



**T.C
ERCIYES ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI**

**MÜŞTERİ MEMNUNİYETİNİN KALİTE BOYUTLARI
BAKIMINDAN METİN MADENCİLİĞİ İLE KEŞFİ**

**Hazırlayan
Metin ERGÖKTAŞ**

**Danışman
Prof. Dr. Mehmet Sıtkı İLKAY**

Doktora Tezi

Ağustos 2024, KAYSERİ

**T.C.
ERCIYES ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI**

**MÜŞTERİ MEMNUNİYETİNİN KALİTE BOYUTLARI
BAKIMINDAN METİN MADENCİLİĞİ İLE KEŞFİ
(Doktora Tezi)**

**Hazırlayan
Metin ERGÖKTAŞ**

**Danışman
Prof. Dr. Mehmet Sıtkı İLKAY**

Ağustos 2024, 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.

Metin ERGÖKTAŞ



T.C.

ERCIYES ÜNİVERSİTESİ

Sosyal Bilimler Enstitüsü Müdürlüğü

Öğrenci No : 4030541726

Anabilim Dalı : İşletme

Program Adı : Doktora

Tez Başlığı : Müşteri Memnuniyetinin Kalite Boyutları Bakımından Metin Madenciliği ile Keşfi

Yukarıda bilgileri verilen tez çalışmasının a) Giriş, b) Ana bölümler ve c) Sonuç kısımlarından oluşan (Kapak, Ön söz, Özet, İçindekiler ve Kaynakça hariç) toplam 127 sayfalık kısmına ilişkin 23.08.2024 tarihinde **Turnitin** intihal programından aşağıda belirtilen filtreleme uygulanarak alınmış olan özgünlük raporuna göre tezin benzerlik oranı: %5 dir.

Filtrelemeye **alıntılar dahil** edilmiştir. Filtrelemede **yedi (7) kelimedenden daha az**

örtüşme içeren metin kısımları hariç tutulmuştur.

Erciyes Üniversitesi Sosyal Bilimler Enstitüsü Tez İntihal Raporu Uygulama Esaslarını inceledim ve bu uygulama esaslarında belirtilen azami benzerlik oranlarına göre tez çalışmasının herhangi bir intihal içermediğini, aksinin tespit edilmesi durumunda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini bilgilerinize arz ederim. 23.08.2024

Danışman: Prof. Dr. Mehmet Sıtkı İLKAY

Öğrenci: Metin ERGÖKTAŞ

KILAVUZA UYGUNLUK

“Müşteri Memnuniyetinin Kalite Boyutları Bakımından Metin Madenciliği ile Keşfi” başlıklı Doktora Tezi, Erciyes Üniversitesi Sosyal Bilimler Enstitüsü, Lisansüstü Tez Yazım Kılavuzuna uygun olarak hazırlanmıştır.

Hazırlayan

Metin ERGÖKTAŞ

Danışman

Prof. Dr. Mehmet Sıtkı İLKAY

İşletme ABD Başkanı

Prof. Dr. Ayşe CİNGÖZ

KABUL VE ONAY TUTANAĞI

Prof. Dr. Mehmet Sıtkı İLKAY danışmanlığında **Metin ERGÖKTAŞ** tarafından hazırlanan “**Müşteri Memnuniyetinin Kalite Boyutları Bakımından Metin Madenciliği ile Keşfi**” adlı bu çalışma jürimiz tarafından Erciyes Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı’nda **Doktora** tezi olarak kabul edilmiştir.

05/08/2024

(Tez Savunma Sınav Tarihi yazılacak)

JÜRİ:

Danışman : Prof. Dr. Mehmet Sıtkı İLKAY
Üye : Prof. Dr. İbrahim Doğan
Üye : Prof. Dr. Yasemin YAVUZ
Üye : Doç. Dr. Yılmaz DELİCE
Üye : Doç. Dr. Burcu ORALHAN

ONAY:

Bu tezin kabulü Enstitü Yönetim Kurulu'nun /.../..... tarih ve.....sayılı kararı ile onaylanmıştır.

.... /...../

Prof. Dr. Atabey KILIÇ

Enstitü Müdürü

ÖN SÖZ

Uzun süren tez çalışmam süresince ilgisini, desteğini ve sabrını esirgemeyen, tecrübe ve bilgi birikimleri ile kendilerinden çok şeyler öğrendiğim kıymetli hocam ve danışmanım Prof. Dr. Mehmet Sıtkı İLKAY'a çok teşekkür ederim.

Çalışmalarına yön veren çok değerli katkı ve yönlendirmeleri dolayısıyla kıymetli hocalarım Prof. Dr. Yasemin YAVUZ ve Prof. Dr. İbrahim DOĞAN'a ayrı ayrı teşekkürü bir borç bilirim.

Metin ERGÖKTAŞ

Ankara-2024

MÜŞTERİ MEMNUNİYETİNİN KALİTE BOYUTLARI BAKIMINDAN METİN MADENCİLİĞİ İLE KEŞFİ

Metin ERGÖKTAŞ

Erciyes Üniversitesi, Sosyal Bilimler Enstitüsü

Doktora Tezi, Ağustos 2024

Danışman: Prof. Dr. Mehmet Sıtkı İLKAY

ÖZET

Bu çalışmanın amacı satın alınan ürünle ilgili müşteri yorumlarının, ürün kalitesinin sekiz ürün kalite boyutu dikkate alınarak metin madenciliği ve duygu analizi teknikleri ile analiz edilmesidir. Çalışmada veri odaklı bir model kullanılmıştır. Bu çalışma üretici firmalara ürünleriyle ilgili iyileştirmeye açık yanları iyileştirmek için önemli bir bilgi sağlarken, müşteriler açısından da satın alma kararına konu olan ürünün katlanılan maliyete değip değmeyeceğini her bir kalite boyutu bakımından ortaya koyacaktır.

Çalışmada internet üzerindeki müşteri ürün yorumları metin madenciliği teknikleri kullanılarak, kalite boyutları bağlamında analiz edilmiş ve tüketicilerin ürünle ilgili algılamalarının duygusal analiz sonuçları ortaya konmuştur.

Ürün yorumlarının hangi kalite boyutu kapsamında değerlendirilmesi gerektiğini tahminleyen bir sınıflandırma modeli dört farklı sınıflandırma algoritması ile tasarlanmış ve model başarımları ölçümlenmiştir. Yorum verilerinin kalite boyutlarına atanmasından sonra, her bir kalite boyutu özelinde ilgili yorumların içerdiği duygusal kutbun pozitif veya negatif olarak iki kutuplu bir ölçekle tespit edilmesi amacıyla duygu analizi sınıflandırma modeli oluşturulmuştur. İki farklı sınıflandırma yöntemini kullanan bu model içinde her bir teknik için başarımlar ölçütleri ortaya konmuştur.

Son olarak tahminlenen duygusal kutupsallık verilerinden yola çıkarak kalite boyutlarının pozitif ve negatif yorum sayıları ortaya çıkarılmış ve genel müşteri algısıyla ilgili bir perspektif sunulmuştur.

Anahtar Kelimeler: Makine Öğrenmesi, Ürün Kalite Boyutları, Müşteri Memnuniyeti, Duygu Analizi, Metin Madenciliği

DISCOVERY OF CUSTOMER SATISFACTION THROUGH TEXT MINING IN TERMS OF QUALITY DIMENSIONS

Metin ERGÖKTAŞ

Erciyes University, Graduate School of Social Sciences

PhD Thesis, August 2024

Supervisor: Prof. Dr. Mehmet Sıtkı İLKAY

ABSTRACT

The purpose of this study is to analyze customer reviews about the purchased product by using text mining and sentiment analysis techniques by considering eight product quality dimensions of product quality. A data-driven model was used in the study. This study will provide manufacturers with important information to improve the aspects of their products that are open to improvement, while for customers, it will reveal whether the product subject to the purchase decision is worth the cost in terms of each quality dimension.

In this study, customer product reviews on the internet were analyzed in the context of quality dimensions using text mining techniques and the results of sentiment analysis of consumers' perceptions about the product were revealed.

A classification model that predicts the quality dimension under which product reviews should be evaluated is designed with four different classification algorithms and model performance is measured. After assigning the review data to the quality dimensions, an sentiment analysis classification model was created to determine the emotional polarity of the related reviews in each quality dimension with a bipolar scale as positive or negative. Within this model, which utilizes two different classification methods, success criteria for each technique are established.

Finally, based on the estimated sentiment polarity data, the number of positive and negative comments of the quality dimensions is revealed and a perspective on the overall customer perception is presented.

Keywords: Machine Learning, Product Quality Dimensions, Customer Satisfaction, Sentiment Analysis, Text Mining

İÇİNDEKİLER

MÜŞTERİ MEMNUNİYETİNİN KALİTE BOYUTLARI BAKIMINDAN METİN MADENCİLİĞİ İLE KEŞFİ

BİLİMSEL ETİĞE UYGUNLUK.....	i
TEZ ÖZGÜNLÜK SAYFASI.....	ii
KILAVUZA UYGUNLUK	iii
KABUL VE ONAY TUTANAĞI	iv
ÖN SÖZ	v
ÖZET	vi
ABSTRACT	vii
İÇİNDEKİLER.....	viii
SİMGELER VE KISALTMALAR LİSTESİ	xi
TABLOLAR LİSTESİ	xiii
ŞEKİLLER LİSTESİ	xiv
GİRİŞ.....	1

BİRİNCİ BÖLÜM

KALİTE KAVRAMI VE ÜRÜN KALİTE BOYUTLARI

1.1. Kalite Tanımı ve Kavramsal Çerçeve.....	6
1.1.1. Kalite Kavramının Tanımı, Önemi ve Kavramsal Çerçeve.....	6
1.1.2. Ürün Kalitesinin Tanımı ve Diğer Kalite Türlerinden Ayrımı.....	8
1.2. Ürün Kalite Boyutları	11
1.2.1. Performans.....	14
1.2.2. Güvenilirlik.....	16
1.2.3. Dayanıklılık	17
1.2.4. Servis	19
1.2.5. Estetik	21

1.2.7. Algılanan Kalite.....	24
1.2.8. Uygunluk	26
1.3. Ürün Kalitesi, Müşteri Beklentileri ve Müşteri Memnuniyeti Arasındaki İlişkinin İncelenmesi.....	27
1.3.1. Müşteri Beklentilerinin Karşılansının Müşteri Memnuniyeti Üzerindeki Etkisi.....	28
1.3.1.1. Kano Modeli	29
1.3.1.2 Beklenti Çıkarım Teorisi	31
1.3.1.3 Nedensel Eylem Teorisi.....	32
1.3.1.4 Müşteri Memnuniyet İndeksi.....	34
1.3.2. Kalite Boyutları ve Müşteri Beklentileri Arasındaki İlişki	35
1.3.3. Ürün Kalitesi ve Müşteri Memnuniyeti Arasındaki İlişki	39
1.4. Literatür Taraması	41

İKİNCİ BÖLÜM

VERİ VE METİN MADENCİLİĞİ, DUYGU ANALİZİ KAVRAMSAL ÇERÇEVE

2.1 Veri Madenciliği Genel Tanım ve Kavramsal Çerçeve	48
2.2 Veri Madenciliği Uygulama Alanları	50
2.3. Veri Madenciliği Süreçleriyle İlişkili Disiplinler	56
2.3.1 İstatistik Veri Madenciliği İlişkisi	56
2.3.2 Makine Öğrenmesi Veri Madenciliği İlişkisi	60
2.3.3 Veri Tabanı Yönetim Sistemleri Veri Madenciliği İlişkisi	62
2.3.4 Matematik Veri Madenciliği İlişkisi.....	64
2.3.5 Bilgisayar Bilimleri Veri Madenciliği İlişkisi	67
2.3.6 Yapay Zekâ Veri Madenciliği İlişkisi	69
2.3.7 Görselleştirme Veri Madenciliği İlişkisi	70
2.4 Metin Madenciliği	73
2.4.1 Metin Madenciliği Tanım ve Kavramsal Çerçeve.....	73
2.4.2 Metin Madenciliği Süreçleri.....	76

2.4.2.1 Veri Toplama	76
2.4.2.2 Metin Ön İşleme	78
2.4.2.3 Özellik Çıkarımı	80
2.4.2.4 Sınıflandırma Algoritmaları	83
2.4.2.4.1 Karar Ağaçları	84
2.4.2.4.2 Rastgele Ormanlar	86
2.4.2.4.3 Destek Vektör Makineleri	88
2.4.2.4.4 K-en Yakın Komşu	89
2.5 Duygu Analizi	91

ÜÇÜNCÜ BÖLÜM

MÜŞTERİ YORUMLARININ SINIFLANDIRILMASI VE DUYGU ANALİZİ

3.1 Müşteri Yorumlarının Sınıflandırılması	95
3.1.1 İkili Sınıflandırma Modelleri	96
3.1.1.1 Verinin Okunması, Düzenleme ve Doküman Formatına Dönüştürme....	98
3.1.1.2 Metin Ön İşleme Süreçleri.....	100
3.1.1.3 Anahtar Kelime Çıkarımı ve Vektör Dönüşümü	102
3.1.1.4 Sınıflandırma Algoritmalarının Uygulanması ve Başarım Ölçümü	103
3.1.2 Çok Sınıflı Sınıflandırma Modeli	106
3.2 Duygu Analizi Modelinin Oluşturulması	108

SONUÇ	123
--------------------	------------

KAYNAKÇA.....	128
----------------------	------------

SİMGELER VE KISALTMALAR LİSTESİ

API	:	Application Programming Interface
BERT	:	Bidirectional Encoder Representations from Transformers
BoW	:	Bag of Words
CLV	:	Customer Lifetime Value
CRISP-DM	:	Cross-Industry Standard Process for Data Mining
CSI	:	Customer Satisfaction Index
DBMS	:	Database Management System
DMAIC	:	Define, Measure, Analyze, Improve, Control
ETL	:	Extract, Transform , Load
FMEA	:	Failure Mode and Effects Analysis
HDMI	:	High-Definition Multimedia Interface
HTTP	:	Hyper Text Transfer Protocol
KNN	:	K Nearest Neighbor
LDA	:	Latent Dirichlet Allocation
MUSA	:	Multi-criteria satisfaction analysis
NLP	:	Natural Language Processing
NoSQL	:	Not Only Structred Query Language
POS	:	Part of Speech

REGEX	:	Regular Expression (Kurallı İfadeler)
RF	:	Random Forest
SPC	:	Statistical Process Control
SQL	:	Structured Query Language
SVM	:	Support Vector Machine (Destek Vektör Makineleri)
TF-IDF	:	Term Frequency-- Inverse Document Frequency
URL	:	Uniform Resource Locator
USB	:	Universal Serial Bus
WWW	:	World Wide Web
XML	:	Extensible Markup Language
YBS	:	Yönetim Bilişim Sistemleri

TABLÖLAR LİSTESİ

Tablo 1.1. Ürün kalite boyutları.....	12
Tablo 1.2. Garvin'in kalite yaklaşımı ve klasik kalite anlayışları arasındaki farklar	14
Tablo 2.1. Gövdeleme ve kök bulma işlem karşılaştırması	80
Tablo 3.1. Karışıklık matrisi	105
Tablo 3.2. Kalite boyutları sınıflandırma modeli başarımları ölçümleri.....	106
Tablo 3.3. Çok sınıflı sınıflandırma başarımları ölçümleri.....	107
Tablo 3.4. Her bir kalite boyutu için duygu analizi modeli doğruluk oranları	119
Tablo 3.5. Tüm yorumlar için duygu analizi modeli doğruluk oranları	120
Tablo 3.6. Kalite boyutları bazında veri satırı sayıları.....	120
Tablo 3.7. Kalite boyutları bazında tahminlenen duygusal kutup sayıları	121

ŞEKİLLER LİSTESİ

Şekil 1.1. Sürekli iyileştirme döngüsü.....	41
Şekil 2.1. İris veri seti karar ağacı	86
Şekil 3.1. Müşteri yorumları veri seti.....	96
Şekil 3.2. Örnek bir kalite boyutu veri seti	97
Şekil 3.3. Sınıflandırma modeli.....	98
Şekil 3.4. Veri okuma ve doküman dönüşümü iş akışı	98
Şekil 3.5. Table creator konfigürasyon penceresi	99
Şekil 3.6. Metin ön işleme süreçleri	100
Şekil 3.7. Anahtar kelime çıkarımı ve vektör dönüşümü	102
Şekil 3.8. Anahtar kelime çıkarımı konfigürasyon penceresi	103
Şekil 3.9. Sınıflandırma iş akışları	104
Şekil 3.10 Uygunluk boyutu sınıf. modeli karışıklık matrisi	105
Şekil 3.11. Duygu analizi sınıflandırma modeli.....	109
Şekil 3.12. Kelime çantası düğümü çıktı tablosu	110
Şekil 3.13. Duygu Analizi Sınıflandırma Modeli (Boyut İndirgeme ile).....	112
Şekil 3.14. Hesaplanan TF değerlerini gösteren çıktı tablosu.....	113
Şekil 3.15. Category to class çıktı tablosu	114
Şekil 3.16. Bölümleme düğümü yapılandırma seçenekleri	115
Şekil 3.17. Random forest learner düğümü yapılandırma penceresi.....	118
Şekil 3.18. Kalite boyutları pozitif kutup oranları karşılaştırması	121

GİRİŞ

Pazarlama karması elemanlarının dijital enstrümanlarla kurgulanmaya başlaması ve E-Ticaret olgusu dünyayı global, kolay erişilebilir ve kompakt bir pazar haline getirmiştir. 2021 yılında 4,9 trilyon dolar olan dünya e-ticaret hacminin 2025 yılında yaklaşık 7,4 trilyon dolar olacağı tahmin edilmektedir (Alma, 2022). Bununla birlikte ülkemizde ve tüm dünyada internet penetrasyonu ve mobil erişim yıllar geçtikçe artmaktadır.

Bunun sonucu olarak işletmeler için yüksek rekabet edebilirlik hiç olmadığı kadar kritik bir önem kazanmıştır. Bu bağlamda, işletmeler için müşteri sadakatinin sağlanması ve yeni müşteri kazanımları sağlayabilmek oldukça önemlidir. Bu amaca yönelik olarak değişen ve gelişen müşteri taleplerinin anlaşılabilmesi için önemli bir veri kaynağı da çevrimiçi ürün yorumları, ürün değerlendirme siteleri, bloglar ve sosyal paylaşım ağlarıdır. Müşteriler satın alma sürecinde sıklıkla bu alanları ziyaret ederek geçmiş kullanıcı tecrübelerinden, incelemelerinden ve deneyimlerinden faydalanmak suretiyle satın alma kararlarını vermektedirler. Müşterilerin çok büyük bir kısmının e-ticaret alışverişlerinde 4 yıldız ve üzeri olarak değerlendirilen ürünleri tercih ettiği artık bilinen bir gerçektir.

Bu gerçek, çevrimiçi müşteri yorumlarının müşteriler için deneyimledikleri ürünler bakımından ifade ettiği değeri göstermektedir. Dolayısıyla işletmelerin e-ticaret kanalıyla sattıkları ürünleriyle ilgili olarak müşteri değerlendirmelerini, hızlı ve doğru okuyup yorumlaması gerekmektedir. Bu amaçla yalın girişim (lean startup) gibi bir ürünün olabilecek en basit (yalın) formuyla piyasaya sürülüp müşteri dönütlerine göre nihai formunu aldığı yaklaşımlar da göz önüne alındığında, ürünle ilgili olarak müşteri değerlendirmelerinin işletmeler için ifade ettiği önem anlaşılmaktadır. Bu tez çalışması bir yönüyle çevrimiçi ürün yorumları ve incelemelerinin işletmeler açısından ürünlerinde iyileştirmeye açık yanları ortaya çıkarmayı amaçlarken diğer yandan da müşteriler

açısından ürün yorumlarının değerlendirmesi anlamında bir vizyon sunacak ve katlanılan maliyete değip değmediği hususunda fikir verecektir. Çalışma çevrimiçi yorum verilerinin kalite boyutları bağlamında metin madenciliği ve duygu analizi teknikleri kullanılarak analize tabi tutulması yönüyle yapılmış ilk çalışmadır.

Çalışmada ürün kalitesi 8 kalite boyutunda değerlendirilmiştir. Sekiz ürün kalitesi boyutu şunlardır: Performans, özellikler, güvenilirlik, uygunluk, dayanıklılık, servis, estetik ve algılanan kalite (Sebastianelli & Tamimi, 2002, s. 445).

Çalışma kapsamında müşterilerin çevrimiçi ürün yorumlarının metin madenciliği ile analizi sonucu yukarıda belirtilen ürün kalite boyutları bakımından aşağıdaki soruların cevaplanması amaçlanmıştır.

1. Performans - Ürünün beklenen performansına ilişkin genel müşteri memnuniyeti düzeyi nedir?
2. Uygunluk - Tanımlı spesifikasyonlara ve standartlara uyum konusunda genel müşteri algısı nedir?
3. Güvenilirlik - Bir ürünün ömrü boyunca performans özelliklerinin tutarlılığına ilişkin genel müşteri bakış açısı nedir?
4. Dayanıklılık - Müşterilerin bir ürünün uzun ömürlülüğü ve sağlamlığına ilişkin beklentileri, ürünün gerçek dayanıklılığı ve güvenilirliği ile ne ölçüde örtüşmektedir?
5. Servis - Ürünle ilgili sorunların ve şikayetlerin çözülmesindeki verimlilik ve çözüm hızı açısından müşteri memnuniyeti derecesi nedir?
6. Estetik - Ürünün çekiciliği ve duyuları harekete geçirme kapasitesi hedef demografinin tercihleriyle ne ölçüde örtüşmektedir?
7. Özellikler - Çekiciliği artıran ve birincil işleve ikincil eklemeler olarak hizmet eden ürün özelliklerine ilişkin müşteri algıları ne ölçüde olumluya veya olumsuzaya doğru yönelmektedir?
8. Algılanan Kalite - Bir şirketin veya ürünün itibarına ilişkin, ürün imajı, marka imajı ve şirket imajı gibi çeşitli faktörlerden etkilenen genel müşteri algısı nedir?

Bununla birlikte, internet ortamındaki binlerce müşteri yorumunun ve ürün incelemesinin oluşturduğu devasa metnin hem boyut olarak büyüklüğü hem de belirli bir standartta

olmayışı bu verilerin manuel olarak analiz sürecini işletmeler açısından maliyet, insan kaynağı, zaman ve emek boyutlarında önemli ölçüde artırmaktadır. Metin madenciliği çalışmaları bu anlamda adı geçen yapılandırılmamış devasa boyuttaki metin verisinden işletmeler için rekabet edebilirlik, inovasyon ve müşteri ilişkileri yönetimi gibi konularda katma değer katacak anlamlı kurallar, örüntüler ve nitelikli bilgi keşfi süreçlerini ifade etmektedir.

Çalışmada metin madenciliği teknikleri kullanılarak, internetteki ürün yorumları ve değerlendirmelerinin sekiz ürün kalite boyutunun her birisi için ifade ettiği anlam tespit edilmiş ve bu boyutlar açısından müşteri deneyimlerinin ifade ettiği duygusal kutbun pozitif veya negatif olarak hangi tarafa yakın olduğu değerlendirilmeye çalışılmıştır.

Bu analitik çalışması her şeyden önce işletmelere müşterilerin ürünle ilgili genel algı seviyelerini ölçebilmek için maliyet, emek ve işgücü olarak etkin bir sistematik sunmaktadır. Dolayısıyla işletmelerin ürünle ilgili değişen müşteri talepleri karşısında daha esnek ve adaptif olabilmesi mümkün olacak ve rekabet edebilirlik düzeyleri artacaktır.

Bununla birlikte kalitenin sekiz boyutu bağlamında ürün yorumları ve incelemelerinin her bir boyuttan hangileri ile ilgili olduğu kategorize edilerek, müşterilerin sıklıkla hangi kalite boyutlarıyla ilgili konuştukları anlaşılmaya çalışılmıştır.

Çalışma bu yönüyle de işletmenin ürünle ilgili yoğunlaşması gereken noktalara işaret eden bir vizyon sunmayı hedeflemektedir. Bu doğrultuda Çevrimiçi müşteri yorumları, işletme için “kullanıcı deneyimi (user experience)” bağlamında önemli bir “bilgi” ye dönüşecektir. Bununla birlikte analitik çalışma sonucunda kalitenin her bir boyutuyla ilgili olarak izlenmesi gereken strateji daha net ve yalın olarak ortaya konabilecektir.

E-Ticareti ve dijital pazarlama kanallarını aktif olarak kullanan işletmeler için ürün yorumlarının analizi sonucu şu soruların cevabını büyük ölçüde verebilen bir model oluşturulması amaçlanmıştır:

- Ürünün performans özellikleri müşteriler tarafından nasıl görülüyor?
- Hangi özellikleri yüzünden ürün tercih ediliyor veya edilmiyor?
- Ürün kendisinden beklenen fonksiyonları ne derece yerine getirebiliyor?
- Ürünün standartlara uygunluğu hususunda müşterilerin genel algısı nasıl?

- Ürünün kullanım ömrü müşteri beklentileriyle ne kadar örtüşüyor?
- Ürünün bakım ve arıza durumlarındaki servis imkânları hız, maliyet ve müşteri memnuniyeti açısından ne kadar tatmin edici bulunuyor?
- Ürün estetik görünüm itibarıyla müşteri açısından ne derece albenili bulunuyor, ürünün, markanın veya firmanın hedef kitle gözündeki kurumsal imajı ve marka itibarı ne derecedir?
- Tüm bu soruların bileşimi olarak müşteri ürünü katlanmak zorunda kaldığı maliyete değer buluyor mu?

Bu çalışma kapsamında ulaşmak istediğimiz sonuç; müşteri görüşlerinin, kalitenin her bir boyutu bakımından olumlu/olumsuz hangi kutba yakın olduğunun anlaşılmasına çalışılması ve sonuçta kalite boyutları bakımından ürün kalitesi konusunda müşteri görüşünü ortaya koymaktır. Çalışma kapsamında KNIME platformu kullanılmıştır. KNIME platformu veri ve metin madenciliği çalışmaları için tasarlanmış, akış-diyagramı mantığıyla çalışan ve görselleştirme imkânı sunan ücretsiz bir platformdur.

Çalışma üç ana bölümden oluşmaktadır. Birinci bölüm kalite kavramı, ürün kalite boyutları ve müşteri memnuniyeti ile ilgili teorik bir çerçeve sunmaktadır. İkinci bölüm veri madenciliği, metin madenciliği ilişkisi, bağlı disiplinler ve duygu analizi bağlamında teorik bir değerlendirme yapmaktadır. Üçüncü Bölüm de hem ikili sınıflandırma hem de çoklu sınıf sınıflandırma yaklaşımına göre yorum verileri kalite boyutlarına göre sınıflandırılmıştır. İkili sınıflandırma modelinde her bir kalite boyutu özelinde oluşturulan modelin başarımlarını performansları ölçülerek sunulmuştur. Çoklu sınıf sınıflandırma yaklaşımıyla aynı zamanda yorum verileri sekiz kalite boyutuna göre sınıflandırılarak model başarımlarını performansları ortaya çıkarılmıştır.

Üçüncü bölümde duygu analizi modeli oluşturularak, her bir kalite boyutu özelinde ve tüm yorumlar bakımından modelin başarımlarını performansları ölçülerek pozitif veya negatif kutupsallık ölçümü yapılmıştır.

Çevrimiçi platformlardaki yapısal olmayan veri miktarı ve buna olarak metin madenciliği uygulamalarının her geçen gün yaygınlaştığı düşünüldüğünde, işletmelerin müşteri davranışlarını anlayarak bu doğrultuda ürün geliştirme ve pazarlama stratejileri

geliştirebilmeleri ve böylece rekabet avantajı sağlayabilmeleri açısından bu çalışma pazarlama alanında veri odaklı karar verme vizyonu sunmayı hedeflemektedir.



BİRİNCİ BÖLÜM

KALİTE KAVRAMI VE ÜRÜN KALİTE BOYUTLARI

Bu bölümde kalite kavramının tanımı, kavramsal çerçevesi detaylı olarak izah edilecek, ürün kalite boyutlarını oluşturan bileşenlerin detaylı açıklamalarına yer verilecektir. Bunun yanında ürün kalite boyutları ile müşteri memnuniyeti arasındaki ilintiye değinilerek, müşteri odaklılık, kalite ve müşteri sadakati, kalitenin maliyeti ve rekabet avantajı gibi temel kavramlarda detaylandırılacak ve konuyla ilgili yapılmış çalışmalar literatür taraması başlığı altında değerlendirilecektir.

1.1. Kalite Tanımı ve Kavramsal Çerçeve

1.1.1. Kalite Kavramının Tanımı, Önemi ve Kavramsal Çerçeve

Kalite, ürünlerin müşteri ihtiyaçlarını karşılayan ve böylece müşteri memnuniyeti sağlayan özellikleri anlamına gelmektedir (Juran & Godfrey, 1999, s. 26). Buradan hareketle kalite bir ürün veya hizmetin belirli gereksinimlerle uyumlu olma ve müşterilerin beklentilerini karşılama ve böylece müşteri memnuniyetinin sağlanmasına katkıda bulunma derecesi olarak tanımlanabilir.

Kalitenin tarihsel gelişimi, endüstrinin ve teknolojik yeniliklerin ilerlemesiyle doğrudan ilişkilidir. Geçmişte kalite, sanayi devrimi sırasında öncelikle üretim süreçlerinin denetlenmesine odaklanmışken günümüzde kaliteye ilişkin çağdaş bakış açıları, ürün tasarımından satış sonrası desteğe kadar daha geniş bir kapsamı ele almaktadır. Kalitenin tarihsel süreçteki gelişimi, müşteri odaklılık, sürekli inovasyon ve operasyonel verimlilik gibi faktörlerin önemini daha da artırmıştır.

Kalite kavramının tarihsel süreç içerisinde geçirdiği değişim oldukça ilgi çekicidir. 20. yüzyılın başlarında Frederick Taylor'ın bilimsel yönetim ilkeleriyle başlayan süreç daha sonra ticari faaliyetlerin etkinliğini ve verimliliğini optimize ederek artırmaya yönelik

alıřmalarla devam etmiřtir. İlerleyen srete Shewhart ve Deming gibi isimler istatistiksel sre kontrol ve kalite kontrol teorilerinin geliřmesine katkıda bulunmuřlardır. Bu alıřmaların ıřıřında aędař kalite ynetimi ilkelerinin temelleri atılmıřtır. Gnmzde kalite kavramı, iřletmeler iin rekabet avantajı saęlama ve mřteri baęlılıęını artırma anlamında bařat faktr olarak tanımlanmaktadır.

Kalite kavramı bir řirketin etkili bir řekilde rekabet etme, mřteri memnuniyeti saęlama, uzun sreli srdrlebilirlięin saęlanması aılarından ok nemlidir. Kalite, mřteri gvenini oluřturmak, rn veya hizmetlerin pazardaki konumunu korumak ve iřletmelerin genel itibarını arttırmak iin gerekli bir faktrdr. Kalite kavramı beraberinde giderlerin en aza indirgenmesi, operasyonel verimlilięin artırılması ve srekli iř geliřtirme abalarını da getirir.

Bir řirketin rn veya hizmetlerinde kalite gvencesi, mřteri beklentilerinin karřılanması ve sonuta yksek dzeyde mřteri memnuniyeti saęlanması anlamlarına gelir. Kalite, iřletmenin rn veya hizmetlerinin karřılařtıkları rekabet ortamında nemli bir avantaj saęlar. Rekabeti pazar ortamında tketiciler, kaliteli rn veya hizmetler sunan řirketleri semeye daha yatkındır, bu da potansiyel gelir artıřına ve iřletmenin pazarda daha iyi konumlanmasına yol aar.

Yksek kalite standartlarını srdrmek, aynı zamanda bir řirketin marka itibarının řekillenmesinde rol oynayan nemli faktrlerden biridir. Bir řirket kaliteli rn veya hizmetler sunduęunda tketiciler nezdinde gvenilir ve tercih edilen bir marka olarak grlr. Olumlu bir itibar, bir řirketin mřteri tabanının geniřlemesine ve mřteri kaybının en aza indirilmesine katkıda bulunabilir.

Bir iřletmede yksek kalite standartlarını srdrmenin sonucu olarak operasyonel verimlilik artar. İřletmeler, mřteri beklentilerini karřılayan rn veya hizmetler sunarak zaman ve kaynak kullanımını optimize edebilir. Bu da iř srelerinin daha etkin ynetilmesini kolaylařtırarak maliyetlerin azalmasına yol aacaktır. Tm bu faktrler dikkate alındıęında kalite kavramının bir iřletmenin bařarısını saęlamada kritik bir bileřen olduęu sylenabilir. Kalite, bir taraftan iřletmenin rekabet avantajı elde etmesine, kurumsal itibarını artırmasına, giderleri azaltmasına ve operasyonel verimlilięi optimize etmesine olanak tanırken dięer taraftan da mřteri memnuniyetini artırmada ok nemli

bir rol oynar. Sonuç olarak, işletmelerin kaliteye öncelik vermesi ve kalite kontrol süreçlerini dinamik olarak iyileştirmesi zorunludur.

Bununla birlikte, kalite yönetimi alanı, özellikle yeni metodolojiler ve araçların oluşturulmasına imkân veren teknoloji ve dijitalleşmenin entegrasyonu ile birlikte sürekli bir gelişim ve yenilik sürecinden geçmektedir.

Toplam Kalite Yönetimi (TKY) ise, bir işletmede müşteri memnuniyetini yüksek düzeyde artırmayı ve sürdürülebilir kılmayı amaçlayan tüm organizasyonel prosedürleri ve varlıkları kapsayan bir yaklaşımdır. Toplam kalite yönetiminin ilkeleri; müşteri odaklılık, sürekli iyileştirme, eğitim, tam katılım, stratejik planlama, yönetim sorumluluğu, ekip çalışması ve kurum kültürü olarak ifade edilebilir (Kazan vd., 2017, s. 45). Toplam kalite yönetimi, kalitenin sadece ürün veya hizmetin sonunda değil, tüm süreç boyunca sağlanması gerektiğini vurgular. Toplam kalite kontrol, her kuruluştaki her düzeyde performansın iyileştirilmesine yönelik, tamamıyla entegre olmuş çabalarla, yöneticiden işçiye kadar herkesi kapsayan düzenli sürekli iyileştirme faaliyetleridir (Imai, 1997, s. 25).

Toplam kalite yönetimi, müşteri memnuniyetini artırmak ve sürekli olarak iyileştirmek için bir işletmenin tüm fonksiyonlarını kapsayan kapsamlı bir stratejidir. Toplam kalite yönetimini benimseyen bir işletme önemli ölçüde rekabet avantajı kazanacaktır. TKY yaklaşımı işletmenin genel iş performansını artırmak hedefiyle sürekli iyileştirme ve eğitim faaliyetleri üzerine odaklanmaya dayanır. TKY yaklaşımında ürün ve hizmet kalite kalitesinin iyileştirilmesi amacıyla sürekli kalite kontrolü ve müşteri geri bildirimi gibi veriye dayalı teknikler kullanılır. Bu müşteri odaklı yaklaşım, kalıcı müşteri sadakatini sağlar ve işletmeye karşı müşteri güvenini kazandırır.

1.1.2. Ürün Kalitesinin Tanımı ve Diğer Kalite Türlerinden Ayrımı

Kalite, müşteri memnuniyetini ve iş başarısını doğrudan etkilediğinden herhangi bir ürün veya hizmet için çok önemli bir faktördür. Ürün kalitesini farklı şekillerde tanımlamak ve değerlendirmek mümkündür, ancak genel olarak ürün kalitesi ürünlerin müşteri ihtiyaçlarını karşılayan ve böylece müşteri memnuniyeti sağlayan özellikleridir (Juran vd., 1988, s. 21). Ürün kalitesi, ürünün özellikleri, güvenilirliği, dayanıklılığı, performansı ve estetiği gibi farklı metriklerle ölçülebilir. Müşteri hizmetleri, satış sonrası destek ve genel kullanıcı deneyimi gibi diğer yönler de ürün kalitesinin algılanmasına

katkıda bulunur. Bunun yanında ürün kalitesini diğer kalite türlerinden ayırmak önemlidir. Ürün kalitesi özellikle fiziksel bir ürünün özelliklerini ve niteliklerini ifade ederken hizmet kalitesi veya süreç kalitesi gibi diğer kalite türleri, sağlanan hizmetlerin kalitesine veya bu hizmetlerin sunulmasında yer alan süreçlerin verimliliğine ve etkinliğine odaklanır. Örneğin hizmet kalitesi, bir hizmetin müşteri beklentilerini karşılama veya aşma derecesidir. Müşteri memnuniyeti, kalite güvencesi sisteminin en temel ilkelerinden birisini oluşturmaktadır. Dolayısıyla hizmet kalitesinin belirlenmesinde, ölçülmesinde temel ölçüt olarak kullanılmaktadır (Özkan & Alp, 2020, s. 97). Hizmet kalitesi, hizmet sağlayıcının yanıt verme yeteneği, zamanlılığı, nezaketi ve yeterliliği gibi faktörlerle beraber, müşteri beklentilerinin karşılanması, hizmetin doğruluğu, güvenilirlik, empati, fiziksel çevre ve iletişim gibi çeşitli boyutları da kapsamaktadır. Bununla birlikte hizmet kalitesi müşteri beklentileri ile algılamaları arasındaki farklılıklar olarak ifade edilmektedir (Parasuraman vd., 1985, s. 42). Güvenilirlik, müşterilere hizmetlerin tutarlı ve zamanında teslimi hususunda güvence duygusu aşılarken, güvenlik hizmet etkileşimleri sırasında müşteriler arasında güven duygusunu geliştirmek için gereklidir. Empati, müşterilerin bireysel ihtiyaçlarını kişiselleştirilmiş ilgiyle anlama ve karşılama kapasitesiyle ilgiliyken, sorun çözme becerisi, müşteri şikayetlerini etkili bir şekilde dinlemeyi ve memnuniyetsizlik durumunda çözüm sunmayı içerir. Fiziksel çevrenin kalitesi, hizmet ortamındaki temizlik, organizasyon ve konfor gibi faktörleri kapsar. İletişim ise müşteri ve hizmet sağlayıcı arasındaki etkili iletişimi ifade eder.

Hizmet kalitesinin değerlendirilmesi tipik olarak müşteri geri bildirimlerinin, anketlerin ve denetimlerin toplanmasını ve analizini içerir. Müşteri memnuniyeti oranları, hizmet sunumundaki hata oranları, teslimat süreçlerinin verimliliği ve çalışan performansı gibi çeşitli ölçümler bu amaçla yaygın olarak kullanılmaktadır. Öte yandan ürün kalitesi, genellikle kalite kontrol prosedürleri, deneme üretimleri ve yerleşik standartlara uygunluğun değerlendirilmesi yoluyla değerlendirilir. Ürün kalitesinin belirlenmesi dayanıklılık, işlevsellik ve genel performans gibi faktörlerin değerlendirilmesine dayanır. Hizmet kalitesi ile müşteri deneyimi arasında doğrudan çok kuvvetli bir ilişki vardır ve çalışanlar bu ilişkiyi etkileyen birincil faktördür. Hizmet kalitesinin artırılması için çalışanlara yönelik düzenli olarak iş hedeflerine ve müşteri memnuniyeti politikalarına uygun eğitim programlarının uygulanması gerekmektedir. Hizmet kalitesinin sürekli

olarak artırılmasında çalışanların rol ve sorumluluk anlayışlarının artırılması zorunludur. Bununla birlikte ürün kalitesinin artırılması da müşteri memnuniyeti ve marka itibarı açısından göz ardı edilemez. Kalite kontrol prosedürleri, ürünlerin belirtilen standartları karşıladığını doğrulamak için uygulanır ve kalite kontrol departmanı, tüm ürün yaşam döngüsü boyunca kalite yönetimini denetler. Standartlardan herhangi bir sapmayı derhal gidermek ve ürün kalitesini sürekli olarak yüksek düzeyde tutmak için düzenli kalite testleri gerçekleştirilir ve sonuç olarak müşteri memnuniyeti ve sadakatinin kaybolmaması hedeflenir. Hizmetin temel özelliklerinden dolayı, müşteriler için hizmet kalitesinin değerlendirilmesi mamul kalitesini değerlendirmekten daha zordur (Savaş & Kesmez, 2014, s. 3)

Hizmet kalitesi ve ürün kalitesi arasındaki temel farklar, öncelikli olarak sunulan hizmetin özelliklerine ve tüketici üzerindeki etkisine göre kendisini göstermektedir. Bunun yanında hizmet kalitesinin değerlendirilmesi, müşterinin deneyiminin, müşteriye sunulan değerin ve müşteri memnuniyetin yanı sıra sunum, iletişim ve problem çözme yetenekleri gibi unsurların dikkate alınmasını da gerektirir. Sonuç olarak hizmet kalitesinin değerlendirilmesi ürün kalitesine göre daha sübjektiftir. Tersine, ürün kalitesi öncelikle bir ürünün fiziksel özellikleri ve işleyişiyle ilgilidir. Ürünün uzun ömürlülüğü, kullanılabilirliği, estetiği ve malzeme bileşimi gibi unsurlar ürünün kalitesini etkiler. Ürün kalitesi, ürünün belirli performans standartlarını karşılayıp karşılamadığını tespit edebilen testlerle daha objektif değerlendirme ve ölçüme tabidir.

Hizmet kalitesi ve ürün kalitesinin iş başarısındaki önemi göz ardı edilemez. Müşteriler her iki açıdan da tatmin edici bir deneyim beklemektedir. Bu da işletmeler tarafından bu niteliklerin sürekli izlenmesini ve geliştirilmesini gerektirmektedir. Ancak işletmelerin hizmet ve ürün kalitesi arasındaki farkları iyi anlaması ve bu farklılıklara göre doğru önceliklerin belirlenmesi büyük önem taşımaktadır. İşletmelerin bu doğrultuda hizmet sunma ve üretim stratejileri oluşturup müşteri memnuniyetini merkeze alan bir yaklaşımla hizmet ve ürün kalitelerini geliştirmekle ilgili stratejilerini belirlemeleri gerekmektedir.

Özetle, ürün kalitesi fiziksel bir ürünün niteliklerini ve özelliklerini ifade ederken, diğer kalite türleri hizmetlerin veya süreçlerin kalitesine odaklanır. Başka bir deyişle, ürün kalitesi, özellikle bir ürünün fiziksel özellikleriyle ilgiliyken, diğer kalite türleri, sağlanan hizmetlerin kalitesine veya bu hizmetlerin sunulmasında yer alan süreçlerin etkinliğine

odaklanır. Ürün kalitesini diğer kalite türlerinden ayırt etmek için, ürün kalitesinin özellikle fiziksel bir ürünün niteliklerine ve özelliklerine atıfta bulunduğunu, diğer kalite türlerinin ise sağlanan hizmetlerin kalitesine veya teslimatta yer alan süreçlerin etkinliğine atıfta bulunduğunu anlamak önemlidir. Ürün kalitesi, ürüne ait özellikler, güvenilirlik, dayanıklılık, performans ve estetik gibi bileşenlerden oluştuğu için ürün kalitesini, hizmet kalitesi veya süreç kalitesi gibi diğer kalite türlerinden ayırmak önemlidir.

1.2. Ürün Kalite Boyutları

Ürün kalitesi, tüketicinin beklentilerini karşılayan veya aşan ürün özelliklerinin toplamıdır (Buzzell & Gale, 1987; V. Zeithaml vd., 1990). Ürün kalite boyutları, bir ürünün kalite değerini belirlemek için kullanılan kriterlerin bir bileşimi olarak tanımlanabilir. Algılanan kalite, fonksiyonellik (özellikler), dayanıklılık, güvenilirlik, estetik, performans, uygunluk ve servis hizmetleri bu kriterlere dahil edilmektedir. Literatürde çeşitli modeller ve yaklaşımlar bu kavramları açıklamaktadır.

Kalite anlayışı ilerleyen süreç içerisinde önemli bir dönüşüm geçirerek daha detaylı ve çok yönlü bir hale gelmiştir. 1980'ler, kalite yönetiminin gelişiminde, kalite ilkelerinin anlaşılmasını ve yönetimini geliştirmeyi amaçlayan araştırma çabalarındaki ve teorik ilerlemelerdeki artışla karakterize edilen çok önemli bir döneme işaret etmektedir. 1980'lerde Harvard Business School profesörlerinden David Garvin, kalite kavramını daha detaylı olarak ele almak ve daha kapsamlı bir çerçeve sağlamak amacıyla sekiz farklı boyut tanımlamıştır (Garvin, 1987a, ss. 104-107). Garvin'in araştırması, işletmelerin kaliteyi daha bütüncül ve kapsamlı bir bakış açısıyla değerlendirmesine olanak tanıyan önemli bir çerçeve sunmuştur. Bu çalışmaya göre bir ürünün değerlendirme kriterleri performans, özellikler, güvenilirlik, uygunluk, dayanıklılık, servis kolaylığı, estetik ve algılanan kalite olarak sınıflandırılabilir. Performans, ürünün temel işlevlerini ne kadar etkili bir şekilde yerine getirdiğini ifade ederken, özellikler ürünün beklenen faydasını artıran her türlü ek özelliği ifade etmektedir. Güvenilirlik, ürünün uzun bir süre boyunca tutarlı ve hatasız çalışma olasılığıyla ilgilidir. Uygunluk, ürün tasarımının ve özelliklerinin belirlenmiş ve kabul görmüş endüstri standartlarıyla ne ölçüde uyumlu olduğunu gösterir. Dayanıklılık, ürünün kullanım ömrü boyunca arıza yapma sıklığıyla birlikte kalitesini koruyabilme yeteneğini tanımlar. Servis kolaylığı, ürüne ait bakım ve

onarım faaliyetlerinin ne ölçüde kolay yapılabilceğinin bir ölçütü olarak kabul edilir. Estetik, ürünün görünümü, ölçüleri, boyutları ve dokusunun öznel değerlendirmelerini içerir. Algılanan kalite, marka itibarı ve pazarlama çabaları gibi faktörlerden etkilenen tüketiciler tarafından ürüne atfedilen değeri ifade eder. Bu parametreler, ürünlerdeki ve tüketici taleplerindeki değişimlere uyum sağlayabilmek için değişim ve dönüşüme uğrar. İşletmeler, bu değişim ve dönüşümün dinamiklerine bağlı kalarak kalite yönetim sistemlerini ve kalite iyileştirme stratejilerini geliştirmek için ürün kalite boyutlarını referans almalıdırlar.

Tablo 1.1. Ürün kalite boyutları

Kalite Boyutu	Açıklama
Performans	Ürünün veya hizmetin temel işlevlerini yerine getirme düzeyi.
Özellikler	Ek işlevler veya nitelikler.
Güvenilirlik	Ürünün veya hizmetin tutarlılığı ve güvenilirliği.
Uygunluk	Ürünün veya hizmetin belirli standartlara uyma derecesi.
Dayanıklılık	Ürünün veya hizmetin ömrü.
Servis	Ürünün veya hizmetin bakım ve onarım kolaylığı.
Estetik	Ürünün veya hizmetin görünüşü ve hissiyatı.
Algılanan Kalite	Müşterilerin ürüne veya hizmete dair algıları.

1980'lere kadar tarihsel süreç içerisinde kalite yönetimi stratejileri, kaliteyi ağırlıklı olarak tek bir boyuta veya sınırlı bir boyut dizisi içerisinde değerlendirmiştir. Bu metodolojiler öncelikli olarak ürünlerin teknik standartlara uygunluğu veya hataların olmaması veya minimizasyonu gibi sınırlı sayıda kriterlere odaklanmışlardır. Garvin ise kaliteyi sekiz farklı boyuta ayırarak kalite değerlendirmesinin kapsamını genişletmiştir. Bu çok yönlü yaklaşım, kalitenin daha kapsamlı, çeşitli ve analitik bir şekilde incelenmesine olanak tanıyarak işletmelere kalite yönetim stratejileri için daha kapsamlı bir uygulama ve değerlendirme altyapısı sunmuştur.

Garvin'in yaklaşımının klasik yaklaşımlardan farklılaştığı bir diğer nokta müşteri odaklılıktır. Önceki klasik yaklaşıma göre, kalite yönetim stratejileri öncelikle üretim süreçlerine ve teknik standartlara vurgu yaparken, müşteri beklentileri ve algılarına daha az vurgu yapmaktaydı. Kalite değerlendirmesi ağırlıklı olarak işletme içi operasyonların etkinliğine ve verimliliğine dayanmaktaydı. Garvin'in kalite boyutları ise müşteri odaklılığın önemine atıf yapmaktadır. Özellikle algılanan kalite ve estetik gibi boyutlar,

müşterilerin ürüne yönelik beklenti ve algılarıyla birlikte duygusal tepkilerini de ölçümlemeyi içermektedir. Garvin'in metodolojisi, müşteri memnuniyetini artırmak ve marka sadakatini korumak amacıyla müşteri beklentilerinin daha iyi anlaşılmasını ve bunlara uyum sağlanmasını teşvik etmiştir. Bununla birlikte klasik yaklaşımların odak noktası operasyonel verimliliği artırarak üretim hatalarını azaltmak ve sadece belirli standartlara uyumluluğu sürdürmek için kalite kontrol ve güvence prosedürlerini uygulamak iken, Garvin'in sekiz boyutlu çerçevesi hem operasyonel hem de stratejik iyileştirme ve geliştirmeler için oldukça kapsamlı, somut ve etkin bir çerçeve sunmaktadır. İşletmeler, bu boyutları ürün geliştirme süreçlerine dahil ederek uzun vadeli stratejik hedefleri önceliklendirerek, yenilikçi, müşteri odaklı yaklaşımları benimseyebilirler. Performans, dayanıklılık ve servis kolaylığı gibi temel boyutlar, ürünlerinin kullanım ömrü boyunca istikrarlı bir kalite çizgisi yakalamasında ve sürekli iyileştirme kültürünün yaygınlaştırılmasında önemli bir rol oynamıştır. Klasik kalite yaklaşımlarının bir diğer ortak özelliği kalitenin değerlendirilmesi süreçlerinde çoğunlukla teknik spesifikasyonlara uygunluk ve üretim süreçlerinde hataların bulunmaması gibi kriterlere odaklanmış olmalarıydı. Bu noktada, Garvin'in metodolojisi ise kalite boyutlarının değerlendirilmesinde objektif ölçümlerin yanı sıra estetik ve algılanan kalite gibi subjektif kriterlerin de değerlendirmeye alınmasını kapsamaktadır. Bu kapsamlı yaklaşım, ürünlerin görünüm, malzeme ve doku gibi duyuşsal yönlerinin yanı sıra marka imajı ve işletme itibarı bağlamında müşteri memnuniyet ve sadakatinin değerlendirilmesini içermektedir. Hem subjektif hem de objektif kriterlerin dengeli bir şekilde ele alınması çok boyutlu, daha kapsamlı ve gerçekçi bir kalite değerlendirmesine imkân vermektedir. Öte yandan klasik kalite yönetimi yaklaşımları performansı değerlendirmek için ağırlıklı olarak ölçülebilir standartlar ve ölçümler kullanmaya odaklandığından teknik isterlere uygunluk ve üretim hatalarının belirlenmesi gibi kriterler kalitenin değerlendirilmesinde temel faktörlerdi. Garvin'in çok boyutlu kalite yaklaşımı, kalite yönetimi kapsamında gerçekleştirilen sürekli iyileştirme ve inovasyon süreçleri için sağlam bir çerçevenin oluşturulmasını sağlamıştır. Özellikler ve performans boyutlarının kalite yönetimine entegrasyonu, ürünlere yeni işlevlerin entegrasyonunu ve performans seviyelerinin istikrarlı bir şekilde geliştirilmesini teşvik ederek, devam eden iyileştirme ve ilerleme süreçlerine daha vizyoner bir perspektif katmıştır. Sonuç olarak işletmeler,

gelişen ve değişen pazar dinamiklerine ve değişen tüketici tercihlerine karşı daha adaptif olabilmişlerdir.

