRA AND SMAA-DEMATEL-GRA", **JOURNAL OF THE FACULTY OF ENGINEERING AND ARCHITECTURE OF GAZI UNIVERSITY**, vol.30, pp.627-640, 2015
4. Akgül E., Özmen M., Kizilkaya Aydoğan E., Türksoy H. G., " Optimization of the Murata Vortex Spinning machine parameters by the SMAA-MOORA approach", **JOURNAL OF INDUSTRIA TEXTILA** (basım aşamasında)
5. Kizilkaya Aydoğan E., Özmen M., Delice Y., Toksarı M. D., "COST SENSITIVE AND MULTI OBJECTIVE CHURN PREDICTION: A CASE FOR TELECOMMUNICATIONS SECTOR" (değerlendirme aşamasında)

6. Özmen M., Kizilkaya Aydoğan E., Delice Y., "CBR-PSO: Cost Based Rough Particle Swarm Optimization Approach for High-Dimensional Imbalanced Problems" (değerlendirme aşamasında)

DİĞER DERGİLERDE YAYIMLANAN MAKALELER

1. Kizilkaya Aydoğan E., Özmen M., "Attribute Reduction in Stochastic Information Systems Based on Dominance", **INTERNATIONAL JOURNAL OF INFORMATICS TECHNOLOGIES**, vol.10, pp.211-219, 2017
2. Kizilkaya Aydoğan E., Özmen M., Delice Y., "Cost Sensitive Feature Selection in Decision-Theoretic Rough Set Models for Customer Churn Prediction: The Case of Telecommunication Sector Customers", **International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering**, vol.10, no.4, pp.163-167, 2016

HAKEMLİ KONGRE/SEMPOZYUM BİLDİRİ KİTAPLARINDA YER ALAN YAYINLAR

1. Kizilkaya Aydoğan E., Özmen M., "Determination Of Municipality Management Information System Criteria And Determination Of Weights By Rough-Ahp Method", *Second International Symposium on INDUSTRIAL DESIGN & ENGINEERING*, NEVŞEHİR, TÜRKİYE, 13-15 Eylül 2017, pp.20-24
2. Kizilkaya Aydoğan E., Özmen M., "Classification Of Diabetes' Disease With Data Mining Techniques", *Second International Symposium on INDUSTRIAL DESIGN & ENGINEERING*, NEVŞEHİR, TÜRKİYE, 13-15 Eylül 2017, pp.10-14
3. Kizilkaya Aydoğan E., Özmen M., "Municipality Management Information System Selection Problem: An Application Of Smaa-Vikor", *Second International Symposium on INDUSTRIAL DESIGN & ENGINEERING*, NEVŞEHİR, TÜRKİYE, 13-15 Eylül 2017, pp.15-19
4. Özmen M., Kizilkaya Aydoğan E., "Application of Clustering Algorithms in Churn Management: A Case Study in Telecommunication Sector", *1st International Mediterranean Science and Engineering Congress*, ADANA, TÜRKİYE, 26 Ekim - 28 Kasım 2016, pp.1-1
5. Özmen M., Kizilkaya Aydoğan E., Zaralı F., "LOJİSTİK MERKEZ YER SEÇİMİNDE DEMATEL-MOORA YÖNTEMİYLE KAYSERİ İLİ İÇİN BİR UYGULAMA ", *Yöneylem Araştırması ve Endüstri Mühendisliği (YAEM) 35. Ulusal Kongresi*, ANKARA, TÜRKİYE, 9-11 Eylül 2015, cilt.1, no.1, ss.133-133

6. Özmen M., Kizilkaya Aydoğan E., Zarali F., "Logistics Center Location Selection Using The New Method Smaatodim ", *23rd International Conference on Multiple Criteria Decision Making MCDM 2015 - Bridging Disciplines*, HAMBURG, ALMANYA, 2 Ağustos - 31 Temmuz 2015, pp.1-2
7. Kizilkaya Aydoğan E., Ateş N., Uzal N., Özmen M., "Determining The Priority Waste In Aluminum Manufacturing Sector Using The Smaa-2 Method: A Case Study Of Kayseri", *Joint International Symposium on CIE44 and IMSS'14*, İSTANBUL, TÜRKİYE, 14 Ekim - 16 Eylül 2014, pp.10-10
8. Danacı Şakalar T., Özmen M., Kizilkaya Aydoğan E., Toksari M.D., "AIRLINE FLEET ASSIGNMENT AND ROUTING FOR COMBINATION CARRIERS", *LM-SCM 2014 XII. International Logistics and Supply Chain Congress*, İSTANBUL, TÜRKİYE, 30-31 Ekim 2014, pp.1-1
9. Özmen M., Başak A., Kizilkaya Aydoğan E., Uzal N., Ateş N., "Investigation Of Priority Wastes And Wastewater Treatment Studies In Aluminum Products Manufacturing Sector", *INTERNATIONAL ENVIRONMENTAL SCIENCES SYMPOSIUM OF VAN*, VAN, TÜRKİYE, 4-7 Haziran 2014, pp.10-10
10. Özmen M., İçer S., Kizilkaya Aydoğan E., "Application Of Segmentation Algorithm For The Assessment Of Amniotic Fluid Quantity From Ultrasound Images", *Tıp Teknolojileri Ulusal Kongresi 2014*, NEVŞEHİR, TÜRKİYE, 25-27 Eylül 2014, ss.10-10 (Özet)(Abstract)
11. Erdem S., Kizilkaya Aydoğan E., Atmaca E. , Gencer C., Özmen M., "Özel Eğitim Ve Rehabilitasyon Merkezi Seçiminde Çok Kriterli Bir Yöntem Uygulaması ", *XIVth INTERNATIONAL SYMPOSIUM ON ECONOMETRICS, OPERATIONS RESEARCH AND STATISTICS* , KÜTAHYA, TÜRKİYE, 24-28 Mayıs 2013, pp.22-22
12. Özmen M., Kizilkaya Aydoğan E., "Reverse Logistics Options Selection Problem: An Application Of Smaa-2 ", *IX. INTERNATIONAL LOGISTICS SUPPLY CHAIN CONGRESS*, NEVŞEHİR, TÜRKİYE, 7-9 Kasım 2013, pp.673-680
13. Zeydan M., Özmen M., "Sağlık Sektöründe Gerçek Verimliliğin Veri Zarflama Analizi ve Analitik Hiyerarşi Süreci ile Belirlenmesi", *4. ULUSAL ÜRETİM ARAŞTIRMA SEMPOZYUMU*, KONYA, TÜRKİYE, 8-10 Ekim 2004, ss.1-1