Tablo 1.2. Garvin'in kalite yaklaşımı ve klasik kalite anlayışları arasındaki farklar

Kriter	Garvin'in Kalite Yönetim Anlayışı	Klasik Kalite Yönetim Anlayışları
Kalite Boyutları	Performans, Özellikler, Güvenilirlik, Dayanıklılık, Hizmet Edilebilirlik, Estetik, Algılanan Kalite	Standartlara uyum, hata oranları, sürekli iyileştirme esaslı
Süreç (Process) Yönetimi	Süreçlerin etkin yönetimi ve iyileştirilmesi	Süreç yönetimi önemli ancak çoğunlukla ürün kalitesi odaklı
İnsan Kaynakları	Eğitim ve geliştirme, çalışan memnuniyeti	Çalışanların kalitesi önemli, ancak çok vurgu yapılmaz
Odak Noktası	Ürün, Süreç (Process), Örgütsel Yapı, İnsan Kaynakları, Müşteri Odaklılık	Ürün ve Hizmet Kalitesi
Örgütsel Yapı Kalitesi	Örgütsel yapı ve örgüt kültürü, liderlik	Örgütsel yapıya vurgu daha az
Sürekli İyileştirme	Sürekli iyileştirme ve inovasyon	Sürekli iyileştirme, genellikle Deming döngüsü (PDCA)
Müşteri Odaklılık	Müşteri beklentileri ve gereksinimlerinin anlaşılması	Müşteri tatmini önemsenir ancak daha az vurgulanır
Standartlar ve Uyumluluk	Standartlara uyum ve dış denetimler	Standartlara uyum, ISO gibi uluslararası standartlar

David Garvin'in sekiz kalite boyutu, kalite yönetimi ve kontrolüne ilişkin geleneksel görüşlerden önemli ölçüde ayrışarak farklı bir çerçeve ortaya koymuştur. Bu ayrışmalar sonucu ortaya konan anlayış, daha kapsamlı ve çok boyutlu, müşteri odaklı ve stratejik bir kalite anlayışını benimseyerek ürün ve hizmet kalitesinde sürekli iyileştirmelere odaklanan bir vizyon geliştirmiştir. Garvin'in ortaya koyduğu çok boyutlu çerçeve, kalite yönetimine daha bütünsel, uyarlanabilir ve esnek bir yaklaşımın kazandırılmasında önemli bir rol oynamıştır.

Bundan sonraki bölümde kalite boyutların her biri üzerinde detaylı tanımlar yapılacaktır.

1.2.1. Performans

Performans; bir mamulün temel işlev özellikleridir (Çavdar & Zerdali, 2020, s. 223). Bir ürünün temel işlevsel özellikleri anlamına gelen performans, örneğin bir otomobil için hız, konfor; bir televizyon için renk, ses, görüntü vb. özellikler olabilmektedir (Garvin, 1987b, s. 104) Hizmet işletmelerinde ise performans, servis hızı ve bekleme zamanının azlığı ile ölçülebilir (Türkmen, 2017, s. 7). Nihai tahlilde, ürün kalitesinin performans

boyutu, ürünün belirlenen işlevselliklerini karşılama derecesi ile ilgilidir. Bir ürünün performansının değerlendirilmesi tipik olarak ürünün spesifikasyonlara uygunluğunun ve kullanıcı memnuniyetinin değerlendirilmesini içerir. Örneğin bir aracın performansı hızlanma, yakıt tüketimi ve yol tutuş özellikleri gibi ölçümlerle ölçülebilir. Performans, müşterilerin satın alma kararlarında kritik bir rol oynar ve çoğu zaman ürünün diğer kalite boyutları üzerinde de etkili olabilir.

Ürün kalite boyutlarının her biri kavram ve bağlam olarak birbiriyle doğrudan bağlantılı olduğu için aralarında çok kesin bir ayırmadan söz edilemez. Ancak her bir boyutun odaklandığı farklı noktalar bulunmaktadır. Farklı ürün grupları için konuyu örneklemek gerekirse, bir elektrikli süpürge için, ürünün emiş gücü ve farklı yüzeyler için gösterdiği performansın nasıl olduğu, enerji verimliliği, filtrasyon sisteminin ne derece etkin çalıştığı, bir televizyonun görüntü kalitesi ve ses sisteminin performansı ve kalitesi, bir otomobil için yakıt tüketimi ve motor performansı, frenleme mekanizmasının etkililiği, farklı yol ve hava koşullarındaki sürüş deneyimi, bir bilgisayar için işlemci hızı ve çoklu işlemleri yapabilme yeteneği, uzun süreli kullanımlarda batarya ömrü, bir otomobil lastiği için ise yol tutuşu, frenleme mesafesi, farklı iklim koşullarında gösterdiği yol tutuş performansı, aşınma ömrü gibi sorular performans boyutunda sorulabilecek sorulardır. Görülebileceği üzere ürün performans boyutundaki özellikleri genellikle benzer ürünlerin performans açısından objektif bir şekilde sıralanmasına olanak tanıyan ölçülebilir metrikleri içermektedir.

Ürünün kalite boyutunun performans bileşeni kalite boyutunun diğer bileşenleri üzerinde de önemli bir etki derecesine sahiptir. Örneğin, performansın iyileştirilmesi, genellikle ürünün dayanıklılığını da artırır, çünkü kaliteli bileşenler ve tasarım çözümleri kullanılır. Performans, ürünün istenen işlevleri ne kadar iyi yerine getirdiğiyle ilgilidir; performans yetersizse ürün başarısız olur (Norman, 2004, s. 37). Performansın optimize edilmesi, servis gereksinimlerini azaltır ve müşteri memnuniyetini yükseltir, çünkü daha az bakım ve onarım ihtiyacı doğar. Performans, ürünü dizayn edenler tarafından belirlenen ve üretilme amacına ilişkin temel fonksiyonunu yerine getirmesi ile ilgili özellikleri ifade etmektedir (Özveri & Kabak, 2018, s. 146).

Sonuç olarak, performans boyutu, ürün kalitesinin genel değerlendirilmesinde başat bir rol oynamakta ve kalitenin diğer boyutlarıyla dinamik etkileşimler kurarak ürünün genel

kalitesine önemli ölçüde etki etmektedir. Bu sebepten dolayı performans boyutunu iyileştirmenin, genel ürün kalitesini artırmakla beraber müşteri memnuniyetini ve marka sadakatini de artıracığı söylenebilir.

1.2.2. Güvenilirlik

Ürün kalitesinin boyutlarının önemli bileşenlerinden olan güvenilirlik boyutu, ürünün belirli koşullar altında ve belirli bir süre boyunca kendisinden beklenen işlevlerini yerine getirebilme yeteneğini ifade eder. Tüketicilerin ürünü uzun süre sorunsuz kullanabilmelerini ifade eden bir metrik olarak değerlendirilebilir. Bir diğer ifadeyle, güvenilirlik bir ürünün kullanılmaya başlandığında belirli bir süre içinde arızalanmama olasılığıdır (Chen vd., 2018, s. 23). Ürünlerin güvenilir olması, kullanıcıların arıza veya bozulma endişesi taşımadan ürünlerini kullanmalarını sağlar. Bununla birlikte, güvenilirlik, müşteri memnuniyeti ve marka sadakati açısından önemli bir etkidir. Müşteri memnuniyeti, müşterilerin bir ürün veya hizmetten beklentilerinin karşılanması veya aşılması durumunda ortaya çıkan duygusal tatmin olarak tanımlanır (Kotler, 1997, s. 467). Yüksek güvenilirliğe sahip ürünler, müşteri beklentilerini karşılayarak ve hatta aşarak marka sadakatini artırmaktadır.

Güvenilirlik ve dayanıklılık arasında çok kuvvetli ve doğrudan bir etkileşim vardır. Dayanıklı ürünler, daha güvenilir olarak kabul edilir çünkü bu ürünler uzun süreli kullanıma ve zorlu koşullara dayanabilir. Bu durum, ürünlerin daha uzun ömürlü olmasını sağlayarak müşterilerin ürünlere daha fazla güvenmesine neden olur. Güvenilirlik aynı zamanda ürünlerin bakım ve onarım gereksinimlerini de azaltır. Daha az bakım ve onarım gerektiren ürünler, kullanıcıların zaman ve maliyet tasarrufu yapmasına olanak tanır. Aynı zamanda güvenilirlik, ürünün uzun süre ve bilinen koşullar altında olması gerektiği gibi çalışmaya devam edebilmesi anlamına gelir. Bu durum kalite boyutlarının performans bileşenini doğrudan etkileyen bir faktördür.

Güvenilirlik bir diğer kalite boyutu olan özellikler ile de kuvvetli bir biçimde ilgilidir. Bu özellikler müşterilerin gereksinimleriyle uyumlu olmalı ve bu gereksinimleri karşılamada tutarlı bir güvenilirlik sergilemelidir. Bu bağlamda güvenilirlik müşteri gereksinimlerini karşılayabilmenin bir göstergesi olarak görülebilir. Güvenilir özelliklere sahip ürünler, müşteri memnuniyetini artırarak ürünün genel algılanan kalitesini yükseltir. Diğer yandan, algılanan kalite boyutu, müşteriler tarafından bir ürün veya hizmet hakkında

yapılan değerlendirmedir ve güvenilirlik bu değerlendirmenin şekillenmesinde önemli bir rol oynar. Güvenilirlik derecesi yüksek olarak görülen ürünler genellikle algılanan kalite de yine düzeyi yüksek olan ürünlerdir, çünkü müşteriler bu ürünlerin uzun süreli olarak beklentilerini karşılayacağına inanırlar. Bunun bir sonucu olarak güvenilirlik ve algılanan kalite arasındaki pozitif korelasyon, müşteri sadakatini ve tekrar satın alma niyetini önemli ölçüde pozitif etkilemektedir.

Farklı ürün grupları özelinde konuyu özetlemek gerekirse, bir elektrikli süpürge için arızalanma sıklığı ve motorunun ömrü, parça kalitesinin güvenilirliği, televizyonun ekran ömrü ve piksel hatalarına karşı garanti süresi, uzun vadeli kullanımında görüntü ve ses kalitesindeki değişim, otomobilin motor ve şanzıman arızalanma oranı, bulaşık makinesinin su sızdırmaya karşı dayanıklılığı, makinenin programları ve elektronik bileşenlerinin uzun vadeli güvenilirliği, bilgisayarın donanım bileşenlerinin arızalanma sıklığı ve kullanım ömrü, üretici firmanın sağladığı desteğin süresi ve kalitesi gibi hususlar güvenilirlik boyutu kapsamında sorulabilecek sorulardır.

1.2.3. Dayanıklılık

Dayanıklılık, bir ürünün kendisinden beklenen faydayı uzun bir süre boyunca sürdürebilme kapasitesiyle ilgilidir ve ürün kalitesinin temel boyutlarından biridir. Bu boyut, ürünün fiziksel sağlamlığını, sık kullanıma dayanma yeteneğini ve çeşitli çevresel faktörlere karşı direncini kapsar. Algılanan kalite ile dayanıklılık müşteri memnuniyeti üzerinde en yüksek etkiye sahiptir ve dayanıklılık ile müşteri memnuniyeti arasında direkt bir ilişki vardır (Chen vd., 2018, s. 27).

Ürünlerin dayanıklılığı, üretiminde kullanılan malzemelerin kalitesi, üretim yöntemleri ve tasarım özellikleri gibi unsurlarla bağlantılıdır. Yüksek kaliteli malzemeler tipik olarak daha dayanıklıdır ve daha fazla mukavemete sahiptir, bu da daha uzun süre dayanacakları ve uzun vadede daha güvenilir olacakları anlamına gelir (Yip, 2022).

Bir ürünün dayanıklılığı ya da diğer bir tabirle uzun ömürlülüğü müşteri memnuniyeti üzerinde önemli ölçüde etkilidir. Dayanıklılığı yüksek ürünler, ürünün değiştirme veya onarım ihtiyacını azaltarak genel sahip olma maliyetinin artmasına engel olur. Ayrıca dayanıklılık, ürün güvenilirliğini artırır, tutarlı performans sağlar ve müşteri sadakatini artırır. Dayanıklılık çevresel sürdürülebilirlik açısından da kalite yönetiminin kritik bileşenlerinden biridir. Uzun ömürlü ya da dayanıklı ürünler atık oluşumunu azaltarak

çevreye verilen zararın azaltılmasında önemli bir rol oynamaktadırlar. Bunun sonucunda hem üreticiler hem de müşteriler için sürdürülebilirlik hedeflerine ulaşmak mümkün olabilir. Dayanıklı ürünler, hammadde tedariki ve üretim süreçlerine olan ihtiyacı azaltarak kaynakların korunmasına katkıda bulunur. Dayanıklılığı daha yüksek olan ürünlerin kusurlar veya erken arıza nedeniyle iade edilme olasılığı daha düşüktür, bu da iadelerin işlenmesi, envanterin yeniden stoklanması ve müşteri memnuniyetsizliğinin yönetilmesiyle ilgili maliyetleri azaltır. İşletmeler, dayanıklı ürünler sunarak müşteri memnuniyetini ve sadakatini artırırken, iade ve değişimlerden kaynaklanan mali yükü de en aza indirebilir (Wainwright, 2024).

Dayanıklılık boyutunu endüstriyel olarak farklı bakış açılarında ifade etmek mümkündür. Teknoloji eksenli bir bakış açısıyla dayanıklılık bir ürünün teknolojik eskimeye karşı dayanıklılığı, onun güncellemelere veya yükseltmelere açıklığı ve gelişen teknolojilere uyarlanabilirliğiyle ilintilidir. Aynı zamanda dayanıklılık teknolojik olarak teknolojideki yeniliklere uyum sağlayabilme yeteneğini ve donanımsal ve yazılımsal olarak yeni sistem ve standartlarla entegre edilebilme kabiliyetini ifade eder. Finansal açıdan dayanıklılık, ilk satın alma maliyeti, devam eden bakım ve onarım giderleri ve enerji kullanımı gibi kullanım ömrü boyunca oluşan maliyet unsurlarıyla beraber, ürünün belirli bir kullanım süresi sonrasında ikinci el piyasalardaki değerini koruma potansiyelini de ifade eder.

Dayanıklılık boyutu, sosyal ve kültürel açıdan ise modadaki veya toplumsal tercihlerdeki dalgalanmalara bakılmaksızın ürünün estetik çekiciliğini uzun süre sürdürme kapasitesi olarak ifade edilebileceği gibi, ürün birden fazla nesil tarafından kullanılabildiği için yaş, yetenek veya kullanıcı çeşitliliğini göz ardı etmeden geniş bir kullanıcı kitlesine hitap edebilmesi olarak da ifade edilebilir.

Operasyonel anlamda dayanıklılık boyutu, makine ve ekipmanların yoğun iş süreçleri içerisinde arıza ve duruşlara karşı mukavemetini ifade ettiğinden dolayı etkin kapasite değeri üzerinde direkt etkilidir.

Bilgi ve eğitim ihtiyacı bakımından dayanıklılık, ürünle ilgili edinilen bilgi veya uzmanlığın, gelişen teknolojilere ve metodolojilere karşın konuyla ilgili ve kalıcı kalma derecesi veya değişen koşullar karşısında uzun vadeli etkinliği sürdürme kapasitesi olarak tanımlanabilir.

Hukuksal bağlamda dayanıklılık, ürün ve hizmetlerin mevcut ve gelecekteki muhtemel düzenleyici gerekliliklere uygunluğu sürdürülebilirlik süresi olarak değerlendirilebilirken, etik kavramı üzerinden bakıldığında, etik normlar ve toplumsal beklentilerle ilişkili olarak bir ürün veya hizmetin uzun vadeli sürdürülebilirliğini ve toplumsal kabulünü ifade eder.

Dayanıklılık boyutunu farklı ürün grupları özelinde ifade etmek gerekirse, bir elektrik süpürgesinin motor ömrünü ve emiş gücünü ne kadar süre koruyabileceği, toz haznesi ve filtrelerin dayanıklılığının ne kadar olduğu, televizyon ekranının fiziksel darbelere karşı dayanıklılığı, otomobilin kaporta, şasi ve gövdesinin korozyon ve fiziksel darbelere karşı mukavemeti, bulaşık makinesinin sepetinin uzun zamanlı kullanımda ne ölçüde eskime yapacağı, paslanmaya karşı dayanıklılığı, bilgisayarın dış darbelere karşı fiziksel dayanıklılığı, soğutma sisteminin uzun vadeli kullanımlarda performansını koruyabilme kabiliyeti, batarya ömrünün uzunluğu, otomobil lastiğinin yanaklarının kesilme ve darbelere karşı mukavemeti, sıcaklık değişimlerine verdiği tepki gibi hususlar dayanıklılık boyutu kapsamında değerlendirilebilir.

1.2.4. Servis

Ürün kalitesi bileşeni olarak, ürüne eşlik eden hizmetlerin çeşitliliği ve kalitesini kapsayan "servis" kavramı büyük önem taşımaktadır. Servis kavramı, müşterilere ürün kullanımları boyunca sunulan destek, bakım, onarım ve diğer tamamlayıcı hizmetleri içerir. Ürünle ilgili müşteri destek hizmetleri, teknik yardım, garanti hizmetleri, satış sonrası destek ve müşteri eğitimi gibi unsurlar bu kapsamda değerlendirilir. Servis kolaylığı, tüketicinin onarım hizmeti alma kolaylığını, servis personelinin duyarlılığını ve hizmetin güvenilirliğini içerir (Brucks vd., 2000, s. 361). Aynı zamanda, sunulan servisin kolaylığını, yanıt verebilirliğini ve güvenilirliğini ifade eder. Müşteri tarafından kolayca erişilebilen ve güler yüzlü, yetkin çalışanların çalıştığı bir servis merkezi bu boyutta yüksek oranda değerlendirilecektir (Chowdhury & Menon, 1995).

Bir anlamda üretici tarafından sunulan ürünle ilgili bakım onarım faaliyetlerinin müşteri beklentileriyle ne derece örtüştüğü sorusunun cevabıdır. Servis boyutu, doğası gereği kaliteyle doğrudan bağlantılı olan bir dizi niteliği kapsamaktadır. Bu nitelikler erişilebilirlik, yanıt verme yeteneği, güvenilirlik, yeterlilik ve empatiyi olarak sayılabilir. Erişilebilirlik, müşterilerin ürün ve hizmetlerden beklenen faydayı elde edebilme kolaylığıyla ilgiliyken yanıt verebilirlik, müşteri taleplerine verilen geri bildirimlerin hız

ve etkisini ifade eder. Güvenilirlik, ürün ve hizmetlerin tutarlı ve güvenilir bir şekilde ulaştırılması veya sağlanmasıyla ilgiliyken, yeterlilik, ürün ve hizmet sunumu için gerekli bilgi, beceri ve donanıma sahip olmanın bir ifadesidir. Empati ise, müşteriye özel kişiselleştirilmiş etkileşimler yoluyla müşterilerin bireysel ihtiyaçlarını anlama ve cevap verebilme yeteneğini içerir.

Farklı bakış açılarından servis boyutuna değinmek gerekirse, müşteri bakış açısından servis boyutu, satın alma sürecinin öncesinde, süreç içerisinde ve sonrasında müşteriye sunulan destek ve yardım düzeyini temsil eder. Bu husus müşteri memnuniyetinin sağlanması ve müşteri sadakatının geliştirilmesi açısından önemlidir. Operasyonel açıdan bakıldığında servis boyutu, işletmelerin müşteri taleplerini karşılamak ve memnuniyet düzeylerini artırmak için yapılan tüm eylemleri kapsar. Bu durum servisin hem etkili hem de verimli bir şekilde sunulmasını gerektirmektedir. Servis boyutu stratejik açıdan ise işletmeler için rakiplerine göre bir farklılaşma ve rekabet avantajı oluşturma aracı gibi düşünülebilir. Servis boyutunda müşteriye sunulan güvenlik algısı sunulan ürün ve hizmetlerin tutarlı ve güvenli bir şekilde müşteriye ulaştırılmasını sağlar ve müşteri nezdindeki marka itibarı ve algılanan kalite değerlerini artırır. Müşteriler için ürün değerini artıran ek unsurlar olarak görülen ve ürünün yanında sağlanan tamamlayıcı hizmetler veya imkanlar olarak sayılabilecek ücretsiz hizmetler, müşteri sadakati programları veya müşteriye sunulan özel hizmetler de servis boyutunu kapsamında değerlendirilir.

Servis kalitesi, müşteri memnuniyetini ve güvenini şekillendirmede doğrudan etkili bir unsurdur. Özellikle müşteri sadakatının geliştirilmesi, yeniden satın alma oranlarının artırılması önemli ölçüde olumlu etkiler sunar. Servis kalitesi ile müşteri davranışı arasındaki bu simbiyotik ilişki, yeni müşteri edinim maliyetini azaltır ve zaman içinde müşteriye elde tutma oranlarını artırarak işletmelere önemli ölçüde fayda sağlayabilir. Servis kalitesinin yüksek düzeyde sağlanması, işletmenin rekabetçi konumunu artırır ve rekabet ortamında rakiplerinden öne çıkmasını sağlar. Bu nedenle kalite yönetim sistemleri kapsamında servis kalitesi, sıkı ve periyodik kontroller ile sürekli iyileştirmeler yoluyla yönetilmektedir.

Buradan hareketle işletmelerin servis kalitesinin geliştirilmesi süreçlerini stratejik planlama ve operasyonel çalışmalarına dahil etmeleri zorunludur. Şirketler, servis

mükemmelliğini önceliklendirerek müşteri ilişkilerini güçlendirebilir ve ürün kalitesi, performans ve estetik gibi kalitenin diğer boyutlarını tamamlayabilir.

Farklı ürün grupları özelinde kalitenin servis boyutunu irdelemek gerekirse, elektrikli süpürge'nin garanti süresi ve garanti kapsamında sunulan hizmetler, süpürge'nin arızalanması durumunda yetkili servis ağı'nın büyüklüğü ve ulaşılabilirliği, televizyonun tamir süresi, yazılım güncelleme hizmetlerinin sağlığı ve verilen teknik destek hizmetinin kalitesi, otomobil için periyodik bakım maliyetleri ve parça bulunabilirliği, garanti süresi ve kapsamı, otomobil lastiğinin hasar görmesi durumunda üreticinin sunduğu tamir ve değiştirme imkanları, yedek lastik temini vb. gibi konular kalitenin servis boyutu kapsamında analiz edilebilecek konulardır.

1.2.5. Estetik

Kalitenin estetik boyutu, ürünün görsel çekiciliği, tasarım öğeleri, renk şeması ve genel sanatsal değeri ile ilgilidir. Bu boyut, müşterinin öznel tercihlerini, duyuşsal algılarını ve ürüne yönelik duyuşsal tepkilerini kapsar. Estetik hususlar, salt işlevselliğin ötesinde ürüne zenginlik ve keyif katarak kullanıcı deneyimini geliştirebilir. Estetik boyutu, ürünün tüketiciler üzerinde bıraktığı ilk izlenim ve duyuşsal etki ile yakından ilişkilidir. Ürünlerin estetik özellikleri, ürünler duyuşsal çevrenin bir parçası haline geldikçe hem kullanıcılar hem de kullanıcı olmayanlar üzerinde yıllarca etki yaratabilir. Estetik tepkiler, ürünün performansından veya işlevsel özelliklerinden ziyade tasarımından ve duyuşsal özelliklerinden kaynaklanmaktadır. (Bloch, 1995, s. 20).

Bir ürüne ait kalitenin estetik boyutu kapsamında değerlendirilebilecek görsel çekicilik, tasarım, renk ve şekil, malzeme seçimi, yüzey ve doku yapısı gibi olgulardan bahsetmek mümkündür. Görsel çekicilik bir ürünün estetik çekiciliği, olumlu bir ilk izlenim oluşturma ve potansiyel tüketicilerin ilgisini çekmede önemli bir rol oynaması ile ilgilidir. Tasarım, ürün tasarımının estetiği ve tüketicilerin ürünle nasıl etkileşim kurup nasıl algıladıkları ile bağlantılıdır. Öte yandan renk ve şekillerin uyumu, müşterinin ürün kalitesi algısının şekillenmesinde önemli bir rol oynamaktadır.

Yüksek kaliteli ve görsel olarak çekici malzemelerin kullanılması, bir bütün olarak ürünün algılanan kalitesini artırma potansiyeline sahiptir. Ek olarak, tüketicilerin ürün yüzeyi ile etkileşime girdiklerinde yaşadıkları dokunma hissi, onların estetik algısına katkıda bulunabilmektedir.

İşletmeler, ürünlerini pazarda farklılaştırmak ve hedef müşteri kitlesine hitap etmek için estetik tasarımdan yararlanmaktadır. Görsel açıdan çekici ve zengin bir tasarım, markaların rekabet ortamında öne çıkmasına yardımcı olmaktadır. Bunun yansısı, kaliteli ve çekici bir tasarım tüketicilerin markaya sadakatini artıracığından dolayı bir markanın algılanan değeri ve imajının, estetik boyutuyla çok yakından ilişkili olduğunu söylemek mümkündür. Estetik açıdan üst düzey tasarımlar, ürünün daha yüksek kalitede olduğu izlenimini oluşturarak, müşterinin ürünle ilgili kalite algısını yükseltir. Stratejik bir araç olarak ürün estetiği, marka imajını ve sembolik anlamı güçlendirerek tüketici inançlarını etkileyebilir (Jiang vd., 2023, s. 3). Bu durum, ürünün güvenilirlik algısını ve müşteri güvenini artırır. Ürünün Pazar rekabet koşulları altında farklılaşması adına estetik boyutu tüketici ilgisini çekmede ve ürünü rakiplerinden ayırmada önemli bir rol oynamaktadır. Görsel olarak çekici ve ayırt edici bir tasarım, bir markanın veya ürünün müşteriler tarafından algılanma derecesini yükseltir. Aynı zamanda estetik boyut açısından gelişmiş tasarımlar, müşteride ürünün en son teknoloji ve trendlere uyumlu olduğu izlenimini uyandıracığından kullanıcı deneyimi ve marka itibarına olumlu katkı sağlar.

Buradan hareketle, estetik boyutun dikkate alınması, tüketici taleplerinin karşılanmasında ve memnuniyetlerinin garanti edilmesinde; güvenilirlik, kullanıcı dostu olma ve işlevsellik gibi ürün mükemmelliğinin diğer yönleriyle birleştirildiğinde çok önemli bir rol oynar. Bir ürünün estetik boyutu sadece ürünün kendisini etkileyen bir olgu değil, aynı zamanda ürünü sunan markanın genel imajını da etkileyerek üstün kalite algısını artırır.

Yine farklı ürün grupları özelinde kalitenin estetik boyutunun irdelenmesi istenirse, örneğin, elektrikli süpürge tasarımının ev dekorasyonuna uyumlu olup olmaması, kullanılan renk seçeneklerinin müşteri beklentileriyle ne derece örtüştüğü, tasarımının ergonomik gereksinimleri ne derece karşıladığı, televizyonun ekran ve çerçeve tasarımındaki uyum, kumandasının ve ara yüzünün ne derece estetik algılandığı ve kullanıcı dostu olup olmaması, otomobilin aerodinamik tasarımının etkinliği, iç mekan ve kokpit tasarımının konfor ve estetik açıdan başarısı, renk ve döşeme desen seçeneklerinin kişiselleştirilebilir olup olmaması, bulaşık makinesinin genel mutfak dekorasyonu ile uyumu, kontrol panelinin estetik ve kullanıcı dostu olması, bilgisayarın ekran klavye düzeni ve ergonomisi, otomobil lastiğinin yanak tasarımı ve desenlerin estetik açıdan nasıl algılandığı, aracın genel imajına katmış olduğu değer gibi hususlar analiz edilebilir.

1.2.6. Özellikler

Kalitenin özellikler boyutu, belirli gereksinimleri karşılaması amaçlanan bir ürünün benzersiz işlevsel yönlerine, niteliklerine ve yeteneklerine odaklanarak ürün kalitesini değerlendirmede önemli bir rol oynar. Bu boyut, ürünün teknik özelliklerini ve bunun müşteri tercihleri ve beklentileriyle uyumunu açıklar. Ürün özelliklerinin doğru tanımlanması ve etkili yönetimi, bir ürünün performansını, uzun ömürlülüğünü, görsel çekiciliğini ve güvenilirliğini etkilemede çok önemli bir yer tutar. Özelliklerin yetersiz veya kusurlu olması ürün kalitesinden ödün verme riskini doğurur ve bu da potansiyel olarak müşteri memnuniyetinin azalmasına neden olabilir. Dolayısıyla başarılı bir kalite yönetimi yaklaşımının temel bileşeni olarak özelliklerin ürün kalitesi üzerindeki etkisinin anlaşılması ve doğru şekilde ele alınması zorunludur.

Özellikler boyutu, araştırma ve geliştirme (Ar-Ge) süreciyle yakından ilgilidir. Araştırma ve geliştirme süreci, yeni ürünler geliştirmek veya mevcut ürünleri iyileştirmek için ihtiyaç duyulan yenilikçi ve teknolojik çözümleri sağlar. Ar-Ge süreçleri bir ürünün sahip olacağı özelliklerin belirlenmesi, geliştirilmesi ve test edilmesi fazlarından oluşur. Ar-Ge süreci teknik performansı, işlevselliği ve ürün uygunluğunu müşteri gereksinimlerine göre optimize ettiğinden Ar-Ge süreçlerinin etkinliği özellikler boyutu ile doğrudan ve kuvvetli olarak ilgilidir. Yeni ürünleri piyasaya sürerken, ürün özelliklerini müşteri ihtiyaçları ve tercihleriyle doğru bir şekilde tanımlamak ve uyumlu hale getirmek çok önemlidir. Araştırma ve geliştirme (Ar-Ge) ekipleri, ürün özelliklerinin hem ileri hem de rekabetçi olmasını sağlamak için pazar analizine, teknolojik değerlendirmeye ve kullanıcı girdilerine güvenir. Bu yinelenen süreç, ürün özelliklerinin müşterilere maksimum değer sağlayacak şekilde optimize edilmesini sağlar.

Cooper ve Kleinschmidt'in çalışması (1995), yeni ürün geliştirme çabalarının başarılı bir şekilde yürütülmesinde araştırma ve geliştirmenin (Ar-Ge) önemini vurgulamaktadır. Kanıtlar, Ar-Ge'nin, ürünlerin müşteri beklentilerini aşan ayırt edici özelliklere sahip olmasını sağlamada çok önemli bir rol oynadığını göstermektedir. Sonuç olarak, Ar-Ge operasyonlarının yetkin yönetimi, ürün özelliklerinin kalibresinin belirlenmesinde ve kuruluşların rekabet avantajı elde etmesini kolaylaştırmada çok önemlidir.

Kalitenin özellikler boyutu, bir ürünün tüketici talep ve tercihlerini karşılamasını sağlayan farklı özellikleriyle ilgili olduğundan bu boyutun odağı, analizi yapılan ürünün türüne

göre deđiřecektir. Örneđin bir televizyon için özellikler boyutu kapsamında deđerlendirilebilecek spesifikasyonlar, ekran kalitesi, çözünürlük, kontrast oranı, ses kalitesi, akıllı TV platformları, HDMI, USB, bluetooth bağlantı seçenekleri, enerji verimliliđi, gibi spesifikasyonlar olabilecekken, çamařır makinesi için yıkama kapasitesi, program sayısı ve çeřitleri, gürültü seviyesi, devir hızı, tambur ömrü, buharlı yıkama özelliđi, çocuk kilidi, bir tencere için ise malzeme kalitesi, sap ve kapak ergonomisi, bulařık makinesinde yıkanabilme ve temizlik kolaylıđı, kaplama malzemenin niteliđi gibi spesifikasyonlar özellik boyutu kapsamında deđerlendirilebilir.

1.2.7. Algılanan Kalite

Algılanan kalite, bir ürün veya hizmetin tüketiciler tarafından öznel olarak deđerlendirilmesi ve algılanması anlamına gelir. Bu deđerlendirme, ürünün özelliklerine, işlevselliđine, performansına, estetik unsurlarına, güvenilirliđine ve diđer çeřitli faktörlere dayanmaktadır. Diđer bir deyiřle algılanan kalite tüketicinin ürünle ilgili algısının, geçmiş etkileřimlerinin, beklentilerinin ve ürüne iliřkin genel izlenimlerinin bileřiminden oluşur. Bu nedenle, algılanan kalite, objektif kalite ölçütlerinden farklı olarak daha çok subjektif ve kiřisel bir deđerlendirmedir (V. A. Zeithaml, 1988a, s. 3). Tüketicilerin bir ürüne olumlu bakıř açısı, o ürünün genel kalitesinin önemli bir göstergesidir. Algılanan kalitenin önemi, tüketici satın alma davranıřı üzerindeki etkisinde yatmaktadır; çünkü olumlu algılar, yeniden satın alma niyetlerinin artmasına neden olur ve işletmenin itibarını artırır.

Algılanan kalite, diđer kalite boyutlarıyla karmařık bir řekilde iç içe geçmiş ve dinamik iliřki içerisinde. Bu iliřki, müşterinin çeřitli kalite boyutlarına iliřkin algısı ve deneyiminden etkilenir. Bunun dışında algılanan kalite boyutunu etkileyen başka faktörlerden de bahsetmek mümkündür. Marka imajı, tüketicilerin bir markayı algılama ve ona deđer verme biçimlerini ifade eder. Bir markanın müşterilerin algısı ile oluşan kalitesi veya üstünlüđünün tamamına algılanan kalite denmektedir (Aaker, 1991b; Keller, 1998; Yasin vd., 2007). Algılanan kaliteyi oluřturan řey, markanın gerçek kalitesi deđil, müşterinin markaya iliřkin öznel deđerlendirmesidir (Zeithaml, 1988, s. 5). Algılanan kalite, çeřitli řekillerde deđer üretme yeteneđine sahiptir. Algılanan kalite, müşterileri belirli bir markayı satın almaya teřvik eder.

Bunun yanında tüketiciler, reklamlar yoluyla sağlanan bilgilerle bir ürün hakkında bilinçli kararlar verirler; bu, algılarını etkileyebilir ve sonuçta ürünün kalitesine olan algılarını değiştirebilir. Dolayısıyla reklam ve pazarlama stratejileri algılanan kalite üzerinde doğrudan etkilidir. Algılanan kalite üzerinde etkili olgulardan bir diğeri fiyattır ve genellikle daha yüksek fiyatlı ürünlerin daha kaliteli olduğu düşüncesi yaygındır. Tüketici algısı diğer tüketicilerin yorumlarından, inceleme ve analizlerinden de etkilenebilir. Olumlu yorumlar, algılanan kaliteyi artırırken, olumsuz yorumlar negatif yönde etki etmektedir. Algılanan kaliteye etki eden faktörlerden biri de tüketicilere ürünü deneyimleme imkanı sunmaktır. Tüketicilere ürünleri deneyimleme fırsatı sunmak, kalite algılarını olumlu yönde etkiler. Ürün tanıtımları, ürün numunelerini, denemeleri, iade politikalarını, çevrimiçi inceleme platformlarını veya bir firmanın tüketicilerin yeni bir ürünün değerini öğrenmesine olanak tanıyan diğer araçları içerebilir (Boleslavsky vd., 2016, s. 1). Diğer taraftan ürünün satışa sunulduğu mağaza veya ofis ortamı algılanan kaliteyi kuvvetli bir şekilde ve doğrudan etkileyen bir başka faktördür. Tüketiciler çoğunlukla prestijli ve iyi organize edilmiş mağazalarda satılan ürünleri daha kaliteli olarak algılama eğilimindedirler. Ürünün ambalajı, tüketicilerin algısını büyük ölçüde etkileyebilir. Underwood (2003) ürün ambalajının, markayı ve öz kimliği etkileyebilecek, aynı zamanda tüketicilerle markalar arasındaki ilişkiyi güçlendirebilecek, ürünlerle ilgili önemli bir özellik olduğuna dikkat çekmektedir.

Algılanan kalite üzerinde etkin bir diğer faktör sosyal medya ve çevrimiçi platformlardır. Sosyal medya ve çevrimiçi platformlar tüketicilerin algılarını hızla etkileme yeteneğine sahiptir. Bir ürüne ilişkin olumlu ya da olumsuz içeriğin varlığı, tüketicilerin ürünle ilgili algıları üzerinde manipülatif bir etkiye sahiptir. Ödüller, sertifikalar ve bağımsız test sonuçları gibi üçüncü taraf onayları da algılanan kalite üzerinde etkin başka bir etmendir. Bunun yanında özellikle gıda ve sağlık ürünleri gibi özellikli ürünlerde ürün güvenliği algılanan kalite üzerinde etkili kilit faktörlerden biridir.

Örnek olması açısından, yine farklı ürün grupları özelinde algılanan kalite boyutu irdelenmek istenirse; elektrikli süpürge için, ürünle ilgili genel müşteri memnuniyet derecesi, ürünün marka itibarı, malzeme kalitesi ve tasarımı hakkındaki genel değerlendirmeler, televizyon için, görüntü kalitesi ve ses için genel müşteri deneyimi ve algısı, uzun süreli kullanımlardaki dayanıklılığı hakkında müşteri düşünceleri, otomobil için, sürüş deneyimi, konforu ve güvenlik donanımları hakkındaki genel müşteri algısı,

marka ve modelin sürücüler arasındaki genel prestij ve itibarı, bulaşık makinesi için, yıkama performansı ve ses seviyesi hakkındaki kullanıcı değerlendirmeleri, uzun süreçli dayanıklılıkla ilgili oluşan genel düşünce, müşteri memnuniyeti açısından markanın sahip olduğu ün, otomobil lastiği için ise, yol tutuş ve frenleme performansı ile ilgili müşteri geri bildirimleri, markanın güvenilirlik anlamında sürücüler arasında bırakmış olduğu izlenim gibi hususlar analiz edilebilir.

1.2.8. Uygunluk

Kalite değerlendirmesinde uygunluk boyutu, bir ürünün belirlenen gereksinimlere veya standartlara ne ölçüde uygunluk gösterdiğiyle ilgilidir. Bir ürünün belirlenen özelliklerinin ve performans kriterlerinin ne kadar karşılandığını ifade eder. Uygunluk boyutu kapsamında bir ürün veya hizmetin tasarım ve üretim aşamalarında belirlenen spesifikasyonlara uygunluğu değerlendirilir. Bu yönüyle, ürün veya hizmet kalitesine ilişkin genel algıya katkıda bulunduğu için müşteri beklentileri, güvenilirlik ve dayanıklılık gibi diğer kalite unsurlarıyla da bağlantılıdır. Ayrıca, müşteriler satın alma tercihlerinin beklentilerine uygun olup olmadığını değerlendirdikleri için uygunluk boyutu genel müşteri memnuniyetini doğrudan etkiler.

Ürünün yaşam eğrisi boyunca kendisinden beklenen performansı tutarlı bir şekilde devam ettirmesi önemli uygunluk göstergelerinden birisidir. Uygunluk boyutunu doğrudan etkileyen hususlardan bir diğeri de üretim süreçlerindeki kalite kontrol noktaları ve değişkenlerinin doğruluğudur. Üretim sürecindeki kalite yönetim politikalarının yetkinliği ve iyi uygulamalar uygunluk boyutu üzerinde son derece pozitif etkiye sahiptir. Müşteri taleplerinin ve beklentilerinin doğru anlaşılması ve ürünlerin bu beklentilerle örtüşecek spesifikasyonlarda üretilmesi de uygunluk boyutundaki pozitif etmenlerden biridir. Bu faktörlerin dışında, uygunluk boyutunun odak noktası, ürünün kendisi için belirlenen teknik ve endüstriyel standartlara uyumluluğudur. Bu uyum ve uygunluk ürünün aynı zamanda başta güvenilirlik olmak üzere diğer kalite boyutlarıyla da yakından ilgilidir.

Kalitenin uygunluk boyutunun müşteri memnuniyeti üzerinde önemli etkileri bulunmaktadır. Kendisi için belirlenen standart ve özelliklere uyumlu olarak üretilen bir ürün müşteride daha fazla güven duygusunun gelişmesini sağlayarak, marka aidiyetini ve

sadakatine katkı sağlar. Bu standart ve spesifikasyonlar çoğu zaman müşteri beklenti ve taleplerini için belirleyici ve yol gösterici olurlar. Bu açıdan uygunluk oldukça önemlidir.

Belirlenmiş standartlara uygunluk daha az hata ve işlevsel bozukluk anlamında geldiğinden dolayı uygunluk boyutu aynı zamanda ürünle ilgili şikâyet ve olumsuz geri bildirimlerin de önüne geçen bir faktördür. Bu durum müşteri nezdinde marka ve ürünle ilgili değer algısını yükseltir. Sonuç olarak müşterinin ürün için katlandığı maliyet ve elde ettiği fayda arasındaki dengeyi olumlu olarak katkı sunması açısından uygunluk boyutu kritik önemdedir.

1.3. Ürün Kalitesi, Müşteri Beklentileri ve Müşteri Memnuniyeti Arasındaki İlişkinin İncelenmesi

Pazarlama disiplindeki önemli araştırma konularından birisi de her zaman ürün kalitesini oluşturan bileşenler ile müşteri memnuniyeti ve müşteri beklentilerinin açıklanması olmuştur. Kalite süreçleri ürünün üretilmesine ait proje fikrinden, tasarımına, üretim süreçlerine ve nihai olarak satış sonrası hizmetlere kadar tüm süreçleri kapsar. Bu bağlamda ürün kalitesi, müşteri beklentileri ve müşteri memnuniyeti arasındaki karşılıklı bağımlılığın ürünün pazardaki başarısını belirlemede kilit rol oynadığı söylenebilir. Buradan hareketle, işletmelerin yaptığı kalite iyileştirme çalışmaları müşteri memnuniyetini, artan müşteri memnuniyetinin müşteri sadakatini, oluşan müşteri sadakatinin de satışları artıracakları anlaşılmaktadır (Kurnuç vd., 2015, s. 25).

Müşteri beklentisi, müşterilerin bir ürün veya hizmetten bekledikleri performans, kalite, hizmet düzeyi ve genel değer algılarının toplamıdır. Kuşluvan (1999)'a göre bu beklentiler, bireysel ihtiyaç ve istekler, geçmiş deneyim, dışsal iletişim gibi kaynaklara bağlı olarak şekillenmektedir.

Müşteri memnuniyeti, bir müşterinin satın aldığı ürün veya hizmetten elde ettiği deneyim sonucunda, beklentilerinin ne derece karşılandığını gösteren bir ölçüttür. Müşteri memnuniyeti, müşteriler tarafından satın alınan ve tüketilen ürünlerin performansına ilişkin müşterilerden gelen bir yanıttır. Müşteri memnuniyeti, müşterilerin ihtiyaçlarının, arzularının ve beklentilerinin karşılanabildiği bir düzeydir ve bu da tekrar satın alma veya devam eden sadakat ile sonuçlanacaktır (Oktavanny & Sulistiadi, 2022, s. 312).

Kalite kavramı, müşteri memnuniyeti ve müşteri beklentilerinin bir fonksiyonu olarak ifade edilebilir. Bununla birlikte, bir ürün veya hizmetin algılanan kalitesi büyük ölçüde

müşterinin isteklerini karşılama yetkinliğine göre belirlenir. Bu istekler, müşterinin ürün veya hizmetten beklediği performansı, faydayı, özellikleri ve değerleri içerir ve geçmiş kullanıcı deneyimleri, pazarlama iletişimleri ve marka imajı gibi çeşitli unsurlardan etkilenir.

Sonuç olarak, kalite kavramı, müşteri memnuniyeti ile müşteri beklentilerinin kesişim alanında bulunmaktadır. Yüksek kaliteye ulaşmak için müşteri beklentilerinin hem doğru kavranması hem de doğru karşılanması zorunlu bir durumdur. Bu nedenle müşteri memnuniyeti ve müşteri beklentileri kalitenin temel unsurlarıdır.

1.3.1. Müşteri Beklentilerinin Karşılanması Müşteri Memnuniyeti Üzerindeki Etkisi

Müşteri memnuniyeti, bir müşterinin bir ürün veya hizmetin beklentilerini karşılama derecesine ilişkin yaptığı subjektif değerlendirme olarak tanımlanabilir. Bu beklentiler, ürün kalitesi, fiyatlandırma, bulunabilirlik, müşteri hizmetleri ve marka bilinirliği gibi çeşitli faktörler tarafından şekillenmektedir. Bu faktörlerin müşterinin beklentileriyle uyumlu hale getirilmesi memnuniyet düzeyinin artmasına hizmet eder.

Müşteri memnuniyetinin sağlanması, marka sadakatının artmasına, olumlu ağızdan ağıza yönlendirmelere yol açar ve yeni müşterilerin kazanılmasını kolaylaştırır. Buna ilaveten, memnun müşteriler tekrar satın alma eğilimi gösterir ve bu da işletmeler için tutarlı ve öngörülebilir gelir akışlarına yol açar. Bu nedenle, müşteri beklentilerinin karşılanması ve aşılması, stratejik ticari başarıya ulaşmak için çok önemlidir.

Bugüne kadar yapılmış olan çok sayıda çalışma, müşteri memnuniyeti ile satın alma davranışı arasında sürekli olarak anlamlı bir ilişki olduğunu ortaya koymuştur. Araştırmalar, yüksek düzeydeki müşteri memnuniyetinin, aynı marka veya şirkete devam etme olasılığının artmasıyla bağlantılı olduğunu ortaya çıkararak müşteri aidiyetini kuvvetlendirdiğini ortaya koymuştur. Bununla birlikte, memnun müşteriler olumlu geri bildirimde bulunma ve ürün veya hizmetleri başkalarına tavsiye etme eğilimindedir. Bunun sonucunda, müşteri beklentileri karşılandığı sürece artan müşteri aidiyeti ve pozitif değerlendirmeler nedeniyle işletmelerin pazar payı ve gelirlerini artırdıkları görülmektedir.

Müşteri memnuniyetini sağlamak için kullanılan yaklaşımlar ve taktikler farklı sektörlerde farklılık göstermektedir. Hizmet sektöründe, müşteri beklentilerinin

aşılmasını savunan bir düşünce ekolü vardır, zira sadece beklentileri karşılamak yeterli olmayabilir. Genellikle kolayca standardize edilmesi mümkün olmayan hizmetlerin benzersiz ve kişiselleştirilmiş doğasından dolayı, bunun ötesine geçmek müşteri bağlılığını ve genel memnuniyetini artırabilmektedir.

Müşteri memnuniyetini sağlayan önemli bir husus, satın alma sonrasında sağlanan hizmet ve desteğin düzeyidir. İyi işleyen bir müşteri hizmetleri ve destek sistemi, ortaya çıkabilecek sorunların derhal ve tatmin edici bir şekilde ele alınmasında önemli bir rol oynar ve bu da müşteri memnuniyetinin ve ardından müşteri sadakatinin artmasını sağlayacaktır.

Müşteri memnuniyetinin yüksek düzeyde tutulması için sürekli geri bildirim mekanizmaları şarttır. Müşteri geri bildirimleri, müşteri beklentilerinin doğru anlaşılmasında ve bu beklentilere hızlı yanıt verilmesinde önemli bir rol oynar. Geri bildirim, hizmet ve ürün geliştirme süreçleri için önemli bir girdi olarak değerlendirilmelidir.

Müşteri beklentileri ve memnuniyeti arasındaki ilişkiyi araştıran bazı çalışma ve modeller mevcuttur. Bunlardan bazıları şu şekildedir.

1.3.1.1. Kano Modeli

Japon bilim adamı Noriaki Kano tarafından 1980'lerde tasarlanan Kano Modeli, müşteri memnuniyetini değerlendirmek ve ürün geliştirme çabalarını müşteri beklentileriyle uyumlu hale getirmek için bir çerçeve görevi görmek üzere tasarlanmıştır. Müşteri perspektifinden bakıldığında Kano modeli, ürünün kalite özelliklerini belirleyerek ve sınıflandırarak müşteri ihtiyaçlarını anlamak için kullanılır. Kano (2001) tarafından ortaya konan Kano Modeli, müşteri ihtiyaçlarını ve beklentilerini beş farklı kategoriye ayırır: Temel Gereksinimler, Performans Gereksinimleri, Heyecan Verici Gereksinimler, Kayıtsız Gereksinimler ve Ters Gereksinimler.

Temel Gereksinimler, müşterilerin ürünü satın alma kararının merkezinde olan özelliklerdir. Diğer bir deyişle müşterilerin bir üründe gerçekten görmek istediği özelliklerdir. Bu gereksinimler karşılanmadığında müşteri memnuniyetsizliği oluşur, ancak karşılandığında müşteri memnuniyeti üzerinde büyük bir etkisi olmaz (Matzler & Hinterhuber, 1998, s. 28).

Performans Gereksinimleri, Performans gereksinimleri, müşterilerin bir üründen ne kadar iyi performans gösterdiğine ilişkin beklentilerini ifade eder. Performans ihtiyaçları müşterinin tanımlayabileceği ve üreticinin tartışabileceği ihtiyaçlardır. Bu ihtiyaçlar "daha fazlası daha iyidir" kuralına tabidir. Burada karşılanan ihtiyaçlar, bir ürün veya hizmeti diğerinden ayıran ihtiyaçlardır. Bu kategori rakipler arasındaki ayrımı sağlayan kategoridir (Rotar & Kozar, 2017, s. 343). Bu gereksinimlerin iyi karşılanması, daha yüksek müşteri memnuniyetine yol açar. Bu gereksinimlerin karşılanmaması durumunda ise memnuniyetsizlik oluşur.

Heyecan Verici Gereksinimler, müşterilerin dahil edildiğinde çok tatmin edici bulduğu beklenmedik özelliklerdir. Bu özelliklerin bulunmaması müşterileri mutsuz etmese de bunlara sahip olmak müşteri memnuniyetini büyük ölçüde artırmaktadır.

Kayıtsız Gereksinimler, müşterilerin bir üründe umursamadığı özellikler olarak ifade edilebilir. Bu özelliklerin ürün spesifikasyonlarına dahil olup olmaması müşteri memnuniyetini etkileyen bir durum değildir.

Ters Gereksinimler, Mevcudiyeti müşteri memnuniyetsizliğine neden olurken, yokluğu müşteri memnuniyeti yaratan özelliklerdir (Yılmaz, 2021, s. 845).

Kano Modeli, Müşteri anketlerinin hazırlanması ve dağıtılması, verilerin toplanması, veri analizi ve sınıflandırma, ürün geliştirme stratejilerinin geliştirilmesi, uygulama ve izleme gibi fazlardan oluşmaktadır.

Müşteri ihtiyaçlarının tespiti için hazırlanan ankette, her bir gereksinim için bir pozitif bir de negatif olmak üzere iki soru sorulur. Hazırlanan anketlerin mümkün olduğunca geniş bir kitleye ulaştırılması önemlidir. Müşterilerden alınan bildirimler Kano modeline göre beş kategoride sınıflandırılır. Burada amaç müşteri memnuniyeti üzerindeki etkinin analizi edilmesidir. Sınıflandırılan müşteri geri bildirimlerinden hareketle ürün geliştirme stratejileri geliştirilir. Bu stratejilerin geliştirmesindeki temel amaç, müşteri memnuniyetinin artırılmasıdır. Stratejinin uygulamaya konması ve sonuçların izlenerek ihtiyaç duyulduğunda sürekli iyileştirme faaliyetlerinin yürütülmesi ile süreç devam eder. Müşteri geri bildirimleri sürekli iyileştirme faaliyetleri için önemli girdiler sunar.

Kano Modeli, müşteri memnuniyetini artırmada önemli bir çerçeve sunar ve bu çerçeve, işletmelerin müşterilerinin hem gizli hem de temel tercihlerini ayırt etmesine ve yerine

getirmesine olanak tanıdığından dolayı etkin bir şekilde uygulandığında ürün geliştirme alanında önemli avantajlar sağlamaktadır.

1.3.1.2 Beklenti Çıkarım Teorisi

Beklenti-Çıkarım Teorisi, müşteri memnuniyeti kavramını açıklayan ve müşteri beklentilerinin etkin bir şekilde karşılanmasına yönelik stratejilerin ana hatlarını çizen önemli bir çerçeve sunar. Bu teorik çerçeve, müşteri beklentileri ile bir ürün veya hizmetin performansına ilişkin müşteri algıları arasındaki ilişkiyi araştırır. İlk olarak Oliver (1980a)'ın çalışmasından yola çıkan bu teori, müşteri memnuniyetinin başlangıçtaki beklentiler ile gerçek algılanan performans arasındaki varyansa bağlı olduğunu öne sürmektedir.

Teorinin öne sürdüğü gibi müşteri memnuniyeti, bir ürün veya hizmetin müşteri beklentileriyle uyumlu veya bu beklentileri aşan algılanan performans düzeyine bağlıdır. Tam tersine, algılanan performans beklentilerin altına düşerse memnuniyetsizlik ortaya çıkar. Teori, beklentiler, algılanan performans ve çıkarım olmak üzere üç temel kavram üzerinden açıklanmaktadır. Beklentiler, müşterilerin ürün veya hizmetten elde etmeyi umdukları fayda veya ürün ve hizmet hakkında daha önceden elde etmiş oldukları düşüncelerdir. Algılanan performans ise müşterilerin ürün veya hizmetle etkileşim süreci sonucunda edinmiş oldukları, kişisel deneyimlerine bağlı olarak gelişen tecrübelerdir. Beklentiler ve algılanan performans arasında var olan varyans veya sapma miktarı ise çıkarım olarak tarif edilir.

Beklenti Çıkarım Teorisine göre müşteri beklentilerinin belirlenmesinde şöyle bir yol izlenebilir:

1. Müşteri beklentilerinin anlaşılması, müşteri anketleri, geri bildirim formları ve pazar araştırması enstrümanları kullanılarak gerçekleştirilebilecek bir görev olan müşteri beklentilerinin doğru bir şekilde anlaşılması aşamasıdır.
2. Müşteri beklentileriyle uyumlu ürün ve hizmet sunumu, müşteri beklentileri kapsamlı bir şekilde anlaşıldıktan sonra, bu beklentilere yalnızca uyum sağlayan değil aynı zamanda bu beklentileri aşan ürün ve hizmetler sunmak zorunlu hale gelmektedir. İşletmelerin şu anda müşterilerinin ihtiyaç ve beklentilerini anlamaları ve gelecekte müşteri beklenti ve ihtiyaçlarının nasıl gelişeceğini öngörmeleri gerekmektedir (Tukiran vd., 2021, s. 482).

3. Beklenti yönetimi ve müşteri iletişimi aşaması gerçekçi olmayan beklentilerin oluşmasını önlemek ve müşterilere ürün ve hizmetlerle ilgili doğru bilgilerin sunulması açısından önemlidir. Yanıltıcı veya abartılı bilgilerin yayılması müşteri beklentilerinin artmasına yol açarak algılanan performans ile beklenen sonuçlar arasındaki eşitsizliğin artmasına neden olur.
4. Geri bildirim toplama sürecinin sürekliliği ve iyileştirme aksiyonları: müşteri geri bildirimlerinin toplanması bir ritüel olarak kabul edilmeli, toplanan bu geri bildirimler, ürün veya hizmet geliştirme süreçleri için bir girdi teşkil etmelidir. Memnuniyetsizliğe sebep olan ve iyileştirilmesi gereken faktörler belirlenmelidir.
5. Çalışan eğitimi ve gelişimi: çalışanlara müşteri hizmetleri konusunda eğitim vererek gelişimlerinin sağlanması, müşteri beklentilerinin karşılanması için kritik önem seviyesindedir. Çalışanlara verilecek eğitimlerle çalışanların müşteri memnuniyetinin nasıl sağlanabileceği, müşterilere nasıl davranılması gerektiğinin önemi öğretilerek, müşteri memnuniyeti amaçlı çalışanların gelişimi sağlanabilir (Şahin, 2017a, ss. 1178-1179).

Beklenti-Çıkarım Teorisi müşteri memnuniyetini artırmak için sistematik bir çerçeve sunar. Bu teori, müşteri beklentilerini doğru bir şekilde tanımlamanın, karşılamanın veya aşmanın önemini vurgulamaktadır. Bu beklentileri etkili bir şekilde karşılamak için onları doğru bir şekilde tespit etmek, açık iletişim kurmak, sürekli geri bildirim istemek ve bu geri bildirimi iyileştirme girişimlerine dahil etmek çok önemlidir.

1.3.1.3 Nedensel Eylem Teorisi

Nedensel Eylem Teorisi, belirli bir sonuç üretmek için nedenler ve eylemler arasındaki karşılıklı bağlantıyı araştırır. Bu teorik çerçeve, psikoloji ve sosyal bilimler disiplinlerinde, bireylerin veya toplumların arzu edilen sonuçlara ulaşmak için takip ettikleri davranış kalıplarının yanı sıra, onların seçimlerini yönlendiren normları saptamak için sıklıkla kullanılmaktadır. Ajzen ve Fishbein (1975) bireylerin ve toplumların eylemlerini anlamak ve tahmin etmek için geliştirdikleri teorik çerçeve üzerinde modellenir. Bu teori, bireylerin davranışlarının, onların niyetleri tarafından yönlendirildiğini ve bu niyetlerin de tutumlar ve öznel normlar tarafından şekillendirildiğini öne sürer.

Nedensel Eylem Teorisi, tutumlar, öznel normlar ve davranışsal motivasyon olmak üzere 3 ana kavram üzerinde modellenmiştir. Tutumlar, bireylerin belirli davranışlara yönelik subjektif değerlendirmeleridir. Bu tutumlar, kişinin davranışın sonuçlarına ilişkin inançlarından ve bu sonuçlara verdiği öneme göre şekillenmektedir. Öznel normlar, bireylerin belirli bir davranış kalıbının aile, arkadaşlar, meslektaşlar gibi önemli referans grupları tarafından onaylaması veya reddetmesine yönelik tutumlarını ifade eder. Bu normlar bireyin bulunduğu sosyolojik çevre içindeki toplumsal beklentileri temsil eder. Davranışsal motivasyon, bireyin belirli bir davranışta bulunma dürtüsüyle ilgilidir. Tutumlar ve öznel normlar gibi faktörler bu eğilimleri önemli ölçüde etkilemektedir.

Nedensellik teorisi, müşteri memnuniyetini artırmak ve müşteri beklentilerini anlamak için kapsamlı bir çerçeve sunar ve pazarlama iletişimi ve müşteri deneyiminin geliştirilmesi amacıyla ürün ve hizmet süreçlerinin nasıl iyileştirilmesi gerektiğini anlamak için önemli içgörüler sağlar. Teori genel olarak müşterilerin satın alma kararlarının arka planındaki içsel motivasyonu anlamaya çalışır.

Ürün kalitesinin artırılması, müşteri hizmetlerinin ve pozitif marka itibarının geliştirilmesi yoluyla gerçekleştirilebileceği gerçeğinden hareketle işletmelerin müşteri memnuniyetinin artırılması amacıyla müşteri tutumları anlayarak olumlu tutumlar geliştirmeleri zorunludur. Müşteri davranışlarının anlaşılmasında önemli faktörlerden biri sosyal çevrenin etkisinin anlaşılmasıdır. Sosyal çevrenin tüketici satın alma davranışı üzerindeki etkisi önemlidir. Öznel normlar, bireylere empoze edilen toplumsal beklentileri ve bunların karar alma süreçleri üzerindeki etkilerini yansıtır. Bu nedenle işletmelerin tüketicilerin sosyal ağlarını hedef alan pazarlama stratejileri geliştirmeye odaklanması gerekmektedir. Sosyal medyanın günlük hayatımızın ayrılmaz bir parçası olduğu günümüzde yeni bir pazarlama yöntemi olan influencer pazarlama, tüketicilerin satın alma kararları üzerinde giderek daha etkili olmaya başlamıştır. Böylelikle influencer pazarlama, markalar açısından uygun fiyatlı ve yüksek ciroyla gün geçtikçe daha popüler hale gelmektedir (De Veirman vd., 2017; Freberg vd., 2011).

Teorinin hedeflerinden biri de davranışsal motivasyonların ortaya çıkarılmasıdır. Müşteri memnuniyetini ve beklentilerini kapsamlı bir şekilde anlayabilmek için müşterilerin davranışsal motivasyonlarını derinlemesine incelemek zorunludur. Bu incelemeler,

müşterilerin gelecekteki satın alma kararlarının öngörülmesi için önemli bir altyapı oluşturur.

Nedensel Eylem Teorisi, müşteri memnuniyetini artırmak ve müşteri beklentilerini anlamak için değerli bir çerçeve sunar. Bu teori, müşterilerin tutumları, öznel normları ve davranışsal motivasyonları arasındaki etkileşimi açıklayarak işletmelerin müşteri odaklı stratejiler oluşturmalarına katkı sağlar. Müşteri memnuniyetini etkili bir şekilde sağlamak ve beklentilerle uyum sağlamak için müşterilerin tutumlarını ve sosyal bağlamlarının etkilerini dikkate almak çok önemlidir.

1.3.1.4 Müşteri Memnuniyet Endeksi

Müşteri Memnuniyeti Endeksi, müşterilerin bir şirketin ürünüyle ilgili deneyimlediği memnuniyet düzeyini değerlendirmek için kullanılan bir yöntemdir. Bu endeks, müşteri memnuniyetiyle ilgili çeşitli ölçümleri içerecek şekilde yapılandırılmıştır ve işletmelerin farklı alanlardaki performansını analiz etmek ve belirli bir periyot içindeki eğilimleri izlemek için kullanılır.

Müşteri Memnuniyeti Endeksi genellikle, müşterilerin satın alma deneyimleri, ürün veya hizmetlerin kalitesi, müşteri hizmetlerinin etkinliği, fiyatlandırma rekabetçiliği, marka algısı ve müşteri sadakati dahil olmak üzere müşterilerin işletmelerle etkileşimlerinin çeşitli boyutları hakkında sorgulandığı anketler aracılığıyla toplanan verilerden elde edilir. Toplanan yanıtlar, genellikle 0'dan 100'e kadar sayısal bir ölçekte sunulan bir memnuniyet skoru oluşturmak için toplanır. Bu skor, müşterilerin genel memnuniyet düzeyini ve sadakatini belirlemek için kullanılabilir.

Müşterilere sunulan anketlerle, müşterilerden belirli bir ürün veya hizmet ile ilgili memnuniyet düzeylerinin değerlendirilmesi istenir. İlk aşamada, sunulacak anketin tasarımı yapılır ve çoğunlukla 1'den 5'e veya 1'den 7' ye kadar bir ölçek kullanılarak yanıtlar toplanır. Toplanan yanıtlar veri analitiği sürecine tabi tutularak analiz edilir. Daha önceden belirlenmiş olan ölçekler üzerinden puanlanır. Son aşamada memnuniyet puanlarının ortalamaları elde edilerek genel bir memnuniyet skoru ortaya konmuş olur. Bu skor çoğunlukla 100 puan üzerinden ifade edilir. Bir ile beş arasında ölçeklendirilmiş bir anket için müşterilerin ortalama 3,6 puan verdiği farz edildiğinde, müşteri memnuniyet indeksi $(3,6 * 20)$ 72 olarak hesaplanacaktır.

1.3.2. Kalite Boyutları ve Müşteri Beklentileri Arasındaki İlişki

Ürün kalite boyutlarının her biri kalitenin farklı bir yönüne odaklandığından dolayı her bir kalite boyutu özelinde müşteri beklentilerini şekillendiren farklı dinamiklerden bahsetmek mümkündür. Bu kısımda her bir kalite boyutu özelinde müşteri beklentilerini şekillendiren farklı parametrelere değinilecek ve bu parametrelerin müşteri beklentilerini ne şekilde etkilediği üzerinde durulacaktır.

Performans boyutunda; ürünün temel fonksiyonlarını yerine getirmedeki etkinliği müşteri beklentilerini önemli ölçüde etkilemektedir. Müşteriler, ürünün kullanım amacına bağlı olarak değişen düzeylerde olmak üzere üründen üstün performans beklemektedir. Oyun oynayan veya yüksek düzey grafik işlemci gerektiren tasarım programları kullananlar kullanıcıların bilgisayarın işlemci hızına ve grafik özelliklerine özellikle önem vermesi gibi. Kotler'e (2000) göre müşteri memnuniyeti, tüketicinin beklentileriyle ilişkili olarak ürünün algılanan performansı hakkında tüketicinin hoşnutluk veya hoşnutsuzluk duygusudur. Ayrıca, yüksek kaliteli ürünler, ağızdan ağıza tavsiyeler yoluyla yeni müşteri çekmede hayati bir rol oynamaktadır. Müşteriler ürünün performansından memnun kaldıklarında bu sadece beklentilerini karşılamakla kalmaz, aynı zamanda müşterilerde marka sadakatini de artırır.

Ürün kalitesinin özellikler boyutunda ise müşteri beklentilerini şekillendiren hususlar şu şekilde sıralanabilir: Fonksiyonellik, ürünün müşteri gereksinimlerine ne ölçüde uyum sağladığının ve kendisinden beklenen işlevi etkin bir şekilde yerine getirme derecesinin incelenmesini ifade eder. Yenilikçilik, ürünün kendisini rakiplerden ne ölçüde farklılaştırdığı veya ek işlevler sağladığını gösterir. Kişiselleştirme ise ürünün müşteri beklentilerine göre özelleştirilebilir olup olmama derecesini gösterirken, kullanım kolaylığı ürünün her çeşit müşteri kitlesi tarafından kullanılabilme durumunu ve anlaşılır ve erişilebilir olma durumunun göstergesidir. Çok yönlülük ürünün birden fazla işlevi gerçekleştirip gerçekleştirmediğinin ifadesidir. Ek özellikler de ürünün standart fonksiyonları dışında sunulan ilave işlevleri ifade eder. Bunların dışında, kalite ve dayanıklılık, ürünün tasarımı, ürünün diğer ürün veya sistemlerle olan entegrasyon kabiliyeti, çevreye duyarlı oluşu gibi faktörler müşteri beklentilerini şekillendiren diğer hususlar olarak yazılabilir. Tüm bu faktörler müşterileri beklentilerini doğrudan etkileyen ve satın alma kararları üzerinde önemli belirleyici olan faktörlerdir. Bir ürünün

özellikleri, tüketicinin satın alma tercihlerini etkileyebilecek çeşitli faktörlerle etkileşim halindedir. Tüketiciler yalnızca kendi özel gereksinimlerini karşılayan özellikleri değil, aynı zamanda inançlarına, yaşam tarzlarına ve bireysel zevklerine uygun özellikleri de dikkate almaktadırlar. Bu beklentiler çeşitli tüketici grupları arasında farklılık gösterebilir ve bu da işletmelerin hedef pazarlarını kapsamlı bir şekilde anlamalarını gerektirir. Bu perspektif, ürünün tasarımı, geliştirme, pazarlama ve dağıtım süreçlerini şekillendirerek işletmelere rekabet edebilirlik avantajı sağlar.

Bir ürünün güvenilirliği ve istikrarı, müşterinin ürüne olan güvenini ve memnuniyetini artırmaya katkıda bulunan temel faktörlerden olduğu için, ürün kalitesinin güvenilirlik boyutu da müşteri memnuniyetinin belirlenmesinde kritik boyutlardan biridir. Bir ürünün güvenilirliği, ürünün zaman içindeki performansının tutarlılığını ve güvenilirliğini yansıttığı için müşteri beklentilerinin kritik bir yönüdür. Güvenilirlik, belirli koşullar altında çalışan bir ekipman parçasının belirli bir süre boyunca tatmin edici bir performans gösterme olasılığıdır (N & Seetharam, 2014, s. 219). ve dolayısıyla bakım, onarım maliyetleri üzerinde önemli bir etkiye sahiptir. Tüketicilerin ürünlerin uzun bir süre boyunca sorunsuz bir şekilde çalışacağına dair yüksek beklentileri olduğundan dolayı, müşteri beklentileri ürün kalitesinin güvenilirlik boyutuyla yakından ilişkilidir. Bir ürünün güvenilirliği, kalıcı müşteri sadakatini geliştirmede çok önemli bir faktör olan müşteri sadakatini artırmaya hizmet eder. Güvenilir olmayan ürünler, müşteri memnuniyetsizliğine ve marka itibarının zedelenmesine neden olabilir. Ürünün güvenilirliği müşteri şikâyetlerinin azalmasına yol açar ve sonuç olarak müşteri hizmetleri harcamalarının azalmasına neden olacaktır.

Ürünlerin önceden tanımlanmış standart ve regülasyonlara uygunluğunu ifade eden kalitenin uygunluk boyutu da müşteri beklentilerinin şekillenmesinde önemli bir etkidir. Uyumluluk, ürünün belirli pazar segmentlerinde kullanılabilirliğini artırır ve bu durum, müşteri tabanını genişletir. Tüketiciler, bir ürünün kullandığı sistemleriyle veya diğer ürünlerle sorunsuz çalışabileceği beklentisine sahiptirler. Bu beklenti, ürünün kullanıcı dostu olması ve diğer teknolojilerle entegrasyon sağlayabilme kabiliyeti ile yakından bağlantılıdır. Uyumsuzluk, müşteri hoşnutsuzluğuna neden olabilir ve sonuçta marka aidiyetine zarar verebilir.

Müşteri beklentilerini şekillendiren bir diğer önemli kalite boyutlarından biri de dayanıklılıktır. Dayanıklılık boyutunun müşteri beklentileri üzerindeki etkisi, malzeme ve üretim süreci, ürün kullanım ortamı ve kullanım şartları, garanti ve servis imkanları, müşteri geri bildirimleri, marka itibarı ve güvenilirlik faktörleri üzerinden incelenebilir. Bu faktörlerin her birini detaylandıracak olursak; üst düzey malzemelerin kullanımı ve titiz üretim teknikleri, uzun ömürlülüğü artırır ve müşteri beklentilerini daha fazla karşılar. Üretiminde yüksek kalite kauçuk malzemesi kullanılmış bir otomobil lastiğinin kullanım ömrünün ve aşınma direncinin yüksek olması dayanıklılık boyutunda malzeme kalitesinin belirlemiş olduğu müşteri beklentisiyle ilgilidir. Ürünler sıkı testlere tabi tutulduğunda ve öngörülen kullanım ve çevre parametrelerine göre dayanıklılıkları açısından akreditasyon aldığı anda, tüketicilerin bu tür ürünlerin güvenilirliğine güvenme olasılığı daha yüksek olacaktır. Suyla maruz kaldığında bazı suyu cep telefonlarının su geçirmemesi birçok kullanıcı için önemli bir dayanıklılık fonksiyonudur. Uzatılmış garantiler ve etkili müşteri hizmetleri, müşterilere ürünün uzun ömürlülüğü konusunda güven sağlayarak müşteri beklentilerini pozitif anlamda şekillendirir. Ömür boyu garanti sunan bir ürünün dayanıklılığı konusunda kullanıcılarında büyük bir beklenti oluşturacağı kesindir. Öte yandan ürünün dayanıklılığına ilişkin müşterilerden gelen olumlu geri bildirimler ve değerlendirmeler, potansiyel yeni müşterilerin beklentilerini önemli ölçüde etkileyecektir. Ürün incelemesinde oldukça pozitif yorumlar alan bir bilgisayar modeli için kullanıcıda oluşan uzun süre problemsiz ve yüksek performanslı kullanım beklentisi bu konuya bir örnektir. Saygın işletmelerin ürünlerinin genellikle uzun ömürlü olarak değerlendirilmesi de müşteri beklentilerinin şekillenmesinde etkili bir faktördür.

Ürün kalitesinin servis boyutu ile ilgili olarak müşteri beklentilerini etkileyen faktörler ise şu şekilde sıralanabilir:

Hızlı ve etkin bir müşteri hizmeti, müşteriler soru ve sorunlarına hızlı ve etkili yanıtlar beklerler. Müşteri hizmetleriyle zahmetsiz ve akıcı iletişim, müşteri beklentilerini önemli ölçüde artırır.

Servis personelinin ürün bilgisinin derinliği ve müşteri hizmetleri temsilcilerinin sergilediği problem çözme becerileri ile müşteri beklenti düzeyi de pozitif yönde ilişkilidir. Müşterilere sunulan hizmetlerin müşterilerin öznel gereksinim ve tercihlerine göre özelleştirilebilmesi de müşteri beklentilerini artırma potansiyeline sahiptir.

Ürünle ilgili satış sonrası destek süreçleri de servis boyutuyla ilgili kalite beklentisini şekillendiren bir diğer parametredir. Bunun yanında servis ulaştıran kişilerin müşteriyle iletişimlerinde açık, net, şeffaf ve dürüst olmaları da servis boyutu kalitesinin önemli bir parçasıdır.

Kalitenin estetik boyutu itibarıyla konuya yaklaşıldığında bir ürünün görsel estetiği, tasarımı ve fiziksel çekiciliği, müşterilerin algı ve beklentilerinin şekillenmesinde önemli rol oynadığı söylenebilir. Görsel olarak çekici ürünler, müşteri algıları üzerinde olumlu bir etkiye sahiptir ve müşterilerin belirli bir ürüne yönelik tercihlerini artırabilir. Bir ürünün estetiği, marka itibarını geliştirmede ve sonucunda müşteri beklentilerini etkilemede çok önemli bir rol oynar. Görsel olarak hoş olan ürünler, daha yüksek düzeyde müşteri memnuniyetine ve dolayısıyla ürünün müşterilerce tavsiye edilme oranını önemli ölçüde artırır.

Bir diğer kalite boyutu olan algılanan kalite boyutuyla ilgili olarak müşteri beklentilerinin odaklandığı belirli faktörler vardır. Bunlardan birisi marka imajı ve itibarıdır. Müşterilerin kalite algısı, marka imajı ve itibarından büyük ölçüde etkilenmektedir. İyi bilinen ve saygı duyulan bir marka genellikle üstün algılanan kaliteyle ilişkilendirilir.

Bu faktörlerden bir diğer olan ürün özellikleri ve işlevsellik bağlamında ise bir ürünün özellik ve yeteneklerinin, algılanan kalite düzeyi üzerinde doğrudan etkiye ettiği söylenebilir. Müşteri beklentileri, her zaman daha iyi, daha kaliteli ve daha kaliteliyi daha uygun fiyata alma eğilimindedir. Çünkü her müşteri almak için ödediği bedelin karşılığını alma amacındadır (Şahin, 2017b, s. 1176). Kullanım kolaylığı ve ergonomi ise bir ürünün rutin kullanımda sunduğu pratiklik ve konfor düzeyini ortaya koymaktadır. Tüketiciler, kullanıcı dostu ve ergonomik tasarımlara sahip ürünlere üstün kalitenin göstergesi olarak değer verirler. Bununla birlikte ürünün dayanıklılığı ve güvenilirliği, algılanan kalitesinin belirlenmesinde çok önemli bir rol oynamaktadır. Müşteriler, dayanıklı ve arızalanmayan ürünlerin yüksek kaliteli olduğunu düşünürler.

Algılanan kaliteyi artıran bir diğer faktör ürün tasarımı ve estetiğidir. Estetik açıdan hoş ve iyi tasarlanmış bir ürün, algılanan kalitenin daha yüksek olmasını sağlar. Müşteriler, güzel ve şık tasarımlara sahip ürünleri daha kaliteli olarak algırlar. Diğer müşterilerin yorum, inceleme ve değerlendirmeleri de algılanan kalite üzerinde son derece etken bir faktör olarak sayılabilir. E-Ticaret sitelerinde satılan ürünlerdeki kullanıcı puanları ve

yorumları ürünlerin satın alma sayısını büyük ölçüde etkilemektedir. Çevrim içi alışveriş platformlarında kullanıcılar, bir ürünü satın alırken bu incelemeler ürün alıcılarına fikir verme açısından oldukça önemlidir (Aslan, 2023, s. 876). Olumlu yorumlar ve yüksek puanlar, ürünün kaliteli olduğu algısını güçlendirir. Ürünün fiyat ve değeri de algılanan kalite özelinde etkin ve etkili bir güce sahiptir. Kalite arttıkça tüketicinin ödeme isteği de artmaktadır (Anton vd., 2023, s. 2). Algılanan kalite bağlamında müşteri beklentilerini şekillendiren bir diğer faktör de satış sonrası hizmet süreçleridir. Satış sonrası hizmet şartları ve garanti kapsamının genişletilmesinin, ürünün algılanan kalitesini artırdığı görülmüştür. Pazarlama gücünün en önemli kısmını satış sonrası hizmet desteği oluşturmaktadır (Herbig & Palumbo, 1993, s. 339). İnsanlar ürün satın aldıklarında satın alma sonrası desteğin belirli seviyelerde olmasını beklemektedir. Bu beklentileri tanımlamak pazarlama başarısında önemlidir ve pazarlar olgunluğa girdiğinde daha da önemli olacaktır (Özgören, 2014, s. 4).

Sonuç olarak kalitenin her boyutu, işletmelerin müşteri beklentilerini ürünün farklı yönleri bağlamında karşılama yeteneğinin bir ölçüsü olarak ifade edilebilir. Müşteri beklentileri ile kalite boyutları arasındaki etkileşim, müşteri memnuniyetini etkileyen ve sürekli değerlendirme ve iyileştirmeyi gerektiren döngüsel bir süreçtir. Müşteri memnuniyetini artırmak için şirketlerin kalitenin her bir boyutunu kapsamlı olarak analiz etmesi ve ürünlerini sürekli olarak geliştirmesi zorunludur.

1.3.3. Ürün Kalitesi ve Müşteri Memnuniyeti Arasındaki İlişki

Bir ürünün kalitesi, müşteri deneyimini ve müşteri tatminini şekillendirmede önemli bir rol oynar. Oliver (1980b) tarafından yapılan çalışmada müşteri tatmini, bir ürün veya hizmetin müşteriye sunduğu değer açısından tüketici beklentilerini ne ölçüde karşıladığının veya aştığının göstergesi olarak ifade edilmiştir. Ürün kalitesi bir ürünün belirlenen müşteri ihtiyaçlarını karşılama kabiliyetine ilişkin özellikleridir. Ürün kalitesi, pazarlamacının en önemli konumlandırma araçlarından biridir. Kalite, ürün veya hizmet performansını etkiler; dolayısıyla müşteri değeri ve memnuniyetiyle yakından bağlantılıdır (Kotler & Armstrong, 2017, s. 249). Bununla birlikte müşteri memnuniyeti veya tatmini, bir işletmenin ticari hedeflerine ulaşması ve sürdürülebilirliği açısından da önemlidir. Bu sebeple müşterilerin kalite düzeyi yüksek bir üründen memnuniyeti, marka

sadakatının ve güveninin artmasına yol açarak, sürdürülebilir iş başarısı için kritik bir temel görevi görür.

Müşterilerin beklentileriyle uyumlu olan kaliteli ürünler, müşterilerin markaya ve ürüne olan güvenlerini güçlendirir. Ürün kalitesi düşük olduğunda müşteri memnuniyeti azalır ve bu da müşteri kaybına neden olabilir. Bir ürünün kalitesi, müşterinin markaya olan bağlılığını etkilemede de önemli bir rol oynar. Müşteriler, yüksek kaliteli ürünler sunan markalara karşı daha üst seviyede bir sadakat sergilerler ve bu markaları başkalarına tavsiye etme eğilimlerinde artış gözlenir. Bu sadakat, işletmelere sadece gelir artışını sağlamakla kalmaz, aynı zamanda pazarda sürdürülebilir bir rekabet avantajı da sağlamaktadır.

Müşteri memnuniyetini etkileyen diğer değişkenler arasında fiyatlandırma, satış sonrası hizmetler, müşteri hizmetlerinin mükemmelliği ve marka itibarı yer alır. Ancak bu faktörler arasında en kritik belirleyici olanı ürün kalitesidir. Müşteriler, fiyatına göre ürün değerlemesi yaparken ürünün kalitesine büyük önem vermektedirler. Kaliteli bir ürün, yüksek bir fiyatla sunulsa bile, müşteriler bu ürünü satın almaya daha yatkın olabilirler çünkü kalite, uzun vadede daha ekonomik bir seçenek olarak değerlendirilebilir.

Ürün kalite boyutlarının her biri, müşterinin karar vermesini ve memnuniyet düzeylerini etkilemede önemli bir rol oynar. Uzun vadeli müşteri memnuniyetinin sağlanmasında dayanıklılık ve güvenilirlik gibi faktörler esastır; estetik özellikler ve tasarım ise ilk satın alma kararlarında daha büyük etkiye sahip olabilir.

Marka algısı müşteri memnuniyetini etkilemede önemli bir rol oynamaktadır. Olumlu bir marka imajı tüketiciler arasında güveni artırır ve artan memnuniyet düzeylerine katkıda bulunur. Müşteriler, tanınmış ve güvenilir markalardan alışveriş yapmayı tercih ederler, çünkü bu markalar kalite konusunda daha güvenilir olarak algılanmaktadır.

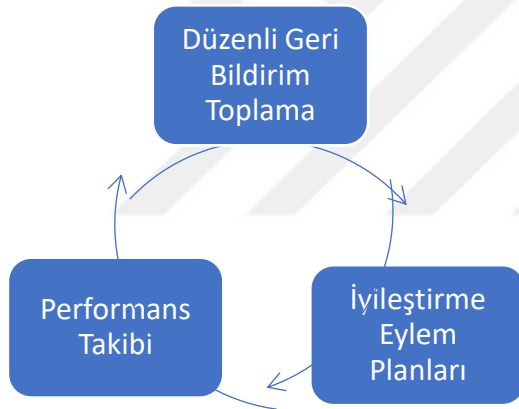
Ürün kalitesi, müşteri memnuniyetinin maksimize edilmesinde çok önemli bir faktördür ve işletmelerin müşteri tatminini geliştirmek için hedef kitlelerine yüksek kaliteli ürünler sunmaları gerekmektedir. Ürün kalitesinin sürdürülebilir şekilde geliştirilmesi ve müşteri taleplerinin karşılanması, müşteri memnuniyeti ve bağlılığının artırılması yoluyla bir işletmenin kalıcı başarısına katkıda bulunur.

Müşteri memnuniyetinin artırılması için kalite boyutlarını besleyen faktörler bağlamında müşteri geri bildirimleri ve buna bağlı sürekli iyileştirme faaliyetlerinden oluşan döngüsel

bir sürecin yürütülmesi ve uzun vadeli stratejiler olarak planlanması gerekmektedir. Bu döngü genel çerçevesi itibariyle üç adımdan oluşmaktadır.

Bu döngüsel sürecin ilk aşaması düzenli geri bildirim toplama sürecidir. Müşterilerden sürekli geri bildirim almak ve toplanan bilgileri değerlendirmek olarak tanımlanabilecek bir süreçtir. Bu süreci, bu değerlendirmeleri baz alarak optimizasyon ve inovasyon eylem planlarının oluşturduğu, iyileştirme eylem planları aşaması takip etmektedir. Son aşama ise, bu inovasyon faaliyetlerinin müşteri memnuniyeti üzerindeki etkilerinin ve performansı sürekli değerlendirildiği performans takibi aşamasıdır. Bu aşamalar müşteri beklentilerini ve memnuniyetini analiz etmek ve geliştirmek için sağlam bir çerçeve sunar. Müşteri memnuniyetini çeşitli kalite boyutlarında artırmak hem ürün geliştirme stratejilerine hem de müşteri etkileşimine önemli ölçüde fayda sağlamaktadır.

Şekil 1.1. Sürekli iyileştirme döngüsü



1.4. Literatür Taraması

Bu bölümde metin madenciliği, duygu analizi teknikleri, ürün kalite boyutları ve müşteri memnuniyeti kullanılarak yapılmış çalışmaların amacı, yöntemi, içeriği ve elde edilen sonuçları kapsamında yapılmış olan literatür taraması çalışmasının sonuçları paylaşılmıştır.

Sarioğlu (2023) tarafından yapılan çalışma beyaz eşya sektöründe satış sonrası hizmet kalitesinin müşteri sadakati üzerindeki etkisini, müşteri memnuniyeti ve elde tutma gibi aracı faktörlere odaklanarak incelemektedir. Veri analizi, satış sonrası hizmet kalitesi, müşteri memnuniyeti ve sadakati arasında pozitif bir ilişki olduğunu ortaya koymakta ve müşteri memnuniyeti ile müşteriye elde tutmanın aracı rollerini vurgulamaktadır.

Çalışma, müşteri ilişkileri yönetimi stratejileri ve akademik araştırmalar için değerli bilgiler sunarak, rekabetin yoğun olduğu beyaz eşya pazarında satış sonrası hizmet kalitesinin marka bağlılığını artırmadaki önemini vurgulamaktadır.

Tabaei ve Fathian (2014) tarafından yapılan çalışmada, elektronik perakende sektöründe müşteri memnuniyetini değerlendirmek için çok kriterli memnuniyet analizi modelini (Multi-criteria satisfaction analysis - MUSA) kullanan bir vaka çalışmasını incelemektedir. MUSA modeli, müşteri beklentileri ile algılanan performans arasındaki farklılıkların ve memnuniyet düzeylerinin değerlendirilmesine olanak sağlar. Çalışma, ürün kalitesinin çeşitli yönlerinin memnuniyet üzerindeki etkisini araştırmakta ve müşteri ilişkileri yönetimi için kullanılabilecek verilere ilişkin içgörüler sağlamaktadır. Sonuç olarak bulgular müşteri memnuniyetini artırmada ürün kalitesi ve müşteri ilişkileri yönetiminin önemini vurgulamaktadır.

Öge & Kayaalp (2021) tarafından yapılan çalışmada, Naive Bayes, Lojistik Regresyon, Destek Vektör Makinesi (SVM), Karar Ağacı, Rastgele Orman (Random Forest) ve Yapay Sinir Ağı (ANN) dahil olmak üzere altı sınıflandırma algoritmasının duygu analizindeki etkinliği üzerine karşılaştırmalı bir analiz yapılmıştır. Çalışma kapsamında Bag of Word (BoW), TF-IDF, FastText ve Word2Vec gibi metin temsil yöntemleri kullanılmıştır. IMDB üzerindeki kullanıcı yorumları kullanılarak yapılan çalışma sonucunda Lojistik Regresyon ve Destek Vektör Makineleri en yüksek doğruluk oranlarını almıştır.

Kınık ve Güran (2021) tarafından yapılan çalışma kapsamında ardışık kelime kombinasyonlarının belirlenmesi yoluyla sınıflandırma doğruluğunun artırılması amaçlanmış ve bu doğrultuda Türkçe metinlerin kategorize edilmesinde TF-IDF ve Doc2Vec teknikleri kullanılmıştır. Naive Bayes, K-En Yakın Komşu (KNN), Lojistik Regresyon ve Karar Destek Makineleri gibi algoritmalar uygulanarak ardışık kelime gruplarının tespitiyle algoritma performansı iyileştirilmiştir. Temel makine öğrenimi sınıflandırıcılarının ve topluluk öğrenimi algoritmalarının, hem TF-IDF hem de Doc2Vec vektörleştirme teknikleri kullanılarak dönüştürülen metin belgeleri üzerindeki etkinliğini değerlendirmek için karşılaştırmalı bir analiz yapılmıştır. Sınıflandırıcılar uzunlukları farklı haber makalelerinden oluşan dört ayrı Türkçe veri seti üzerinde uygulanmıştır. Çalışma sonucunda ardışık kelimelerin tek bir kelime olarak değerlendirilmesiyle

vektörleştirilmesi sonucunda sınıflandırıcıların model başarımlarının arttığını göstermiştir.

Çalış vd. (2013) yapılan çalışmada ise, reklam içerikli e-postaları belirlemek için metin madenciliği teknikleri kullanılmıştır. Analizde kullanılan sınıflandırma modelleri Destek Vektör Makinesi (SVM), k-En Yakın Komşu (KNN) ve Naive Bayes'ten (NB) oluşmuştur. Araştırmada reklam e-postaları ikili (binary), frekans ve TF-IDF ağırlıklandırma teknikleri ile vektörel olarak temsil edilmiştir. Çalışma, Türkçe yazılmış e-postalar ile gerçekleştirilmiş, seçilen toplamda 800 e-postanın 400'ü normal, 400'ü ise reklam içerikli olarak belirlenmiştir. Sonuç olarak reklam içerikli e-postaların kNN algoritması ile %96,5 gibi yüksek bir doğruluk derecesi ile tespit edildiği gözlenmiştir.

Kılınç vd. (2016) tarafından yapılan çalışma kapsamında, bilimsel makaleler kNN algoritması ve R programlama dili kullanılarak kategorize edilmiştir. Çalışma verileri olarak Research Gate üzerinden elde edilen bilimsel makale özetleri kullanılmıştır. Amaç, metin madenciliği metodolojilerinin uygulanması yoluyla makaleleri iki ayrı kategoride sınıflandırarak, farklı metin temsil tekniklerinin değerlendirilmesini ve kNN algoritmasının etkinliğinin gözlemlenmesidir. Çalışma sonucunda %96,67 oranında doğruluk (ACC) değeri bulunarak yayınların hangi sınıfa ait olduğu tespit edilmiştir.

Makine öğrenmesine dayalı veri sınıflandırma teknikleri ile glokom hastalığının erken teşhisini tahminlemeye çalışan çalışma Şatır vd. (2016) tarafından yapılmıştır. Glokomlu bireylerin verilerinin analizi için çeşitli sınıflandırma algoritmaları ve metin madenciliği metodolojileri uygulanmış ve yapılan değerlendirmeler sonucunda en etkili algoritmalar ortaya konmuştur. Veri seti olarak bir yüksek lisans tezinden alınan veri kümesi kullanılmıştır. Veri kümesindeki nitelik sayısı düşürüldükten sonra veriler karar ağaçları ve yapay sinir ağları algoritmaları ile sınıflandırılmıştır. Son aşamada ise çapraz doğrulama ile sınıflandırma performansları ölçülmüştür. En düşük başarı oranı %91,67 +/- %3,95 ile CART algoritmasına, en yüksek başarı oranı ise %93,45 ile C4.5 algoritmasına ait olarak ölçülmüştür.

Diler ve Demir (2023)'in yaptıkları çalışmanın amacı ise, veri madenciliği sınıflandırma algoritmalarının çalışılan veri kümesi içerisinde veri kalitesini bozan etmenlerden biri olarak kabul edilen sağdan sansürlü veriler bulunması durumunda sınıflandırma performansının nasıl etkilendiğinin ortaya konmasıdır. Çalışma süresince Naive Bayes,

Lojistik Regresyon ve k-nn algoritmaları kullanılmıştır. Çalışma sonucunda lojistik regresyon algoritmasının sansürlü veriler üzerinde en etkin çalışan algoritma olduğu tespit edilmiştir. Bununla birlikte örneklem büyüklüğüyle doğru orantılı olarak sınıflandırma performanslarının da iyileştiği ortaya konmuştur.

Çelik vd. (2015) tarafından yapılan çalışmanın amacı ise birliktelik analizi algoritmalarını metin madenciliği disiplini içerisinde uygulayarak, Yönetim Bilişim Sistemleri (YBS) alanında ele alınan konuların önem sırasının belirlenmesidir. Buna bağlı olarak yüksek kalitedeki bilgilerin YBS alanında yazılmış bilimsel makalelerdeki anahtar kelimelerden çıkarılması hedeflenmiştir. Sonuç olarak çoklu disiplin olan YBS'nin çalışma alanlarının önem sıralaması ve bilgi haritaları ortaya konulmuştur.

A. S. Yüksel ve Tan (2018) tarafından yapılan çalışmada amaç sosyal ağlar üzerinden daha yararlı bilgilere ulaşmak, kişilerin yaşadıkları çevrede gitmeyi planladıkları yerlerle ilgili olarak görüşlerini otomatik olarak yorumlamak amacıyla, metin madenciliği ve duygu analizi teknikleri kullanılarak bir karar destek sisteminin geliştirilmesidir. Olumlu ve olumsuz görüşlerin gerçek zamanlı belirlenerek çalışma kapsamında duygu analizleri yapılmıştır. Metin madenciliği uygulamaları için kullanılan veri setleri, bireylerin dünya çapındaki şehirlerdeki yemek, alışveriş ve gezi için çeşitli işletmeleri keşfetmesine olanak tanıyan Foursquare sosyal ağ platformundaki kullanıcı yorumlarından oluşturulmuştur. İkili sınıflandırmada %81,97 olan başarı oranı üçlü sınıflandırmada ise %84,49 olarak ölçülmüştür.

Gizli Dirichlet Ayırımı (GDA), ürün özelliklerini kullanıcı değerlendirmelerinden elde etmek için kullanılan en yaygın konu modelleme yöntemidir. Ekinci ve İlhan Omurca (2017) tarafından yapılan çalışma Türkçe otel yorumlarının analizi, GDA'nın özellik çıkarmada başarılı olduğunu göstermiştir. Çalışma aynı zamanda özellik tabanlı duygu analizinin en zor görevlerinden biri olan ve duyguların ifade edildiği başlık olarak tanımlanan özelliklerin çıkarılması konusunda önemli bir çalışma olmuştur. Son dönemdeki araştırmalarda konu modelleme, duygu analizine benzer şekilde yaygın olarak kullanılan denetimsiz bir yöntem olarak ortaya çıkmaktadır. Bu çalışma, etkili özellik çıkarımı amacıyla duygu analizi ve konu modelleme tekniklerinin başarılı bir entegrasyonunu sağlamıştır.

Göker ve Tekedere (2017) tarafından yapılan çalışma ile FATİH projesine yönelik çevrimiçi görüşlerin metin madenciliği teknikleri kullanılarak otomatik olarak tespit edilmesi amaçlanmıştır. Çalışma iki aşamada yürütülmüştür. Birinci aşamada, internet ortamında dağıtık olarak bulunan ve yapısal olmayan verilerin belirli ön işleme süreçlerinden geçirilerek yapısal hale dönüştürülmesi amaçlanmıştır. İkinci aşamada, yapısal veri kümesine dönüştürülen veriler üzerinde makine öğrenmesi algoritmaları kullanılarak yorumların otomatik kategorize edilmesi amaçlanmıştır. Deneyde FATİH projesine ilişkin toplam 444 görüşe ait metinler tf-idf ağırlıklandırma tekniği kullanılarak vektörlere dönüştürülerek analiz edilmiştir. Çalışma daha sonra sınıflandırma algoritmalarını model performans başarımları ölçütlerine göre karşılaştırılarak sürdürülmüştür. Sonuçlar, kullanılan tüm metin sınıflandırma algoritmalarının başarımlarının %80'i aştığını ancak Sıralı Minimal Optimizasyon Algoritmasının değerlendirilen algoritmalar arasında %88,73 ile en yüksek model başarımlarına ulaştığını göstermiştir.

Aytekin ve Mayda (2013) yaptıkları çalışmada sosyal medyadaki rekabet analizi bağlamında karşılaştırma görevine özel olarak fikir madenciliği uygulamaları için bir model oluşturmuşlardır. Veri toplamak için karşılaştırma siteleri, YouTube ve teknoloji forumları gibi çeşitli kaynaklardan karşılaştırma ifadeleri içeren 100 yorum manuel olarak toplanmıştır. Toplanan bu yorumlar ile model için bir test veritabanı oluşturma işleminden sonra modelin etkililiği duyarlılık kriterlerine göre değerlendirilmiş ve değerlendirme sonuçları analiz edilmiştir. Çalışma kapsamında tespit edilen hususlardan biri yazım denetimi ve gövde analizi gibi hususların fikir madenciliği çalışmalarının performansını olumsuz etkileyen faktörler olduklarıdır. Yorumların içerdikleri kelime sayısının da yazım denetimi gibi model başarımları üzerinde olumsuz bir etkiye sahip olduğu belirtilmiştir. Bu bağlamda aşırı uzun ve kısa yorumların model performansı üzerinde olumsuz etkiye sahip olduğu ifade edilmiştir.

Kaşıkçı ve Gökçen (2014) yaptıkları çalışma ile kullanıcı tarafından belirlenen web sitelerinin e-ticaret sitesi olup olmadıklarını belirlemek amacıyla içeriklerini analiz edebilmek amacıyla metin madenciliği tekniklerini kullanan bir yazılım uygulaması geliştirmişlerdir. Araştırmada metin sınıflandırma amacıyla çeşitli kaynaklardan veriler toplanmış ve işlenmiştir. Verilere k-NN En Yakın Komşu ve Naive Bayes sınıflandırma algoritmalarının uygulamalarından elde edilen sonuçlar değerlendirilerek

karşılaştırılmıştır. Daha başarılı sonuçlar veren algoritma seçilerek Java programlama dili kullanılarak geliştirilen masaüstü uygulama arayüzüne entegre edilmiştir. Çalışmada, kullanılan metin sınıflandırma algoritması Naïve Bayes'in doğruluk ve etkililik açısından k-NN En Yakın Komşu algoritmasından daha iyi performans gösterdiği gözlenmiştir.

Sebastiani (2002) tarafından yapılan çalışmada makine öğrenimi alanında metin kategorizasyonu için kullanılan temel metodolojiler incelenmektedir. Belge gösterimi, sınıflandırıcı oluşturma ve sınıflandırıcı değerlendirme olmak üzere üç temel bileşenle ilgili zorlukların detaylı olarak incelenmiştir. Araştırma, çeşitli sınıflandırma algoritmaları ve metin madenciliği yöntemleri kullanılarak metinleri sınıflandırmaya çalışarak bu algoritmaların etkinlik oranlarının ve hassasiyet ölçümlerinin karşılaştırılmasını içermektedir.

Keivandarian ve Carvalho (2023) tarafından yapılan çalışma duyguları sınıflandırmaya yönelik metodolojileri ve süreçte karşılaşılan zorlukları incelemektedir. Hem makine öğrenmesine dayalı yaklaşımların hem de sözlük tabanlı yaklaşımların kapsamlı bir analizi sağlanmaktadır. Çalışma hem makine öğrenimini hem de derin öğrenmeyi içeren çeşitli sınıflandırma yöntemlerinin üzerinde durarak duygu analizini derinlemesine incelemektedir. Bu incelemeyi yaparken belge, cümle ve özellik düzeyleri de dahil olmak üzere farklı duygu analizi düzeylerinin de üzerinde durulmuştur. Çalışma aynı zamanda veri toplama ve özellik çıkarma tekniklerinin yanı sıra duygu analizinde karşılaşılan ironi, alaycılık ve öznel duygu yoğunluğu gibi engelleri de ele almaktadır. Bulgular, SVM, RF ve Naive Bayes gibi makine öğrenimi algoritmalarının etkinliğini vurgularken aynı zamanda LSTM ve BERT gibi derin öğrenme modellerinin de tercih edildiğine de dikkat çekmektedir. Duygusal sınıflandırma alanının hala keşfedilmemiş bir alan olduğu ifade edilmektedir.

Hossain vd. (2021) tarafından yapılan çalışmanın amacı Bangladeş'teki gazete manşetlerinin örüntüsünü ve önceliklerini metin analitiği tekniklerini kullanarak incelemektir. Çalışmanın kapsamı, The Daily Star gazetesinin 2018 ve 2019 yıllarındaki ön sayfa manşetleriyle sınırlıdır. Çalışmada verileri analiz etmek için kelime bulutları, duygu analizi ve küme analizi kullanılmıştır. Çalışma sonucunda, yapılan duygu analizinde olumsuz duyguların her iki yılda da olumlu duygulardan daha sık ortaya çıktığı gözlenmiştir. Küme analizi, belirli kelimelerin benzer olduğunu ve aynı gruba ait

olduğunu göstererek dönemin sosyal ve siyasi bağlamına işaret etmektedir. Genel olarak bu çalışma, metin madenciliği tekniklerinin bir ülkenin sosyal, siyasi ve kanun ve düzen durumu hakkında nasıl içgörü sağlayabileceğini göstermektedir.

Hu ve Liu (2004) tarafından yapılan çalışma da çok fazla sayıda müşteri yorumunu veri kümesi olarak kullanmışlardır. Bu veri kümesinin bir özetini çıkaran bir modeli veri madenciliği ve doğal dil işleme tekniklerini kullanarak özellik tabanlı olarak geliştirmişlerdir. Uygulama için bir den çok teknik önerisinde bulunmuşlardır. Yapılan çalışma tavsiye edilen yaklaşımların görevi başarmada etkin olduğunu göstermiştir. Özet modelin oluşturulması iki fazda gerçekleştirilmiştir. Birinci fazda müşterilerin fikir beyan ettikleri ürün özelliklerinin belirlenmesi ve bu özellikler hakkında yapılan değerlendirmelerin sıklığının ölçülmesi gerçekleştirilmiştir. İkinci fazda ise her özellik için kaç müşteri değerlendirmesinin olumlu veya olumsuz olduğunu belirlenmesi işlemi gerçekleştirilmiştir. Özellik çıkarımı ve fikir madenciliği esasen çalışmanın temelini oluşturmaktadır. Müşteri yorumlarından oluşan veri setlerinde özellik çıkarımı teknikleri uygulanmış ve hangi ürün özelliklerine odaklanılması gerektiğine dair bir çıktıya ulaşılmaya çalışılmıştır. Fikir madenciliği ile de bu özellikler özelindeki pozitif veya negatif kutuplar tespit edilmeye çalışılmıştır.

İKİNCİ BÖLÜM

VERİ VE METİN MADENCİLİĞİ, DUYGU ANALİZİ

KAVRAMSAL ÇERÇEVE

Bu bölümde veri madenciliği ve onun bir alt çalışma alanı olarak metin madenciliğine dair genel kavramsal çerçeve çizilecektir. Bu kavramsal çerçeve içerisinde çalışma alanına ait nirengi noktası sayılabilecek, veri madenciliği süreçleri, yapısal olmayan verinin önemi ve analitiği, metin madenciliği süreçleriyle birlikte, bu çalışmada yapılan metin madenciliği sınıflandırma algoritmalarına yer verilecektir. Metin madenciliği süreçleri, ön işleme, özellik çıkarımı, sınıflandırma fazlarının teorik altyapısı detaylandırılacak ve doğal dil işleme, bilgi çıkarımı, bilgi çekme gibi metin madenciliğine direk temas eden kavramlar ele alınacaktır. Bölüm sonunda duygusal polarite ölçümü amacıyla kullanılan duygu analizi yöntemlerine değinilecektir.

2.1 Veri Madenciliği Genel Tanım ve Kavramsal Çerçeve

Veri madenciliği, büyüyen veri hacimleri ve çeşitliliğiyle başa çıkma ihtiyacı karşısında ortaya çıkan ve verilerden bilgi çıkarmayı amaçlayan bir süreçtir. Bu sürecin aşamaları arasında işletme probleminin tanımlanması, verilere erişim, verilerin yapılandırılması ve veri ambarının oluşturulması, uygun model ve algoritmanın seçimi, uygulama ve sonuçların değerlendirilip sunumu yer almaktadır. Veri madenciliği teknolojide yaşanan gelişmelere paralel olarak işletmelerde depolanan veri sayısının artmasıyla birlikte her geçen gün değişik uygulamalarıyla literatürde yerini almaktadır (Çınar & Silahtaroglu, 2012, s. 310).

Veri madenciliği, gizli örüntüleri, korelasyonları ve ilişkileri ortaya çıkarmak için büyük veri kümelerinin sistematik olarak araştırılması ve analizidir. Veri madenciliği, verilerden değerli bilgiler elde etmek için makine öğrenimi algoritmalarının, istatistiksel yöntemlerin ve veri görselleştirme tekniklerinin kullanılmasını kapsayan süreçleri içerir.

Tahmine dayalı analiz ise gelecekteki olaylar veya davranışlar hakkında bilinçli tahminler yapmak için geçmiş verileri kullanmayı içerir.

Veri madenciliği, verileri farklı kategoriler veya sınıflar halinde organize etmeyi ve benzer niteliklere sahip verileri birleştirmeyi içerir. Sınıflandırma, verilerin belirli nitelikleri baz alınarak gruplandırılması anlamına gelir ve genellikle karar ağaçları, destek vektör makineleri veya yapay sinir ağları gibi algoritmaların kullanılması yoluyla yürütülür. Kümeleme veya keşifsel veri analizi olarak adlandırılan denetimsiz sınıflandırmada, etiketli veriler mevcut değildir (Everitt vd., 2001; Jain & Dubes, 1988). Kümelemenin amacı, aynı olasılık dağılımından üretilen gözlemlenmemiş örneklerin doğru bir karakterizasyonunu sağlamak yerine, sonlu ve etiketsiz bir veri kümesini sonlu ve ayırık bir doğal gizli veri yapıları kümesine ayırmaktır (Baraldi & Alpaydin, 2002; Cherkassky & Mulier, 1998).

Bununla birlikte veri madenciliği aynı zamanda veri kümeleri içindeki anormal veya gürültülü veri noktalarının tanımlanması ve düzeltilmesi, anomali ve sahtekarlık tespiti, kalite kontrol ve ağ güvenliği dahil olmak üzere çeşitli alanlarda çok önemli bir süreçtir. Veri temizleme, veri kümelerindeki hataların, eksikliklerin ve gürültünün tanınmasını ve düzeltilmesini içerir. Veri temizleme veya ayıklama olarak da adlandırılan veri temizleme, verilerin kalitesini artırmak için verilerdeki hataların veya tutarsızlıkların tespit edilmesi ve giderilmesi anlamına gelir. Dosyalar ve veri tabanları gibi tekil veri koleksiyonlarında veri kalitesi sorunları mevcuttur (Jinlin Wang Xing Wang, 2020, s. 1053).

İş zekâsı kavramı perspektifinden bakıldığında veri madenciliği, işlenmemiş verilerin stratejik karar alma sürecine bilgi sağlamak için değerli içgörülere dönüştürülmesini içeren bir tekniktir. İşletmeler, müşteri davranışını ve pazar eğilimlerini analiz ederek operasyonel verimliliği artırabilir, satış sonuçlarını tahmin edebilir ve pazarlama yaklaşımlarını geliştirebilir. Sonuçta veri madenciliği, veri analizinden elde edilen pratik bilgiler sunarak iş zekâsının daha geniş hedefleriyle uyumludur.

Veri tabanı yönetimi sistemleri (DBMS) bakış açısından veri madenciliği, büyük veri tabanlarından değerli bilgi ve öngörülerin çıkarılmasına yönelik tekniklerin bir koleksiyonu olarak kabul edilir. Verileri verimli bir şekilde almak ve analiz etmek için indeksleme, sorgu optimizasyonu ve veri ambarı yöntemlerinin kullanılmasını içerir. Veri

madenciliği araçları, çok büyük veri tabanlarından karar vericiler için yeni bilgiler üretir. Bu üretimin çeşitli mekanizmaları arasında verilerin soyutlanması, toplanması, özetlenmesi ve karakterize edilmesi yer almaktadır (Carbone, 2000, s. 1872).

2.2 Veri Madenciliği Uygulama Alanları

Kapsamlı veri yığınlarından değerli bilgilerin ve modellerin çıkarılmasını kapsayan veri madenciliği uygulaması, çok sayıda akademik ve profesyonel alan içerisinde önemli ölçüde uygulama imkânı bulmuştur. İşletme yönetimi disiplini, veri madenciliği uygulamasıyla çeşitli yönlerden kapsamlı bir şekilde iç içe geçmiştir. Veri madenciliği yöntemleri, müşteri davranışlarının analiz edilmesi, ürün pazarlama stratejilerinin optimize edilmesi, envanter seviyelerinin yönetilmesi ve finansal risklerin tahmin edilmesi dahil olmak üzere temel işletme fonksiyonlarının optimizasyon ve yönetiminde önemli bir rol oynamaktadır. Veri madenciliği teknikleri kullanılarak müşterilerin satın alma kalıpları ve davranışları analiz edilerek müşteriye özel pazarlama stratejileri formüle edilebilir. Tıp ve sağlık bilimleri alanı da veri madenciliğinin önemli uygulama alanlarından biridir. Veri madenciliğinde kullanılan teknikler, hastalık teşhisi, ilaç keşfi, epidemiyolojik araştırmalar ve hasta takip sistemleri dahil olmak üzere bu alanda önemli bir rol oynamaktadır. Veri Madenciliği, sağlık ve tıp alanındaki büyük veritabanlarından değerli bilgileri ortaya çıkartarak hem tıp açısından hem de hizmet kalitesinin artırılması açısından büyük katkılar sağlamaktadır (Yıldırım vd., 2008, s. 433).

Epidemiyoloji ve halk sağlığı araştırmalarında çeşitli analitik tekniklerin kullanılması, hastalıkların yayılma modellerinin incelenmesine yardımcı olur. Sağlık verilerinin zaman içinde sürekli izlenmesi ve analiz edilmesi yoluyla, hasta bakım ortamlarında komplikasyonlar önceden belirlenebilir. Doğal dil işleme kullanılarak değerli bilgiler elde etmek için elektronik sağlık kayıtları çıkarılırken, örüntü tanıma yöntemleri de ilaç geliştirme sürecinde yeni ilaç adaylarının keşfedilmesine yardımcı olmaktadır. İlaç etkinliği ve yan etkilerine ilişkin tahminler, kimyasal bileşik özelliklerinin kümelenmesiyle yapılır ve hastane kaynakları, lojistik regresyon ve optimizasyon algoritmaları kullanılarak optimize edilir. Bulaşıcı hastalık salgınlarının erken tespiti, anormallik tespiti ile kolaylaştırılıyor ve veri görselleştirme araçları, karmaşık sağlık verilerini daha anlaşılır hale getirerek karar alma süreçlerini kolaylaştırmaktadır. Benzer şekilde genomik ve biyoinformatiğin akademik disiplinlerinde, veri madenciliği

metodolojilerinin kullanılması, biyolojik verilerin, özellikle DNA dizileri ve gen ifade verileriyle ilgili analiz ve yorumlanmasında oldukça işlevsel bir rol oynamaktadır.

Bunun yanında veri madenciliği sosyal bilim araştırmacıları tarafından karmaşık sosyal olguları anlamak ve incelemek için kullanılan önemli bir araçtır. Sosyal bilimlerde veri madenciliği seçmen davranışı, iletişim ağları ve ekonomik eğilimler gibi sosyal olgulardaki girift kalıpları belirlemek için kullanıldığı gibi insan davranışı, sosyolojik bağlantılar ve toplumsal çerçevelerle ilgili verilerden de önemli çıkarımlar yapmak için kullanılmaktadır. Bu çıkarımlar toplumsal ölçekte yapılacak, pazarlama, sağlık hizmetleri ve sosyal bilim araştırmaları alanındaki diğer alanlar gibi çeşitli alanlarda oluşturulacak politikalar geliştirmek için önemli bir enstrümandır.

Sosyal medya madenciliği, çeşitli toplumsal boyutlara ilişkin verilere hızlı ve kapsamlı erişim sağlar. Bu veriler sosyal bilimler alanındaki araştırmalar için belirli bir zaman dilimi içinde toplumsal değişimleri gözlemlemek ve kavramak amacıyla kullanılmaktadır. Örneğin, sosyal medya verileri seçim süreçlerindeki kamuoyu tercihlerini, halk sağlığıyla ilgili acil durumları ve popüler kültürdeki eğilimleri incelemek için kullanılmaktadır. Bu tür veriler, araştırmacılara geleneksel araştırma yöntemleri veya veri toplama teknikleriyle elde edilemeyecek benzersiz çıkarımlar yapma imkânı vermektedir.

Bilhassa halk sağlığı alanında, araştırmacılar hastalık bulaşışını izlemek, sağlık hizmetlerine erişimdeki eşitsizlikleri ortaya çıkarmak ve sağlıkla ilgili davranışları etkileyen belirleyicileri analiz etmek için veri madenciliği tekniklerini kullanmaktadırlar.

Kentsel planlama ve gelişime dair içgörü kazanmak amacıyla, mekânsal verilerin incelenmesi, hareket modellerinin analiz edilmesi, trafik dinamiklerine ve metropol alanlardaki kaynak tahsisinin optimizasyonu gibi konularında veri madenciliği süreçlerinden faydalanılmaktadır.

Diğer yandan veri madenciliği uygulamaları pazarlama disiplini içerisinde de yaygın bir kullanım alanına sahiptir. Örneğin veri madenciliği yöntemleri kullanılarak müşteri segmentasyonu, demografik veriler, satın alma modelleri ve diğer çeşitli değişkenler analiz edilerek müşteriye göre özelleşmiş pazarlama stratejilerinin geliştirilmesi bunlardan biridir. Diğer bir deyişle makine öğrenimi algoritmaları ile müşterilerin alışveriş alışkanlıkları ve tercihlerinin analizi yapılır. Bu analizler sonucunda genellikle

amaç bir müşteri profilinin oluşturulmasıdır. Müşteri profili, pazarlamacının bu müşterinin ihtiyaçlarını karşılamak için doğru strateji ve taktiklere karar verdiği bir müşteri modelidir. Veri madenciliği tekniklerinin sistematik olarak uygulanması, bilgi yönetimi sürecini geliştirecek ve pazarlamacıları müşterilerine daha iyi hizmet sunmalarını sağlayacak daha iyi bilgilerle donatacaktır (Shaw vd., 2001, s. 127).

Kişiselleştirilmiş tavsiye sistemleri veri madenciliği tekniklerinin pazarlama uygulamaları içinde önemli bir konuma sahiptir. Amazon.com gibi e-ticaret platformları, müşterilerine özelleştirilmiş ürün önerileri sunmak için veri madenciliği metodolojilerini kullanır. Tavsiye motoru adı verilen bu uygulamalar, müşterilerin geçmiş etkinliklerini ve tercihlerini analiz ederek, karşılaştırılabilir kullanıcıların tercihlerinden içgörüler elde ederek öneriler üretir. Bu şekilde müşterilere ilgi duyabilecekleri ek ürünler veya daha pahalı alternatifler önerilir. Tavsiye Sistemleri tipik olarak insan bilgisayar etkileşimi veya bilgi edinme (information retrieval) gibi diğer komşu alanlardan teknikler ve metodolojiler uygulamaktadır (Amatriain & Pujol, 2015, s. 227).

Müşteri sadakati ve tutundurma stratejilerinin belirlenmesi de pazarlama disiplini içinde veri madenciliği tekniklerinin etkin kullanım alanlarından biridir. Zira işletmeler müşteri memnuniyeti, müşteri deneyimi, müşteri ilişkileri yönetimi ve müşteri değeri için etkili stratejiler uygulayarak daha yüksek müşteri sadakati seviyelerine ulaşabilir, bu da müşteriyi elde tutma, karlılık ve rekabet avantajı ile sonuçlanabilir (Rane vd., 2023, s. 428).

Müşteri sadakati ve elde tutma stratejileri, mevcut müşterilerin uzun soluklu olarak devamlılığını sağlamak amacıyla geliştirilmektedir. Bu stratejilerin etkinliğini değerlendirmek ve geliştirmek için veri madenciliği yöntemleri kullanılmaktadır. Örneğin, müşteri kaybı (churn) analizi yoluyla hangi müşterilerin terk etme olasılığı taşıdığını görmek mümkündür. Bu analiz, müşterilerin satın alma alışkanlıkları, şikâyet geçmişi ve müşteri hizmetleriyle olan etkileşimleri gibi çeşitli veri noktalarının incelenmesini içerir. İşletmeler, erken uyarı sistemlerini kullanarak, terk etme potansiyeli olan müşterilere hedeflenen teşvikler ve indirimler sunarak müşteri kaybını önlemek için proaktif önlemler alabilir. Ayrıca müşteri sadakati girişimlerinin başarısı, veri madenciliği teknikleriyle değerlendirilerek bu programların optimizasyonuna olanak sağlanmaktadır. İşletmeler, pazarlama kampanyalarını ve sadakat ödülleri bireysel

müşterilere göre düzenleyerek memnuniyet ve sadakat düzeylerini artırabilir. Sonuçta veri madenciliği, müşterilerin tercihlerinin ve beklentilerinin anlaşılmasında önemli bir rol oynayarak müşteri ilişkilerinin güçlenmesine ve müşteri sadakatinin artmasına katkı sunar.

Pazarlama disiplini içerisinde veri madenciliği süreçlerinin etkin kullanıldığı alanlardan biri de birliktelik kuralları algoritmalarının yoğun olarak kullanıldığı pazar sepeti analizi çalışmalarıdır. Pazar sepeti analizi, müşterilerin bir diğerinin yanında birlikte satın aldıkları ürünlerin incelenmesini içerir. Bu yöntem, müşterilerin alışveriş alışkanlıklarındaki kalıpları belirlemek için veri madenciliği algoritmalarının kullanıldığı perakende sektöründe yaygın olarak kullanılmaktadır. Örneğin süpermarketler, yaygın olarak eşleştirilmiş ürünleri belirlemek ve bunları stratejik olarak bir araya getirmek veya paket halinde sunmak için pazar sepeti analizini kullanabilir. Bu yaklaşım müşterilerin alışveriş deneyimini zenginleştirir ve satışları artırır. Bununla birlikte pazar sepeti analizi, sıklıkla ilişkilendirilen ürünler için çapraz satış girişimleri gibi promosyon stratejilerinin hazırlanmasında da etkilidir. Veri madenciliği, büyük veri kümelerini analiz ederek değerli içgörüler ve bağlantıları ortaya çıkarır ve pazarlama stratejilerini optimize eder. Sonuçta pazar sepeti analizi müşteri davranışını daha derinlemesine anlamak için değerli ve efektif bir araç olarak değerlendirilebilir.

Rakip analizi çalışmaları da veri madenciliği yöntemleri için yaygın çalışma alanlarından biridir. Rakip analizi veya rekabet analizi adı verilen bu çalışmalar, rakip işletmelerin pazarlama taktiklerinin ve müşterilerden gelen geri bildirimlerin incelenmesi yoluyla rekabet avantajı elde etmeyi amaçlar. İşletmeler, sosyal medyayı ve diğer çevrimiçi kaynakları değerlendirmek için veri madenciliği tekniklerini kullanarak rakiplerinin güçlü ve zayıf yönlerini tespit edebilmektedir. Bu analiz, sosyal medya geri bildirimlerine ve çevrimiçi müşteri yorumlarına dayalı olarak müşteri hizmetlerinde ve ürün kalitesinde iyileştirilecek alanların belirlenmesi gibi iş stratejilerinin geliştirilmesine yardımcı olur. Ayrıca, rekabet analizi, fiyatlandırma stratejilerinin geliştirilmesinde ve rakipler tarafından gözden kaçırılan müşteri segmentlerinin özel kampanyalar aracılığıyla hedeflenmesi gibi, kullanılmayan pazarlama fırsatlarının belirlenmesinde önemli bir rol oynar. Veri madenciliğinin kullanılması, işletmelerin kapsamlı veri kümelerini hızlı ve etkili bir şekilde analiz etmelerini sağlar ve sonuç olarak pazar payının ve rekabet gücünün artmasına yol açar. Nihai tahlilde rekabet analizi, bir firmanın faaliyet gösterdiği

çevreyi anlaması için gereklidir. Firmanın rekabet avantajı elde etmek ve bunu sürdürmek ve sonucunda performansı artırmak için benimsediği stratejiler üzerinde önemli bir etkiye sahiptir (Deshpandé & Gatingon, 1994, s. 271).

Veri madenciliği yöntemlerinin özellikle tahminleme algoritmalarının yoğun kullanılarak yürütülen bir başka çalışma alanı da satış tahminleridir. Satış tahmini, daha kesin ve güvenilir tahminlere olanak tanıyan veri madenciliği tekniklerini kullanarak geçmiş satış verilerinin analizine dayanarak gelecekteki satışları tahmin etmeyi içerir. Örneğin zaman serisi analizi, geçmiş satış verilerini inceleyerek gelecekteki satış eğilimlerini belirlemek için kullanılabilir. Bu tahminler, etkili envanter yönetimi ve üretim planlamasının yanı sıra pazarlama kampanyalarının planlanması için de çok önemlidir. Veri madenciliği, büyük veri kümelerinden değerli kalıpların ve ilişkilerin çıkarılmasına yardımcı olarak satış tahminlerinin doğruluğunu artırır. Ayrıca satış tahmini, belirli dönemlerde talepleri artan ürün gruplarının belirlenmesini ve stok seviyelerinin de talebe göre optimize edilmesine olanak sağlayarak müşteri talebine ilişkin daha derin bir anlayış elde edilmesine yardımcı olmaktadır. Sonuç olarak, satış tahmini, geçmiş verilere, pazar analizine ve diğer ilgili bilgilere dayanarak bir ürün veya hizmetin gelecekteki satışlarını tahmin etme sürecidir. Satış tahmini, bir şirketin iş gücünü, nakit akışını ve kaynaklarını nasıl yönetmesi gerektiği konusunda fikir verir (Cheriyen vd., 2018, s. 53).

Müşteri Yaşam Boyu Değeri (MYD), müşterilerin işletmeyle olan tüm ilişkileri boyunca işletme için üretmeleri beklenen toplam değeri ifade eder. MYD analizi veri madenciliği teknikleri yoluyla bu değerin tahmin edilmeye çalışılmasıyla ilgili süreçleri ifade eder. MYD hesaplamaları, müşterilerin geçmiş satın alma kayıtlarının, demografik verilerinin ve müşteri hizmetleriyle olan etkileşimlerinin incelenmesinden elde edilir. Bu tahminler müşteri segmentasyonunda ve pazarlama planlarının oluşturulmasında kullanılır. MYD'si yüksek olan müşterilere odaklanmak, uzun vadede işletmenin genel karlılığını artırma potansiyeline sahiptir. Ayrıca MYD analizi, müşteri bağlılığını ve elde tutulmasını artırmayı amaçlayan stratejilerin geliştirilmesinde de etkilidir. Veri madenciliği, kapsamlı veri kümelerini tarayarak önemli eğilimleri ve korelasyonları ortaya çıkarır, böylece MYD tahminlerinin doğruluğunu artırır. Ayrıca, pazarlama girişimlerinin etkinliğini değerlendirmek için de MYD analizinden yararlanılmaktadır. Örneğin, belirli kampanyaların MYD üzerinde nasıl bir etkisi olduğu analiz edilerek, bu kampanyaların optimize edilmesi sağlanabilir. Sonuç olarak, MYD analizi, müşteri ilişkileri yönetiminde

doğru müşterilerin kazanılması, büyütülmesi ve elde tutulması için bir ölçüt olarak hızla kabul görmektedir (Venkatesan & Kumar, 2004, s. 106).

Çevrimiçi müşteri yorumlarından ve müşteri geri bildirimlerinden müşteri memnuniyet derecelerinin anlaşılması da veri madenciliği uygulamaları için önemli çalışma alanlarından biri olarak öne çıkmaktadır. Müşteri görüşleri analizi çalışmalarında müşteri beklentilerini ve sorunlarını anlamak amacıyla sosyal medya ve müşteri anketi verilerinin veri madenciliği metodolojilerinden yararlanılarak veri analitiği süreçlerine tabi tutulması söz konusudur. Bu analitik çalışmaları ürün ve hizmetlerin kalitesinin artırılmasına yardımcı olur. Örneğin, olumsuz geri bildirimlerin ardındaki nedenler araştırılarak bu sorunları çözmeye yönelik stratejiler geliştirilebilir. Ayrıca müşteri geri bildirimlerinin analizi, müşteri memnuniyetinin artırılması açısından önemlidir. Ayrıca, pazarlama girişimlerini ve satış sonrası hizmet süreçlerini geliştirmek için müşteri geri bildirimlerinden yararlanılmaktadır. Kapsamlı veri kümelerinin işlenmesi yoluyla veri madenciliği, önemli kalıpları ve bağlantıları ortaya çıkararak müşteri yorumlarının daha verimli bir şekilde analiz edilmesini sağlar. Sonuç olarak, Metin madenciliği kullanılarak yapılan müşteri geri bildirim analizi, büyük ölçüde metne gömülü duyguları otomatik olarak tahmin etmek için daha doğru modeller geliştirmeye odaklanmıştır. Bununla birlikte, müşteri geri bildirim verilerini düzenli olarak yönetebilen şirketler, rakiplerine kıyasla ortalama %5 daha üretken ve %6 daha kârlıdır (Mishra & Vegad, 2024, s. 636).

Pazardaki yeni satış ve satın alma trendlerinin ve eğilimlerin belirlenmesi süreçlerinde de veri madenciliği uygulamaları önemli bir rol oynamaktadır. Büyük veri setlerinin kapsamlı analizi yoluyla pazarda ortaya çıkan eğilimleri ve tüketici davranışlarını belirleme uygulaması, pazarlama trendi tespiti olarak bilinir. Veri madenciliği tekniklerinden yararlanmak bu süreci geliştirerek işletmelerin proaktif stratejiler oluşturmaya olanak tanımaktadır. Örneğin, işletmeler sosyal medya gibi çevrimiçi platformları inceleyerek popüler ürün ve hizmetleri belirleyebilir ve bunları daha sonra yeni ürün geliştirme ve pazarlama kampanyaları için önemli bir girdi olarak kullanabilir. Veri madenciliği, geniş veri kümelerinden değerli kalıpların ve bağlantıların çıkarılmasını kolaylaştırır, böylece pazarlama stratejilerinin verimliliğini artırır. Pazarlama trendlerinin anlaşılması müşteri beklentilerinin karşılanması ve müşteri memnuniyetini artırılması için çok önemlidir.

2.3. Veri Madenciliği Süreçleriyle İlişkili Disiplinler

Veri madenciliği çeşitli akademik alanlarla bağlantılıdır ve her disiplin, veri madenciliğinin metodolojilerini ve uygulamalarını geliştirmede önemli bir rol oynar.

2.3.1 İstatistik Veri Madenciliği İlişkisi

Veri madenciliği ve istatistik arasındaki karmaşık ilişki, veri analizine ve matematiksel tekniklerin kullanımına olan ortak odaklanmalarından kaynaklanmaktadır. İstatistik, birçok veri madenciliği tekniğinin altında yatan teorik temeli ve analitik araçları sağlar. İstatistik ve matematiksel yöntemlerin yardımıyla veri madenciliği giderek önemli ve kullanışlı hale gelmiştir. Veri analizi yapmak için kullanılan başlıca istatistiksel ölçütlerden bazıları ortalama, medyan, varyans, standart sapma, varyans analizi, korelasyon ve regresyondur (Ribeiro vd., 2017, s. 156).

Veri madenciliği alanı, geleneksel istatistiksel tekniklerin ilkeleriyle tutarlı bir şekilde uyumlu olmayan kapsamlı ve karmaşık veri setlerinin analiziyle ilgili zorlukların üstesinden gelebilmek amacıyla geliştirilmiştir. Veri madenciliği ve istatistik arasındaki kesişim, makine öğrenimi algoritmalarının geliştirilmesinde ve kullanılmasında çok önemli bir rol oynamaktadır. Makine öğrenimi algoritmalarında, verinin altında yatan bilgi yinelenen kalıpları araştırmak veya gizli nedensel ilişkileri keşfetmek için araştırılan ve analiz edilen verilerin kendisinden çıkarılır. Tahmin modeli bilgiyi tümevarımsal bir süreçle çıkarır; girdi veriler ve muhtemelen beklenen çıktının ilk örneğidir, makine daha sonra aynı sonucu elde etmek için izlenecek algoritmayı öğrenecektir (Ciaburro & Iannace, 2021, s. 1).

Veri madenciliği yöntemleri regresyon analizi, hipotez testi, korelasyon analizi ve örnekleme teknikleri gibi istatistiksel yöntemlerden yoğun olarak faydalanmaktadır. Doğrusal regresyon y olarak isimlendirilen sayısal bir bağımlı değişkenle x olarak ifade edilen bir veya daha fazla bağımsız değişken arasındaki ilişkiyi modelleme yaklaşımıdır (Kılıç, 2013, s. 90). Bu yaklaşım veri madenciliği projelerinde tahmin yapma amacıyla kullanılabilir. Ek olarak veri madenciliği algoritmaları tarafından üretilen sonuçların tutarlılık analizi için istatistiksel hipotez testleri kullanılmaktadır.

İstatistik disiplini veri madenciliği süreçlerinin her aşamasında olduğu gibi veri ön işleme aşamasında da bazı işlemlere sahiptir. Bunlardan en temel olanları eksik verinin yerine ortalama, medyan ya da mod gibi istatistiksel verilerin kullanılması, bir önce ve bir

sonraki verinin kaydırılması ya da eksik veri kaydının silinmesidir (Jayalakshmi & Santhakumaran, 2011, ss. 91-92). Ayrıca veri temizleme ve ön işleme aşamalarında eksik veri analizi ve outlier tespiti gibi istatistiksel teknikler kullanılır. Ön işleme süreci veriyi daha iyi hale getirir ve daha güvenilir sonuçlar alınmasına yardımcı olur. Veri madenciliği süreçlerinin doğruluğunu ve güvenilirliğini test etmek için de istatistiksel metotlardan faydalanılmaktadır. Çapraz doğrulama ve ROC eğrileri bunlara örnektir. Çapraz doğrulama, tahmine dayalı bir modelin genelleme yeteneğini değerlendirmek ve aşırı uyumu önlemek için en yaygın kullanılan veri yeniden örnekleme yöntemlerinden biridir (Berrar, 2018, s. 6).

Veri madenciliği ve istatistik disiplinleri arasında çift yönlü bir ilişki mevcuttur. Veri madenciliği bilimi istatistiksel metotlardan yararlandığı gibi istatistik bilimi de veri madenciliği yöntemlerinden faydalanmaktadır. Veri madenciliği metodolojileri, istatistiksel analiz prosedürlerinin verimliliğini artırır ve kapsamlı veri kümelerinin incelenmesini basitleştirir. Örneğin, kümeleme, verilerin benzer nesnelerden oluşan gruplara bölünmesidir. Küme olarak adlandırılan her grup, kendi aralarında benzer olan ve diğer gruplardaki nesnelere kıyasla farklı olan nesnelerden oluşur (Abbas, 2008, s. 1). Kümeleme algoritmaları, büyük veri setlerinde doğal grupları belirlemek için kullanılır. Bu durum istatistiksel analizlerin daha verimli yapılmasını sağlar. Ayrıca veri madenciliği, veri görselleştirme yöntemlerinin uygulanması yoluyla karmaşık veri yığınlarının anlaşılabilirliğini artırır, böylece istatistiksel çıkarımların yorumlanmasını kolaylaştırır. Özellikle büyük veri analizi ve makine öğrenimi uygulamaları alanında her iki disiplininin kesişim alanlarının oldukça arttığı görülür. Kullanılan analizler incelendiğinde; Kümeleme teknikleri, diskriminant analizi, korelasyon ve regresyon analizi gibi teknikler veri madenciliği problem tiplerinde kullanılan önemli istatistiksel tekniklerden bazılarıdır (Tüzüntürk, 2010, s. 73). Ayrıca karar ağaçları ve lojistik regresyon gibi teknikler hem istatistiksel analizlerde hem de veri madenciliği projelerinde kullanılmaktadır. Veri madenciliği disiplini istatistiksel metodolojilerin entegrasyonundan büyük ölçüde yararlanmıştır. Bu istatistiksel metodolojiler, Naive Bayes sınıflandırıcıları ve k-en yakın komşu algoritması dahil olmak üzere çeşitli veri madenciliği algoritmalarının temelini oluşturur. Ayrıca, veri madenciliği çalışmalarında kullanılan pek çok doğrulama yönteminin istatistiksel tekniklere dayanması veri madenciliği algoritmalarının güvenilirliğini ve kesinliğini de artırmaktadır.

Veri madenciliği ve istatistik disiplinleri arasındaki etkileşim, güvenilir ve doğru bilgi keşfi süreçlerinin oluşturulmasında hayati bir role sahiptir. Veri Madenciliği problem tiplerinin çözümünde istatistiksel teknikler oldukça önemli işlevler görür. Özünde, veri madenciliği tekniklerinin çoğu istatistiksel teknikleri kullanır. Birçok alanda çok sık kullanılan regresyon analizi, korelasyon analizi bu tekniklerden bazılarıdır (Tüzüntürk, 2010, s. 77).

Veri madenciliği tekniklerinin sunduğu güçlü keşif ve tahmin yetenekleri, istatistiksel analizlerle desteklenerek, anlamlı ve güvenilir sonuçlara ulaşılabilir. Örneğin, veri madenciliği teknikleri kullanılarak oluşturulan sınıflandırma modellerinin etkinliği ve performansı, istatistiksel testler yoluyla değerlendirilmektedir. Benzer şekilde, istatistiksel analizden elde edilen çıkarımlar, veri madenciliği algoritmalarının tasarımı ve optimizasyonunu geliştirebilir. İstatistiksel metodolojiler, veri madenciliği modellerinin etkinliğini artırmak için değişkenlerin seçilmesi, öznetelik mühendisliği ve model parametrelerinin ayarlanması gibi görevlerde önemli katkılar sunmaktadır.

Veri madenciliği ve istatistik arasındaki bu güçlü korelasyon çeşitli disiplinlerde açıkça görülmektedir. Örneğin biyoloji çalışmalarında veri madenciliği süreçleri gen verilerinden yola çıkarak gen-hastalık ilişkisini analiz ederken, istatistiksel yöntemler bu ilişki analitiği çalışmalarının ortaya koyduğu modellerin doğruluğunu denetleyici çalışmalarda kullanılmaktadır. Tıp bilimindeki çalışmalarda veri madenciliği çalışmaları hastalık belirtileri ve risk faktörleri arasındaki bağlantı faktörlerinin analizini yaparken istatistik metotlar bu bağlantı faktörlerin tutarlılığını test etmektedir. Benzer şekilde finans sektöründe, portföy optimizasyonu ve risk yönetiminde veri madenciliği tahmin modellerinin kullanılması, örüntü ve kalıpların istatistiksel analiz yoluyla değerlendirilmesini içerir. Veri madenciliği teknikleri kullanılarak geliştirilen hisse senedi fiyat hareketi tahmin modellerinin ve risk tahmin modellerinin doğruluğunu ve güvenilirliğini değerlendirmek için istatistiksel yöntemler kullanılmaktadır. Veri madenciliği ve istatistik disiplinlerinin birlikte kullanımı finans sektöründeki piyasa risklerinin ölçümü ve yönetimi için de oldukça önemlidir.

Bununla birlikte her iki disiplininin de kapsam ve uygulama teknikleri itibarıyla birbirinden farklılaştığı noktalar mevcuttur. Örneğin, istatistik genellikle yapılandırılmış ve görece küçük veri kümeleriyle ilgilenirken, veri madenciliği daha büyük,

yapılandırılmamış ve karmaşık veri kümelerini işlemek için tasarlanmıştır. İstatistiksel metodolojiler sıklıkla verilerin stokastiklik sergilediği ve belirli olasılık dağılımlarına bağlı olduğu önermesine dayanır. Veri madenciliği ise, veri yığınları içindeki örüntüleri tespit etmek için makine öğrenimi gibi teknikler kullanılır. İstatistik öncelikle kabul görmüş hipotezleri test etmeye ve sağlanan verilerle uyumlu modeller oluşturmaya odaklanır. Bunun tersine, veri madenciliği, mevcut verilere dayalı olarak yeni hipotezlerin formüle edilmesine ve kapsamlı veri kümelerinden elde edilen bilgileri kullanarak modellerin geliştirilmesine odaklanır. Diğer bir deyişle istatistiksel analizler genellikle verileri önceden kabul görmüş bir modele uydurmaya odaklanırken, veri madenciliği kalıpları ve ilişkileri daha veri odaklı bir şekilde ortaya çıkarmak için veri kümelerinden bilgilerin otomatik olarak çıkarılması fikrinden hareket eder. İstatistik alanı matematiksel ilkelere, özellikle de verileri analiz etmek için temel bir çerçeve görevi gören olasılık teorisine dayanmaktadır. Bunun aksine veri madenciliği, istatistiksel öğrenme teorisi gibi yenilikçi kavramları tanıtarak, uyarlama, optimizasyon ve bilgisayar bilimlerinden teknikleri birleştirerek istatistik disiplininin bu teorik çerçevesini genişletir.

Temel olarak, istatistiksel analiz tipik olarak verileri belirli bir olasılık dağılımı veya model bağlamında yorumlama konusunda rehberlik sunarken veri madenciliği, model tarafından tasvir edilen ilişkilerin belirlenmesine odaklanarak, modelden bağımsız olarak verilerdeki örüntü ve anomalileri ortaya çıkarmaya çalışmaktadır. İstatistikte sonuçların yorumlanması süreci, bir model veya testin ardından istatistiksel ölçümlerden anlam çıkarılmasını gerektirir. Bu durum genellikle parametre tahminlerinin, güven aralıklarının, p değerlerinin ve diğer istatistiksel test ölçümlerinin analiz edilmesini içerir. Regresyon analizi gibi istatistiksel analizler yapılırken tutarlı bir hipotez testi kapsamında yorumlar yapılır ve sonuçlar belirli varsayımlara göre değerlendirilir.

Veri madenciliği alanında, sonuçların yorumlanması süreci genellikle daha geniş bir araştırma bağlamında yürütülür. Temel amaç, kullanılan spesifik modele bakılmaksızın, veri yığınınındaki mevcut temel kalıpları, bağlantıları veya düzensizlikleri ortaya çıkarmak ve kavramaktır. Yapılan analizlerin merkezinde genellikle ilgili verisel problemin üstesinden gelmek için pratiğe aktarılabilir somut pratik çözümler bulabilmek vardır. Örneğin müşteri segmentasyonu veya ürün öneri sistemleri gibi uygulamalarda, belirlenen modeller kurumsal stratejiler ve karar alma mekanizmaları üzerinde doğrudan etkiye sahip olabilir.

İstatistikte model tabanlı yaklaşımlar genellikle teorik ilkelere dayanırken, veri madenciliğindeki yorumlamalar öncelikle veri özellikleri, gerçek dünya uygulamaları ve tahminlere göre belirlenmektedir. Veri madenciliğinde vurgu, oluşturulan modellerin veya kalıpların iş karar verme süreçlerini nasıl bilgilendirebileceğini ve yönlendirebileceğini anlamakta yatmaktadır.

2.3.2 Makine Öğrenmesi Veri Madenciliği İlişkisi

Makine öğrenimi ve veri madenciliği, kapsamlı veri kümelerinden değerli bilgiler elde etmeyi amaçlayan birbiriyle bağlantılı disiplinlerdir. Bunlar, benzer metodolojileri ve araçları kullanan veri analitiğinin ayrılmaz bileşenleridir. Benzerliklerine rağmen her birinin farklı odak noktaları ve pratik uygulamaları bulunmaktadır. Özünde, makine öğrenimi ve veri madenciliğinin her ikisi de verilerden örüntülerin, ilişkilerin ve bilginin çıkarılmasını içerir. Makine öğrenimi algoritması, tam anlamıyla programlanmadan istenen bir görevi gerçekleştirmek için girdi verilerini kullanan bir hesaplama sürecidir (El Naqa & Murphy, 2015, s. 4). Öte yandan veri madenciliği, makine öğrenimi, içgörü ve veritabanı sistemlerinin kesişme noktasındaki stratejileri içeren önemli veri kümelerindeki tasarımları bulmaya yönelik bir yoldur (Kaithwas & Borkar, 2019, s. 511).

Makine öğrenimi, büyük veri yığınları içindeki kalıpların, eğilimlerin ve örüntülerin otomatik olarak tanımlanmasını kolaylaştırmak için gerekli algoritmaları ve istatistiksel modelleri sunduğundan, veri madenciliği alanında çok önemli bir bileşendir. Veri madenciliğinde sınıflandırma, kümeleme, regresyon ve boyut indirgeme gibi çeşitli makine öğrenmesi yöntemlerinin kullanılması, manuel analiz yoluyla ortaya çıkarılması çok zor olabilecek belirsiz içgörülerin keşfini kolaylaştırmaktadır. Bununla birlikte veri madenciliği de makine öğrenmesi tekniklerinin ilerlemesini ve geliştirilmesini kolaylaştıran büyük veri kümeleri ve problem alanları sunarak makine öğreniminin gelişimi için çok önemli bir altyapı sunar.

Makine öğrenmesi algoritmalarının veri madenciliği çalışmaları açısından önemi birkaç kavram etrafında özetlenebilir. Bunlardan biri olan otomatik modellemedir. Makine öğrenmesi algoritmaları, büyük veri yığınları içindeki karmaşık örüntüleri ve kalıpları otonom bir şekilde ayırt etme yeteneğine sahiptir. Bu sayede makine öğrenimi sorunu daha doğru ve daha hızlı bir şekilde çözmeye yardımcı olan bir tekniktir (Dahiya vd., 2022, s. 1).

Bir diğ er kavram olan tahminleme yeteneđi kapsamında, makine   renmesi algoritmalarının gelecekteki olayları veya davranışları   ng r mek i in ge miř verileri kullanma yeteneđine sahip olması pazarlama stratejileri, m řteri ve pazar b l mlemesi ve satıř tahmini gibi alanlarda  nemli faydalar sađlamaktadır. Makine   renmesi algoritmalarının veri madenciliđi a ısından  nemli hale geldiđi bir diğ er kavram desen tanımadır. Makine   renmesi, veri k meleri i indeki gizli kalıpları tanımlama ve kategorize etme konusunda son derece yetkin bir y ntemler b t n  sunmaktadır. B ylece veri madenciliđi s re lerinde daha dođru ve anlamlı sonu lar elde edilebilmektedir.

Birbiriyle fonksiyonel olarak  ok yakın iliřkileri bulunan iki disiplin arasında odaklandıkları noktalar ve kullanılan teknikler itibariyle de bazı farklılařmalar mevcuttur. Veri madenciliđi, verilerden deđerli i  r lerin sistematik olarak arařtırılmasını ve  ıkarılmasını i erir. Makine   renimi, eđitim ve   renmeye dayalı tahmin  zerine yođunlařır. Veri madenciliđi ise bir ok makine   renimi y ntemi kullanır; makine   renimi ayrıca daha iyi   renme ve dođruluk i in veri madenciliđi y ntemlerini  n iřleme s re lerinde kullanmaktadır (Wang, 2017, s. 55).

Veri madenciliđi, veri  n iřleme, veri arařtırması, modelleme ve skorumlama gibi  eřitli ařamalarını kapsarken, makine   renmesi yaklařımı  ncelikle   renme ve model oluřturma s re lerine odaklanmaktadır. Veri madenciliđi veri yıđınları arasındaki anlamlı ve deđerli iliřkileri ortaya koyarken, makine   renimi veri yıđınları arasındaki iliřkiyi temsil eden bilgilerden yararlanarak modellere iliřkin tahmini sonu lar bulur (Ers z &  ınar, 2021, s. 402).

Makine   renmesi, yapay zek , robot bilimi, dođal dil iřleme ve g r nt  iřleme gibi kullanım alanlarına sahipken, veri madenciliđi genellikle iř zek sı, m řteri iliřkileri y netimi ve pazarlama gibi iř uygulamalarına odaklanır. Makine   renimi sıklıkla veri yıđınlarını modellemek ve tahminlerde bulunmak i in kullanılırken veri madenciliđi, veri temizleme,  n iřleme, keřif ve analiz dahil olmak  zere daha uzun soluklu veri iřleme s re lerini kapsar. Sonu  olarak makine   renimi, veri madenciliđi alanında kullanılan g  l  bir enstr mandır. Veri madenciliđi, makine   renimi algoritmalarından yararlanarak b y k veri yıđınlarından kapsamlı bir řekilde i  r ler ortaya  ıkarabilir ve tahminleme analizlerinin dođruluđunu artırabilir.

2.3.3 Veri Tabanı Yönetim Sistemleri Veri Madenciliği İlişkisi

Veritabanı Yönetim Sistemleri (DBMS), verilerin yapılandırılmış olarak depolanmasını, yönetilmesini ve işlenmesini kolaylaştıran özel yazılım araçlarıdır. DBMS kapsamlı veri kümelerini işlemek, veri bütünlüğünü korumak ve veri işleme süreçlerini kolaylaştırmak için özel olarak tasarlanmıştır. Dosya sistemleri veri depolamak için temel bir yöntem olarak hizmet verirken, DBMS'nin ortaya çıkışı sayısız avantajı beraberinde getirmiş ve veri işleme ortamını dönüştürmüştür (Kochar, 2023). DBMS'nin temel misyonu, veritabanı olarak adlandırılan veriler için merkezi ve ilişkisel bir havuz oluşturmaktır. Bu veri tabanları, tablolar, satırlar, sütunlar ve diğer veritabanı nesneleri halinde yapılandırılmış ve birbiriyle ilişkilendirilmiş bilgilerin kapsamlı bir derlemesi olarak işlev görerek verilerin etkin bir şekilde depolanması, işlenmesi ve düzenlenmesi süreçlerini yönetmektedir.

Veritabanı Yönetim Sistemleri, büyük hacimli verilerin depolanması ve yönetilmesi için merkezi bir veri havuzu sağlayarak veri madenciliğinde çok önemli bir rol oynamaktadır. Veritabanı Yönetim Sistemleri, etkili veri madenciliği işlemleri için gerekli olan veri bağımsızlığı, veri bütünlüğü, güvenlik ve eşzamanlılık kontrolü gibi avantajlar sağlar. Veritabanı Yönetim Sistemleri ve veri madenciliği tekniklerinin sentezlenmesi sonucunda işletmelerin veri tabanlarındaki büyük hacimli verilerden değerli içgörüler, modeller ve bilgiler çıkarması mümkün olur. Böylece işletmeler karar destek sistemleri için kaliteli girdiler oluşturma imkânı bulur.

Veri madenciliği işlemleri sırasında verilerin depolanmasını ve işlenmesini kolaylaştırmak için Veritabanı Yönetim Sistemlerinin kullanılması önemlidir. Zira ilişkisel veritabanları ve NoSQL veritabanları, büyük veri yığınlarının depolanması için platform görevi görmektedir. Bu veritabanları, verilerin etkin bir şekilde sorgulanması ve işlenmesi için optimize edilmiştir (Han vd., 2011). Veri madenciliği süreçlerinde verinin gürültü faktörlerinden arındırılması ve ardından belirli ön işleme işlemlerinden geçirilmesi zorunludur. Veritabanı Yönetim Sistemlerinin sunmuş olduğu pek çok avantaj bulunmaktadır. Arşiv olarak tek bir birleşik veri kaynağı sağlaması başta olmak üzere, özniteliklerin tanımlanması, iş operasyonlarının kayıtlarının tutulmasını basitleştirmek için gerçek dünya varlıkları arasında ilişkiler kurulmasına yardımcı olması, gereksiz veri fazlalığının azaltılması ile depolama maliyetlerinden tasarruf edilmesini sağlaması

bunlardan bazılarıdır. Birden fazla eşzamanlı kullanım, işletmelere verilerini çeşitli departmanlar arasında paylaşma fırsatı vermiştir. Ayrıca, DBMS kurumsal veri güvenliği, erişim kontrolü, veri doğruluğu, geçerliliği ve bütünlüğünü sağlamaktadır (Islam, 2013, s. 22). Bu faktörler veri madenciliği modellerinin doğruluğunu artırır. Veritabanlarından veri çekmek, veriyi modifiye etmek ve silmek için kullanılan SQL (Structured Query Language) dili aynı zamanda veri madenciliği uygulamalarında da verinin temizlenmesi, filtrelenmesi, gruplanması gibi işlemler için yetkin ve güçlü bir enstrüman olarak öne çıkmaktadır.

Ayrıca, gelişen veritabanı teknikleri ve veritabanı yönetim sistemlerinin uygulanmasıyla birlikte dünyada veri depolama hızla artmaktadır (Zhao vd., 2021, s. 752). Bunun yanında veri tabanı yönetim sistemleri veri madenciliği süreçlerini tamamlayan veri ambarı ve çevrimiçi analitik işleme gibi özellikler barındırmaktadır. Dolayısıyla veritabanı yönetim sistemleri, verilerin depolanması, yönetilmesi ve işlenmesi için güçlü ve özelleştirilebilir bir çerçeve sunarak, veri madenciliği profesyonellerinin, verinin organizasyonu, konsolidasyonu ve erişimiyle ilgili karmaşık işlemlere boğulmadan veri madenciliği süreçlerindeki analitik görevlere konsantre olmalarını sağlar.

Veri madenciliği ile veritabanı yönetim sistemleri arasındaki temel yaklaşım ve uygulama farklılıklarına değinilecek olursa, veritabanı yönetim sisteminin temel amacı, verileri güvenli ve tutarlı bir şekilde depolamak ve yönetmek iken veri madenciliğinin amacı, veri yığınları içindeki gizli örüntü ve içgörülerin ortaya çıkarılmasıdır. Diğer taraftan veritabanı yönetim sistemleri tipik olarak yapılandırılmış sorgulama dillerini ve veri işleme operasyonlarını kullanırken, veri madenciliği süreçleri istatistik yaklaşımlar ve makine öğrenmesi algoritmalarından faydalanmaktadır. Veritabanı Yönetim Sistemi, bir veritabanının oluşturulmasını, bakımını ve kullanımını kontrol eden bir dizi bilgisayar programı iken (Chaplot, 2015, s. 24) veri madenciliği, örüntüleri ve faydalı bilgileri çıkarmak için büyük miktarda veriyi analiz eden hesaplama süreci olarak tanımlanmaktadır (Gullo, 2015, s. 18).

Veritabanı Yönetim Sistemleri ile veri madenciliği arasındaki simbiyotik ilişki, birincisi tarafından ikincisi için gerekli veri kaynaklarının sağlanmasını gerektirir. Buna karşılık, veri madenciliğinden elde edilen bilgiler, veritabanı yönetim sistemlerinin etkinliğini, performansını ve yapısını iyileştirmek için kullanılır. Bu süreç, veritabanı şemasını

optimize etmek, sorgu performansını geliřtirmek ve veri depolama ve indeksleme tekniklerini iyileřtirmek iin veri madencilięi yoluyla keřfedilen kalıplardan ve iliřkilerden yararlanmayı iermektedir. Veritabanı ynetim sistemleri ve veri madencilięi, byk veri yığınlarının iřlenmesini geliřtirmek iin etkileřim iinde alıřan bileřenlerdir. DBMS, veri madencilięi faaliyetleri kapsamında verilerin depolanması, zelleřtirilmesi ve iřlenmesi iin temel platformu sunar. Bu disiplinlerin birleřmesi, verilerin daha verimli ve etkin iřlenmesini ve analizini kolaylařtırarak, iřletmelere ve karar vericilere rasyonel seimler yapma gc vermektedir.

2.3.4 Matematik Veri Madencilięi İliřkisi

Veri madencilięi ve matematik arasındaki simbiyotik iliřki, byk veri yığınlarından anlamlı igrlerin ıkartılması srecinin ayrılmaz bir parasıdır; nk her iki disiplin de ilerlemeler iin byk lde birbirlerine baęımlıdır. Veri madencilięinin temel amacı, karmařık ve kapsamlı veri kmeleri iindeki gizli kalıpları, baęlantıları ve bilgiyi ortaya ıkarmaktır. Veri Madencilięi, verileri analiz etmek iin statik, olasılık, cebirsel fonksiyonlar, lmleme, grafik uyumu, kme teorisi ve mantıksal akıl yrtme gibi matematiksel yaklařımları kullanır. Bu matematiksel yaklařımlar, algoritmaları daha verimli hale getirmek iin deęiřtirilebilir, filtrelenebilir, derlenebilir ve birleřtirilebilir (Pandey vd., 2017, s. 1073).

Veri madencilięi temel olarak matematiksel kuram ve algoritmalara dayanır. Veri madencilięi yntemlerinin oęunda istatistik byk nem tařımaktadır. Veri daęılımlarını anlamak ve incelemek iin kullanılır. Ayrıca veri seti boyutlarının indirgenmesinde ve veri dnřmlerinin yrtlmesinde lineer cebirden yararlanılmaktadır. rneęin, temel bileřen analizi (Principal Component Analysis PCA), kovaryanslarını korurken verilerin boyutluluęunu azaltan ok deęiřkenli bir analizdir (Elhaik, 2022, s. 1) ve nemli bir boyut indirgeme teknięidir. Verileri daha ynetilebilir bir hale getirir. Bunlarla birlikte olasılık teorisinin veri madencilięine uygulanması, modellerin belirsizlikleri giderme ve tahminlerde bulunma yetkinliklerini artırmaktadır. Bayesian aęları ve Markov zincirleri, veri madencilięinde olasılık teorisinin kullanıldıęı rneklerdir. Ayrıca veri madencilięi algoritmalarının etkinlięini artırmak iin optimizasyon yntemlerinden yararlanılmaktadır. Bu yntemler, zellikle byk veri setlerinde hesaplama kaynaklarını etkin bir řekilde kullanmak iin nemlidir. Veri

madenciliği süreçlerindeki modelleme ve tahmin uygulamalarının odak noktasında yine matematiksel kuramlar vardır. Veri madenciliği operasyonlarında istatistiksel modeller ve makine öğrenmesi algoritmaları sıklıkla kullanılmaktadır. Regresyon analizi, bağımlı ve bağımsız değişkenler arasındaki bağlantıyı aydınlatan modeller oluşturmak için kullanılmakla birlikte daha sonra veri madenciliği çalışmalarında tahmin amaçlı olarak da kullanılmaktadır. Verileri önceden belirlenmiş gruplara göre tasnif etmek için sınıflandırma teknikleri kullanılır. Örneğin karar ağaçları, lojistik regresyon ve destek vektör makineleri, veri sınıflandırması için yaygın olarak kullanılan algoritmalarıdır. Kümeleme algoritmaları ise verileri içsel benzerliklerine ve farklılıklarına göre farklı kümelere bölmek için kullanılır. Kümeleme, birçok özelliğe sahip büyük veri kümelerini analiz etmek için kullanılabilecek denetimsiz bir yaklaşımdır. Verilerinizin net bir görüntüsünü sağlayan bir veri modelleme tekniğidir (Gupta vd., 2021, s. 412). Bu amaçla K-means ve hiyerarşik kümeleme gibi teknikler kullanılmaktadır.

Birliktelik kuralları, ürünlerin birbirleriyle birlikte satın alınma olasılığını değerlendirerek veri madenciliğinde çok önemli bir rol oynar. Bu yöntemler, özellikle perakende sektöründe tüketici satın alma modellerini incelemek açısından çok önemlidir. Veri madenciliği çabalarında verilerin şifresini çözmek ve anlamlı içgörüler çıkarmak için matematiksel modellerin ve algoritmaların kullanılması zorunludur.

Veri madenciliği uygulamalarında veri kalitesinin önemi göz ardı edilemez. Matematiksel teknikler bu kaliteyi garanti etmede temel araçlar olarak hizmet vermektedir. Veri ön işleme ve temizleme adımları, veri kümesini analiz için hazırlamayı amaçlayan bir dizi işlemi kapsarken, eksik veri analizi, eksik veri girişlerini yönetmek için belirli bir sistematik sunar. Eksik verinin tamamlanması ile ilgili farklı yaklaşımlar mevcuttur. Bunlardan en temel olanları eksik verinin yerine ortalama, medyan ya da mod gibi istatistiksel verilerin kullanılması, bir önce ve bir sonraki verinin kaydırılması ya da eksik veri kaydının silinmesidir (Jayalakshmi & Santhakumaran, 2011, ss. 91-92). Bir veri yığınınındaki aykırı değerlerin tanımlanması ve düzeltilmesi, analiz öncesinde veri güvenilirliğinin sağlanması açısından önemlidir. Verilerin normalleştirilmesi, verileri değişen ölçeklerden tekdüze bir ölçeğe getirmeyi içerir ki bu makine öğrenmesi algoritmasının performansını artırmak için çok önemli bir adımdır. Logaritmik veya karekök dönüşümleri gibi veri dönüşümleri, daha normal bir veri dağılımını teşvik ederek

analiz sürecini basitleştirir. Bu matematiksel teknikler, veri madenciliği süreçlerinin optimize edilmesinde ve verilerin analize hazırlanmasında önemli bir rol oynamaktadır.

Öte yandan veri madenciliği yöntemlerinde en verimli çözümleri belirlemek için matematiksel tekniklerin kullanılmasını içeren optimizasyon kuramı çeşitli veri madenciliği algoritmalarında çok önemli bir rol oynamaktadır. Optimizasyon teknikleri, özellikle büyük veri setlerinde hesaplama verimliliğini artırmak için kritik öneme sahiptir. Örneklendirmek gerekirse, K-means kümeleme algoritması, veri noktalarını etkili bir şekilde kümelere ayırmak ve küme merkezlerinin belirlenmesini optimize etmek için küme içi varyansı en aza indirecek şekilde tasarlanmıştır. Destek vektör makineleri (SVM), doğruluğu artırmak ve daha sağlam modeller oluşturmak için veri sınıflandırma sırasında sınıflar arasındaki marjı en üst düzeye çıkarmayı amaçlamaktadır. Doğrusal regresyon, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi doğru bir şekilde modellemek için karesel hataların toplamını en aza indirmeyi amaçlar ve sonuçta en uygun regresyon çizgisi elde edilir. Genetik algoritmalar ve sinir ağları gibi daha karmaşık modeller de öğrenme sürecinde optimizasyon tekniklerinden yararlanır. Bu matematiksel optimizasyon yöntemleri, veri madenciliği algoritmalarının verimliliğini ve doğruluğunu artırarak büyük veri yığınlarının analizinde en iyi sonuçları sağlamaktadır. Doğrusal programlama, doğrusal olmayan programlama ve tamsayı programlama gibi matematiksel optimizasyon teknikleri, veri madenciliğinde karşılaşılan karmaşık optimizasyon zorluklarının üstesinden gelmek için yapılandırılmış bir yaklaşım sunar. Bu yöntemler, kaynakların verimli bir şekilde tahsis edilmesini, maliyetlerin en aza indirilmesini ve istenen sonuçların en üst düzeye çıkarılmasını sağlar.

Matematik ve veri madenciliği arasındaki etkileşim karşılıklıdır; veri madenciliği matematiksel anlayışın ilerletilmesinde rol oynar. Veri madenciliği yalnızca matematiksel prensiplerden faydalanmakla kalmaz, aynı zamanda mevcut teori ve tekniklerin sınırlarını zorlayan yeni araştırma soruları ve problem formülasyonlarının üretilmesi yoluyla matematiksel bilginin gelişimini de katalize eder. Örneğin, büyük ölçekli, yüksek boyutlu veri kümelerini ele alma ihtiyacı optimizasyon yöntemleri ve manifold öğrenme teknikleri gibi matematiksel yaklaşımları ortaya çıkarmıştır. Manifold öğrenme algoritmaları, verilerin genellikle yapay olarak yüksek boyutlu olduğu önermesine dayanır. Aslında, yüksek boyutlu uzaya gömülmüş çok daha düşük boyutlu

bir manifold tarafından iyi bir şekilde temsil edilebilecekleri ifade edilmektedir (Bhatia vd., 2012, s. 514).

Ek olarak, gerçek dünya verilerinin kompleks ve değişken yapısı, bu verilerde mevcut olan belirsizliği ve doğrusal olmamayı doğru bir şekilde temsil edebilen yeni matematiksel modellerin ve algoritmaların araştırılmasını teşvik etmiştir. Bulanık mantık, Bayesian ağları ve grafik teorisi gibi matematiksel ilkeleri kullanan veri madenciliği, daha önce karmaşık veya belirsiz kabul edilen zorlukların üstesinden gelebilmiştir. Bu karşılıklı ilişki aynı zamanda matematikte ilerlemelere de yol açarak verilerde gözlemlenen kalıpları ve bağlantıları etkili bir şekilde yorumlamak ve analiz etmek için yenilikçi teorilerin ve metodolojilerin geliştirilmesini teşvik etmektedir.

Lineer cebir, olasılık teorisinin ilkeleri ve istatistiksel çıkarım yöntemleri veri madenciliğinin temel bileşenleri ve algoritmalarının temel yapı taşlarını oluşturmakla birlikte, bilgi erişimine (information retrieval) yönelik bazı yöntemlerde, veri sıkıştırma ve erişim optimizasyonu için doğrusal lineer cebir teknikleri (tekil değer ayrıştırma vb.) kullanılmaktadır (Eldén, 2006, s. 332).

Sonuç olarak, matematik ve veri madenciliği arasında karşılıklı simbiyotik bir ilişki vardır. Çünkü veri madenciliğindeki ilerlemeler aynı zamanda matematiksel anlayışta da ilerlemeyi sağlayabilir. Veri madenciliği sıklıkla mevcut matematiksel teorilerin ve metodolojilerin geliştirilmesine ve genişletilmesine ilham veren yeni araştırma soruları üretmektedir.

2.3.5 Bilgisayar Bilimleri Veri Madenciliği İlişkisi

Bilgisayar bilimindeki ilerleme, veri madenciliğinin gelişmesinde ve genişlemesinde çok önemli bir faktör olduğundan, bilgisayar bilimi ile veri madenciliği arasındaki bağ çok kuvvetlidir. Veri madenciliğini kapsayan veri bilimi, ham verileri pratik içgörülere dönüştüren metodolojiler geliştirmek için bilgisayar bilimi gibi çeşitli disiplinleri birleştiren disiplinler arası bir alandır. Veritabanı teknolojisinin hızla gelişmesi ve bunun sonucunda veri depolamada yaşanan artış, veri madenciliğini giderek daha önemli ve genişleyen bir alan haline getirmiştir (Zhao vd., 2021, s. 750). Uygulamaların karmaşıklığı ve işlem süresi, veri madenciliği metodolojilerinin ilerlemesini önemli ölçüde etkilemiş ve bilgisayar bilimi alanı bu engellerin aşılmasında çok önemli bir rol oynamıştır. Bilgisayar bilimlerindeki ilerlemelerin kolaylaştırdığı derinlemesine veri

analizi ve bulut bilişimin entegrasyonu, çeşitli platformlardan ve cihazlardan veri madenciliği hizmetlerine erişim sağlayan çözümlere olanak sağlamıştır. Bilgisayar teknolojisindeki gelişmeler veri depolama sistemlerinin yaygınlaşmasını sağlayarak büyük miktarlardaki verilerin manyetik ortamlarda saklanması kolay ve ucuz hale getirmiştir. Bunun sonucunda veri madenciliğinin temeli olan büyük veri kümelerinin depolanması ve işlenmesi kolaylaşmıştır (Özbay, 2015, s. 263).

Devasa veri yığınlarının depolanması ve yönetimi, bilgisayar bilimi metodolojilerinin uygulanmasıyla kolaylaştırılabilir. Veritabanı yönetim sistemleri (DBMS), verilerin etkili bir şekilde depolanması ve işlenmesi süreçlerinde kritik bir işlev görmektedir. Büyük Veri platformu, üretilen veri ve bilgilerin yapılandırılması ve işlenmesi süreçlerini önemli ölçüde hızlandırabilir ve işletmelerin verimliliğini, performansını ve çevikliğini artırabilir (Pamucar vd., 2024, s. 1). Hadoop ve Spark gibi platformlar, devasa veri kümelerinin eş zamanlı olarak işlenmesini ve incelenmesini mümkün kılmaktadır.

Bilgisayar bilimi alanı, Hadoop ve bulut bilişim gibi teknolojilerin yenilikçiliği yoluyla, genellikle "Büyük Veri" olarak bilinen devasa veri yığınlarını yönetme ve analiz etme zorluğunun üstesinden gelmede çok önemli bir rol oynamıştır. Bunun yanında altyapı ve hesaplama yetenekleri üzerindeki etkisinin yanı sıra bilgisayar bilimi, veri analizinde kullanılan algoritmaları ve teknikleri de önemli ölçüde şekillendirmiştir. Hesaplamalı istatistikler hem istatistiksel çıkarımı hem de veri analizini kolaylaştıran algoritmalar oluşturarak veri biliminin ilerlemesinde çok önemli bir faktör olmuştur.

Bilgisayar bilimi alanı, veri madenciliği süreçlerine yardımcı olan yazılım araçları ve geliştirme platformlarının inşa edilmesinden sorumludur. Python, R ve MATLAB gibi programlama dilleri, verileri analiz etmek ve modellemek için kapsamlı bir geliştirme kütüphanesi ve altyapısı sağlamaktadır. TensorFlow ve PyTorch gibi derin öğrenme platformları karmaşık veri madenciliği görevlerinin gerçekleştirilmesini sağlamaktadır.

Bilgisayar bilimi alanında, verilerin görselleştirilmesine yardımcı olarak sonuçların anlaşılabilirliğini artırmak için de çeşitli araçlar geliştirilmektedir. Veri görselleştirme tekniklerinin uygulanmasıyla bilgiler hem basitleştirilmiş hem de kolayca yorumlanabilir bir şekilde sunulmaktadır. Microsoft Power BI ve Tableau, işletmelerin verileri etkili bir şekilde analiz etmesini ve görselleştirmesini sağlayan iki önemli platform olarak

kullanıcıların verilerini yapılandırmalarına ve dönüştürmelerine olanak tanıyan veri modelleme özellikleri sunmaktadırlar (Parthe, 2023, s. 2).

Sonuç olarak bilgisayar bilimi ve veri madenciliği arasındaki sinerji, birbirlerinin veri işleme, analiz ve modelleme alanlarındaki yeteneklerini geliştirir. Bu süreçte bilgisayar bilimlerinin sağlamış olduğu algoritma ve yazılımlar, veri depolama teknikleri, hesaplama gücü ve buna bağlı donanımsal altyapı önemli rol oynamaktadır.

2.3.6 Yapay Zekâ Veri Madenciliği İlişkisi

Yapay Zekâ (Artificial Intelligence / AI) ve veri madenciliğinin entegrasyonu, son yıllarda yapay zekanın veri madenciliği süreçlerinin yeteneklerini ve üretkenliğini artırmadaki yeteneklerini göstermesi nedeniyle daha yaygın hale gelmiştir. Yapay zekâ, büyük veri yığınlarının işlenmesini kolaylaştırarak, işletmelerin değerli çıkarımlar elde etmesini ve karar verme süreçlerini geliştirmesini sağlayarak veri madenciliği alanında hayati bir unsur olarak ortaya çıkmıştır.

Yapay Zekâ (AI), bilgisayarlara insanlara benzer bilişsel problem çözme yetenekleri kazandırmak için tasarlanmış bir dizi algoritma ve metodolojiyi kapsar. Buna karşılık Veri Madenciliği, devasa veri kümelerinden anlamlı bilgiler çıkarma sürecini ifade eder. Yapay zekâ ve veri madenciliğinin entegrasyonu her iki sistemin kapasitesini artırır; Yapay zekâ, veri madenciliğinin bilgileri daha etkili bir şekilde yorumlayıp analiz etmesini sağlarken, veri madenciliği eğitim amaçlı ek veriler sağlayarak yapay zeka algoritmalarının sürekli iyileştirilmesine katkıda bulunmaktadır. Makine öğrenimi, bir bilgisayar programını uygulamak için kullanılan yapay zekanın bir alt kümesidir. Bu programlar, bazı algoritmalar ve istatistikler kullanılarak bir dizi kapsamlı veri tarafından yönlendirilir (Wazid vd., 2022, s. 313).

Yapay zekâ, veri madenciliğinin derinlemesine ve hassas veri analizi yapma yeteneğini geliştirmektedir. Özellikle derin öğrenme tekniklerinin önemli bir faydası, büyük miktarda denetimsiz verinin analiz edilmesi ve öğrenilmesidir; bu da onu ham verilerin büyük ölçüde etiketlenmemiş ve kategorize edilmemiş olduğu büyük veri analitiği için değerli bir araç haline getirmektedir. Derin Öğrenme algoritmaları, hiyerarşik bir öğrenme süreci aracılığıyla veri temsilleri anlamında üst düzey karmaşık çıkarımlar yapmaktadır (Najafabadi vd., 2015, s. 1).

Örnek olarak, yapay zekâ ve veri madenciliği tekniklerinin entegrasyonu sonucu, sağlık hizmetlerinde hastalık prognozu ve finans alanında sahtekarlığın tespiti dahil olmak üzere çeşitli alanlarda yapılan önemli çalışmalar gösterilebilir.

Yapay zekâ ve veri madenciliği birçok açıdan karşılıklı olarak birbirini geliştiren ve güçlendiren disiplinlerdir. Örneğin yapay zekâ, veri madenciliği çalışmalarında veri analizi ve modelleme süreçlerinin etkinliğini artırırken veri madenciliği de yapay zekanın öğrenme algoritmaları için büyük veri setleri sağlayarak modelleme ve tahmin algoritmalarının geliştirilmesini sağlar. Büyük veri yığınlarının kullanılması, yapay zekanın öğrenme ve uyum sağlama yeteneğini önemli ölçüde artırmaktadır. Yapay zekâ, veri madenciliği süreçlerini otomatikleştirerek analiz süreçlerini hızlandırmaktadır.

Yapay zekâ genellikle problem çözme, doğal dil işleme, görüntü tanıma ve robotik gibi alanlara odaklanırken veri madenciliği veri analizi, örüntü çıkarma ve tahmine dayalı modellerin oluşturulmasına odaklanmaktadır. Örneğin, yapay zekanın doğal dil işleme tekniklerinin gelişmesi metin madenciliği ve duygu analizi gibi veri madenciliğinin alt alanlarına doğrudan katkı sağlamaktadır.

Sonuç olarak, yapay zekanın veri madenciliği süreçlerine entegrasyonu yalnızca teknik faydalar sağlamakla kalmamış, aynı zamanda veriye dayalı karar verme misyonu ve metodolojisinde de bir değişikliğe yol açmıştır. Yapay zekâ odaklı veri madenciliği, yalnızca geçmiş verilere ve önceden belirlenmiş algoritmalara bağlı kalmak yerine, işletmelerin yeni içgörüler keşfetmesine ve verilerinin keşfedilmemiş yönlerini araştırmasına olanak tanımaktadır. Bu durum, görüntü ve dil tanımanın yanı sıra üretim ve tedarik zinciri yönetimi gibi alanlarda, bilgi yönetimi, strateji oluşturma ve planlama da dahil olmak üzere farklı uygulamalar için çeşitli olasılıkların önünü açmıştır.

Yapay zekanın yetkinliği ilerledikçe, veri madenciliğine bağlı olarak daha geniş araştırma ve keşif çabalarının geleceğini şekillendirmede çok temel bir rol oynayacağını göstermektedir.

2.3.7 Görselleştirme Veri Madenciliği İlişkisi

Veri görselleştirme, araştırmacılar, analistler ve karar vericilerin kompleks verileri anlamlı bir şekilde yorumlamasına olanak tanıdığı için veri madenciliği sürecinde çok önemli bir rol oynamaktadır. Karmaşık veri yığınlarını kolayca anlaşılır görsel formatlara dönüştürerek, değerli içgörülerin ortaya çıkarılmasına ve yeni bilgilerin keşfedilmesine

yardımcı olan önemli bir enstrümandır. Veri madenciliği ve veri görselleştirme birbiriyle birlikte çalışan iki temel alandır. Veri madenciliği, büyük veri yığınlarından önemli kalıpları ve öngörülerini ortaya çıkarmaya odaklanırken, veri görselleştirme, çıkarılan bu öngörülerin görsel olarak sunumuna odaklanır. Bu iki disiplin arasındaki etkileşim, verilerin daha anlaşılır ve kullanışlı hale gelmesini sağlar. Veri görselleştirme, işletmeler tarafından her gün üretilen büyük miktarda verinin etkili bir şekilde anlaşılması ve yorumlanmasında çok önemli bir unsur haline gelmiştir. Veri görselleştirme, çeşitli kaynaklardan gelen verileri kolayca erişilebilir ve anlaşılabilir hale getirmek için bilgisayar tarafından oluşturulan grafiksel gösterimleri kullanır ve karar vericilerin verileri görsel bir biçimde analiz etmesine olanak tanır. Bu da onların örüntüleri tanımlamalarını, bilgileri anlamalarını ve bilinçli kararlar almalarını sağlar. Ayrıca, veri görselleştirme, bilgi görselleştirme veya bilimsel görselleştirme olarak da bilinir (Chavan, 2023, s. 1338).

Veri görselleştirme, veri madenciliği alanındaki keşfedici veri analizinde çok önemli bir rol oynar. Araştırmacılar, verileri çizelgeler veya dağılım grafikleri gibi çeşitli formatlar aracılığıyla görsel olarak ortaya koyarlar ve böylece orijinal veri setinde kolayca görünmeyebilecek kalıpları, aykırı değerleri ve ilişkileri etkili bir şekilde tespit edebilir. Bu görsel inceleme, hipotezlerin geliştirilmesine yardımcı olur ve uygun veri madenciliği metodolojilerinin seçimini kolaylaştırır.

Ayrıca, veri görselleştirmenin kullanılması hata tanımlama ve doğrulama için çok önemlidir. Görsel temsiller, verilerdeki yanlışlıkları veya düzensizlikleri anında ortaya çıkarma yeteneğine sahip olup, araştırmacıların yanıltıcı yorumlar yapmadan önce sorunları tespit etmesine ve çözmesine imkân verir. Etkili görselleştirmeler karar vericilerin analitiği görsel biçimde görmelerini sağlayarak kalıpları tanımlamalarını, bilgileri anlamalarını ve bir fikir oluşturmalarını kolaylaştırır (Chavan, 2023, s. 1338). Görsel veri keşfinin avantajı, kullanıcının veri madenciliği sürecine doğrudan dahil olmasıdır. Büyük veri setlerinin keşfedilmesini desteklemek için son yıllarda geliştirilen çok sayıda bilgi görselleştirme tekniği bulunmaktadır.

Veri görselleştirme araçlarının veri madenciliği süreçlerine sunduğu en kritik katkı desen tanımadır. Görselleştirme araçları, veri madenciliği sürecinde veri içindeki kalıpların ve desenlerin tanımlanmasını kolaylaştırmada çok önemli bir rol oynar. Bu araçlar, geniş

veri kümelerinden elde edilen bilgileri görsel olarak temsil ederek analizin verimliliğini artırmaktadır. Bununla birlikte video görselleştirme veri kümelerindeki aykırı değerlerin ve anomalilerin belirlenmesinde önemli kolaylıklar sağlamaktadır. Ayrıca veri görselleştirme, karar vericiler için veri madenciliği bulgularının netliğini artırır. Verilerin görsel sunumu anlamayı kolaylaştırır ve karar destek sistemlerine olumlu katkı sağlar.

Veri görselleştirme amacıyla yaygın olarak kullanılan bir kısmı Tableau, Power BI, D3.js, R, Python, QlikView gibi ortamlar ve kütüphaneler kullanılmaktadır. Tableau, verileri verimli bir şekilde görselleştirmek için geniş seçenek yelpazesine sahip, kullanıcı dostu bir ara yüz sağlar. Sürükle bırak işlevi, kullanıcıların veri kümelerinde sorunsuz bir şekilde gezinmesine ve analiz etmesine olanak tanır. İş zekâsı, veri analitiği ve raporlama, büyük veri yığınlarını analiz etmek ve etkileşimli gösterge tabloları geliştirmek için kullanılır. Microsoft tarafından geliştirilen Power BI, verilerin görsel olarak temsili ve incelenmesi için sağlam bir çerçeve sunar. Excel ile detaylı entegrasyon imkânı sunması önemli bir avantajdır. İş zekâsı, veri analitiği ve raporlamanın yanı sıra kurumsal verilerin analizi için sıklıkla kullanılmaktadır. JavaScript tabanlı bir kütüphane olan D3.js ise, web üzerinde veri görselleştirmelerinin geliştirilmesinde sıklıkla kullanılmaktadır. Yüksek esnekliği ve uyarlanabilirlik kabiliyeti ile öne çıkmaktadır. Özellikle etkileşimli grafik ihtiyacı duyulan web tabanlı projelerde sıklıkla kullanılmaktadır. Veri analizi ve istatistiksel hesaplamalar için yaygın olarak kullanılan bir programlama dili olan R, karmaşık veri görselleştirmeleri oluşturmak için çok uygun olan ggplot2 gibi güçlü görselleştirme kütüphaneleri sunmaktadır. Veri madenciliği süreçleriyle birlikte istatistiksel analiz ve raporlama alanında yaygın kullanımı bulunmaktadır. Bu alanda yaygın kullanılan bir diğer programlama dili olan Python, veri görselleştirmesi için geniş bir kütüphane sunmaktadır. Statik grafikler oluşturmak için Matplotlib ve Seaborn yaygın olarak kullanılırken, etkileşimli grafikler geliştirmek için Plotly tercih edilmektedir. QlikView, kullanıcıların verileri araştırmasını ve yorumlamasını sağlayan güçlü bir iş zekâsı aracıdır. Veri konsolidasyonu ve etkileşimli göstergelerin oluşturulmasında kullanılmaktadır.

2.4 Metin Madenciliği

2.4.1 Metin Madenciliği Tanım ve Kavramsal Çerçeve

Sürekli gelişen ve her geçen gün önemi artan bir alan olarak metin madenciliği, hızlı büyüyen yapılandırılmamış veri yığınlarından önemli bilgi ve öngörüler elde etmek için dönüştürücü bir enstrüman olarak ortaya çıkmıştır. Metin madenciliği, yapılandırılmamış haldeki metinsel verilerin analiz edilmesi, yorumlanması ve anlamlı desenler çıkarılması amacıyla karmaşık algoritma ve tekniklerin kullanılmasından oluşan bir süreçtir. Bu süreçte, doğal dil işleme (Natural Language Processing, NLP), bilgi çıkarımı, özellik çıkarımı, örüntü tanıma ve makine öğrenmesi gibi yöntemler kullanılır.

Metin madenciliği alanının önemini her geçen gün artıran hususların bazı başlıklar altında toparlanması gerekirse, bunlardan en önde geleni büyük veri miktarındaki hızlı artış ve yaşanan bilgi patlaması gerçeğidir. İnternet ve dijital teknolojilerin yaygınlaşması, mevcut metinsel veri hacminde devasa bir artışa yol açmıştır. Dijital platformların yaygınlaşması ve bilginin dijital formata dönüştürülmesiyle birlikte metin verilerinin miktarının ve erişilebilirliğinin hızla artması, daha önce görülemeyen bilgilerin keşfedilmesi için oldukça önemli imkanlar sunmaktadır. Bu verilerin analiz edilmesi, işletmeler ve araştırmacılar için kritik bir hale gelmiştir. Ayrıca, metinsel verilerin artan karmaşıklık ve çeşitliliği, bu verilerin doğasında olan tutarsızlık ve düzensizlikler ile, devasa metinsel verilerin etkin bir şekilde analiz edilmesi için gelişmiş yöntemlere olan ihtiyacı artırmıştır. Metin madenciliği çalışmalarının önemini artıran bir diğer saik rekabet avantajı sağlamasıdır. İşletmeler, müşteri geri bildirimleri, sosyal medya yorumları ve diğer metin tabanlı verileri analiz edebilirler. Metin madenciliği teknikleri rekabet güçlerini, müşteri kitlelerini ve pazarlama stratejilerini analiz etmelerine yardımcı olabilir(Fan vd., 2006, s. 82). Genel anlamda metin madenciliği teknikleri müşteri içgörülerini oluşturmak, pazarlama stratejilerini geliştirmek ve ortaya çıkan eğilimleri tespit etmek için pazarlama alanında kullanılmaktadır. Bir diğer deyişle işletmeler müşteri algısını belirlemek, pazar hareketlerini tanımak ve rakipleri hakkında daha derin öngörüler elde etmek için metin madenciliği tekniklerini kullanmaktadırlar. Bununla birlikte metin madenciliği tekniklerinin kullanımı, karmaşık problemlerin anlaşılmasına, gizli kalıpların tespit edilmesine ve sonuç olarak daha bilinçli karar verme süreçlerinin kolaylaştırılmasına yardımcı olabilir. Karar Destek Sistemi (KDS), karar verme sürecinin

tüm aşamalarında yöneticilerin becerilerini desteklemek için iş verilerini toplayan, düzenleyen ve analiz eden etkileşimli yazılım tabanlı bir sistemdir (Daniela, 2019, s. 77). Metin Madenciliği tabanlı karar destek sistemleri, yönetim olarak iyi yapılandırılmış vatandaş merkezli kararlar üretmek üzere yapılandırılmamış metinsel verileri tahmine dayalı analitiklerle bütünleştirmektedir (Rao & Dey, 2011, s. 74). Bu sayede metin madenciliği, politika yapıcılar ve yöneticiler için daha bilinçli kararlar almayı sağlayan değerli içgörüler sunar.

Metin madenciliği veri madenciliği ile çok yakın benzerlikler taşımasına ve her iki disiplin arasında oldukça kapsamlı ilişkiler bulunmasına rağmen veri madenciliğinden farklılaştığı bazı noktalar bulunmaktadır. Geleneksel veri madenciliği tipik olarak kurallı bir arayüze sahip yapılandırılmış veri tipleriyle ilgilenir. Metin madenciliği ise belgeler, e-postalar, sosyal medya gönderileri ve web sayfaları gibi yapılandırılmamış metinsel verilere odaklanmaktadır. Bu veriler, makaleler, web sayfaları, sosyal medya paylaşımları gibi, doğal dil yapısı içinde oluşturulmuş veriler olduğu için yapılandırılmış veya yarı yapılandırılmış olabilir.

Genel olarak, veri madenciliği yapılandırılmış veritabanı ile ilgilenir. Bu nedenle, veri madenciliği tekniklerinin metin madenciliğine kolay bir şekilde uyarlanabilmesi için yapılandırılmamış metnin yapısal veri tabanına dönüştürülmesi önemlidir (Sirsat vd., 2014, s. 54). Öte yandan metin madenciliği ise, belirsizliği yönetmek, ilgili özellikleri belirlemek ve gelişmiş analitik süreçlere hazırlık amacıyla yapılandırılmamış metinsel bilgileri düzenlemek de dahil olmak üzere doğal dil işlemenin karmaşıklıklarıyla başa çıkmayı amaçlamaktadır.

Tipik olarak geleneksel veri madenciliği süreçlerinde, veri temizleme yöntemleri genellikle eksik değerlerin atanması, anomalilerin belirlenmesi ve veri normalizasyonunun gerçekleştirilmesi görevlerini kapsamaktadır. Metin madenciliğinde ise ön işleme daha karmaşıktır ve metinlerin tokenize edilmesi, durak kelimelerin çıkarılması, kök forma dönüştürme (stemming) ve lemmatizasyon gibi adımları içerir. Bu süreçler boyunca ham metin verilerinden yapılandırılmış belge temsilleri oluşturmak için farklı ön işleme teknikleri birlikte kullanılır (Feldman & Sanger, 2007, s. 57).

Geleneksel veri madenciliği uygulamalarında, özellik çıkarımı tipik olarak yaş ve gelir düzeyine benzer şekilde müşteri demografik özelliklerinin analiz edilmesi gibi

özelliklerin veri kümesinden doğrudan çıkarılmasını içerir. Buna karşılık metin madenciliği, metnin anlamsal ve içeriksel yanlarını temsil eden özelliklerini ortaya çıkarmak için kelime frekansları, n-gramlar, kelime gövdeleri ve konu modelleme gibi daha kompleks teknikleri kullanır. Bu konuyla bağlantılı olarak bahsedilebilecek bir diğer farklılık, veri madenciliğinde büyük veri yığınlarından yüksek kaliteli bilgi elde etme süreci hedeflenir ve veri madenciliği, homojen ve erişimi kolay olan yapısal verilerden bilginin keşfedilmesi süreci iken, metin madenciliği heterojen ve yapısal olmayan verilerden metin keşfi sürecidir ve metindeki sözcüksel, sözdizimsel ve anlamsal bir özelliği keşfetmek için metin yöntemlerinin kullanılmasına odaklanır (Ersoz, 2019, s. 4394).

İlave olarak her iki alan için yapılan çalışmalarda da kullanılan analitik yöntem ve yaklaşımlar da birbirinden ayrışabilmektedir. Veri madenciliği, regresyon, sınıflandırma, kümeleme gibi istatistiksel ve sayısal yöntemlere dayanırken, metin madenciliği daha çok doğal dil işleme ve duygu analizi ve metin sınıflandırma gibi yaklaşımlar kullanmaktadır. Bununla birlikte geleneksel veri madenciliğinde kullanılan teknikler, regresyon analizi, karar ağaçları, kümeleme ve sınıflandırma algoritmaları gibi modelleme yaklaşımlarını içerirken, metin madenciliği süreçleri doğal dil işleme, duygu analizi ve derin öğrenme tekniklerinin uygulamalarını kapsamaktadır. Metin madenciliği yoğun bir şekilde Doğal Dil İşleme (NLP) tekniklerine dayanır. NLP teknikleri, metinsel kaynaklardan varlıkları ve aralarındaki ilişkileri çıkarmak için kullanılır (Zhong vd., 2023, s. 1).

Geleneksel veri madenciliği süreçlerinde analize tabi olan veriler daha ziyade yapılandırılmış veriler olduğu için ölçeklenebilirlik görece problemsizdir. Buna karşın metin madenciliği devasa miktardaki yapılandırılmamış metinsel veri ile çalışmak durumunda olduğundan verinin ön işleme ve analitik işlemleri çok daha fazla zaman ve işlemci gücü gerektirmektedir. Verinin yorumlanması bağlamında geleneksel veri madenciliği süreçleri daha kolay ve nettir. Lojistik regresyon ve karar ağaçları analize tabi veriler üzerindeki kararların nasıl alındığını net bir şekilde ortaya koyabilir. Diğer taraftan derin öğrenme ve doğal dil işleme yöntemlerini sıklıkla kullanan metin madenciliği modelleri yorumlanabilirlik bakımından daha kısıtlı bir yapıya sahiptir. Derin öğrenme yöntemlerinin ve büyük veri analitiğindeki diğer karmaşık yöntemlerin çoğunun önemli bir sınırlaması, yorumlanabilirlik eksikliğidir (Chai & Li, 2019, s. 1).

Geleneksel veri madenciliği duygusal kutupsallık veya metin anlamı üzerinde doğrudan çalışmazken, metin madenciliği içeriğin (Örneğin, müşteri yorumlarının) duygusal analizi ve semantik keşfi alanlarında çalışmaktadır.

Sonuç olarak metin madenciliği, anlamlı çıkarımlar yapmak için kapsamlı metinsel verilerin sistematik analizini ifade etmektedir. Veri yığınlarının büyüklüğü, rekabet avantajı kazanma potansiyeli, karar alma süreçlerine sağladığı destek, gibi faktörler nedeniyle önemi giderek artmaktadır. Geleneksel veri madenciliği yöntemlerinden farklı olarak metin madenciliği, yapılandırılmamış verilerle ilgilenmekte ve metinlerden anlamlı bilgiler çıkarmak için doğal dil işleme yöntemlerini kullanmaktadır. Bu özellikler, metin madenciliğinin veri analitiği alanında önemli bir konuma gelmesini sağlamıştır.

Bu ayrımların ışığında, işletmelerin operasyonları ve müşterileri hakkındaki bilgilerini geliştirmek için yapısal ve yapısal olmayan veriler de dahil olmak üzere çok çeşitli verilerden yararlanmayı amaçlamaları nedeniyle, veri madenciliği ve metin madenciliği arasındaki iş birliği gün geçtikçe daha yoğun bir hal almaktadır.

2.4.2 Metin Madenciliği Süreçleri

2.4.2.1 Veri Toplama

Veri toplama aşaması metin madenciliği süreçleri için önemli aşamalardan biridir. Metin madenciliği uygulamaları için veri toplama süreçlerinde her biri farklı avantaj ve işleyişe sahip çok sayıda yöntem ve teknik kullanılmaktadır.

Metin madenciliği süreçlerinde uygulanan veri toplama tekniklerinden birisi web kazıma (web scraping) yöntemidir. Web kazıma, web sitelerinden verilerin otomatik olarak çıkarılması ve yapılandırılması işlemidir. Bu yöntem, istenilen bilgilerin elde edilmesi için web sayfalarının HTML içeriğinin ayrıştırılmasını ve önceden tanımlanmış bir formatta saklanmasından ibarettir. Bir diğer ifadeyle, web kazıma genellikle web sitelerine otonom olarak HTTP istekleri gönderen ve ardından analiz için verileri çıkaran yazılımlar kullanılarak gerçekleştirilir. Web kazıma genellikle Python gibi programlama dilleri kullanılarak gerçekleştirilir ve BeautifulSoup, Scrapy gibi kütüphaneler bu süreçte yaygın olarak kullanılır. Bu yaklaşım, verilerin manuel yolla toplanmasının aksine büyük veri yığınlarının otomatik olarak elde edilmesini kolaylaştırarak görece büyük ölçekli

veri projelerindeki verimliliği artırarak ve iş yükünü azaltmaktadır. Web kazıma, farklı kaynaklardan gelen verilerin konsolidasyonu ve analizi için de etkili bir yöntemdir.

Veri toplama aşamasında kullanılan bir diğer yöntem API (Application Programming Interface) kullanımıdır. Bir uygulama programlama ara yüzü olan API, yazılım uygulamaları arasında veri ve fonksiyonların alışverişine imkân vererek çeşitli yazılım bileşenleri arasında kesintisiz iletişim ve iş birliğini mümkün hale getirmektedir. İşte bu yüzden, özellikle veri toplama süreçlerinde, API'ler büyük önem taşır çünkü belirli bir hizmetin verilerine doğrudan ve güvenli bir şekilde erişim sağlar (Aydın, 2017). API kullanımı hizmet sağlayıcılar tarafından da resmi olarak desteklenmektedir. API'ler, RESTful veya SOAP gibi protokollerin parametreleri dahilinde çalışarak HTTP istekleri aracılığıyla veri alışverişini sağlarlar. Örneğin, Twitter API'si belirli anahtar kelimelerle ilgili tweet'leri toplamak için kullanılabilirken, Google Haritalar API'si coğrafi konum bilgilerine erişime izin vermektedir. API kullanımı veri elde edilme sürecinin hem güvenli hem de daha hızlı bir şekilde yürütülmesini sağlar. Bununla birlikte, API kullanımı genellikle belirli kotalar ve erişim sınırlamaları ile birlikte gelir, bu da kullanıcıların veri çekme işlemlerini belirli bir düzen içinde yapmasını gerektirir. Bununla birlikte bir kimlik doğrulama mekanizması içerdiğinden veri güvenliği de dikkate alınmış olmaktadır. API'lerin kullanımı, veri entegrasyonu ve otomasyon prosedürleri açısından da önemlidir. Örneğin, çevrimiçi bir alışveriş platformu, Stripe veya PayPal API'lerinden yararlanarak ödeme işlemlerini kolaylaştırabilir, böylece verilerin çeşitli sistemler arasında sorunsuz aktarımı gerçekleşir ve manuel müdahale ihtiyacı ortadan kalkar. Buradan hareketle, API'lerin, modern yazılım geliştirme ve veri toplama süreçlerinin vazgeçilmez bir parçası olduğu söylenebilir.

Metin madenciliği süreçlerinde kullanılan veri toplama yöntemlerinden birisi de dosya okumadır. Süreç, metin dosyaları, PDF'ler ve Word belgeleri gibi hem yapılandırılmış hem de yapılandırılmamış veri kaynaklarından veri çıkarmayı içerir. Python'da bu dosyalardaki bilgilere erişmek ve bilgileri okumak için PyPDF2 ve docx gibi kütüphanelerden yararlanılmaktadır. Verilerin elde edilmesi süreci için uygulanabilecek bir diğer yöntem de veri tabanlarından bilgi çekme tekniğidir. Bu teknik, hem İlişkisel Veritabanı Yönetim Sistemleri (RDBMS) hem de NoSQL veri tabanları gibi çeşitli veritabanı sistemlerinden bilgilerin sorgulanmasını ve çıkarılmasını kapsar. SQL komutları ile belirli koşullara dayalı olarak ilişkisel veri tabanlarından veriler

alınabilmektedir. MongoDB ve Cassandra gibi NoSQL veritabanları, ise büyük ölçekli veri depolama ve uyarlanabilir veri yapıları için yaygın olarak kullanılmaktadır. Veritabanlarından veri elde etme süreci, orijinal veriler genellikle analitik süreçler için uygun olamayabileceğinden veri temizleme ve dönüştürme gibi ek eylemleri de kapsayabilmektedir.

Metin madenciliği süreçlerinde veri toplama aşamasında kullanılabilecek enstrümanlardan biri de anket ve formlardır. Analitik sürecine tabi tutulacak geri bildirimler anketler ve dijital anketler olan formlar aracılığıyla toplanabilir ve bu bilgiler metin madenciliği yöntemleri aracılığıyla incelenebilmektedir. Aynı zamanda sosyal medya platformları üzerinden paylaşılan kullanıcı görüşlerinin toparlanması da metin madenciliği için yapılan veri toplama yöntemlerinden biridir. Sosyal medya paylaşımlarının toplanması için ilgili platformların API'leri veya web kazıma yöntemleri tercih edilmektedir. Veri toplama süreçlerinde kullanılan bazı tekniklerden bahsetmek gerekirse, tarayıcı otomasyonu önemli tekniklerden biri olarak ifade edilebilir. Selenium, bir dizi görevi yerine getirmek üzere web tarayıcılarını otomatikleştirmek için kullanılan bir Python kütüphanesi ve aracıdır (Pfalzgraf, 2020).

Selenium gibi araçlar kullanılarak, web sayfalarının otomatik olarak gezilmesi ve verilerin çekilmesi sağlanır. Kurallı ifadeler (Regular Expression, RegEx) ise, metinsel verilerden tanımlanmış kalıpların çıkarılması için kullanılır ve dosyalar ve web sayfaları gibi kaynaklardan bilgi çıkarılmasında yoğun olarak kullanılmaktadır. API istekleri, özellikle RESTful veya SOAP yapısındaki API'lere yapılan istekler yoluyla hizmet sağlayıcıdan veri çekme sürecini ifade eder. Bu amaçla Python'da requests kütüphanesi yaygın olarak kullanılmaktadır. Bunun yanında JSON, XML gibi veri yapılarının ayıklanarak okunaklı bir hale getirilmesi işlemleri veri ayrıştırma (data parsing) olarak ifade edilir ve veri toplama aşamalarında sıklıkla uygulanan tekniklerden biridir.

2.4.2.2 Metin Ön İşleme

Metin madenciliğinde ön işleme, yapılandırılmamış haldeki metin verilerinin daha sonraki analizler için optimize edilmesi amacıyla sistematik olarak temizlenmesini ve düzenlenmesini içerir. Temel amaç, gereksiz bilgileri azaltmak, gürültülü veriyi temizlemek ve analitik algoritmaları daha anlamlı örüntü ve ilişkileri tespit edecek şekilde güçlendirmektir. Ön işleme süreçlerinin amaçlarını birkaç noktada toplamak mümkündür.

Bunlardan birisi veri tutarlılığıdır. Veri tutarlılığı büyük/küçük harf dönüşümü, noktalama işaretlerinin kaldırılması, boşlukların ve özel karakterlerin düzenlenmesi ve yazım düzeltmeleri gibi prosedürlerin uygulanması yoluyla sağlanır. İkinci bir nokta olan gürültü azaltma işlemi için tek başına bir anlam ifade etmeyen durak kelimelerin (stop words) ve özel bazı karakterler metin verisinden temizlenir. Verinin standart bir yapıya kavuşturulması için, kelimenin anlamsal köküne indirgeme (lemmatization) veya gövdeleme (stemming) işlemleri uygulanır. Buradaki amaç analize konu olan metinsel verinin farklı formlarını ortak bir notasyon ile temsil edebilmektir.

Metin madenciliği süreçlerinin ilk adımı daha önceki bölümlerde de değinildiği üzere veri toplama aşamasıdır. Bu aşama analiz edilecek metin verilerinin toplanmasıdır. Bu veriler, sosyal medya, e-postalar, anketler ve çeşitli diğer kaynaklardan gelebilir.

Verinin temizlenmesi, gürültü faktörlerinden ve anomali içeren veri gruplarından arındırılmasını ifade eder. Diğer bir ifadeyle bu aşama, metinsel verilerdeki gereksiz bilgilerin kaldırılmasını ve gürültünün azaltılmasını içerir. Web kazıma veya diğer kaynaklardan elde edilen metin verileri, NLP algoritmalarının doğruluğunu ve verimliliğini olumsuz yönde etkileyebilecek HTML etiketleri, metin dışı karakterler ve noktalama işaretleri gibi çeşitli istenmeyen öğeler içerebilir (Qais & Veena, 2023, s. 1). Bu sebeple HTML etiketlerinin ve özel karakterlerin çıkarılması ve metinlerin normalleştirilmesi gibi işlemler yapılır.

Tokenizasyon işlemi, metin verilerinin, her bir kelimenin veya ifadenin ayrı bir belirteç olarak bağımsız olarak ele alınmasını kolaylaştırmak amacıyla, tek tek kelimeler veya cümleler gibi daha küçük parçalara (token) indirgenmesini ifade etmektedir.

Kök bulma, aynı köke sahip tüm sözcükleri ortak bir biçime indirgeme işlemidir; lemmatizasyon ise çekim eklerini kaldırır ve bir sözcüğün temel veya sözlük biçimini döndürür (Balakrishnan & Ethel, 2014, s. 174). Kök kısaltma (stemming) kelimenin kök haline en yakın formunu bulur. Gövdeleme işleminin kök bulma işleminden en bariz farkı kelimenin anlamlı olup olmadığı dikkate alınmaksızın sadece son eklerin kesilmesi ile gerçekleştirilir. Gövdeleme algoritmaları sözcüğün kök formuna ulaşmak için görece basit kurallar üzerine inşa edilmişlerdir. Gövdeleme metinsel indeks oluşturma ve arama motorları gibi uygulamalarda genellikle tercih edilir. Kök bulma (lemmatization) işlemi sözcüklerin dilbilimsel olarak kök formuna ulaşma hedefine odaklandığından sözcüklerin

anlam ve bağlamını dikkate alarak çalışmaktadır. Lemmatizasyon işlemi sonucunda sözcüklerin dilbilimsel formasyonları korunmuş olur. Lemmatizasyon işlemi konu modelleme, duygu polaritesini ölçümü ve doğal dil işleme gibi işlemlerde çoğunlukla tercih edilir. Tablo 2.1’de gövdeleme (stemming) ve kök bulma (lemmatizasyon) işlemleri sonucu sözcüklerin nihai formu karşılaştırmalı olarak gösterilmiştir.

Tablo 2.1. Gövdeleme ve kök bulma işlem karşılaştırması

Kelime	Stemming	Lemmatizasyon
Çalışıyordum	Çalış	çalışmak
Kitapları	Kitap	kitap
Koşuyorduk	koş	koşmak
Güzellikleri	güzel	güzel
Öğretmenlerimizin	öğretmen	öğretmen
Bilgisayarları	bilgisayar	bilgisayar
Gördüğümüz	gör	görmek
Yazılımlarımızın	yazılım	yazılım
Müşterilerinizle	müşteri	müşteri

Bir diğer önemli aşama durak kelimeler olarak tanımlanan sözcük gruplarının veri setinden ayıklanmasıdır. Durak kelimeler (stop words), cümle içerisinde kullanım frekansı yüksek olan ancak tek başlarına bilgi değerleri çok düşük olan bağlaçlar, edatlar, zamirler, belirtili isimler gibi sözcük veya sözcük kalıplarıdır. Bunlar analitik süreçlerine dahil edilmezler. Bu sözcük gruplarının analitik sürecinden çıkarılması hem veri setinin önemli ölçüde küçültülmesini sağlar, hem de veri setindeki gürültüyü azaltarak analiz kalitesinin artırılmasına pozitif katkı sağlar.

2.4.2.3 Özellik Çıkarımı

Özellik çıkarımı (feature extraction), ham metinsel verilerden ilgili ve kullanışlı özelliklerin çıkarıldığı metin madenciliğinde çok önemli bir adımdır. Bu dönüşüm,

verileri daha kısa ve bilgi değeri yüksek bir yapıya evrilterek makine öğrenimi algoritmalarının daha etkin çalışmasını sağlar. Özellik çıkarımı metinsel verilerin anlamsal bütünlüğünü korurken boyutunu azaltır ve bu şekilde verilerin analitik algoritmaları için anlaşılabilir ve uygulanabilir bir sayısal formata dönüştürülmesini amaçlamaktadır. Özellik çıkarımı, derlemi en iyi temsil edebilecek özellikler ve niteliklerin ortaya çıkarılma sürecini ifade eden metin madenciliğinin önemli bir bileşenidir. Bu amaçla, kelime frekansı–ters doküman frekansı (term frequency- inverse document frequency- TF-IDF) terim ağırlıklandırma yöntemi en çok kullanılan yöntemlerdendir (Kınık & Güran, 2021b, s. 324). Bu yöntemler uygulanarak metinlerdeki önemli kelimeler belirlenerek öznitelik vektörleri oluşturulur.

Özellik çıkarımı işleminde kelime çantası (bag of words), n-gramlar, TF-IDF (Term Frequency-Inverse Document Frequency), kelime gömme (word embeddings) ve konu modelleme (topic modeling) gibi farklı teknikler kullanılmaktadır.

Kelime çantası modeli, doğal dil işlemede uygulanan basitleştirici bir temsildir. Bu modelde, her cümle, kelime sırası göz ardı edilerek kelimelerinin kümesi olarak temsil edilir (Peter vd., 2016, s. 293). Uygulamasının basitliği yanında, kelimelerin sırası ve bağlamı hakkında bilgi vermez. Özellik çıkarımı işleminde uygulanan diğer bir teknik olarak n-gram yaklaşımları, kelimelerin tek tek analizi yerine kelime gruplarının analizi yaklaşımına odaklanır. N-gramlar, kelime dizilerinin incelenmesini kolaylaştırır; burada bir N-gram, belirli bir metinde bulunan bir dizi N ardışık kelimeyi belirtmektedir. Örneğin, bir bigram (2 gram) ardışık iki kelimedenden oluşurken, bir trigram (3 gram) ardışık üç kelimedenden oluşmaktadır. Örneklendirmek gerekirse; “Doğal dil işleme teknikleri gelişmektedir” cümlesinden çıkarılabilecek bigramlar, “doğal dil”, “dil işleme”, “işleme teknikleri”, “teknikleri gelişmektedir” olarak söylenebilir. Trigram olarak ise, “doğal dil işleme”, “dil işleme teknikleri” ve “işleme teknikleri gelişmektedir” ifadeleri çıkartılabilmektedir.

N-gramların kullanılması, dil modellerinin geliştirilmesinde çok önemli bir husus olan kelime sıklıklarının ve kelimeler arasındaki bağlantıların belirlenmesinde etkin bir yaklaşımdır. Bu şekilde N-gram yaklaşımı kelime sırası ve bağlamı dikkate alarak analitik kaliteyi geliştirmeyi hedeflemektedir. Bu avantajının yanında N-gram yaklaşımının birtakım dezavantajları da bulunmaktadır. Daha uzun kelime dizilerinin kullanılması

anlamında gelen N'nin boyutu büyüdükçe, modelin karmaşıklığı ve hesaplama maliyetleri de artacaktır. Ayrıca, büyük N-gram modellerinin kullanılması, veri kümesi küçük boyutlu N-gramları kapsadığında aşırı uyum (overfitting) ve gürültü sebebi olabilir. Bu nedenle, N-gram modellerinin titiz bir şekilde dengelenmesi ve geliştirilmesi oldukça önemlidir.

Özellik çıkarımı işlevinde uygulanan bir diğer yöntem TF-IDF (Term Frequency-Inverse Document Frequency) yöntemidir. TF-IDF yöntemi metin madenciliğinde kelimelerin önemini ortaya çıkarmak için kullanılır. Başlangıçta, belirli bir belgedeki her bir kelimenin belgede geçme sıklığı, terim frekansını (TF) belirlemek için hesaplanır. Daha sonra ters belge frekansı (IDF), tüm belge koleksiyonunda kelimenin geçtiği belge sayısının ters logaritması alınarak hesaplanır. Son olarak TF-IDF puanı her kelime için TF ve IDF değerleri çarpılarak elde edilir ve bu da kelimenin belge içindeki önem düzeyini göstermektedir. Terim frekansı (TF) bir kelimenin belirli bir belge içindeki kullanım sıklığını gösterirken, TF-IDF değeri kelimenin bir dizi belge içinde ne derece nadir geçtiğine işaret etmektedir. İlgili kelime ne kadar az sayıda belge dizisi içinde geçiyorsa o kelimenin yüksek düzeyde ayırt edici ve bir o kadar da önemli olduğu varsayımına dayanmaktadır.

$TF-IDF = TF * IDF$ şeklinde formüle edilmiştir.

Özellik çıkarımında kullanılan tekniklerden biri olan kelime gömme, kelimeleri sayısal vektörlere dönüştürmeyi amaçlayan bir yöntemdir. Bu sayısal gösterimler, kelimelerin anlamlarını yakalamak ve aralarındaki semantik benzerlikleri ölçümlemek için kullanılır. Tipik olarak sözcük yerleştirme modelleri, sözcükleri düşük boyutlu, yoğun, sürekli vektörler olarak gösterir. Word2Vec, GloVe ve FastText algoritmaları kelime gömme tekniği uygulamalarında sıklıkla kullanılan algoritmalarlardır.

Kelime gömme tekniklerinin önemli faydalarından birisi de kelimeler arasındaki semantik bağlantıları ve benzerlikleri sayısal olarak temsil edebilme yeteneklerinde yatmaktadır. Örneğin "kral" ve "kraliçe" yi temsil eden vektörlerin yakınlığı onların anlamsal benzerliğini yansıtmaktadır. Bu vektörel gösterimler, kelime anlamlarının ve bağlantılarının anlaşılmasını geliştirerek, makine öğrenimi modellerinin yetkinliğini geliştirmektedir. Kelime gömme yöntemleri, metin kategorizasyonu, duygusal analiz, bilgi edinimi de dahil olmak üzere çok çeşitli doğal dil işleme görevlerinde sıklıkla

kullanılmaktadır. Bununla birlikte kelime yerleştirme modelleri, kapsamlı metin derlemi üzerinde eğitildiğinde daha yetkin bir performans sergiler. Bununla birlikte, bir kez eğitildikten sonra çok çeşitli dil işleme uygulamalarında etkin bir şekilde kullanılabilir ve üstün sonuçlar elde edilebilir.

Konu modelleme (topic modeling), özellik çıkarımı yöntemlerinden biri olarak, metin madenciliği ve doğal dil işleme alanlarında, bir metin gövdesi içindeki önemli temaların otomatik algılanması için yaygın olarak kullanılan bir yöntemdir. Her bir konu, sırayla, altta yatan bir konu olasılıkları kümesi üzerinde sonsuz bir karışım olarak modellenir (Blei vd., 2003, s. 993). Böylece, büyük metin veri kümelerinden anlamlı özellikler (öznitelikler) çıkarılarak verilerin daha iyi anlaşılması sağlanır. Latent Dirichlet Allocation (LDA) algoritması konu modelleme yöntemi kapsamında en yoğun kullanılan algoritmaların başında gelmektedir. Gizli Dirichlet Tahsisi (LDA) algoritması belgelerin farklı konuların birleşiminden oluştuğu ve her konunun da belirli bir kelime koleksiyonuyla karakterize edildiği varsayımına dayanmaktadır. Algoritma, belgeler ve kelimeler arasındaki örtülü bağlantıları belirlemek için istatistiksel yöntemler kullanmaktadır. Öncelikle belgelerdeki belirli konuların sıklığını ve ardından bu konular kapsamındaki kelimelerin frekansını hesaplamaktadır.

Konu modellemenin yönteminin kullanılması, kapsamlı ve karmaşık veri yığınlarından önemli niteliklerin tespit edilmesini kolaylaştırır. Örneğin, bir haber makalelerinden oluşan bir derlem içerisinde "siyaset", "ekonomi" ve "spor" gibi temalar, Gizli Dirichlet Tahsisi (LDA) algoritmasının uygulanması yoluyla otomatik olarak ayırt edilebilmektedir.

2.4.2.4 Sınıflandırma Algoritmaları

Bu bölümde çalışmamız kapsamında müşteri yorumlarının kategorizasyonu amacıyla kullandığımız dört farklı sınıflandırma algoritmasına ilişkin teorik çerçeve çizilecek olup, algoritmaların dayandıkları temel varsayımlar, modeller ve teoremler açıklanacaktır.

2.4.2.4.1 Karar Ağaçları

Karar ağaçları, veri madenciliği ve makine öğreniminde sınıflandırma ve regresyon amacıyla yaygın olarak kullanılan bir yöntemdir. Bu yöntem, veri setlerinin bir ağaç yapısı içindeki bir dizi kural ve karar aracılığıyla bölümlere ayrılmasını ve verilerde tanımlanan kalıplara ve ilişkilere dayalı olarak yapılan tahminleri içerir. Bir karar ağacı kök düğümden başlayarak birden fazla karar düğümü ve yaprak düğümü boyunca ilerler. Her karar düğümü, belirli bir özellik değerinin değerlendirmesine karşılık gelir ve sonuçlarına göre farklı dallara ayrılmaktadır. Karar ağaçları algoritması, veri madenciliği teknikleri içinde önemli bir sınıflandırma yöntemidir. Karar ağacı, kök düğümü, dalları ve yaprak düğümleri olan ağaç yapısında sınıflandırma ve regresyon modelleri oluşturur (Kocarik Gacar & Deveci Kocakoç, 2020, s. 251). Karar ağaçları karmaşık verileri anlamak, önemli özellikleri belirlemek ve olayları sınıflandırmak için değerli bir vizyon sunmakta ve çeşitli uygulamalarda yaygın olarak kullanılmaktadır. Kolayca yorumlanabilen sonuçları ve bir dizi özelliğe dayalı tahminlerde bulunmadaki yüksek doğrulukları nedeniyle tercih edilmektedir. Karar ağaçları, karmaşık veri kümelerini hızlı ve doğru bir şekilde işleme ve anlamlı içgörüler üretme yetenekleri nedeniyle trend analizi, finansal tahmin, pazarlama stratejisi geliştirme, hastalık teşhisi ve daha fazlası gibi görevler için birçok disiplin tarafından kullanılmaktadır. Karar ağaçları veri madenciliğinde kullanılan yöntemlerden biridir. Ağaç görünümüne sahip heterojen ve büyük veri setlerini benzer nitelikleri dikkate alarak homojen gruplara ayırma özelliği taşımaktadır (Berry & Linoff, 2004; Şehribanoğlu & Diler, 2018).

Karar ağaçları, bilgi kazanımı, gini indeksi ve varyans azaltma gibi farklı kriterlere göre belirlenen dallanma süreçlerini kullanır. Bilgi kazanımı, belirli bir özelliğin verileri etkili bir şekilde ayırma gücünü değerlendiren önemli bir ölçümdür. Gini indeksi ise bir karar ağacındaki her düğümden en uygun bölme özniteliğini seçmede ve gelişmiş kategorizasyon hassasiyeti elde etmede bölme öznitelikleri için bir bölme ölçüsü olarak kullanılır (Park & Kwon, 2011, s. 855). Karar ağaçlarının oluşturulması genellikle her kırılım veya dallanmada en iyi bölünmenin seçildiği yukarıdan aşağıya bir metodolojiyi takip eder. Karar ağaçlarına aşırı uyum (overfitting) sorununu çözmek için uygulamada sürecinde budama yöntemleri tatbik edilmektedir. Budama, ağaç yapısından belirli dalları kaldırarak modelin performansını artırmak için uygulanmaktadır.

Karar ağaçlarıyla ilgili temel kavram ve tanımlardan bahsetmek gerekirse, kök düğümler karar ağacının en tepesindeki düğümlerdir. İç düğümler bir özelliği veya özniteliği temsil eden ve veri kümesini alt kümelerine ayıran düğümlerdir. Dal, iç düğümden gelen ve özelliğin potansiyel değerini gösteren bir bağlantıyı temsil etmektedir. Yaprak düğüm ise karar ağaçlarının veri noktalarının sınıf etiket değerini ifade eden bileşenlerine verilen addır. Kirlilik bir düğüme ait verilerin ne kadar net olarak ayrıştığını ifade eden bir ölçüdür. Bir özellik veya öz niteliğe göre dallanma sonucunda kirliliğin ne kadar azaldığının ölçüsüne ise bilgi kazanımı adı verilmektedir. Karar ağaçlarındaki aşırı uyum (overfitting) problemini gidermek için yapılan boyut azaltma işlemine ise budama adı verilmektedir.

Karar ağaçlarının sınıflandırma süreçlerinin kalitesini artıran bazı güçlü yönleri şu şekilde sıralanabilir:

1. Yorumlama Kolaylığı; karar ağaçları, kuralların yapısını oluşturan çok sayıda veri kümesini basit, açık ve anlaşılması kolay bir şekilde sınıflandırmak için denetimli bir yaklaşımdır (Ramani & Shanthi, 2012, s. 2). Aynı zamanda grafiksel olarak hiyerarşik yapılar olarak temsil edilirler, bu da onları diğer tekniklerden daha kolay yorumlanabilir hale getirir (Rokach & Maimon, 2008, s. 12).
2. Esneklik hem kategorik hem de sayısal veriler üzerinde çalışma esnekliği sunmaktadır.
3. Hazırlık Kolaylığı; diğer bir kısım makine öğrenimi algoritmalarına kıyasla daha az miktarda veri ön işleme gerektirmektedir.
4. Veri Bağlantılarını Ortaya Çıkarma; Veri yığınları içindeki bağlantıları ve önemli özellikleri ortaya çıkarma konusunda yetkin bir çerçeve sunar.

Karar ağaçlarının uygulama tarafında bazı zayıf yönleri ise şu şekilde sıralanabilir:

1. Aşırı Uyuma Eğilim; Aşırı uyum, bir modelin belirli bir veri yığını üzerinde eğitildiğinde yüksek doğruluk gösterdiği, ancak yeni ve gözlemlenmemiş veriler üzerindeki sonuçları doğru bir şekilde tahmin etmekte zorlanmasını ifade eden bir olgudur. Bu, modelin eğitim verileri içindeki gürültüyü ve kendine has özellikleri aşırı derecede yakaladığı ve dolayısıyla etkili bir şekilde genelleme yeteneğini

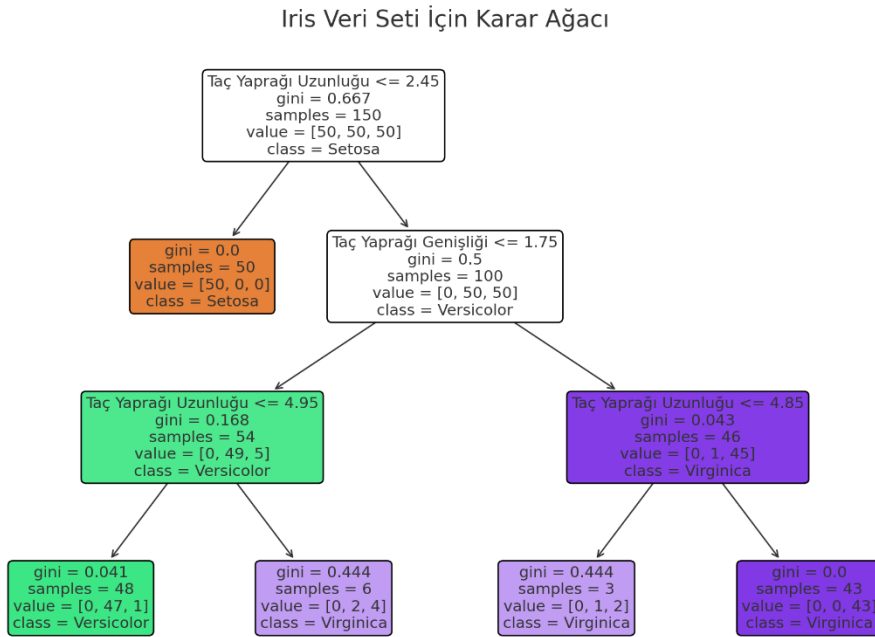
engellediği anlamına gelmektedir. Karar ağaçları bilhassa derin ağaçlarda eğitim verilerine aşırı uyum gösterebilme eğilimindedir.

2. Dallanma Karmaşası; yüksek boyutlu ve birden fazla kategoriye sahip veri kümelerinde dallara ayırma işlemi kaotik hale gelebilir ve bu da çarpık karar ağaçlarının oluşmasına yol açabilir.
3. Hassasiyet; verilerdeki minör değişiklikler bile karar ağacında majör değişikliklere neden olabilir.

Karar ağaçları, kredi riski değerlendirmesi, tıbbi teşhis, metin analizi ve arama motoru geliştirme gibi çeşitli alanlarda yaygın olarak kullanılmaktadır. Yorumlanabilirliği, uyarlanabilirliği ve esnekliği nedeniyle karar ağaçları, veri bilimcilerin araç setinde değerli bir enstrüman olarak hizmet etmektedir.

Örnek bir karar ağacı Şekil 2.1’de gösterilmiştir.

Şekil 2.1. İris veri seti karar ağacı



Kaynak: (Breiman vd., 1984)

2.4.2.4.2 Rastgele Ormanlar

Rastgele ormanlar algoritması, çok yönlülüğü ve çeşitli uygulamalardaki yüksek performansı ile tanınan, yaygın olarak kullanılan bir makine öğrenme yöntemidir. Temel

olarak algoritma, modelin doğruluğunu ve dayanıklılığını artırmak için birden fazla karar ağacını bir araya getirerek topluluk öğreniminden yararlanır.

Rastgele ormanlar, her bir ağacın bağımsız olarak ve ormandaki tüm ağaçlar için aynı dağılımla örneklenen rastgele bir vektörün değerlerine bağlı olacağı şekilde ağaç tahmincilerinin bir kombinasyonudur (Breiman, 2001, s. 5).

Rastgele Ormanlar, her bir ağacın eğitim verilerinin farklı bir alt kümesi üzerinde eğitildiği çok sayıda karar ağacının birleşiminden türetilmektedir. Her bir ağaç, veriler içindeki örüntüleri tanımlamaya çalışarak bir tahminlemede bulunur. Nihai tahmin, sınıflandırma için çoğunluk oyu veya regresyon için ortalama alma gibi teknikler yoluyla tüm ağaçların tahminlerinin toplanmasıyla belirlenir. Birkaç rastgele karar ağacını birleştiren ve tahminlerini ortalama alarak toplayan yaklaşım, değişken sayısının gözlem sayısından çok daha fazla olduğu ortamlarda mükemmel performans göstermiştir (Biau & Scornet, 2016, s. 197).

Algoritma, tipik olarak diğer birçok sınıflandırma algoritmasıyla karşılaştırıldığında üstün hassasiyet sunmaktadır. Aynı zamanda önyükleme toplama (bootstrap aggregating) ve rastgele alt uzay tekniklerinin kullanılarak aşırı uyum olasılığının azaltılması algoritmanın güçlü yönlerinden biridir. Önyükleme toplama (bootstrap aggregating ya da kısaca bagging) tekniğindeki temel amaç, birden fazla modelin oluşturulması ve tahminlerinin sentezi yoluyla bir makine öğrenme algoritmasının değişkenliğini en aza indirmektir. Rastgele ormanlar tekniği tahminleme sürecinde en önemli özellikleri belirlemek için kullanılabilir. Bu amaçla özneliklerin model üzerindeki etkisinin değerlendirilmesi, her bir özneliğin önem skorunun hesaplanmasıyla gerçekleştirilebilmektedir.

Öznelik seçimi bağlamında, her düğümdeki en uygun bölünme noktası, tüm olası özellikleri dikkate almak yerine, rastgele seçilen bir özellik alt kümesi kullanılarak tanımlanmaktadır. Bu yaklaşım, karar ağaçlarının birbirlerinden bağımsızlığını arttırmak için uygulanır. Oylama bakımından ise sınıflandırma görevlerinde nihai sınıf tanımı, bağımsız karar ağaçlarının tahminlerini birleştiren bir oylama süreci aracılığıyla belirlenir. Regresyon görevlerinde nihai sonuç, ağaçların ürettiği tahminlerin ortalaması alınarak elde edilmektedir.

Algoritmanın güçlü yönlerinden birisi genelleme yetkinliğidir. Rastgele Ormanlar, iyi sınıflandırma ve genelleme yeteneği nedeniyle çeşitli alanlarda başarı elde etmiştir (H. B. Li vd., 2010, s. 160). Bunun yanında çoklu karar ağaçlarının birleştirilmesi model içindeki çeşitliliği arttırmakta ve çeşitli hata kaynaklarının dengelenmesini sağlamaktadır. Diğer taraftan çoklu karar ağaçlarının oluşturulması ve ardından gelen oylama süreci, gerekli hesaplama kaynaklarında önemli bir artışa neden olabilmektedir. Ek olarak, çok sayıda karar ağacının bir araya getirilmesiyle üretilen topluluk modeli, yorumlanabilirliği açısından yüksek derecede karmaşıklığa sahiptir ve tekli karar ağaçlarıyla karşılaştırıldığında daha az anlaşılırdır.

Sonuç olarak, rastgele ormanlar algoritması, çeşitli veri madenciliği uygulamalarında etkinliğini gösteren sağlam ve uyarlanabilir bir araçtır. Bununla birlikte, hesaplama karmaşıklığı ve sonuçların yorumlanmasındaki zorluklar da dahil olmak üzere belirli kısıtlamaların kabul edilmesi çok önemlidir.

2.4.2.4.3 Destek Vektör Makineleri

Destek Vektör Makinesi (DVM), sınıflandırmaya dayalı geleneksel bir makine öğrenimi yöntemidir. Büyük boyutlu problemlerin ikili formunu çözme fikrinden türetilmiştir, Böylece sınıflandırıcı yapısal risk minimizasyonu ilkesine ulaşmak için yalnızca az sayıda destek vektörüne dayanır. (Gaye vd., 2021, s. 1). DVM'nin temeli, bir dizi pozitif örneği bir dizi negatif örnekten maksimum marjla ayıran bir hiper düzlem bulmak üzerine kurulmuştur (C. Li vd., 2011, s. 19). Hiper düzlem, verileri iki sınıfa ayıran bir karar sınırı olarak düşünülebilir. Destek vektör makineleri algoritması özellikle küçük ve orta büyüklükteki veri yığınlarıyla çalışırken daha yüksek başarı oranına sahiptir.

Destek vektör makinelerinin teorik altyapısı marjın en çoklanması üzerine bina edilmiştir. Marj, değinildiği gibi, hiper düzlem ile en az mesafedeki veri kümelerini ifade etmektedir. Bu marjın maksimizasyonu veri kümeleri arasındaki ayırımın en net olduğu anlamına gelmektedir. Marjın maksimize edilmesine destek veren veri kümelerine ise destek vektörleri adı verilmektedir.

Destek vektör makineleri, yüksek başarı oranları, aşırı öğrenmeye (overfitting) karşı defansif yeteneği ve oldukça geniş uygulama alanı ile ön plana çıkmaktadır. Ayrıca, destek vektör makineleri doğrusal ve doğrusal olmayan veri kümelerini etkili bir şekilde modelleme yeteneğine sahiptir. DVM fazla öğrenmeyi (overfitting) minimize ederek iyi

bir genelleme yeteneđi sunar. Bu, modelin yeni veriler üzerinde de iyi performans göstermesini sağlar. Aynı zamanda farklı çekirdek fonksiyonlarının kullanımı, veri setinin yapısına uygun şekilde modeli optimize etmeyi mümkün kılar. Çeşitli veri setleri için sınıflandırma probleminin çözümüne ilişkin Destek vektör makinelerinin uygulanması sürecinde çekirdek fonksiyonu seçimi ve parametre optimizasyonu önemli rol oynamaktadır. Dolayısıyla tek bir çekirdek fonksiyonu ile destek vektör makinelerinin performansına ilişkin genel bir yargıya varmak oldukça güçtür (Ayhan & Erdođmuş, 2014, s. 176).

Diđer taraftan destek vektör makinelerinin görece zayıf kaldıđı bazı noktalar bulunmaktadır. Destek vektör makinelerinin nispeten daha büyük veri yığınlarıyla çalışırken hesaplama maliyeti ve işlem süresi bakımından verimsiz olması sebebiyle çok büyük denilebilecek veri kümeleriyle yapılan çalışmalarda problem olabilmektedir. Diđer taraftan yanlış parametre seçimleri model başarımını olumsuz yönde etkileyebileceğinden, modelin etkinliđi uygun parametrelerin seçimiyle doğrudan bağlantılıdır. Ayrıca DVM, sınıflandırma ve regresyon problemlerinde çalıştırılabilir. DVM regresyon metodundaki temel fikir, eldeki eğitim verilerinin karakterini mümkün olduğunca gerçeğe yakın bir şekilde yansıtan ve istatistiksel öğrenme teorisine uyan doğrusal ayırıcı fonksiyonun bulunmasıdır (Yakut vd., 2014, s. 143). Destek vektör makinelerinin etkinliđi, veri kümesinin uygun şekilde ölçeklendirilmiş olması ile doğrudan ilişkilidir. Verilerin gerektiđi şekilde ölçeklendirilmemesi (bir veya birden fazla sınıfın diđerlerine nazaran çok daha fazla yaygın olması) durumu modelin doğruluđu üzerinde olumsuz bir etkiye sahip olacağından, bu durum da DVM'nin zayıf yönleri arasında değerlendirilmektedir.

Sonuç olarak, çok sayıda sınıflandırma çalışmasında DVM'ler etkili bir şekilde uygulanabilmektedir. Ancak, bununla birlikte özellikle büyük veya ölçeklendirilmemiş dengesiz veri yığınlarıyla yapılan çalışmalarda, olası kısıt ve performans kayıplarının farkında olmak önemlidir.

2.4.2.4.4 K-en Yakın Komşu

k-En Yakın Komşu (k-NN) algoritması, makine öğrenmesi ve veri analitiđi disiplinlerinde yoğun bir uygulama alanı bulan denetimli öğrenme yöntemidir. Temel olarak, benzer veri noktalarının büyük olasılıkla aynı sınıfa dahil olacağı veya benzer

hedef değerleri paylaşacağı kabulüne dayanır. Algoritma, önce veri setindeki yeni henüz sınıflandırılmamış bir veri noktasına en yakın olan k veri noktasının belirlenmesi ardından yeni noktanın bu en yakın k komşular arasındaki baskın sınıf veya değere göre kategorize edilmesi prensibine göre çalışmaktadır.

İyi bilinen k-en yakın komşu (kNN) sınıflandırıcısının doğruluğu büyük ölçüde k değerinin seçimine bağlıdır (Bhattacharya vd., 2015, s. 1). Dolayısıyla k parametresinin seçimi, k-En Yakın Komşular (k-NN) algoritmasının etkililiği üzerinde önemli bir etkiye sahip olabilecek çok önemli bir tasarım seçimidir. K sayısı yeni veri noktasına en yakın nokta olarak alınacak nokta sayısını ifade eder. Küçük bir k değeri daha esnek bir modele yol açar, gürültüyü yakalamaya ve yüksek varyans sergilemeye yatkındır, bu da aşırı uyuma neden olmaktadır. Tersine, büyük bir k değeri daha yumuşak bir karar sınırına yol açar, potansiyel olarak modeli aşırı basitleştirir ve yetersiz uyumla sonuçlanmaktadır (Sriramoji, 2024).

Bu bağlamda belirli veri kümeleri ve çalışma alanları özelinde en uygun k değerinin hesaplanabilmesi amacıyla bazı dinamik yöntemler geliştirilmiştir.

Komşuluk algoritma özelinde kullanılan önemli kavramlardan biridir. Her bir veri noktasının uzayda bir noktayı temsil ettiği varsayımında k-NN yeni veri noktasına en yakın “k” adet veri noktasını ortaya çıkarmaktadır.

Bahsedilen veri noktaları arasındaki mesafe farklı yaklaşımlarla hesaplanabilmektedir. Bunlardan en yaygın olarak bilinen Öklid mesafesidir. Öklid mesafesi iki nokta arasındaki düz bir çizgi mesafesini hesaplar ki iki boyutlu düzlemde bu Pisagor teoremine denktir. X ve y iki farklı veri noktası ve x_i ve y_i ise bu veri noktalarının farklı düzlemlerdeki değerleri iken, öklid mesafesinin matematiksel formülü şu şekildedir:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Bir diğer mesafe ölçütü Manhattan mesafesidir. Manhattan mesafesi, iki nokta arasındaki mesafeyi koordinatlarındaki mutlak farkları toplayarak hesaplayan bir ölçümdür. Dik

açılı bir ızgara üzerinde çizilen yolların uzunluğu olarak ifade edilebilir. Matematik formülü şu şekildedir.

$$d(x, y) = \sum_{i=1}^{\{n\}} |x_i - y_i|$$

Minkowski mesafesi Öklid uzayı'nda bir metrik'tir, Öklid mesafesi ve Manhattan mesafesi'nin bir genelleştirilmesi ile oluşturulur (Atcılı, 2022). Matematiksel olarak şu şekilde formüle edilmektedir.

$$d(x, y) = \left(\sum_{i=1}^{\{n\}} |x_i - y_i|^p \right)^{\frac{1}{p}}$$

Burada p mesafenin hesaplanma yöntemini gösteren bir değişkendir. P'nin 1 olması Manhattan mesafesi, 2 olması ise Öklid mesafesinin elde edileceği anlamını taşımaktadır. P değişkeni bu şekilde farklı değerler alarak farklı problem ve veri tipleri için çözüm esnekliği sağlamaktadır. Dolayısıyla analiz için en uygun "p" değerinin seçimi, veri yığınının belirli özelliklerine ve analitik süreçten beklenen amaçlara bağlı olarak değişebilmektedir.

k-En Yakın Komşular (k-NN) algoritmasında mesafe ölçüm tekniğinin seçimi, algoritmanın etkinliğini ve model başarımını önemli ölçüde değiştirebilen önemli bir belirleyicidir. Uygun mesafe ölçüm tekniğinin seçilmesi sürecinde üzerinde çalışılan verinin türü, özellikler arasında belirgin bir korelasyonun olup olmaması, veri yığınlarının boyutu, aykırı verilerin veri seti içindeki yoğunluğu gibi hususlar göz önünde bulundurulmalıdır.

Sonuç olarak, k-En Yakın Komşu algoritması, çeşitli avantaj ve dezavantajlara sahip, çok yönlü ve sağlam bir denetimli öğrenme yöntemidir. Bu teknikle ilgili temel ilkelerin ve tasarım bileşenlerinin detaylı bir şekilde kavranması ile k-NN algoritması etkin olarak kullanılabilir.

2.5 Duygu Analizi

Fikir madenciliği (opinion mining) olarak da adlandırılan duygu analizi, doğal dil işleme alanında, yazılı metindeki öznel içeriğin çıkarılması ve incelenmesine odaklanan bir disiplin olarak ortaya çıkmıştır. Duygu analizi (görüş madenciliği), metin belgeleri

içerisinde yer alan nesnelere, ürünlere, servislere ya da organizasyonlara ilişkin görüş, duygu, tutum gibi öznel bilgilerin, makine öğrenmesi, istatistik ve doğal dil işleme gibi alanlardan teknik ve yöntemlerin kullanılması ile çıkarılmasını amaçlayan bir araştırma alanıdır (Onan, 2020, s. 374). Duygu analizi sayesinde, işletmeler, tüketicilerin duygularını olumlu, olumsuz veya nötr olarak sınıflayarak tüketicilerin tepkilerini ölçebilmektedir (İnan, 2024, s. 111). Bu teknik, sosyal medya platformlarının hızla yaygınlaşması ve işletmelerin müşterilerinin, paydaşlarının ve daha geniş kitlelerin tutum ve algılarına ilişkin fikir sahibi olma zorunluluklarının bir sonucu olarak son zamanlarda rağbet görmüştür.

Duygu analizi alanı, pozitif, negatif ve nötr ifadeler arasında ayrım yaparak belirli bir metinde mevcut olan duygusal tonların tanımlanmasına ve sınıflandırılmasına odaklanır. Bu veriler, pazarlama stratejileri, ürün geliştirme çalışmaları ve sosyopolitik araştırmalar dahil olmak üzere çeşitli amaçlar için oldukça değerlidir. İşletmeler ve araştırmacılar, çevrimiçi incelemeler, sosyal medya güncellemeleri ve müşteri yorumları gibi materyallerdeki duygusal içeriği analiz ederek, hedef kitlelerinin duygu ve bakış açıları hakkında bilgi sahibi olabilmektedirler. Duygu analizinin önemini artıran en önemli husus karar vericilere hedeflenen grubun duygu ve davranış kalıpları hakkında gerçek zamanlı bilgi sağlama kapasitesidir.

Duygu analizini gerçekleştirmek için kural tabanlı yaklaşımlar ve makine öğrenmesine dayalı yöntemler gibi yaklaşımlar mevcuttur. Bununla birlikte her iki yöntemin üstün taraflarını birleştirmeyi amaçlayan hibrit yaklaşımlar da geliştirilmektedir. Kural tabanlı teknikler, belirli bir metnin duygusal tonunu sınıflandırmak için önceden var olan sözlüklerin veya duygu yüklü terimler içeren sözlüklerin kullanılmasını içerir. Öte yandan, belge tabanlı (document level) yaklaşımlar, genellikle (denetimli veya denetimsiz) makine öğrenimi algoritmalarının doğrudan uygulamalarını içerir. Problem duygu sınıflandırması olarak adlandırılırsa da bir regresyon problemi olarak da ele alınabilir. Çoğu duygu sınıflandırma yaklaşımı, belgeleri yalnızca olumlu, olumsuz ve nesnel türlere ayırırken, duygu sınıflandırması örtüşen anlamlara sahip daha fazla kategori içerir ve doğru bir şekilde gerçekleştirilmesi daha zordur (Liu vd., 2016, s. 597).

Kurala dayalı yaklaşımlar, önceden oluşturulmuş duygu sözlükleri ve dilbilgisi kurallarından yararlanarak metinlerin analiz edilmesi mantığıyla çalışmaktadır. Bir diğer

ifadeyle duygu analizine yönelik kuralla dayalı yaklaşımlar, belirli bir metinde ifade edilen duyguyu kategorize etmek için önceden belirlenmiş sözlükleri ve kuralları kullanır. Duygu analizinde kullanılan bir diğer yöntem olan makine öğrenimine dayalı teknikler ise metinsel veriler içindeki duygusal örüntüleri tanımlamak ve belirli eğitim verileri kullanılarak eğitilmiş algoritmalar kullanarak duyguları kategorize etmek için kullanılır. Bu süreçteki başlıca yöntemler Naive Bayes, Destek Vektör Makineleri ve Derin Öğrenme modelleri olarak sayılabilir.

Bununla birlikte duygu analizi çalışmaları insan dilinin karmaşık ve çok yönlü doğasından kaynaklanan birtakım zorlukları da beraberinde getirir. Bu karmaşıklık, bir metinde aktarılan duyguların doğru bir şekilde tanımlanmasını ve yorumlanmasını engelleyebilmektedir. Örneğin alaycılık, ironi veya günlük konuşma dili örnekleri, duygu analizi algoritmaları için zorluklar oluşturabilir ve potansiyel olarak yanlış yorumlamalara yol açabilir. İlave olarak metnin genel tonu ve analizi yürüten kişinin öznel önyargıları gibi bağlamsal faktörler, duygu analizi sürecini daha da karmaşık hale getirebilmektedir.

Barındırdığı zorluklara rağmen duygu analizi, belirlenmiş grubun duygu ve perspektiflerini anlamaya çalışan işletmeler ve araştırmacılar için önemli bir enstrüman olmaya devam etmektedir.

ÜÇÜNCÜ BÖLÜM

MÜŞTERİ YORUMLARININ SINIFLANDIRILMASI VE DUYGU ANALİZİ

Bu bölümde analiz süreçlerine konu olacak müşteri yorum verilerinin internet ortamından belirli bir dosya notasyonu şeklinde elde edilmesi, bu verilerin metin madenciliği sınıflandırma süreçlerine hazır hale getirilmesi, metin kategorizasyon süreçlerindeki işlem basamakları detaylandırılacaktır. Sınıflandırma modellerinin detaylandırılmasından sonra duygu analizi çalışması için tasarlanan metin madenciliği modeli bir akış mantığı içerisinde ele alınarak detaylandırılacaktır. Sınıflandırma modellerine ait başarımlar performansları ve ilgili diğer metrikler her bir kalite boyutu özelinde ve genel olmak üzere tablolar şeklinde paylaşılacaktır.

Analiz edilecek veriler pazaryeri tabir edilen çok büyük çevrimiçi alışveriş sitelerinden alınan gerçek müşteri yorumlarıdır. Yorumların önemli bir bölümü çalışma konusu olan kalite boyutlarıyla doğrudan veya dolaylı olarak ilgili sayılamayacak çok fazla metin verisi içerdiğinden dolayı çalışmaya dahil edilmeyerek göz ardı edilmiştir. Bahsedilen alışveriş platformlarından elde edilen yaklaşık 1182 satır müşteri yorum verisi çevrimiçi alışveriş platformlarından elde edilmiştir. Veri setinin tamamı belirli bir markaya ait çamaşır makinesi modellerine yapılan müşteri yorumlarından oluşmaktadır. Tek bir model üzerinde yapılacak çalışma veri miktarından dolayı istenen etkinliği sağlayamayacağından dolayı aynı markanın tüm çamaşır makinesi modellerine yapılan müşteri yorumları üzerinde analitik süreçler yürütülmüştür. Bu veriler marka-model özelinde ayrı ayrı elde edilmiş ve daha sonra tek bir veri seti oluşturacak şekilde konsolide edilmiştir.

Çalışma kapsamında yapılan veri ve metin madenciliği süreçleri KNIME analitik platformu üzerinden yürütülmüştür. KNIME platformu bir akış (workflow) mantığında çalışan, veri analizi ve görselleştirme işlemlerinin yapılabildiği, görsel ve modüler bir altyapı sunmaktadır. Bunun yanında çok farklı ve çeşitli veri kaynakları ve R, Python ve

Weka gibi çeşitli geliştirme ortamlarıyla entegrasyon kabiliyetinin üst düzey olması önemli bir avantajdır. KNIME’da yürütülen analiz süreçleri farklı görevler için özelleştirilmiş düğümler (node) vasıtası ile gerçekleştirilir. Sürükle bırak yöntemi ile ETL, Extract (Çıkart), Transform (Dönüştür), Load (Yükle) süreçlerinin gerçekleştirilmesine imkân veren birçok düğüme sahiptir.

3.1 Müşteri Yorumlarının Sınıflandırılması

Müşteri yorumları her bir marka model için ayrı ayrı olarak elde edildikten sonra tek bir veri seti haline getirilmiştir. Yaklaşık 6000 kadar müşteri yorum verisi elde edilmiş, bunlardan büyük çoğunluğunun kalite boyutlarıyla doğrudan veya dolaylı olarak ilgili bir veri içermemesi dolayısıyla yaklaşık olarak 1200 satır yorum verisine indirgenmiştir.

Birçok müşteri yorumu aynı anda birden fazla kalite boyutuyla ilgili olduğu ve birçok yorum içerisinde yine bir den fazla kalite boyutuyla ilgili ifadeler olması çalışmanın zorluklarından biridir. Bunun yanında kalite boyutlarının da kapsam ve ifade edilen değer anlamında birbirilerinden çok net sınırlarla ayrılmıyor olmaları da sınıflandırma başarımını zorlaştıran faktörler olarak gösterilebilir. Kalite boyutlarının birbiriyle çok fazlasıyla ilişkili ve geçirgen bir yapıya sahip olmaları dolayısıyla birkaç kelimelik çok kısa yorum verileri bile birden fazla sınıfa (kalite boyutuna) atanabilmektedir.

Öncelikle müşteri yorumlarının kalite boyutlarına göre kategorizasyon probleminde klasik çok sınıflı (multi-class) yaklaşım kullanılmıştır. Bu yaklaşımın temel varsayımı her bir elemanın tek bir sınıf etiketine sahip olabileceği, diğer bir deyişle her bir yorum cümlesinin sadece tek bir kalite boyutuna ait olabileceği ön kabulüne dayanır. Ancak bahsedildiği gibi, kalite boyutlarının birbirleriyle olan bağlam ve kapsam ilişkisinden dolayı birçok defa tek bir müşteri yorumunun birden fazla kategoriye atanabiliyor olması gerekmektedir. Bu problemin çözümü için her bir kalite boyutu özelinde bir sınıflandırma modeli kurularak, müşteri yorumlarının ilgili kalite boyutuna ait olup olmadığının analizinin yapılacağı bir çözüm geliştirilmiştir. Her bir kalite boyutu için geliştirilen bu model ile, çoklu etiketli (multi label) bir sınıflandırma problemi, tüm yorum satırlarının sekiz kalite boyutuna ait olup olmadığını her bir kalite boyutu için analiz eden bir ikili sınıflandırma (binary classification) problemine indirgenmiş olmaktadır.

Çalışmada hem ikili sınıflandırma hem de çoklu sınıf sınıflandırma yöntemleri sonucu oluşturulan modellerin başarımları da gösterilecektir.

Şekil 3.1. Müşteri yorumları veri seti

Burada her yorum satırının karşısında ilgili yorumun ait olduğu tüm kalite boyutları “1”, ilgili olmadığı kalite boyutları ise “0” olarak işaretlenmiştir. Görüleceği üzere birçok yorum birden fazla kalite boyutu bağlamında değerlendirilebilmektedir.

3.1.1 İkili Sınıflandırma Modelleri

96

için model çalıştırılıp veriler etiketlendikten sonra birden fazla kalite boyutuna atanmış yorum verileri görülebilir.

İkili sınıflandırma süreçlerinde test ve eğitim için kullanılan veri seti ilgili kalite boyutu sütunundaki 1 ve 0 değerlerin string değerlere dönüştürülerek kullanılacak veri seti (örneğin uygunluk boyutu için) şu şekilde gözükecektir.

Şekil 3.2. Örnek bir kalite boyutu veri seti

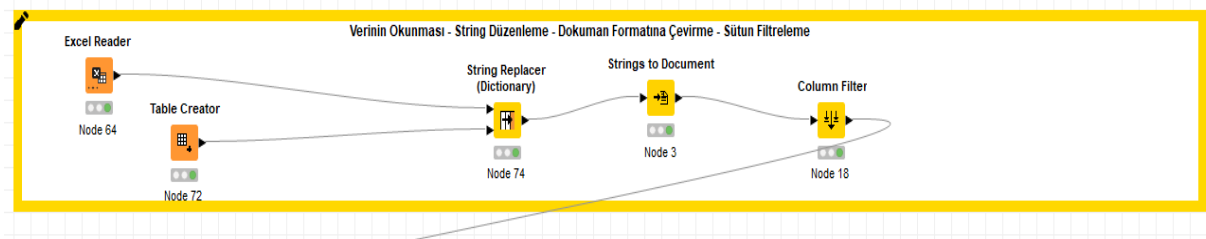
	Yorum	UYGUNLUK	DUYGU	TEK
148	Gayet sessiz ve yıkaması şahane	UYGUNLUK	POZ	UYGUNLUK
149	3 aydır kullanıyorum, gayet iyi	NONE	POZ	GUVENİLİRLİK
150	Kaliteli bir ürün teşekkür ederim	NONE	POZ	ALGILANAN KALITE
151	Su faturam artmadı ancak elektrik faturam yüzde 40 arttı bu ay.	UYGUNLUK	NEG	UYGUNLUK
152	Diğer özellikleri şahane	NONE	POZ	OZELLİKLER
153	elektrik tüketimini göze alarak tercih yaptım.	UYGUNLUK	NEG	UYGUNLUK
154	Makineyi beğendim tavsiye ederim.	NONE	POZ	ALGILANAN KALITE
155	Fiyatına göre çok iyi ürün	NONE	POZ	ALGILANAN KALITE
156	Kaliteli malzemeden yapılmış.	NONE	POZ	DAYANIKLILIK
157	Sıkarken çok ses yapıyor. Her türlü önlemi aldım ayak aldım vs ama sıkma sesini kesemedim	UYGUNLUK	POZ	UYGUNLUK
158	Programlama, teknoloji bu kadar ilerideyken standart kalmış. Öyle aham şaham programları yok. .	NONE	NEG	OZELLİKLER
159	9 KG kısımda doğru. Kazanı büyük.	NONE	POZ	OZELLİKLER
160	Yıkarken 10 yıllık eski makinemden daha fazla ses çıkıyor. Öürültü diyemem ama daha sessiz olmasını bekliyordum.	UYGUNLUK	NEG	UYGUNLUK
161	Yıkaması güzel, özellikleri ve zaman ayarları başarılı.	NONE	POZ	PERFORMANS
162	Fiyatına göre hak ediyor... uygun fiyata kalite budur	NONE	POZ	ALGILANAN KALITE
163	Güzel yıkıyor	NONE	POZ	PERFORMANS
164	çok sessiz değil	UYGUNLUK	NEG	UYGUNLUK
165	Çok sessiz ve kaliteli bir ürün	UYGUNLUK	POZ	ALGILANAN KALITE
166	Bu fiyata bu kalite kaçmaz deyip aldık. Memnunuz	NONE	POZ	ALGILANAN KALITE
167	Sessiz ve kaliteli	UYGUNLUK	POZ	ALGILANAN KALITE
168	Güzel makina biraz sesli sadece rahatsız edici bir ses yok ama	UYGUNLUK	POZ	UYGUNLUK
169	Program sayısı az 50 derece olmaması kötü	NONE	NEG	OZELLİKLER
170	Çok iyi	NONE	POZ	ALGILANAN KALITE
171	Ürün çok iyi enerji tasarrufu hakkında bir şey diyemem daha elektrk faturası gelmedi	UYGUNLUK	POZ	ALGILANAN KALITE
172	çok az sesi var ama rahatsız etmiyor bittiğinde uyarı veriyor	UYGUNLUK	POZ	OZELLİKLER
173	Ürün gayet iyi ve kullanışlı	NONE	POZ	OZELLİKLER
174	Ürün son derece kullanışlı ve ergonomik.	NONE	POZ	ESTETİK

İkili sınıflandırma için oluşturulan modelin görüntüsü Şekil 3.3'deki gibidir.

3.1.1.1 Verinin Okunması, Düzenlemesi ve Doküman Formatına Dönüştürülmesi

Bu bölüme ait KNIME platformu akışı Şekil 3.4'deki gibidir.

Şekil 3.4. Veri okuma ve doküman dönüşümü iş akışı



98

Veri seti Excel formatında olduğu için Excel reader düğümü kullanarak öncelikle veri seti açılmıştır. Müşteri yorumları hem dilbilgisi hem de imla kuralları açısından çok fazla hatalı, yanlış yazılmış anlamsız kısaltma ve ifadelerle dolu olduğu için bu ifadelerin model başarısını minimum etkilemesi açısından düzeltilmesi amacıyla string replacer düğümü kullanılarak müşteri yorum verileri üzerinde belirli oranda bir düzeltme amaçlanmıştır. Bu doğrultuda table creator düğümü ile çok fazla yanlış yazılan ancak kalite boyutları bağlamında etkileyici ve önemli olabilecek kelime ve kelime dizilerinden oluşan bir tablo oluşturulmuştur. Oluşturan tablo konfigürasyon ekranı Şekil 3.5'te gösterilmiştir.

Şekil 3.5. Table creator konfigürasyon penceresi

	S	column1	S	column2
Row0		makina		makine
Row1		makina		makine
Row2		ozellik		özellik
Row3		ozellik		özellik
Row4		guzel		güzel
Row5		şş		ş
Row6		proğram		program
Row7		çook		çok
Row8		çok		çok
Row9		coook		çok
Row10		cook		çok
Row11		calisiyor		çalışıyor
Row12		çalisiyor		çalışıyor
Row13		F/P		fiyat perform...
Row14		f/p		fiyat perform...
Row15		fp		fiyat perform...
Row16		FP		fiyat perform...
Row17		urun		ürün
Row18		ürün		ürün
Row19		hizli		hızlı
Row20		hizli		hızlı
Row21		iyi		iyi
Row22		iyi		iyi

The output table has 58 rows and 2 columns. ☐ Highlight output table

Burada veri seti içinde yaygın olarak yanlış yazılan, ancak model başarımı üzerinde etkili olabilecek sıkça kullanılan ifadeleri tek tipleştirerek düzeltilmiştir. String replacer düğümü tüm doküman içerisinde column1 de bulduğu ifadeleri column2 deki doğrusuyla değiştirerek düzeltecektir.

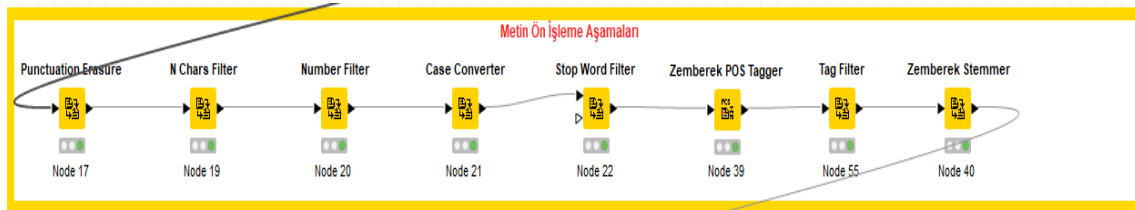
Strings to Document D ğ m  ham metin (string) haldeki metin verilerini ilerleyen s re lerde kullanabilmek i in dok man nesnelere d n  t r r. Metin verilerinin KNIME in metin madencili i d ğ mleriyle  alı abilmeleri i in  n i leme ve analiz s re lerinden  nce dok man formatına d n  t r lmesi gerekmektedir.

Buradaki  nemli parametrelerden birisi Word tokenizer olarak Zemberek Turkish Tokenizer in se ilmesi ve bir sınıflandırma yapaca ımız i in use categories from column se ene i ile ilgili kalite boyutu s tunun se ilmesi gerekti idir. Tokenizer bir karakter dizisini daha kolay analiz edilebilmesi amacıyla daha k   k par alara (token) ayırır. Zemberek ise T rk e do al dil i leme uygulamaları i in tasarlanmış a ık kaynaklı bir yazılım k t phanesidir. T rk enin kendine  zg  dilsel  zelliklerine odaklanılarak geli tirilen Zemberek, T rk e metin verilerinin analizi ve i lenmesi konusunda olduk a yetkindir.  alı manın tamamında dil i leme fonksiyonlarıyla ilgili t m s re lerde zemberek k t phanesi kullanılmıştır. Zemberek KNIME analitik platformunun standart kurulumunda bulunmamakta, eklenti olarak daha sonradan kurulması gerekmektedir. Column Filter d ğ m  ilerleyen a amalarındaki  n i leme ve analitik s re lerinde kullanımı gerekli g r lmeyen veri setindeki s tunların (column) filtrelenebilmesi amacıyla kullanılmaktadır.

3.1.1.2 Metin  n I leme S re leri

Metin  n i leme s re leri metin madencili i s re lerinin en kritik  nemdeki a amasıdır. Burada ama  saf haldeki metin verisini analiz edilmeye hazır, m mk n oldu unca tekd ze, standart, g r lt  fakt rlerinden arındırılmış ve sayısal temsile hazır bir formata d n  t rmektir. İyi bir  n i leme s reci model ba arımını artıran  nemli hususlardan biridir. Model i eresindeki metin  n i leme s re leri  ekil 3.6'da g sterildi i  ekilde y r t lmektedir.

 ekil 3.6. Metin  n i leme s re leri



Punctuation Erasure düğümü giriş portuyla kendisine bağlı dokümandaki alfa nümerik olmayan karakterleri (noktalama işaretleri) dokümandan temizlemektedir. N Chars Filter düğümü ile girdi olarak kullandığımız doküman içerisinde bulunması istenen en küçük karakter sayısına sahip kelimeleri filtrelemek için kullanılmıştır. Burada girdiğimiz 2 değerinden daha düşük karakter sayısına sahip kelimeler düğüm içerisinde filtrelenerek analize dahil edilmemektedir.

Number Filter düğümüyle girdi dokümanındaki nümerik karakter filtrelenmektedir. Case Converter düğümü tüm dokümandaki karakterlerin büyük/küçük harf dönüşümlerini gerçekleştirmek için kullanılan bir düğümdür. Bu model içerisinde tüm karakterler küçük harfe dönüştürülmüştür.

Stop Word Filter düğümü kendi başına bir anlam ifade etmeyip ancak cümle içerisinde anlam kazanan, bağlaç, zamir, edat gibi kelimelerin analiz sürecinden hariç tutulmasını sağlamak için kullanılmaktadır. Etkisiz kelimeler de denen bu kelime grubu bu düğüm içerisinde filtrelenmekte ve doküman üzerinde bir anlamda boyut indirgeme çalışması yapılmış olmaktadır. Buradaki önemli parametre ise düğüm seçenekleri içerisindeki stopwordslist parametresinin turkish olarak yapılandırılması gereğidir.

Zemberek Pos Tagger düğümü zemberek doğal dil işleme kütüphanesi ile Türkçedeki karşılığı cümleleri öğlerine ayırma (Part of Speech Tagging) süreçlerini gerçekleştirmektedir. Çalışma kapsamında tokenlara bölme (tokenization), gövdeleme (stemming), ve kök bulma (lemmatization) işlemlerinde zemberek kütüphanesi kullanılmaktadır.

Tag Filter düğümü bir önceki düğümden işaretlenen bileşenlerin doküman nesnesi içinden filtrelenmesi amacıyla kullanılmaktadır.

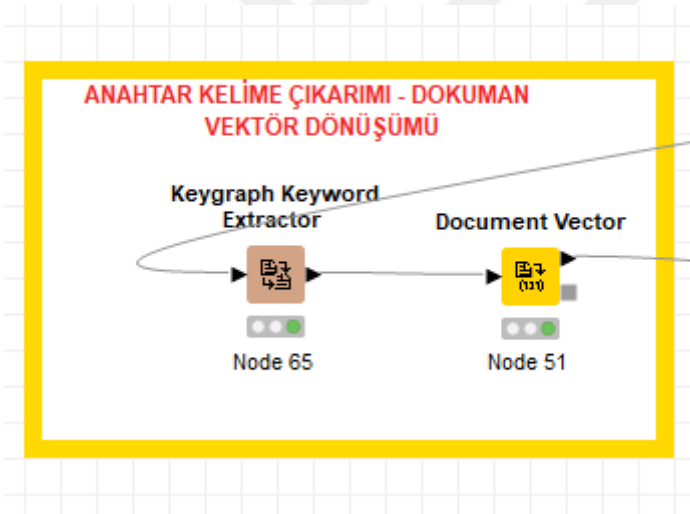
Zemberek Stemmer düğümü metin ön işleme süreçlerinin önemli adımlarından biri olan gövdeleme (stemming) işlemini Türkçe dil kalıp ve kurallarına göre yürütülmesi amacıyla kullanılmaktadır.

Ön işlemenin sonunda metin verisi sayısal olarak temsil edilebilir bir duruma gelmekte, daha standart ve kurallı bir yapıya kavuşmaktadır.

3.1.1.3 Anahtar Kelime Çıkarımı ve Vektör Dönüşümü

Ön işleme süreçlerini takip eden süreç, metin veri kümesini temsil edebilecek anahtar kelimelerin tespiti ve metin verisinin sayısal değerler ifade eden vektörel bir yapıya dönüşümü işlemidir. KeyGraph Keyword Extractor düğümü doküman nesnesini ve metnin ana temasını temsil edebilecek anahtar kelimelerin tespitini ve bu kelime ya da kelime grupları arasındaki birlikte geçme sıklıkları gibi noktaları ortaya çıkarır. Bu yönüyle kelime çantası (bag of words) yaklaşımından farklılık gösterir. Kelime çantası yaklaşımı doküman nesnesi üzerinde kelimelerin bulunma frekanslarına odaklanarak metni kelime frekansları olarak temsil eder. Keygraph Keyword Extractor metindeki bağlamdan kopmamak için kelimelerin birlikte kullanılma durumlarını da dikkate almaktadır. Anahtar kelime çıkarımı ve doküman vektör dönüşüm sürecinin iş akışı Şekilde 3.7’de gösterilmiştir.

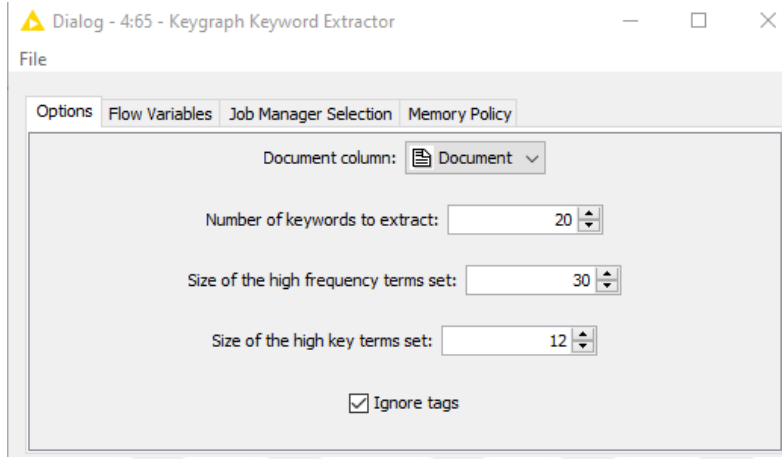
Şekil 3.7. Anahtar kelime çıkarımı ve vektör dönüşümü



Keygraph Keyword Extractor düğümünün seçenekleri içerisinde şekilde görüleceği üzere her bir doküman nesnesi için 20 anahtar kelime çıkarımı ayarlanmıştır. Ignore Tags (Etiketleri yoksay) seçeneği işaretlenerek düğümün terimleri sadece içeriklerine göre karşılaştırması istenmiştir. Düğüm daha önce terimler üzerinde yapılan işaretlemeleri ve meta verileri göz ardı edecektir. Müşteri yorumları büyük çoğunlukla kurallı bir yapıda olmadığından etiket bilgilerinin terimlerin analiz edilme sürecine dahil edilmesi gerekli görülmemiştir. Yüksek frekanslı terimler kümesi boyutu (Size of the high frequency terms set) parametresi düğümün dayandığı makalede genel kural olarak 30 verildiği için aynı değerde bırakılmıştır. Yüksek anahtar terimler kümesinin boyutu (Size of the high

key terms set) aynı şekilde düğümün temel aldığı makaledeki değer olan 12 olarak bırakılmıştır.

Şekil 3.8. Anahtar kelime çıkarımı konfigürasyon penceresi



Document Vector düğümü ile dokümanımız üzerinde sınıflandırma algoritmalarını çalıştırabileceğimiz ve onu terimler uzayında temsil eden bir doküman vektörüne çevirmiş oluyoruz. Yine bu düğümde terimleri karşılaştırırken etiketlerin (tag) hariç tutulması işaretlenmiştir.

3.1.1.4 Sınıflandırma Algoritmalarının Uygulanması ve Başarım Ölçümü

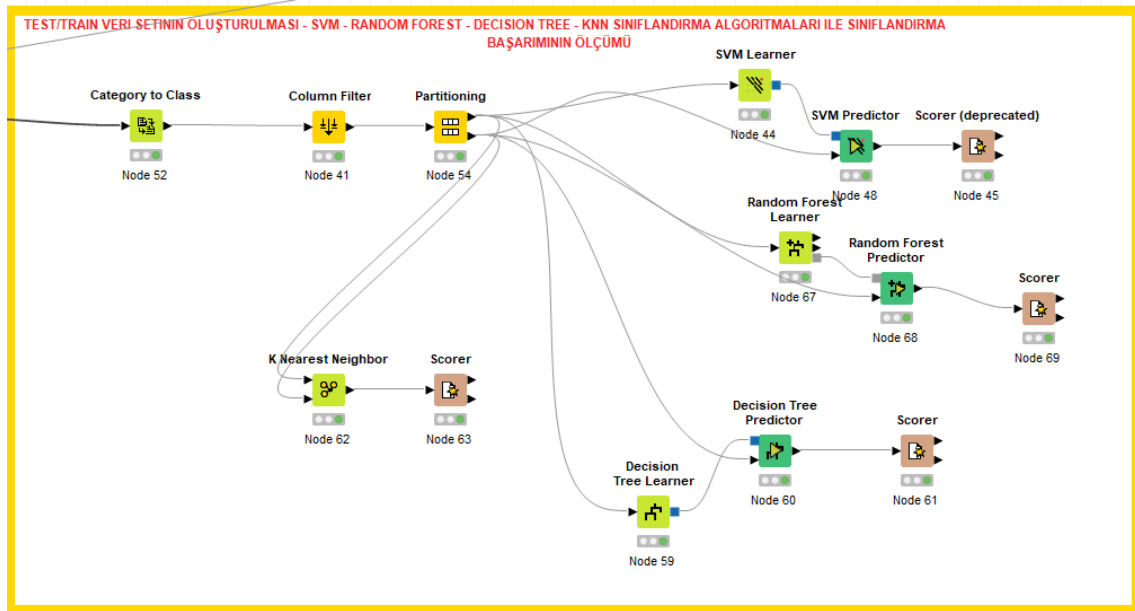
Son aşamada analize hazır hale getirilen veriler sınıflandırma algoritmalarına tatbik edilmektedir. Sınıflandırma sürecinde karar ağaçları (decision tree), destek vektör makineleri (Support vector machine), rastgele orman (random forest) ve K-En yakın komşu (k-nearest neighbor) algoritmaları kullanılarak, her bir kalite boyutu için ayrı ayrı olmak üzere model çalıştırılmıştır.

Test ve eğitim verilerinin oluşturulması, sınıflandırma algoritmalarının uygulanması ve model başarımının ölçülmesi sürecine ait iş akışı şekilde gösterildiği gibi olmaktadır. Buradaki ilk adım, müşteri yorum verilerinin partitioning düğümü ile test ve eğitim verisi olmak üzere iki kısma ayrılmasıdır. Tipik olarak test ve eğitim verileri 70/30 ya da 80/20 olarak seçilmektedir. Bu model için %70 eğitim, %30 test verisi olmak üzere yapılandırılmıştır. Partitioning düğümünden bir çıkış ilgili algoritmanın öğrenmesi için learner düğümüne, bir çıkış da yine aynı algoritmanın öğrenip kurduğu modelin test edilmesi amacıyla predictor düğümüne gitmektedir. Partitioning düğümünde eğitim verisi olarak alınacak verinin her bir kategoriden eşit miktarda alınması ve bu şekilde öğrenme

algoritmasının daha sağlıklı çalışabilmesi amacıyla tabakalı örnekleme (stratified sampling) seçeneği tercih edilmiştir. Predictor düğümü ise son olarak başarımların ölçümünü yapacak scorer düğümüne bağlanmakta ve bu düğüm üzerinden modelin başarımlarıyla ilgili accuracy score (doğruluk skoru) ölçülmektedir. Aynı zamanda model başarımların ölçümü açısından son derece önemli olan karışıklık matrisi (Confusion Matrix) ne de scorer düğümü üzerinden görülebilmektedir. Scorer düğümü model başarımlarıyla ilgili tüm başarımların metriklerinin görülebildiği bir düğümdür.

Sınıflandırma algoritmalarına ait KNIME iş akışlarının genel görünümü Şekil 3.9’da gösterilmiştir.

Şekil 3.9. Sınıflandırma iş akışları



Bu aşamada model başarımlarının doğru tespit edilebilmesi açısından karışıklık matrisinin yapısının anlaşılması oldukça önemlidir. Karışıklık matrisi sınıflandırma sonucunda gerçek değerler ve tahmin edilen değerlerin bir kombinasyonu üzerinden birtakım hesaplamalar gerçekleştirilmektedir. Karışıklık matrisi 2x2 lik bir matris yapısındadır ve tabloda görüldüğü gibi gerçek ve tahmin edilen değerler arasındaki ilişkiyi gösterir.

Tablo 3.1. Karışıklık matrisi

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

Model başarımları metriği içerisindeki en önemli değerlerden biri olan doğruluk oranı doğru pozitif tahminlerle, doğru negatif tahminler toplamının genel toplama oranlanması ile hesaplanmaktadır.

$$\text{Doğruluk (Accuracy)} = (TP + TN) / (TP + TN + FP + FN)$$

Örnek olarak uygunluk kalite boyutu sınıflandırma modeline ait karışıklık matrisi Şekil 3.10'da gösterilmiştir.

Şekil 3.10 Uygunluk boyutu sınıf modeli karışıklık matrisi

Row ID	I NONE	I UYGUN...
NONE	250	9
UYGUNLUK	15	82

Bu matrise göre uygunluk boyutu SVM algoritmasına göre $(250+82)/(250+82+9+15)=0,93$ doğruluk oranı ile tahminleme yapmıştır sonucuna varılmaktadır.

Model başarımları metrikleri arasında bahsedilmesi gereken önemli bir ölçüm olan Cohen Kappa, özellikle kategorik verilerin değerlendirilmesi bağlamında, iki değerlendirici arasındaki anlaşmanın düzeyini değerlendirmek için istatistiksel analizde kullanılan niceliksel bir tekniktir. Katsayının alabileceği değer aralıkları ve ifade ettiği anlam şu şekildedir:

< 0 : Hiç uyuma olmaması

0.0–0.20 : Önemli uyuma olması

0.21–0.40 : Orta derecede uyuma olması

0.41–0.60 : Ekseriyetle uyuma olması

0.61–0.80 : Önemli derecede uyuma olması

0.81–1.00 : Neredeyse mükemmel uyuma olması

İkili sınıflandırma tekniğinin mantığı gereği etiket değerlerinin her birisi (her bir kalite boyutu) özelinde oluşturulan sınıflandırma modeli ayrı ayrı çalıştırılarak, müşteri yorumlarının ilgili etiket değerlerine atanıp atanmadığı test edilmiştir. Başka bir ifade ile sınıflandırma modeli sekiz boyut için çalıştırılıp sınıflandırmaya tabi tutulan metinsel yorum verilerinin her defasında yalnızca tek bir boyut bağlamında ele alınıp alınamayacağı tahminlenmiştir.

Her bir kalite boyutu özelinde modelin doğruluk (accuracy) ve Cohen kappa katsayıları Tablo 3.2’de gösterildiği şekilde ölçülmüştür.

Tablo 3.2. Kalite boyutları sınıflandırma modeli başarımleri ölçümleri

		Decision Tree	SVM	KNN	RF
PERFORMANS	Doğruluk (Accuracy)	87,90%	88,20%	84,30%	88,50%
	Cohens Kappa	0,63	0,63	0,45	0,62
GUVENİLİRLİK	Doğruluk (Accuracy)	89,99%	91,30%	90%	91,30%
	Cohens Kappa	0,30	0,46	0,05	0,14
DAYANIKLILIK	Doğruluk (Accuracy)	93,00%	93,50%	93,30%	93,80%
	Cohens Kappa	0,25	0,40	0,06	0,19
SERVİS	Doğruluk (Accuracy)	96,90%	95,80%	94,10%	94,90%
	Cohens Kappa	0,83	0,78	0,63	0,72
ESTETİK	Doğruluk (Accuracy)	86,00%	90,70%	90,70%	92,70%
	Cohens Kappa	0,14	0,39	0,13	0,42
OZELLİKLER	Doğruluk (Accuracy)	78,40%	82,30%	77,50%	84%
	Cohens Kappa	0,53	0,62	0,52	0,68
ALGILANAN KALİTE	Doğruluk (Accuracy)	86,20%	88,80%	87,10%	87,10%
	Cohens Kappa	0,63	0,71	0,64	0,64
UYGUNLUK	Doğruluk (Accuracy)	93,50%	93,30%	92,40%	94,10%
	Cohens Kappa	0,83	0,82	0,80	0,84

3.1.2 Çok Sınıflı Sınıflandırma Modeli

İkili sınıflandırma veri kümesini iki farklı sınıfa kategorize eder. Bir önceki bölümde anlatıldığı gibi veri setinin sadece tek bir kalite boyutu kapsamına girip girmeyeceğinin

tahminlenmesi ikili sınıflandırma problemine tipik bir örnektir. Hedef değişkenimiz olan müşteri yorumlarının ilgili kalite boyutunun kapsamına girip girmediği tahmin edilmektedir. Çok sınıflı sınıflandırma problemi ise her bir hedef değişkenin dahil olabileceği birden fazla sınıf olması ve hedef değişkenin bu sınıflardan yalnızca bir tanesi tarafından etiketlenebildiği durumları ifade eder.

Müşteri yorumlarının kalite boyutlarına göre sınıflandırılması için oluşturulan model ikili sınıflandırma için kullanıldığı gibi, çok sınıflı sınıflandırma için de kullanılmıştır. Burada her bir yorum tek bir kalite boyutu ile etiketlenebileceği için metin veri kümesinin hazırlanması aşamasında oluşturulan “TEK” sütununda her bir yorumun en ilgili olduğu kalite boyutu işaretlenerek veri kümesi oluşturulmuştur.

Kalite boyutları arasındaki bağlam benzerliği ve fonksiyonel geçişlilik den dolayı sınıflandırma kategorileri arasında oldukça net bir ayrım olmaması bununla birlikte birçok yorum cümlesinde birkaç kelimeyle bile birden fazla kalite boyutuyla ilişkilendirilebilecek ifadelerin olması çok sınıflı sınıflandırma yapmanın zorluk seviyesini artırmaktadır. Özellikle özellik kalite boyutu ile uygunluk ve estetik boyutları arasında, uygunluk kalite boyutu ile özellik ve performans boyutları arasında, dayanıklılık boyutu ile güvenilirlik ve servis boyutları arasında ve algılanan kalite boyutu ile tüm kalite boyutları arasında çok önemli derecede bağlamsal geçişlilik bulunmaktadır. Bu gerçeklerden hareketle, müşteri yorumlarının kalite boyutlarına göre sınıflandırılması ve manuel etiketleme süreçlerinde öznel yorumların ve tercihlerin öne çıktığını söylemek mümkündür. Bu faktörler çoklu sınıflandırma yaklaşımında model başarımını düşüren faktörler olarak değerlendirilebilir.

Çok sınıflı sınıflandırma yöntemine göre yapılan sınıflandırma işleminin model başarımı değerleri Tablo 3.3’de gösterilmiştir.

Tablo 3.3. Çok sınıflı sınıflandırma başarım ölçümleri

		Decision Tree	SVM	KNN	RF
Multi-Class-Sınıflandırma	Doğruluk (Accuracy)	63,20%	69,70%	57,60%	69,70%
	Cohens Kappa	0,56	0,63	0,48	0,63

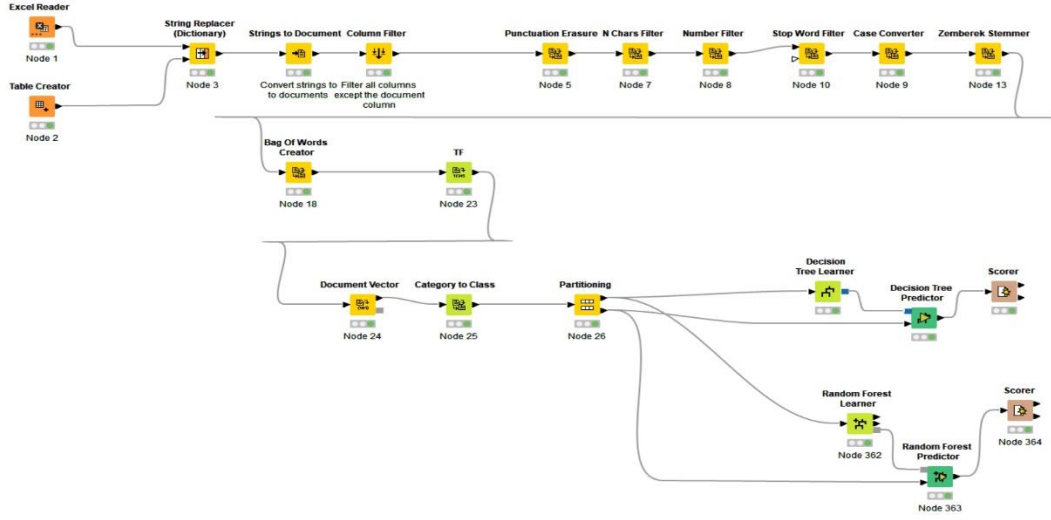
3.2 Duygu Analizi Modelinin Oluřturulması

Duygu analizi ham halinde yapılandırılmamıř bir formatta bulunan hedef deęiřkenlerinin (metin verilerinin) ön iřleme ve vektörel dönüşüm gibi süreçler sonucunda üç (olumlu, olumsuz, nötr) veya çoęu zamanda iki kutuplu (olumlu, olumsuz) duygusal kutba göre sınıflandırılma sürecidir. Metinsel verinin ön iřleme sürecine tabi tutulabilmesi için document nesnesine dönüşümü gereklidir. Konunun teorik çerçevesine daha önceki bölümlerde deęinildięi için burada sadece uygulama ařamasından bahsedilecektir.

Ana veri seti oluřturulurken her bir yorum satırı için ilgili yorumun içerdii duygusal kutupta manuel olarak iřaretlenmiřtir. Çalışma kapsamında duygu analizi süreci pozitif ve negatif olmak üzere iki kutup düzleminde ele alınacaktır. Olumsuz olmayan her müşteri yorumunun aslında olumlu olduęu kabulünden yola çıkılarak iki kutuplu bir analiz çalışması yapılmıřtır. Bununla birlikte müşteri yorumlarının yapıları ve amaçları gereęi pozitif veya negatif fikirler ihtiva ettięi, nötr olarak yorumlanabilecek yorumların sayıca çok az olması da bu fikri desteklemiřtir. Fikir analizi/madencilięi (opinion mining) olarak da adlandırılan duygu analizi (sentiment analysis) çalışmaları doęal dil iřlemenin yapısından, dilin anlamsal ve semantik çeřitlilięinden, analiz edilen kelime/terimler ile bağlam arasındaki iliřkiyi kurmanın zorluęu gibi sebeplerle başarımları düşük olabilen çalışmalardır. Bu çalışma kapsamında bu zorlukların üstesinden gelebilmek için metin ön iřleme süreçleri mümkün olduęunca detaylı olarak ele alınmıř ve bu süreçlerden önce sıklıkla yanlış kullanılan ve önemli olabileceęi düşünölen kelimeler de yine model üzerinde oluřturulan bir “bul ve deęiřtir” mekanizması ile revize edilmiřtir.

KNIME platformunda üzerinde oluřturulan duygu analizi modeline ait iř akıřının görünümü řekil 10’da göröldüęü gibidir.

Şekil 3.10. Duygu analizi sınıflandırma modeli



Model ana hatları itibariyle 3 bölümden oluşmaktadır. Birinci bölüm, verinin okunması ve düzeltme tablosu ile sık yapılan yanlışların düzeltilmesi, ön işleme için gerekli doküman nesnesine dönüşüm adımlarını içermektedir. İkinci bölüm metin ön işleme adımlarını içermektedir. Bunlar sırasıyla, noktalama işaretlerinin temizlenmesi, 2 ve daha az sayıda karakter içeren terimlerin dokümandan temizlenmesi, nümerik karakterlerin temizlenmesi, durak kelimelerinin çıkarılması, küçük harf dönüşümünün yapılması ve zemberek kütüphanesi yardımıyla terim gövdeleme işlemlerinin yapılmasını kapsamaktadır. Üçüncü bölüm kelime çantası yaklaşımı kullanarak terimlerin sayısal dönüşümlerini yapan bag of words creator düğümüyle başlamaktadır.

Kelime çantası (bag of words) yaklaşımı, veri kümesi içindeki her belgenin içerdiği kelime frekanslarına göre temsil edildiği bir yaklaşımdır. Bu yaklaşımın odağında belgedeki kelime frekansları yer almaktadır. Şekil 11’de model içindeki bag of words creator düğümüne ait çıktı tablosu görülmektedir. Her bir belge (yorum cümlesi) ne ait terimler bu tablo içinde görülebilir. Örneğin “*gayet kullanışlı ve tertemiz yıkayan bir makine*” yorumundaki kelimeler (term) kelime frekanslarıyla temsil edilen bir kelime çantasına dönüştürülmüş olduğu çıktı tablosundan gözlemlenebilmektedir.

Şekil 3.11. Kelime çantası düğümü çıktı tablosu

Documents output table - 26:18 - Bag Of Words Creator

File Edit Hilite Navigation View

Table "default" - Rows: 11497 Spec - Columns: 3 Properties Flow Variables

Row ID	Document	Preproc...	T Term
Row0			GAYET []
Row1			KULLANIŞLI []
Row2			VE []
Row3			TERTEMİZ []
Row4			YIKAYAN []
Row5			BİR []
Row6			MAKİNE []
Row7			GÜZEL []
Row8			PAKETLENMİŞ []
Row9			KARGOYLA []
Row10			SORUNSUZ []
Row11			OLARAK []
Row12			GELDİ []
Row13			VE []
Row14			HEMEN []
Row15			AYNI []
Row16			GÜN []
Row17			AKŞAM []
Row18			SAATLERİNDE []
Row19			SAMSUNG []
Row20			SERVİSİ []
Row21			KURULUMUNU []

Metin verilerinin uzunluğundan dolayı, kelime çantası çok sayıda kelime (terim) içerebilir, bu da belge vektöründe yüksek boyutluluğa yol açar ve potansiyel olarak sınıflandırma algoritmasının hızını etkiler. Bu sorunu çözmek için, daha az bilgilendirici veya sınıflandırıcı temsili anlamında da az önemli kelimeleri filtreleyip önemli olanlara odaklanmak yaygın bir uygulamadır. Kelimelerin önemini belirlemenin bir yolu, her bir belgedeki ve veri kümesinin tamamındaki sıklıklarının analiz etmektir. Bu çalışmada veri hacminin çok büyük olmamasından dolayı boyut indirgeme süreçleri yürütülmemiştir. Ancak boyut indirgeme işleminin yapılabileceği bir model Şekil 3.13 de sunulmaktadır.

Bu modelde sırasıyla terim frekanslarının grupta yapılarak belirlenmesi, dokümanı temsil değeri düşük terimlerin belirlenmesi ve bu terimlerin analiz edilecek veri setinden çıkarılması süreçleri yürütülebilmektedir. Yürütülecek bu süreçlere ait akış ve ilgili düğümler şu şekilde sıralanmaktadır:

Term to string düğümü bir önceki düğümde oluşturulan ve “Term” sütununda belirtilen terimleri bir karakter dizesine (string) dönüştürür ve değerleri “Term As String” adında yeni oluşturduğu bir sütunda saklar.

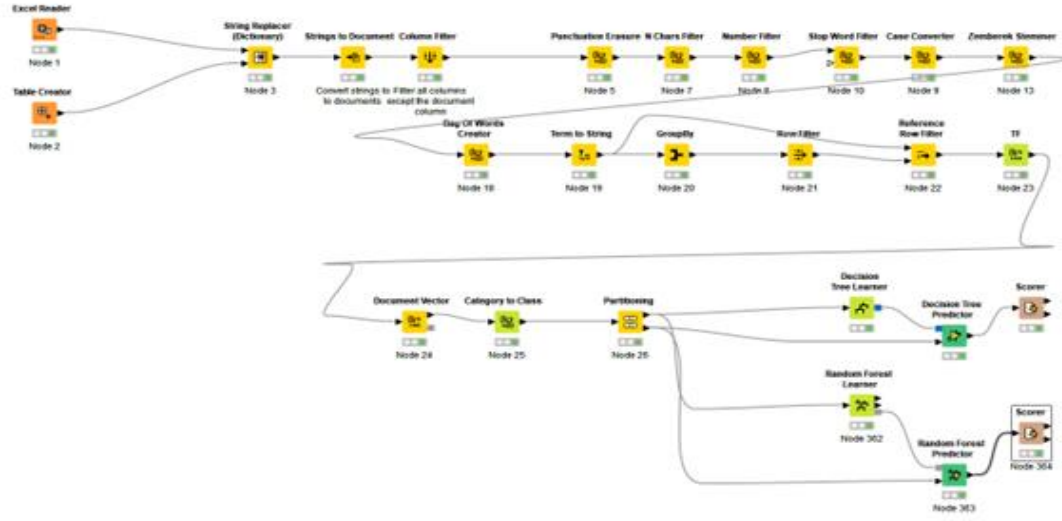
Group By düğümü kendisine geçirilen verileri belirtilen sütunlar bazında gruplama amacıyla kullanılır. Örneğin bir müşteri veri tabanından belirli bir coğrafi lokasyon ve belirli bir yaş aralığındaki müşteriler özelinde bir sorgulama yapmak gerektiğinde veriler coğrafi lokasyon ve yaş aralığına göre gruplanır. Gruplanan değişkenlere ait değerlerin toplamı, ortalaması, en küçük değer, en büyük değer, sayısı vb. gibi değerler ortaya çıkarılabilir. Bu amaçla kullanılan fonksiyonlara birleşik (aggregate) fonksiyonlar adı verilmektedir.

Burada term to string düğümünde oluşturulan karakter dizelerinin gruplanarak kaç adet terim olduğunun hesaplanması için kullanılmaktadır. Karakter dizelerinin sayıları burada hedef değer olduğu için düğüm içerisinde birleşik (aggregate) fonksiyonlardan say (count) fonksiyonu kullanılmıştır.

Row filter düğümü, verileri belirli kriterlere göre seçici olarak sıralamak amacıyla kullanılır. Bu düğümdeki "alt sınır" ve "üst sınır" parametreleri, belirlenen bir sütundaki değerlerin belirlenen aralıklar dahilinde olup olmadığını değerlendirmek amacıyla kullanılmaktadır. Dokuman nesnesindeki tüm belgelere ait kelime çantaları oluşturulup, karakter dizelerine çevrilmesinin ardından Row filter düğümünde “lower bound” (alt sınır) parametresini (örneğin 10) ayarlanarak terim tablosundan frekansı 10’dan düşük terimlerin filtrelenmesi sağlanarak boyut indirgeme işlemi yapılmış olmaktadır. “Lower bound” (alt sınır) parametresi belirlenirken, bir terimin sınıflandırma amacıyla etkin bir özelliği temsil etmesi için tüm belgelerin en az %1’inde bulunması gerektiği varsayımı dikkate alınmalıdır. Bu bir genel kural olarak kabul edilmektedir.

Reference Row Filter düğümüyle kelime çantasında oluşturulan terimlerin boyut indirgeme işlemi frekansı düşük terimlerin filtrelenmesi yoluyla yapılabilecektir.

Şekil 3.12. Duygu Analizi Sınıflandırma Modeli (Boyut İndirgeme ile)



TF (Term Frequency) düğümü her bir belgeye göre her bir terimin göreceli terim sıklığının (TF) belirlenmesini ve hesaplanan TF değerlerinin bulunduğu sütunun çıktı tablosuna dahil edilmesini içerir. Bu değer, bir terimin belge içindeki sıklığının o belgede bulunan toplam terim sayısına bölünmesiyle elde edilir.

Örneğin “Erciyes Üniversitesi'nde yapılan bilimsel çalışmalar, Erciyes Üniversitesi öğrencileri tarafından büyük bir ilgiyle takip edilmektedir. Bilimsel çalışmaların kalitesi, Erciyes Üniversitesi'nde her geçen gün artmaktadır.” İfadesinde belgedeki toplam terim sayısı 22 iken erciyes kelimesi/terimi toplam 3, bilimsel kelimesi 2 kere geçmektedir. Dolayısıyla erciyes terimine ait $TF = \frac{3}{22} = 0,136$ bilimsel terimine ait $TF = \frac{2}{22} = 0,091$ olacaktır. TF düğümünün çıktı tablosu olarak, terimlerin her bir belge başına hesaplanan frekanslarına ait çıktı tablosu Şekil 13.13'te gösterimiştir.

Şekil 3.13. Hesaplanan TF değerlerini gösteren çıktı tablosu

Row ID	Document	Preproc...	Term	TF rel
Row7971	...	...	yıkama	0.167
Row7970	...	...	güzel	0.167
Row7969	...	...	gayet	0.167
Row7968	...	...	denemeden	0.167
Row7967	...	...	sessiz	0.5
Row7966	...	...	başarılı	0.5
Row7965	...	...	herşeyi	0.2
Row7964	...	...	güzel	0.2
Row7963	...	...	kullanışlı	0.2
Row7962	...	...	gayet	0.2
Row7961	...	...	ürün	0.2
Row7960	...	...	harika	0.25
Row7959	...	...	sesi	0.25
Row7958	...	...	tasarımı	0.25
Row7957	...	...	panel	0.25
Row7956	...	...	muazzam	0.333
Row7955	...	...	kalitesi	0.333
Row7954	...	...	samsung	0.333
Row7953	...	...	çalışıyor	0.5
Row7952	...	...	sorunsuz	0.5
Row7951	...	...	yikiyor	0.25
Row7950	...	...	güzel	0.25
Row7949	...	...	ederim	0.25
Row7948	...	...	tavsiye	0.25
Row7947	...	...	memnunum	0.25
Row7946	...	...	alabildiğim	0.25
Row7945	...	...	makine	0.25
Row7944	...	...	fiyata	0.25
Row7943	...	...	ürün	0.5
Row7942	...	...	sağlam	0.5
Row7941	...	...	kırdım	0.25
Row7940	...	...	yüksek	0.25
Row7939	...	...	sesi	0.25
Row7938	...	...	puanı	0.25
Row7937	...	...	çıkarmıyor	0.111
Row7936	...	...	lekeleri	0.111
Row7935	...	...	dolmuyor	0.111
Row7934	...	...	su	0.111

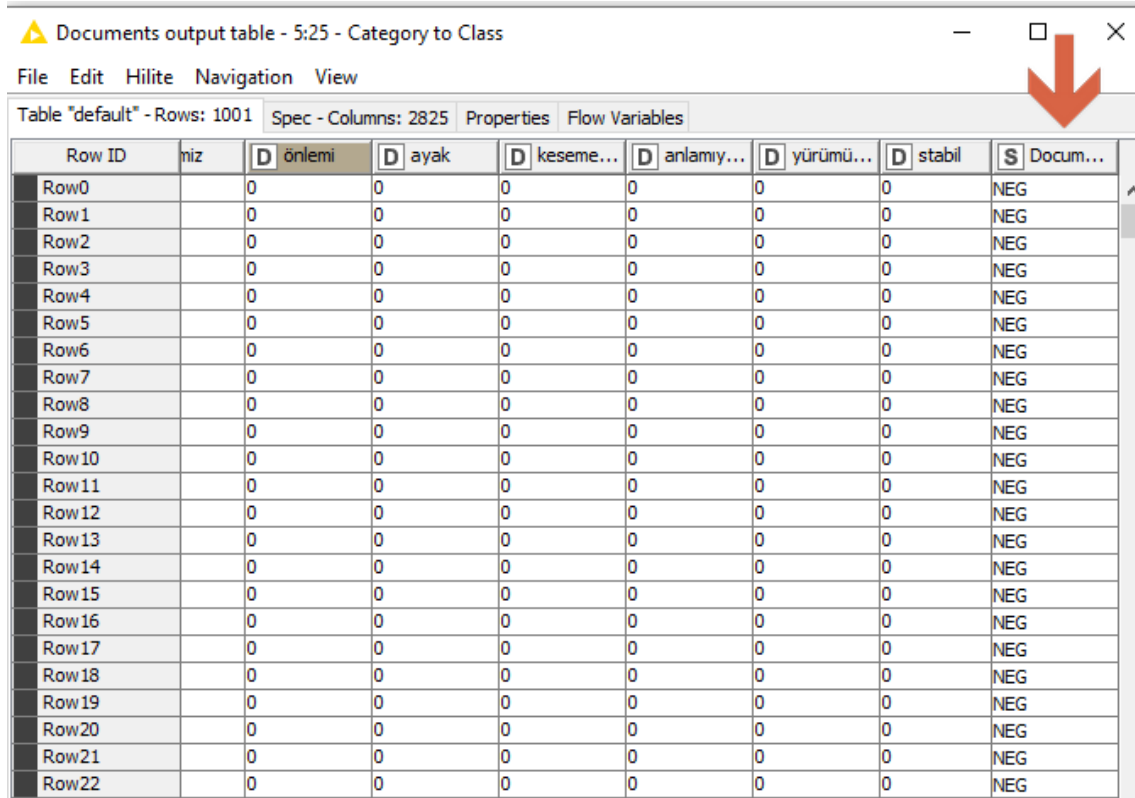
Metin ön işleme sürecinden bu aşamaya kadar yapılan işlemler, kelime çantasının oluşturulması, terimlerin karakter dizelerine dönüşümü, frekansı belirli bir düzeyin altındaki terimlerin belirlenmesi amacıyla terim gruplama ve belirli kriterlere göre filtreleyerek, nihai tahlilde sınıflandırma algoritması için temsil düzeyi en yüksek terimlerin belirlenmesi süresine özellik çıkarımı (feature extraction) adı verilmektedir.

Dokümanlardan çıkarılan kelimeler (özellikler) kullanılarak doküman vektörleri oluşturulmaktadır. Bu vektörler belgelerin sayısal temsilleri olarak işlev görür ve çıkarılan her kelime vektörün bir bileşeni olmaktadır. Document Vector düğümü, bit vektörlerin (0/1) veya sayısal vektörlerin oluşturulmasına imkân vermektedir. Sayısal vektörler, TF veya IDF düğümleri tarafından hesaplanan kelime puanlarına (Word score)

veya frekanslara dayanabilir. Duygu analizi için kurulan modelde TF düğümünün yapılandırılmasında bit vektörler kullanılmıştır.

Category to class düğümü belge hücreesindeki her satıra, belgenin kategorisini temsil eden bir sınıf (dize) sütununun eklenmesini sağlamaktadır.

Şekil 3.14. Category to class çıktı tablosu



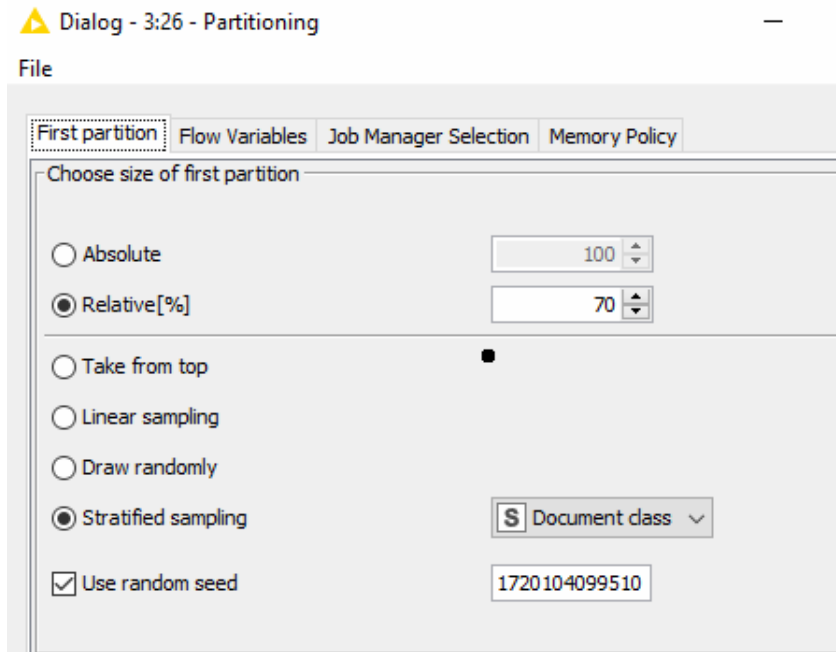
Row ID	niz	D önlemleri	D ayak	D keseme...	D anlamı...	D yürümü...	D stabil	S Docum...
Row0		0	0	0	0	0	0	NEG
Row1		0	0	0	0	0	0	NEG
Row2		0	0	0	0	0	0	NEG
Row3		0	0	0	0	0	0	NEG
Row4		0	0	0	0	0	0	NEG
Row5		0	0	0	0	0	0	NEG
Row6		0	0	0	0	0	0	NEG
Row7		0	0	0	0	0	0	NEG
Row8		0	0	0	0	0	0	NEG
Row9		0	0	0	0	0	0	NEG
Row10		0	0	0	0	0	0	NEG
Row11		0	0	0	0	0	0	NEG
Row12		0	0	0	0	0	0	NEG
Row13		0	0	0	0	0	0	NEG
Row14		0	0	0	0	0	0	NEG
Row15		0	0	0	0	0	0	NEG
Row16		0	0	0	0	0	0	NEG
Row17		0	0	0	0	0	0	NEG
Row18		0	0	0	0	0	0	NEG
Row19		0	0	0	0	0	0	NEG
Row20		0	0	0	0	0	0	NEG
Row21		0	0	0	0	0	0	NEG
Row22		0	0	0	0	0	0	NEG

Bu aşamadan sonra verilerin test ve eğitim verisi olarak ayrıldığı partitioning düğümü gelmektedir. Burada genel kabul gördüğü şekliyle %70 eğitim ve %30 test verisi olacak şekilde düğüm yapılandırılmıştır.

Makine öğrenmesi yaklaşımına göre manuel olarak etiketlenmiş veri setinin belirli bir bölümünün (genellikle %70-%30 veya %80-%20 eğitim-test verisi olacak şekilde) bölümlenmesi amacıyla Partitioning (Bölümleme) düğümü kullanılmaktadır. Partitioning düğümü eğitim ve test verilerinin ayrıştırılmalarıyla ilgili olarak belirli yapılandırma seçenekleri içermektedir. Üzerinde çalışılan verinin türü ve yapılan çalışmanın gerektirdiği veri bölümleme sistematığına göre gerekli görülen bölümleme seçeneğini buradan yapılandırmak gerekmektedir. Düğüm yapılandırma seçeneklerinden görüleceği

üzere çalışma kapsamında stratified sampling (tabakalı örnekleme) seçeneği tercih edilmiştir. Tabakalı örnekleme istatistik biliminde her bir anakütle tabakası için örneklem içerisinde temsilci bulunması mantığına dayanır. Bu durum duygu sınıflandırma modeli için tüm veri kütesi içindeki pozitif ve negatif yorum oranlarının, eğitim ve test verilerinin işaretlenmesinde de dikkate alınacağı anlamını taşımaktadır.

Şekil 3.15. Bölümleme düğümü yapılandırma seçenekleri



Bu aşamada kullanılabilecek diğer bölümleme seçeneklerinden bahsetmek gerekirse, take from top seçeneği veri setinde en üstte kalan satırların eğitim verisi için seçilmesi kalan satırların test verisi olarak kullanılması gerektiğini belirtmektedir. Doğrusal örnekleme (linear sampling) ilk ve son satırın seçilip kalan satırların tüm tablo üzerinden doğrusal bir sistematik ile seçildiği yaklaşımdır. Rastgele çek (draw randomly) seçeneği eğitim veri setinin (örneklem) tamamen rastlantısal olarak seçileceği anlamına gelmektedir. Ayrıca mutlak veya göreceli (absolute, relative) seçenekleri ile eğitim verisinin belirli bir sayı üzerinden veya tüm veri seti üzerinden belirli bir yüzdelik orana göre seçilmesi mümkündür.

Sınıflandırma aşamasında daha önceki bölümlerde değinildiği için burada tekrar edilmeyen KNIME düğümleri kullanılmış, sınıflandırma için karar ağaçları ve rastgele orman algoritmaları yürütülmüştür.

Karar ağaçları algoritmasının yapılandırılması amacıyla, decision tree learner üzerinde kalite ölçümü (quality measure) olarak algoritmanın bölünme hesaplaması için sunmuş olduğu iki yöntemden biri olan gini indeksi tercih edilmiştir. Gini indeks değeri veri集中的 varlıkların oranı olarak tanımlanabilir. İki varlığın gini değeri aynı çıkarsa sonuç dağılımları aynı demektir (Adak & Yurtay, 2014, s. 2). Gini index değeri bir düğümün saflığını temsil eder ve seçilecek olan kök düğümün veri setini olabildiğince çok açıklayabilmesi için bir belirleyici olarak kullanılır. 0 ile 1 arasında bir değer aralığında yer alır ve sıfıra olan yakınlığını veri kümesini dengeli edebilme yetkinliğini göstermektedir.

Gini index değeri şu şekilde hesaplanmaktadır. $Gini(D) = 1 - \sum_{i=1}^n p_i^2$ Burada P_i i sınıfının olasılığını (probability) temsil eden bir değerdir. Görüleceği üzere indeks değeri her bir sınıfın olasılıklarının karelerinin toplamının birden çıkarılmasıyla hesaplanmaktadır. Düşük gini index değeri düğümdeki örneklerin büyük kısmının aynı sınıfa ait olduğunu, dolayısıyla düğümdeki belirsizlik oranının az ve düğümün saflık derecesinin yüksek olduğu anlamına gelmektedir. Tipik olarak karar ağacı analizinde mümkün olduğu kadar homojen düğümler oluşturmak amacıyla dal seçimi gini indeksi en aza indirilerek belirlenir. Gini indeksi her potansiyel bölünme için hesaplanır ve en düşük gini indeksi değeriyle sonuçlanan bölünme tercih edilir. Buradan hareketle ağacın dallarındaki bölünmeye karar verme yöntemi olarak gini indeks metodunu seçilmiştir. Burada seçilebilecek diğer metot gain ratio ise bilgi kazancını normalize eden bir metrik olarak kabul edilmektedir. Bilgi kazancı ise genel olarak bir özelliğin düğümdeki bilgi miktarını ne kadar azalttığını ölçmektedir. Gini indeksi'nin arkasındaki temel mantık, bir düğüm içinde farklı sınıflar arasındaki örneklerin dağılımındaki eşitsizliği ölçmektir; Kazanç oranı (gain ratio) ise bir düğümü bölmede bilgi kazancının etkinliğini göz önüne alır. Bilgi kazancı, bir düğümdeki belirsizliğin azaltılmasıyla ilgilidir. Gini indeks hızlı ve basit bir metottur, özellik sayısının çok olmadığı büyük veri setlerinde kullanılabilir, hesaplama maliyeti düşüktür, kazanç oranı ise daha dengeli ve doğru bölünmeler yapsa da özellik sayısının fazla ve veri miktarının büyük olduğu durumlarda önemli hesaplama maliyeti oluşturabilir.

Çalışma kapsamında pozitif ve negatif olmak üzere sadece iki özellik ile çalıştığımızdan dolayı hesaplama hızı faktörü de dikkate alınarak decision tree learner

yapılandırılmasında kalite ölçüm (quality measure) parametresi olarak gini indeksi tercih edilmiştir. Veri setinin çok büyük olmaması ve özellik sayısının az olması nedeniyle herhangi bir pruning işlemine ihtiyaç duyulmamıştır. Dolayısıyla bu düğüm üzerinde “no pruning” seçilmiştir. Budama (pruning) işlemi karar ağaçlarındaki gereksiz ve fazla dallanmaları önlemek ve ağacın veri setini genelleme performansını optimum düzeyde tutmak için yapılan bir işlemdir. Çünkü aşırı dallanma yapan ağaçların eğitim verilerini ezberleyerek aşırı uyum sağlaması (overfitting) istenen bir durum değildir. Böylece ağacın genelleme performansı üst düzeyde tutulduğu için farklı veri kümeleri üzerinde tahminleme başarımı artmaktadır. Bununla birlikte ağacın fazla dallanmalardan kurtulması hesaplama maliyetlerinin de düşmesi anlamına gelmektedir.

Düğüm başına minimum kayıt sayısı (Min number records per node) parametresi her düğüm için gerekli minimum düğüm sayısını ifade etmektedir. Kayıt sayısının bu sayıdan daha düşük olması durumlarında ağacın büyümesi engellenerek bir anlamda budama işlemi yürütülür.

Decision tree learner düğümünde eğitilen veri kümesi içindeki test verileri model tarafından test edilmek üzere karar ağacı tahminleyici (decision tree predictor) düğümüne gelmektedir. Bu düğümü besleyen portlardan birisi decision tree learner düğümünden diğer ise üzerinde test işleminin yapıldığı verilerin geldiği partitioning düğümüdür. Bu düğümdeki parametre ile verilerin %70’i öğrenme, kalan %30’u ise test verisi olarak işaretlenmiştir.

Duygu analizi sınıflandırma modeli içerisinde kullanılan diğer bir sınıflandırma algoritması rastgele ormanlar (random forest) algoritmasıdır. Bölümleme düğümünden gelen eğitim verileri random forest modelin öğrenmesi amacıyla random forest learner düğümüne bağlanmıştır. Test verileri ise öğrenilen model bilgisini girdi olarak kabul eden rastgele ormanlar tahminleyici (random forest predictor) düğümünün diğer portuyla bağlanmaktadır. Böylece random forest predictor düğümünün öğrenmiş olduğu model bilgisini test verileri üzerinde deneyip tahminleme yapması sağlanmaktadır.

Random forest learner düğümünün yapılandırma penceresi Şekil 3.16’de gösterildiği gibidir.

Şekil 3.16. Random forest learner düğümü yapılandırma penceresi

Dialog - 3:362 - Random Forest Learner

File

Options | Flow Variables | Job Manager Selection | Memory Policy

Target Column: S Document class

Attribute Selection

☐ Use fingerprint attribute

☒ Use column attributes

☒ Manual Selection ☐ Wildcard/Regex Selection

Exclude

Filter

No columns in this list

☒ Enforce exclusion

Include

Filter

☒ Enforce inclusion

Misc Options

☐ Enable Highlighting (#patterns to store) 2.000

☐ Save target distribution in tree nodes (memory expensive - only important for tree view and PMML export)

Tree Options

Split Criterion: Gini Index

☐ Limit number of levels (tree depth) 10

☐ Minimum node size 1

Forest Options

Number of models: 100

☒ Use static random seed 1720199091674 New

Hedef sütun (target column) parametresi, öğrenilecek değeri içeren dokuman nesnesini göstermektedir. Özellik seçimi parametresi olarak, tablodaki sütunların modelin öğrenilmesi amacıyla öznitelik olarak kullanılması amacıyla column attributes seçeneği seçilmiştir. Burada öznitelik seçiminde kullanılacak sütunların manuel olarak veya belirli bir kalıp veya kurallı ifade tanımıyla (wildcard/regex) seçilmesi mümkündür.

Bölme kriteri (Split criteria) olarak karar ağacı algoritmasında olduğu gibi gini index kullanılmıştır. Gini index bahsedildiği gibi bir düğümdeki örneklerin saflığını ölçümleyen istatistiksel bir metriktir. Dolayısıyla burada algoritmanın karar ağaçlarını oluştururken dallanmalara nasıl karar vereceği yapılandırılmış olmaktadır.

Random Forest algoritması, birçok karar ağacı oluşturmakta ve her ağacın öngörülerini birleştirerek nihai sınıflandırmayı yapmaktadır. Gini Index kullanımıyla, her bir ağacın saflığının optimize edilerek genel olarak model başarımının artırılması amaçlanmıştır.

Random forest learner düğümündeki, seviye sınırlaması (limit number of levels) seçeneği, her bir karar ağacının derinliğini ya da seviye sayısını sınırlayarak kontrol altına almayı sağlar. Bir ağacın derinliği, kök düğümünden en uzak yaprak düğüme kadar olan yolun uzunluğudur. Çalışma kapsamında test verisinin çok büyük olmaması ve pozitif ve negatif olmak üzere ikili sınıflandırma yapıldığı için bu kısıtlamaya gerek görülmemiştir.

Her bir kalite boyutu için model ayrı ayrı çalıştırılarak modelin doğruluk oranı hesaplanmıştır. Kalite boyutlarının her biri için karar ağaçları ve rastgele orman algoritmalarının doğruluk değerleri ve Cohen kappa değerlerini karşılaştırmalı olarak Tablo 3.4'te gösterilmiştir.

Tablo 3.4. Her bir kalite boyutu için duygu analizi modeli doğruluk oranları

		Decision Tree	RF
Performans	Doğruluk (Accuracy)	82,70%	86,40%
	Cohens Kappa	0,20	0,00
Güvenilirlik	Doğruluk (Accuracy)	88,60%	85,70%
	Cohens Kappa	0,3	0,00
Dayanıklılık	Doğruluk (Accuracy)	91,30%	87%
	Cohens Kappa	0,82	0,73
Servis	Doğruluk (Accuracy)	54%	69%
	Cohens Kappa	-0,08	-0,04
Estetik	Doğruluk (Accuracy)	83,30%	86,10%
	Cohens Kappa	0,163	0,00
Özellikler	Doğruluk (Accuracy)	77,5%	80,40%
	Cohens Kappa	0,29	0,26
Algılanan kalite	Doğruluk (Accuracy)	83%	93,60%
	Cohens Kappa	0,01	0,37
Uygunluk	Doğruluk (Accuracy)	77,30%	80,40%
	Cohens Kappa	0,3	0,29

Bunula birlikte kalite boyutu gözetilmeksizin tüm müşteri yorumları da aynı model üzerinde analiz edildiğinde oluşan doğruluk oranları Tablo 3.5'te verilmiştir.

Tablo 3.5. Tüm yorumlar için duygu analizi modeli doğruluk oranları

		Decision Tree	RF
Tüm Yorumlar	Doğruluk (Accuracy)	74,70%	81,20%
	Cohens Kappa	0,23	0,21

Her bir kalite boyutu özelinde duygu analizi modeli çalıştırılarak model üzerinden elde edilen tahmin (predicted) verilere göre ilgili boyutlar için yapılan pozitif ve negatif yorum sayıları tespit edilmiştir. Yaklaşık 1100 farklı satırdan oluşan bir veri kümesi içinde manuel etiketleme ile oluşturulmuş her bir boyuta ait satır sayıları Tablo 3.6’da gösterilmiştir.

Tablo 3.6. Kalite boyutları bazında veri satırı sayıları

Performans	Güvenilirlik	Dayanıklılık	Servis	Estetik	Özellikler	Algılanan Kalite	Uygunluk	Toplam
267	114	75	137	117	458	311	322	1801

Aynı yorum satırı çoğu kez birden fazla kalite boyutu ile ilişkili olduğu için toplam sayı veri kümesindeki toplamdan daha fazladır. Her bir boyut için ilgili boyutun dahil olduğu alt veri kümeleri ile çalışılmıştır. Bu veri kümelerinin de %30’luk kısmı eğitilen modelin test edilmesi amacıyla kullanılmış olduğundan analiz edilen veri sayıları kalite boyutuna göre değişebilmekte ve doğal olarak düşmektedir.

Modelin her bir kalite boyutu bazında yaptığı tahminlemelere göre negatif ve pozitif yorum sayıları Tablo 3.7’de gösterilmiştir. Tahmin edilen pozitif kutuplu yorum sayısı toplam yorum sayısına bölünerek bir pozitif oran değeri elde edilmiştir.

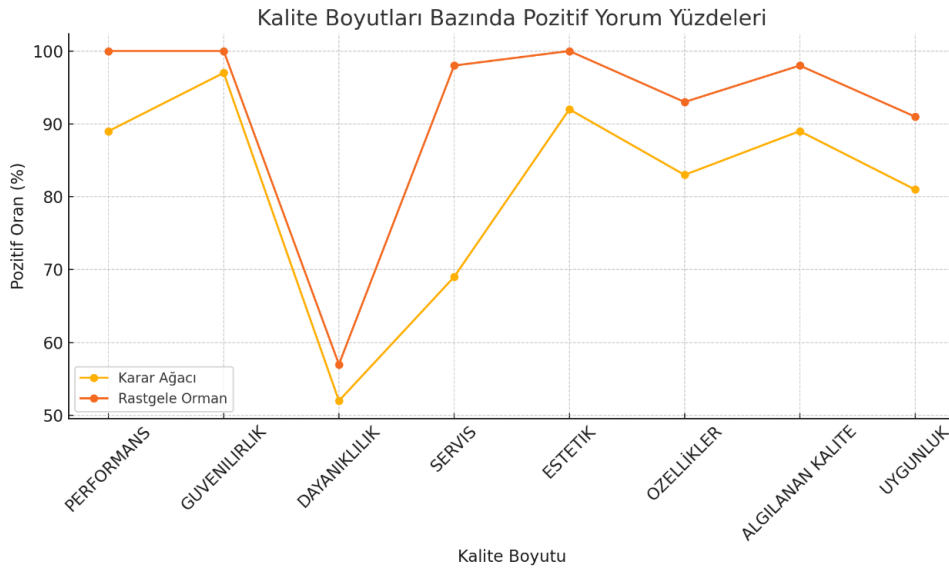
Tablo 3.7. Kalite boyutları bazında tahminlenen duygusal kutup sayıları

	Decision Tree			RF		
	NEG	POZ	Pozitif Oran	NEG	POZ	Pozitif Oran
Performans	9	72	89%	0	81	100%
Güvenilirlik	1	34	97%	0	35	100%
Dayanıklılık	11	12	52%	10	13	57%
Servis	13	29	69%	1	41	98%
Estetik	3	33	92%	0	36	100%
Özellikler	24	114	83%	10	128	93%
Algılanan Kalite	10	84	89%	2	92	98%
Uygunluk	18	79	81%	9	88	91%

Tablodan anlaşılabacağı üzere algılanan kalite, performans ve güvenilirlik boyutlarında yüksek bir pozitif kutup varken, dayanıklılık ve servis boyutlarında müşteri yorumlarının duygusal kutupsallığı daha negatif bir görünüm arz etmektedir.

Tahminlenen değerlere göre oluşturulan tablonun grafiksel gösterimi aşağıdaki gibidir.

Şekil 3.17. Kalite boyutları pozitif kutup oranları karşılaştırması



Görüldüğü üzere algoritmalar arasında genel bir uyum olmakla birlikte dayanıklılık boyutu için pozitif kutup değeri görece olarak daha düşük çıkmıştır. Bunun sebebi olarak dayanıklılık boyutunda müşterilerin genel olarak negatif bir durumla karşılaştıklarında yaşadıkları olumsuzluğu daha net ve görece kısa vadeli olarak ifade etme eğiliminde

olmaları şeklinde değerlendirilmektedir. Duygusal kutbun negatife daha yakın olduğu dayanıklılık ve servis gibi kalite boyutları müşterilerin genellikle ilk sahip olma sürecinde yaşadıkları deneyimlerle ilgilidir. Bu olumsuz deneyimler genel itibarıyla servis ve dayanıklılık boyutuyla ilgili olmasından hareketle ürünün kurulum ve taşıma süreçlerindeki olumsuzluklara işaret etmektedir. Daha uzun vadede deneyimlenen kalite boyutlarıyla ilgili olarak olumlu algıların pozitif kutba yakın olması, müşterilerin olumsuz deneyimlerini aktarmak adına daha hızlı davrandıklarını göstermektedir. Uygunluk boyutundaki olumsuz algılar cihazın ses standartlarına uyumsuzluğu, boyutla ilgili olumsuz deneyimler, enerji tüketimiyle ilgili olumsuz deneyimlerden kaynaklandığı görülmektedir. Bunun yanında algılanan kalite boyutunun tüm kalite boyutlarının optimum bir bileşeni gibi olduğu düşünülürse müşterinin genel marka algısı ve memnuniyet derecesinin üst düzeyde olduğu ve tavsiye edebilecek durumda olduğu gözlenmektedir. Algılanan kaliteye paralel olarak güvenilirlik boyutundaki pozitif kutupsallığın yüksek oluşu, hem markaya ait algılanan kalite algısının yüksek oluşundan hem de müşterinin ürünün kullanım ömrü içindeki performans ve dayanıklılık beklentisinin ve deneyiminin yüksek olmasından kaynaklandığı değerlendirilmektedir.

SONUÇ

Günümüzde dünya üzerinde dolaşımda olan verinin çok büyük bir kısmının yapılandırılmamış haldeki ham metin verileri olduğu gerçeğinden hareketle veri madenciliğinin alt disiplini olarak metin madenciliğinin önemi gün geçtikçe artmaktadır. Çalışma bu olgunun verdiği motivasyonla çevrimiçi ortamda dağıtık ve yapılandırılmamış bir halde bulunan müşteri yorumlarından ürün kalitesinin ürün kalite boyutları bakımından analiz edilmesini hedeflemiştir. Ürün kalitesinin kalite boyutları bakımından metin madenciliği ve duygu analizi teknikleri kullanılarak belirlenmesine yönelik olarak yapılmış ilk çalışmadır. Bu çalışma kapsamında öncelikle müşteri yorum verilerinin sekiz kalite boyutuna göre sınıflandırma süreçleri yürütülmüştür. Bu süreçlerin akabinde her bir boyut için duygu analizi modeli oluşturulmuş ve kalite boyutları bağlamında müşteri algısının duygusal kutupsallığı ölçümlenmiştir.

Yapılandırılmamış metin verileriyle çalışmanın doğal dil işlemenin yapısından kaynaklanan pek çok zorluğu bulunmakla birlikte, analiz yapılan müşteri yorumlarından oluşan veri türünden kaynaklı olarak çalışmanın birçok zorluğu bulunmaktadır. Bunlardan en önemlisi pek çok yorumun uzun ifadeler içermesine ve belirli bir hacimde olmasına karşın, kalite boyutları bakımından anlamsal bir değer sunmamasıdır. Yorumların pek çoğu ürünün özellikleri, sağlamlığı, markanın algılanan kalitesi ve diğer kalite bileşenleri gibi faktörlerden uzak bir kapsama sahiptir. Önemli bir kısmı taşıma, nakliye ve kargolama süreçlerindeki problemlere odaklanmaktadır. Yine çok büyük bir bölümünde yazım yanlışları, eksik/fazla noktalama işaretleri kullanımı, his ve coşku belirtmek için gereksiz fazla harf kullanımı (çooook, sessizzz vb.) bulunmaktadır. Veri kümesi içinde terim frekansı yüksek olabilecek ve dolayısıyla oluşturulacak model başarımı için kritik önemde olan birçok kelimenin birden fazla yazım şekli bulunmaktadır. (Makine yerine, makına, makine, makine, ürün yerine, urun, ürün, ürün, yazılması gibi.) Bununla birlikte mesela performans boyutu sınıflandırma başarımı için çok önemli olan fiyat/performans ifadesi, F/P, fp, FP, f/p, F/p, f/P gibi normal şartlarda metin ön işleme süreçlerinde elimine edilebilecek pek çok değerli bilgi içermektedir. Bu zorluğun üstesinden mümkün olabildiğince tek için tüm model iş akışlarının başında böyle bilgi değeri yüksek kelimeleri tek doğru ve standart versiyonlarının diğer yanlış

versiyonları yerine revize edilmesini sağlayacak bir tablo oluşturulmuştur. Bu tablo bir anlamda sözlük işlevi görmüş ve excel ortamındaki tüm yanlış örnekler doğrusu ile düzeltilerek analiz süreci başlatılmıştır.

Veri seti ile ilgili diğer önemli bir zorluk, yukarda söz edilen gürültülü verinin miktarca çok olmasıyla da ilintili olarak analiz edilebilecek veri miktarının çok fazla olmayışdır. Çalışma kapsamında 6000'e yakın müşteri yorumu incelenmiş ve yaklaşık 1200 satırlık bir veri seti oluşturulmuştur. Bu meyanda bu çalışmanın gelişme alanlarından birisi de analiz edilecek verinin ayıklanma süreçleriyle ilgili metin verisini bağlamından koparmadan cümle cümle veya kelime kelime elde edilmesi ile ilgili detaylı modellerin oluşturulmasıdır. Ek olarak, analiz konusu olarak tek bir markaya ait çamaşır makinesi modelleri seçildiği için üzerinde çalışılacak veri kümesi zaten sınırlandırılmış olmaktadır ve birkaç pazaryeri dışında ilgili müşteri yorumu bulmak mümkün olmadığından kullanılabilir veri hacmi kısıtlıdır. Bu veri kümesinin de kendi içinde sekiz kalite boyutuna atandığı ve her bir kalite boyutunun da kendi içinde pozitif/negatif kutuplara sınıflandırıldığı düşünüldüğünde analiz edilen veri hacmi çok büyük boyutlarda olmamaktadır.

Diğer taraftan analiz edilen verinin atanacağı sınıflar veya işaretleneceği etiketler arasında kategorik denilebilecek net ve kesin ayrımların olmaması, kalite boyutlarının kapsam ve içerik olarak kendi aralarında yüksek geçirgenliklerinin ve etkileşimlerinin olması veri setinin manuel etiketlenmesindeki öznelliği artırmaktadır. Bu durum hiç şüphesiz model başarımını ve sınıflandırma tutarlılığını doğrudan etkileyebilecek önemli bir etkidir. Örneğin özellikle kalitenin dayanıklılık boyutuyla güvenilirlik boyutu arasında, uygunluk ve özellik boyutu arasında, servis ve dayanıklılık arasında önemli bir geçirgenlik ve ifade edilen bağlam anlamında önemli bir kesişim alanı bulunmaktadır.

Tüm bu faktörler ışığında analiz edilen verinin hacmi ve kalitesi bakımından istenilen düzeyde temiz, tutarlı ve kategorik olmasını beklemek güçtür. Bu amaçla metin ön işleme süreçlerinin kritik önemli olduğundan hareketle modelin en başında bilgi seviyesi yüksek kelimeler bakımından revize edilen veri seti bir doküman türüne dönüştürüldükten sonra metin ön işleme süreçlerinden geçirilmiştir. Ön işleme süreçlerinde POS Tagger ve gövdeleme (stemming) işlemlerinde Türkçe doğal dil işleme kütüphanesi zemberek kullanılmıştır.

Çalışmanın birinci bölümünde kalite kavramı, kalite boyutlarının teorik çerçevesi ve tarihsel gelişimine yer verilmiştir. Müşteri beklentileri ile kalite kavramı arasındaki organik ve doğrudan bağlantılar incelenmiş, bu amaca atfen müşteri beklentileri ve müşteri memnuniyeti arasındaki bağlantıyı ortaya koyan bazı modellere ait bir özet sunulmuştur. Kalite boyutlarının temas ettiği alanın net olarak ele alınabilmesi için her bir kalite boyutu özelinde kalite boyutlarının müşteri beklentileriyle ilişkileri ve bunların pratik tezahürleri konusuna ayrıca değinilmiştir. Birinci bölümün son kısmında ise müşteri memnuniyeti ve kalite arasındaki etkileşim alanı incelenmiştir.

İkinci bölüm çalışma konusu ile ilgili olarak veri madenciliği, metin madenciliği ve duygu analizi (fikir madenciliği) ile bunların doğrudan ilişkili oldukları disiplinlere dair teorik bir altyapı oluşturmayı amaçlamaktadır. Bu kapsamda veri madenciliği ile ilişkili disiplinler, metin madenciliği genel teorik çerçevesi, metin madenciliği süreçlerinin öne çıkan kilometre taşı olarak değerlendirilebilecek konulara temas edilmiştir. Çalışma kapsamında kullanılan sınıflandırma algoritmalarının dayandıkları teorik arka plandan da bahsedilerek, ilgili algoritmaların güçlü ve zayıf yönleri belirtilmiştir. Takip eden bölümde duygu analizine dair önemli noktalar özetlenmiştir.

Üçüncü bölümde yaklaşık 6000 müşteri yorumu incelenerek oluşturulan 1200 satırlık veri seti kullanılarak bir sınıflandırma modeli geliştirilmiştir. Her bir satır veri (her belge) birden fazla kalite boyutunun kapsamına girebileceğinden dolayı çok sınıflı sınıflandırma yaklaşımı bu aşamada yeterli görülmemiştir. Çok sınıflı sınıflandırma yaklaşımı bir veri örneğinin (instance) sadece bir sınıfa atanabileceği varsayımından hareket etmektedir. (Bir hayvanın sadece bir türe ait olabileceği gibi) Literatürde çoklu etiketli (multi label) olarak geçen sınıflandırma yönteminin kullanılabilmesi için çalışma kapsamında problem indirgeme tekniği uygulanmış ve problem birden fazla ikili (binary) sınıflandırma probleminin bir kombinasyonu şekline getirilmiştir. Bu amaçla her bir kalite boyutu (her bir etiket değeri) için model ayrı ayrı çalıştırılarak, verilerin o boyuta ait olup olmadıkları tahminlenmiştir. Bu süreçte, karar ağaçları, destek vektör makineleri, k-en yakın komşu ve rastgele ormanlar algoritmaları kullanılmıştır. Buradaki sınıflandırma modelinde kelime çantası yaklaşımından farklı olarak keygraph keyword yaklaşımı kullanılmıştır. Amaç sadece kelimelerin terim frekanslarına değil, birlikte kullanım ilişkilerini de gözleterek bağlamdan koparmamaktır. Bu aşamadaki sınıflandırmaya konu olan kalite boyutlarının karmaşık, kendi aralarında geçirgen ve daha kompleks olmasından dolayı

kelime çantası yaklaşımı kullanılmamıştır. Sekiz kalite boyutuna ait veri setleri ana veri kümesinden çıkarılarak model çalıştırılmış ve dört algoritma için model doğruluk oranları ortaya konmuştur. Buna göre performans boyutu sınıflandırması %88,5 ile rastgele ormanlar, güvenilirlik boyutu sıralaması %91,3 ile destek vektör makineleri ve rastgele ormanlar, dayanıklılık boyutu sınıflandırması %93,8 ile rastgele ormanlar, servis boyutu sınıflandırması %96,9 ile karar ağaçları, estetik boyutu sınıflandırması %92,7 ile rastgele ormanlar, özellik boyutu sınıflandırması %84 ile rastgele ormanlar, algılanan kalite sınıflandırması %88,8 ile destek vektör makineleri, uygunluk boyutu sınıflandırması %94,10 rastgele ormanlar algoritmaları ile en yüksek doğruluk (accuracy) değerlerinde gerçekleştiği ölçümlenmiştir. Her bir kalite boyutu özelinde yapılan ikili sınıflandırmanın haricinde her bir doküman belgesinin (her satır verinin) sadece bir etikete atanabileceği çok sınıflı sınıflandırma da yapılmış ve doğruluk oranları beklendiği üzere düşmüştür. Yine en yüksek doğruluk oranları %69,7 ile rastgele ormanlar ve destek vektör makineleri algoritmaları ile ölçülmüştür.

İkinci aşamada duygu analizi çalışması için makine öğrenmesine dayalı sınıflandırma tekniği kullanılarak iki model oluşturulmuştur. Modellerden biri yüksek hacimli veri setleri ile çalışıldığında temsil düzeyi düşük terimlerin çıkarılmasına imkân veren bir modeldir. Bu model ile kelimelerin terim frekansları hesaplanarak vektörel ifade içerisinde temsil kabiliyeti düşük olacağı düşünülen kelimelerin filtrelenerek boyut indirgeme işlemi yapılması mümkün hale getirilmiştir. Çalışma kapsamında üzerinde çalışılan veri seti çok büyük hacimli olmadığından bu modelin boyut indirgemeye ilgili düğümlerinden arındırılarak oluşturulan model kullanılmıştır. Her bir kalite boyutunun pozitif/negatif olmak üzere içerdiği tüm veri satırları ana veri setinden çekilmiş model bu veri setiyle beslenmiştir. Sonuç olarak bir kalite boyutundaki her bir belgenin pozitif veya negatif olmak üzere kapsadığı uç noktayı tahminleyecek bir model oluşturulmuştur. Model içerisinde sınıflandırma yapılacak etiketlerin ayrışma noktaları çok daha kesin ve net olduğu için, terim frekansına dayalı hızlı ve basit bir çözüm olan kelime çantası (bag of words) tekniği kullanılmıştır. Sınıflandırma süreçlerinde ise karar ağaçları ve rastgele ormanlar algoritmaları kullanılmıştır. Her bir kalite boyutu için model çalıştırılarak doğruluk oranları ortaya konmuştur. Bu aşamada oluşturulan model tarafından tahminlenen verilere göre pozitif ve negatif yorum sayıları ortaya çıkarılmıştır. Sonuç olarak performans, güvenilirlik ve algılanan kalite boyutları olumlu kutupsallık

anlamında yüksek değere sahipken bu anlamda en olumsuz duygusal kutupsallık tarafında dayanıklılık boyutu bulunmaktadır. Bunun nedeni müşterilerin özellikle analize konu olan ürünlerin sağlamlığı, malzeme kalitesi, verdiği arızalar vb. gibi olumsuzlukları net ve detaylı olarak ifade etmemeleri olduğu düşünülmektedir. Uygunlukla ilgili yorumların çok büyük kısmı, ürünün çıkardığı ses seviyesi, ebatları, sahip olduğu enerji seviyesi gibi parametrelerdir. Diğer taraftan müşteri markanın kendisinde bıraktığı olumlu izlenimi ifade ederken çoğunlukla benzer şekillerde kendini ifade etmektedir.

Metin madenciliği ile yapısal olmayan verinin yorumlanması her geçen gün önemini artıran bir konudur. Müşteri yorumları bu yapısal olmayan veri yığınları içerisinde potansiyel bilgi değeri yüksek verilerdir. Bu çalışma kapsamında yapıldığı gibi bu verilerin analize tabi tutularak yorumlanması tüketici davranışlarının anlaşılması bakımından işletmelere önemli bir rekabet avantajı sağlayacaktır. Müşterinin çalışma sonucu çıkan sonuçlarda olduğu gibi ürünün dayanıklılığı ile olumsuz duygulara sahip olduğunun tespiti ürün kalitesinin iyileştirilmesi adına önemli çarpan etkisine sahiptir. Bu çalışma sonucunda kalitenin hangi boyutuyla ilgili olarak genel müşteri algısını ölçümlemek ve bu duruma göre bir strateji geliştirebilmek mümkün olacaktır. Çalışma aynı zamanda veri odaklı karar verme stratejisi anlamında yeni bir vizyon sunmayı da hedeflemektedir.

Çalışma belirli bir marka özelinde sınırlı sayıda yorum verisi üzerinde yapılmıştır. Müşteri istek ve beklentileri hızla değişebilen bir yapıda olduğundan dolayı belirli bir t anındaki müşteri yorumlarının değerlendirilmesi zaman içinde etkisini yitirecektir. Dolayısıyla müşteri yorumlarının zaman içerisindeki değişiminin de gözlemlenerek yorumlardaki duygusal kutupların zamana bağlı değişimlerinin de analize dahil edilmesi gerekmektedir. Müşteri yorumlarının gerçek zamanlı olarak değerlendirildiği ve analiz edildiği sistemlerin geliştirilmesi de bu çalışmanın devamı ve gelişim noktası sayılabilecek çalışmalar olarak değerlendirilmektedir.

KAYNAKÇA

- Aaker, D. A. (1991). *Managing Brand Equity: Capitalizing on the Value of Brand Name*. The Free Press.
- Abbas, O. A. (2008). Comparisons between data clustering algorithms. *International Arab Journal of Information Technology*, 5(3).
- Adak, M. F., & Yurtay, N. (2014). Gini Algoritmasını Kullanarak Karar Ağacı Oluşturmayı Sağlayan Bir Yazılımın Geliştirilmesi. *Bilişim Teknolojileri Dergisi*, 6(3), 1-6.
- Ajzen, I., & Fishbein, M. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Addison-Wesley.
- Alma, F. (2022). (27) *Dünya E-Ticaret İstatistikleri* | LinkedIn. <https://www.linkedin.com/pulse/d%C3%BCnya-e-ticaret-istatistikleri-ferdi-alma/>
- Amatriain, X., & Pujol, J. M. (2015). Data mining methods for recommender systems. *Recommender Systems Handbook, Second Edition*, 227-262. https://doi.org/10.1007/978-1-4899-7637-6_7/FIGURES/7
- Anton, R., Chenavaz, R. Y., & Paraschiv, C. (2023). Dynamic pricing, reference price, and price-quality relationship. *Journal of Economic Dynamics and Control*, 146. <https://doi.org/10.1016/j.jedc.2022.104586>
- Aslan, S. (2023). Doğal Dil İşleme Teknikleri Kullanarak E-Ticaret Kullanıcı İncelemelerinde Özellik Tabanlı Duygu Analizi. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 35(2), 875-882. <https://doi.org/10.35234/fumbd.1335583>
- Atcılı, A. (2022). *Machine Learning Mesafeleri*. <https://medium.com/machine-learning-t%C3%BCrkiye/machine-learning-mesafeleri-8ac88ca393>
- Aydın, M. (2017). API kullanarak veri toplama süreçleri. *Veri Bilimi ve Uygulamaları*, 5(3), 123-134.
- Ayhan, S., & Erdoğan, Ş. (2014). Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 9(1), 175-201.
- Aytekin, C., & Mayda, İ. (2013). *Sosyal Medyada Rekabet Analizi İçin Karşılaştırma Görevine Yönelik Fikir Madenciliği Modeli*. 7(27), 414-425.
- Balakrishnan, V., & Ethel, L.-Y. (2014). Stemming and Lemmatization: A Comparison of Retrieval Performances. *Lecture Notes on Software Engineering*, 2, 262-267. <https://doi.org/10.7763/LNSE.2014.V2.134>

- Baraldi, A., & Alpaydin, E. (2002). Constructive feedforward ART clustering networks—Part I and II. *IEEE Transactions on Neural Networks*, 13(3), 645-677.
- Berrar, D. (2018). *Cross-Validation*. <https://doi.org/10.1016/B978-0-12-809633-8.20349-X>
- Berry, M. J. A., & Linoff, G. S. (2004). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management* (Second Edition). Wiley.
- Bhatia, K. K., Rao, A., Price, A. N., Wolz, R., Hajnal, J., & Rueckert, D. (2012). Hierarchical Manifold Learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7510 LNCS, 512-519. https://doi.org/10.1007/978-3-642-33415-3_63
- Bhattacharya, G., Ghosh, K., & Chowdhury, A. S. (2015). A probabilistic framework for dynamic k estimation in kNN classifiers with certainty factor. *ICAPR 2015 - 2015 8th International Conference on Advances in Pattern Recognition*. <https://doi.org/10.1109/ICAPR.2015.7050683>
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2). <https://doi.org/10.1007/s11749-016-0481-7>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Bloch, P. H. (1995). Seeking the ideal form: Product design and consumer response. *Journal of Marketing*, 59(3), 16-29. <https://journals.sagepub.com/doi/abs/10.1177/002224299505900302>
- Boleslavsky, R., Cotton, C., & Gurnani, H. (2016). Demonstrations and Price Competition in New Product Release. *Management Science*, 63. <https://doi.org/10.1287/mnsc.2016.2449>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://link.springer.com/article/10.1023/A:1010933404324>
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. Wadsworth.
- Brucks, M., Zeithaml, V. A., & Naylor, G. (2000). Price and brand name as indicators of quality dimensions for consumer durables. *Journal of the Academy of Marketing Science*, 28(3), 359-374. <https://doi.org/10.1177/0092070300283005>
- Buzzell, R. D., & Gale, B. T. (1987). *The PIMS principles: Linking strategy to performance*. Simon and Schuster.
- Carbone, P. L. (2000). Expanding the meaning of and applications for data mining. *Smc 2000 conference proceedings. 2000 ieee international conference on systems, man and cybernetics. "cybernetics evolving to systems, humans, organizations, and their*

- complex interactions” (cat. no.0, 3, 1872-1873 c.3.*
<https://doi.org/10.1109/ICSMC.2000.886383>
- Celik, U., Akçetin, E., Herand, D., Yaldir, A., Engin, M., & Ozdemir, S. (2015, Temmuz). *Yönetim Bilişim Sistemleri Alanında Metin Madenciliği ile Bilgi Haritalama.*
<https://doi.org/10.13140/RG.2.1.2896.4242>
- Chai, Y., & Li, W. (2019). Towards Deep Learning Interpretability: A Topic Modeling Approach. *International Conference on Interaction Sciences.*
<https://api.semanticscholar.org/CorpusID:209149776>
- Chaplot, Dr. V. (2015). Databases Management Systems (DBMS). *International Education and Research Journal (IERJ), 1(2).*
<https://ierj.in/journal/index.php/ierj/article/view/11>
- Chavan, A. (2023). Data Visualization Tools Used For Decision Making: A Review. *International Research Journal of Modernization in Engineering Technology and Science.* <https://doi.org/10.56726/irjmets33640>
- Chen, L., Malaysia, M., Bhd, S., & Mansori, S. (2018). The Effects of Product Quality on Customer Satisfaction and Loyalty: Evidence from Malaysian Engineering Industry. *International Journal of Industrial Marketing, 3(1), 20.*
<https://doi.org/10.5296/ijim.v3i1.13959>
- Cheriyen, S., Ibrahim, S., Mohanan, S., & Treesa, S. (2018). Intelligent Sales Prediction Using Machine Learning Techniques. *Proceedings - 2018 International Conference on Computing, Electronics and Communications Engineering, ICCECE 2018.*
<https://doi.org/10.1109/ICCECOME.2018.8659115>
- Cherkassky, V., & Mulier, F. (1998). *Learning From Data: Concepts, Theory, and Methods.* Wiley.
- Chowdhury, J., & Menon, A. (1995). Multidimensional Components Of Quality and Strategic Business Unit Performance: A PIMS Test. İçinde *Source: Journal of Managerial Issues (C. 7, Sayı 4).* <https://www.jstor.org/stable/40604081>
- Ciaburro, G., & Iannace, G. (2021). Machine learning-based algorithms to knowledge extraction from time series data: A review. İçinde *Data (C. 6, Sayı 6).*
<https://doi.org/10.3390/data6060055>
- Cooper, R. G., & Kleinschmidt, E. J. (1995). Benchmarking the firm’s critical success factors in new product development. *Journal of Product Innovation Management, 12(5), 374-391.* <https://doi.org/10.1111/1540-5885.1250374>
- Çalış, K., Gazdağı, O., & Yıldız, O. (2013). Reklam İçerikli Epostaların Metin Madenciliği Yöntemleri İle Otomatik Tespiti. *Bilişim Teknolojileri Dergisi, 6(1), 1-7.*

- Çavdar, E., & Zerdali, N. (2020). Farklı Mamul Grupları İçin Kalite Boyutları Önem Değerlendirmesi. *İşletme Ekonomi ve Yönetim Araştırmaları Dergisi*, 3(2). <https://doi.org/10.33416/baybem.711021>
- Çınar, A., & Silahtaroglu, G. (2012). Veri madenciliği teknikleri ile müşteri memnuniyetine etki eden gizli nedenlerin keşfi. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 34(2), 309-330. <http://search/yayin/detay/141756>
- Dahiya, N., Gupta, S., & Singh, S. (2022). A Review Paper on Machine Learning Applications, Advantages, and Techniques. *ECS Transactions*, 107(1). <https://doi.org/10.1149/10701.6137ecst>
- Daniela, M. (2019). Decision Support Systems To Create A Competitive Advantage. *Internal Auditing and Risk Management*, 55(3), 76-82. <https://ideas.repec.org/a/ath/journal/v55y2019i3p76-82.html>
- De Veirman, M., Cauberghe, V., & Hudders, L. (2017). Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude. *International journal of advertising*, 36(5), 798-828.
- Deshpandé, R., & Gatingon, H. (1994). Competitive analysis. *Marketing Letters*, 5(3), 271-287. <https://doi.org/10.1007/BF00993906>
- Diler, S., & Demir, Y. (2023). Sağdan sansürlü veriler için veri madenciliği algoritmaları performanslarının karşılaştırılması. *İstatistik Araştırma Dergisi*, 13(1), 34-47.
- Ekinci, E., & İlhan Omurca, S. (2017). Ürün Özelliklerinin Konu Modelleme Yöntemi ile Çıkarılması. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 9(1), 51-58.
- El Naqa, I., & Murphy, M. J. (2015). What Is Machine Learning? *Machine Learning in Radiation Oncology*, 3-11. https://doi.org/10.1007/978-3-319-18305-3_1
- Eldén, L. (2006). Numerical linear algebra in data mining. *Acta Numerica*, 15, 327-384. <https://doi.org/10.1017/S0962492906240017>
- Elhaik, E. (2022). Principal Component Analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-14395-4>
- Ersoz, F. (2019). Data mining and text mining with big data: review of differences. *The International Journal of Recent Advances in Multidisciplinary Research*, 6(1).
- Ersöz, F., & Çınar, Y. (2021). Veri Madenciliği ve Makine Öğrenimi Yaklaşımlarının Karşılaştırılması: Tekstil Sektöründe bir Uygulama. *European Journal of Science and Technology*. <https://doi.org/10.31590/ejosat.1035124>
- Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster Analysis*. Arnold.

- Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 76-82.
- Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- Freberg, K., Graham, K., McGaughey, K., & Freberg, L. A. (2011). Who are the social media influencers? A study of public perceptions of personality. *Public relations review*, 37(1), 90-92.
- Garvin, D. A. (1987a). Competing on the eight dimensions of quality. *Harvard Business Review*, 65(6), 101-109.
- Garvin, D. A. (1987b). Competing on the eight dimensions of quality. *Harvard Business Review*, 65(6), 101-109. <https://hbr.org/1987/11/competing-on-the-eight-dimensions-of-quality>
- Gaye, B., Zhang, D., & Wulamu, A. (2021). Improvement of Support Vector Machine Algorithm in Big Data Background. *Mathematical Problems in Engineering*, 2021(1), 5594899. <https://doi.org/https://doi.org/10.1155/2021/5594899>
- Göker, H., & Tekedere, H. (2017). FATİH Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri İle Otomatik Değerlendirilmesi. *Bilişim Teknolojileri Dergisi*, 10(3), 291-299. <https://doi.org/10.17671/gazibtd.331041>
- Gullo, F. (2015). From patterns in data to knowledge discovery: What data mining can do. *Physics Procedia*, 62. <https://doi.org/10.1016/j.phpro.2015.02.005>
- Gupta, A., Sharma, H., & Akhtar, A. (2021). A Comparative Analysis Of K-Means And Hierarchical Clustering. *EPRA International Journal of Multidisciplinary Research (IJMR)*. <https://doi.org/10.36713/epra8308>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Elsevier.
- Herbig, P. A., & Palumbo, F. (1993). Serving the aftermarket in Japan and the United States. *Industrial Marketing Management*, 22(4), 339-346.
- Hossain, A., Karimuzzaman, M., Hossain, M. M., & Rahman, A. (2021). Text Mining and Sentiment Analysis of Newspaper Headlines. *Inf.*, 12(10). <https://doi.org/10.3390/INFO12100414>
- Hu, M., & Liu, B. (2004). *Mining Opinion Features in Customer Reviews*. www.aaai.org
- Imai, M. (1997). *Kaizen: Japonya'nın Rekabetti Başarısının Anahtarı*. Brissa Yayınları.
- İnan, H. E. (2024). Comparison of Machine Learning Algorithms for Classification of Hotel Reviews: Sentiment Analysis of TripAdvisor Reviews. *GSI Journals Serie A: Advancements in Tourism Recreation and Sports Sciences*, 7(1). <https://doi.org/10.53353/atrss.1327615>

- Islam, M. S. (2013). Impacts of Data Mining on Relational Database Management System Centric Business Environments. *International Journal of Computer Applications*, 75, 21-27. <https://doi.org/10.5120/13091-0371>
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice-Hall.
- Jayalakshmi, T., & Santhakumaran, A. (2011). Statistical Normalization and Back Propagation for Classification. *International Journal of Computer Theory and Engineering*, 3(1), 793-8201.
- Jiang, Q., Deng, L., & Yang, C. (2023). The Impact Mechanism of Consumer's Initial Visit to an Automobile 4S Store on Test Drive Intention: Product Aesthetics, Space Image, Service Quality, and Brand Image. *Behavioral Sciences*, 13(8). <https://doi.org/10.3390/bs13080673>
- Jinlin Wang Xing Wang, * Yuchen Yang Hongli Zhang Binxing Fang. (2020). A Review of Data Cleaning Methods for Web Information System. *Computers, Materials & Continua*, 62(3), 1053-1075. <https://doi.org/10.32604/cmc.2020.08675>
- Juran, J. M., & Godfrey, A. B. (1999). *Juran's Quality Handbook*. McGraw-Hill Professional Publishing.
- Juran, J. M., Gryna, F. M., & Bingham, R. S. (1988). *Quality Control Handbook*. McGraw-Hill.
- Kaithwas, K., & Borkar, P. (2019). A review on different data mining algorithms and selection methods. *Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2019*, 511-515. <https://doi.org/10.1109/ISS1.2019.8908048>
- Kano, N. (2001). Life cycle and creation of attractive quality. *4th QMOD Conference*, 18-36.
- Kaşıkcı, T., & Gökçen, H. (2014). Metin Madenciliği İle E-Ticaret Sitelerinin Belirlenmesi. *Bilişim Teknolojileri Dergisi*, 7(1). <https://doi.org/10.12973/bid.2014>
- Kazan, H., Yurtman, G., & Değermen, H. (2017). Toplam Kalite Yönetiminin İç Ve Dış Müşterilerin Performans Değerlendirmesi Üzerine Etkisi. *Yönetim ve Ekonomi Araştırmaları Dergisi*. <https://doi.org/10.11611/yead.283461>
- Keivandarian, N., & Carvalho, M. (2023). A Survey on Sentiment Classification Methods and Challenges. *The International FLAIRS Conference Proceedings*, 36. <https://doi.org/10.32473/FLAIRS.36.133314>
- Keller, K. L. (1998). *Strategic Brand Management: Building, Measuring and Managing Brand Equity*. Prentice Hall International.
- Kılıç, S. (2013). Doğrusal regresyon analizi. *Journal of Mood Disorders*, 3(2).

- Kılınç, D., Borandağ, E., Yücalar, F., Tunalı, V., Şimşek, M., & Özçift, A. (2016). KNN Algoritması ve R Dili ile Metin Madenciliği Kullanılarak Bilimsel Makale Tasnifi. *Marmara Fen Bilimleri Dergisi*, 28(3), 89-94. <https://doi.org/10.7240/mufbed.69674>
- Kınık, D., & Güran, A. (2021a). TF-IDF ve Doc2Vec Tabanlı Türkçe Metin Sınıflandırma Sisteminin Başarım Değerinin Ardışık Kelime Grubu Tespiti ile Arttırılması. *Avrupa Bilim ve Teknoloji Dergisi*, 21, 323-332. <https://doi.org/10.31590/ejosat.774144>
- Kınık, D., & Güran, A. (2021b). TF-IDF ve Doc2Vec Tabanlı Türkçe Metin Sınıflandırma Sisteminin Başarım Değerinin Ardışık Kelime Grubu Tespiti ile Arttırılması. *Avrupa Bilim ve Teknoloji Dergisi*, 21, 323-332. <https://doi.org/10.31590/ejosat.774144>
- Kocarı Gacar, B., & Deveci Kocakoç, İ. (2020). Regression Analyses or Decision Trees? *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, 18(4), 251-260. <https://doi.org/10.18026/cbayarsos.796172>
- Kochar, A. (2023). *Advantages of DBMS Over File System*. <https://www.prepbytes.com/blog/dbms/advantages-of-dbms-over-file-system/>
- Kotler, P. (1997). *Marketing Management: Analysis, Planning, Implementation and Control* (9th bs). Prentice Hall International, Inc.
- Kotler, P. (2000). *Marketing Management* (10th bs). Prentice-Hall.
- Kotler, P., & Armstrong, G. (2017). *Principles of Marketing*. Pearson Education.
- Kurnuç, M., Korucuk, S., & Küçük, O. (2015). Kalite İyileştirme Çalışmalarının Müşteri Memnuniyeti ve Müşteri Sadakatine Etkisi. *The International New Issues in Social Sciences*, 1(1), 21-44.
- Kuşluvan, S. (1999). Turizm İşletmelerinde Hizmet Kalitesi Yönetimi. *Milli Prodüktivite Merkezi Semineri*.
- Li, C., Liu, K., & Wang, H. (2011). The incremental learning algorithm with support vector machine based on hyperplane-distance. *Applied Intelligence*, 34(1), 19-27. <https://doi.org/10.1007/s10489-009-0176-9>
- Li, H. B., Wang, W., Ding, H. W., & Dong, J. (2010). Trees Weighting Random Forest method for classifying high-dimensional noisy data. *Proceedings - IEEE International Conference on E-Business Engineering, ICEBE 2010*, 160-163. <https://doi.org/10.1109/ICEBE.2010.99>
- Liu, B., Zhao, J., Liu, K., & Xu, L. (2016). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. *Computational Linguistics*, 42(3), 595-598. https://doi.org/10.1162/COLI_R_00259

- Matzler, K., & Hinterhuber, H. H. (1998). How to make product development projects more successful by integrating Kano's model of customer satisfaction into quality function deployment. *Technovation*, 18(1), 25-38. [https://doi.org/10.1016/S0166-4972\(97\)00072-2](https://doi.org/10.1016/S0166-4972(97)00072-2)
- Mishra, K., & Vegad, M. (2024). Customer Feedback Analysis Using Text Mining. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 10, 636-641. <https://doi.org/10.32628/CSEIT2410238>
- N, M. K. K., & Seetharam, K. (2014). Reliability prediction using fuzzy logic. 2014 *International Conference on Computing for Sustainable Global Development (INDIACom)*, 219-221. <https://doi.org/10.1109/IndiaCom.2014.6828131>
- Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2(1). <https://doi.org/10.1186/s40537-014-0007-7>
- Norman, D. A. (2004). *Emotional Design: Why We Love (or Hate) Everyday Things*. Basic Books. <https://www.basicbooks.com/titles/donald-a-norman/emotional-design/9780465051366/>
- Oktavanny, A., & Sulistiadi, W. (2022). The Determinant Factors of Customer Satisfaction: Promotion, Service Quality and Brand Image. *IJEED (International Journal of Entrepreneurship and Business Development)*, 5, 312-322. <https://doi.org/10.29138/ijebed.v5i2.1760>
- Oliver, R. L. (1980a). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 460-469.
- Oliver, R. L. (1980b). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of marketing research*, 17(4), 460-469.
- Onan, A. (2020). Evriřimli Sınır Ağı Mimarilerine Dayalı Türkçe Duygu Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, 374-380. <https://doi.org/10.31590/ejosat.780609>
- Öge, B. C., & Kayaalp, F. (2021). Farklı Sınıflandırma Algoritmaları ve Metin Temsil Yöntemlerinin Duygu Analizinde Performans Karşılaştırılması. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 9(6), 406-416. <https://doi.org/10.29130/dubited.1015320>
- Özbay, Ö. (2015). Veri Madenciliği Kavramı Ve Eğitimde Veri Madenciliği Uygulamaları. *Uluslararası Eğitim Bilimleri Dergisi*, 5, 262-272.
- Özgören, Öğr. Gör. F. (2014). Satış Sonrası Hizmetin Müşteri Sadakatine Etkisi Ve Ericsson Örneği. *Marmara Sosyal Araştırmalar Dergisi*, 2, 1-23.
- Özkan, B., & Alp, S. (2020). Toplu Ulaşımda Hizmet Kalitesi: İstanbul'da Yolcu Memnuniyeti Araştırması. *Akıllı Ulaşım Sistemleri ve Uygulamaları Dergisi*, 3(2), 94-111.

- Özveri, O., & Kabak, M. (2018). İşletmelerin Ürün Kalitesi Etkinliğinin Analiz Edilmesi İçin Bulanık Veri Zarflama Analizi Yönteminin Kullanılması. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 18(3), 145-158. <https://doi.org/10.18037/ausbd.552544>
- Pamucar, D., Simic, V., Görçün, Ö. F., & Küçükönder, H. (2024). Selection of the best Big Data platform using COBRAC-ARTASI methodology with adaptive standardized intervals. *Expert Systems with Applications*, 239. <https://doi.org/10.1016/j.eswa.2023.122312>
- Pandey, A., Sharma, A., & Agrawal, K. K. (2017). Developing efficient data mining algorithms. *2017 International Conference on Intelligent Sustainable Systems (ICISS)*, 1073-1076. <https://doi.org/10.1109/ISS1.2017.8389345>
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A Conceptual Model of Service Quality and Its Implications for Future Research. *Journal of Marketing*, 49(4). <https://doi.org/10.1177/002224298504900403>
- Park, H., & Kwon, H.-C. (2011). Improved Gini-Index Algorithm to Correct Feature-Selection Bias in Text Classification. *IEICE Transactions*, 94-D, 855-865. <https://doi.org/10.1587/transinf.E94.D.855>
- Parthe, R. M. (2023). Comparative Analysis of Data Visualization Tools: Power BI and Tableau. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, 07(10). <https://doi.org/10.55041/ijsrem26272>
- Peter, J. T., Wang, W., & Ney, H. (2016). Exponentially Decaying Bag-of-Words Input Features for Feed-Forward Neural Network in Statistical Machine Translation. *54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Short Papers*, 293-298. <https://doi.org/10.18653/V1/P16-2048>
- Pfalzgraf, B. (2020). *How to Use Selenium to Web-Scrape with Example*. <https://towardsdatascience.com/how-to-use-selenium-to-web-scrape-with-example-80f9b23a843a>
- Qais, E., & Veena, M. N. (2023). TxtPrePro: Text Data Preprocessing Using Streamlit Technique for Text Analytics Process. *2023 International Conference on Network, Multimedia and Information Technology, NMITCON 2023*. <https://doi.org/10.1109/NMITCON58196.2023.10275887>
- Ramani, R. G., & Shanthi, S. (2012). Classifier Prediction Evaluation in Modeling Road Traffic Accident Data. *Conference Proceeding of Computational Intelligence & Computing Research (ICCIC)*, 1-4.
- Rane, N., Achari, A., & Choudhary, S. (2023). Enhancing customer loyalty through quality of service: Effective strategies to improve customer satisfaction, experience, relationship, and engagement. *International Research Journal of Modernization in Engineering Technology and Science*, 5, 427-452. <https://doi.org/10.56726/IRJMET538104>

- Rao, K., & Dey, S. (2011). Decision Support for E-Governance: A Text Mining Approach. *International Journal of Managing Information Technology*, 3(3). <https://doi.org/10.5121/ijmit.2011.3307>
- Ribeiro, V., Rocha, A., Peixoto, R., Portela, F., & Santos, M. F. (2017). Importance of statistics for data mining and data science. *Proceedings - 2017 5th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2017, 2017-January*. <https://doi.org/10.1109/FiCloudW.2017.86>
- Rokach, L., & Maimon, O. (2008). *Data Mining with Decision Trees: Theory and Applications*. World Scientific Publishing.
- Rotar, L. J., & Kozar, M. (2017). The Use of the Kano Model to Enhance Customer Satisfaction. *Organizacija*, 50(4). <https://doi.org/10.1515/orga-2017-0025>
- Sarioğlu, C. İ. (2023). Beyaz eşya sektöründe satış sonrası hizmet kalitesinin müşteri sadakati üzerindeki etkisinin aracı değişkenlerle ölçülmesi: Müşteri memnuniyeti ve müşteriye elde tutma. *Business & Management Studies: An International Journal*, 11, 679-695. <https://doi.org/10.15295/bmij.v11i2.2248>
- Savaş, H., & Kesmez, A. G. (2014). Hizmet Kalitesinin Servqual Modeli İle Ölçülmesi: Aile Sağlığı Merkezleri Üzerine Bir Araştırma. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 17, 1-13.
- Sebastianelli, R., & Tamimi, N. (2002). How product quality dimensions relate to defining quality. *International Journal of Quality and Reliability Management*, 19(4), 442-453. <https://doi.org/10.1108/02656710210421599/FULL/PDF>
- Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34(1), 1-47. <https://doi.org/10.1145/505282.505283>
- Shaw, M. J., Subramaniam, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 31(1), 127-137.
- Sirsat, S., Chavan, V., & Deshpande, S. (2014). Mining knowledge from text repositories using information extraction: A review. *Sadhana*, 39, 53-62. <https://doi.org/10.1007/s12046-013-0197-2>
- Sriramoji, P. C. (2024). *Performance of KNN values based on K values*. <https://medium.com/@chandu58627/performance-of-knn-values-based-on-k-values-f7d95d467b32>
- Şahin, A. (2017a). Hizmet Kalitesinin Müşteri Memnuniyeti Üzerine Etkisinin İncelenmesi. *Journal of International Social Research*, 10, 1176-1184. <https://doi.org/10.17719/jisr.2017.1971>
- Şahin, A. (2017b). Hizmet Kalitesinin Müşteri Memnuniyeti Üzerine Etkisinin İncelenmesi. *Journal of International Social Research*, 10(52), 1176-1184. <https://doi.org/10.17719/jisr.2017.1971>

- Şatır, E., Azboy, F., Aydın, A., Arslan, H., & Hacıfendioğlu, Ş. (2016). Veri İndirgeme ve Sınıflandırma Teknikleri ile Glokom Hastalığı Teşhisi. *El-Cezeri*, 3(3). <https://doi.org/10.31202/ecjse.258576>
- Şehribanoğlu, S., & Diler, S. (2018). 2013 Yılı Yaşam Memnuniyeti Araştırmasının C & RT ve CHAID Algoritmaları İle İncelenmesi. *Akademik Sosyal Araştırmalar Dergisi*, 6(67), 132-141.
- Tabaei, Z., & Fathian, M. (2014). Measuring and analysing customer satisfaction with multi-criteria satisfaction analysis (MUSA) model: An electronic retailing case study. *International Journal of Business Information Systems*, 15(2), 189-204. <https://doi.org/10.1504/IJBIS.2014.059253>
- Tukiran, M., Tan, P. H. P., & Sunaryo, W. (2021). Obtaining customer satisfaction by managing customer expectation, customer perceived quality and perceived value. *Uncertain Supply Chain Management*, 9(2). <https://doi.org/10.5267/j.uscm.2021.1.003>
- Türkmen, M. (2017). ÜRETİMDE PARADİGMAL YAKLAŞIMLAR ÜZERİNE DEĞERLENDİRMELER.
- Tüzüntürk, S. (2010). Veri Madenciliği ve İstatistik. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 29(1), 65-90.
- Underwood, R. (2003). The Communicative Power of Product Packaging: Creating Brand Identity via Lived and Mediated Experience. *Journal of Marketing Theory and Practice*, 11, 62-76. <https://doi.org/10.1080/10696679.2003.11501933>
- Venkatesan, R., & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing*, 68(4), 106-125.
- Wainwright, A. (2024, Şubat 26). What Is Product Durability & Why Is It Important? | Zupan. What Is Product Durability. <https://zupan.ai/blog/what-is-product-durability>
- Wang, L. (2017). Data Mining, Machine Learning and Big Data Analytics. *International Transaction of Electrical and Computer Engineers System*, 4(2), 55-61. <https://doi.org/10.12691/iteces-4-2-2>
- Wazid, M., Das, A. K., Chamola, V., & Park, Y. (2022). Uniting cyber security and machine learning: Advantages, challenges and future research. *ICT Express*, 8(3), 313-321. <https://doi.org/10.1016/j.icte.2022.04.007>
- Yakut, E., Elmas, B., & Yavuz, S. (2014). Yapay Sinir Ağları ve Destek Vektör Makineleri Yöntemleriyle Borsa Endeksi Tahmini. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19(1), 139-157.
- Yasin, N. M., Noor, M. N., & Mohammad, O. (2007). Does image of country-of-origin matter to brand equity? *Journal of Product and Brand Management*, 16(1), 38-88.

- Yıldırım, P., Uludağ, M., & Görür, A. (2008). Hastane Bilgi Sistemlerinde Veri Madenciliği. *Akademik Bilişim 2008*, 429-434.
- Yılmaz, E. (2021). Ürün Geliştirme Süreçlerinde Kano Modeli Sınıflandırma Yöntemleri Seçimi. *The Selection Of Classification Methods In Kano's Model For Product Development Processes. Finans Ekonomi ve Sosyal Araştırmalar Dergisi*, 5, 842-859. <https://doi.org/10.29106/fesa.834274>
- Yip, I. (2022, Aralık 19). *Importance of Quality Materials*. <https://qualitysifu.com/the-importance-of-quality-materials/>
- Yüksel, A. S., & Tan, F. G. (2018). Metin Madenciliği Teknikleri İle Sosyal Ağlarda Bilgi Keşfi. *Mühendislik Bilimleri ve Tasarım Dergisi*, 6(2), 324-333. <https://doi.org/10.21923/jesd.384791>
- Zeithaml, V. A. (1988a). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. *Journal of Marketing*, 52(3), 2-22. <https://doi.org/10.1177/002224298805200302>
- Zeithaml, V. A. (1988b). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. *Journal of Marketing*, 52(3), 2-22.
- Zeithaml, V., Parasuraman, A., & Berry, L. (1990). Balancing Customer Perceptions and Expectations. *Ed. Free Press. EE. UU*.
- Zhao, Z., Jian, Z., Gaba, G. S., Alroobaea, R., Masud, M., & Rubaiee, S. (2021). An improved association rule mining algorithm for large data. *Journal of Intelligent Systems*, 30(1), 750-762. <https://doi.org/10.1515/jisys-2020-0121>
- Zhong, S., Scarinci, A., & Cicirello, A. (2023). Natural Language Processing for systems engineering: Automatic generation of Systems Modelling Language diagrams. *Knowledge-Based Systems*, 259. <https://doi.org/10.1016/j.knosys.2022.110071>