



**İSKENDERUN TEKNİK**  
ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**YÜKSEK  
LİSANS  
TEZİ**

**E-TİCARET SİTELERİNDE YER ALAN  
TÜKETİCİ YORUMLARININ  
METİN MADENCİLİĞİ  
YÖNTEMLERİ İLE ANALİZİ**

**Elif UZUN**

**EKONOMİ VE FİNANAS  
ANABİLİM DALI**

**AĞUSTOS 2024**



**E-TİCARET SİTELERİNDE YER ALAN TÜKETİCİ YORUMLARININ  
METİN MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ**

**ELİF UZUN**

**YÜKSEK LİSANS TEZİ**

**EKONOMİ VE FİNANS ANABİLİM DALI**

**İSKENDERUN TEKNİK ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

**AĞUSTOS 2024**

Elif UZUN tarafından hazırlanan “E-TİCARET SİTELERİNDE YER ALAN TÜKETİCİ YORUMLARININ METİN MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile İskenderun Teknik Üniversitesi Ekonomi ve Finans Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

**Danışman:** Dr. Öğr. Üyesi Murat UÇAR  
Bilgisayar Yazılımı Anabilim Dalı, İzmir Bakırçay Üniversitesi  
Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....  
.....

**Başkan:** Doç. Dr. Ümit ATİLA  
Bilgisayar Mühendisliği Anabilim Dalı, Gazi Üniversitesi  
Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....  
.....

**Üye:** Doç. Dr. Yunus AÇCI  
Ekonomi ve Finans Anabilim Dalı, İskenderun Teknik Üniversitesi  
Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

.....  
.....

Tez Savunma Tarihi: 01/08/2024

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....

Doç. Dr. Ersin BAHÇECİ  
Lisansüstü Eğitim Enstitüsü Müdürü

## ETİK BEYAN

İskenderun Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez üzerinde Yükseköğretim Kurulu tarafından hiçbir değişiklik yapılamayacağı için tezin bilgisayar ekranında görüntülediğinde asıl nüsha ile aynı olması sorumluluğunun tarafıma ait olduğunu,
- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

İmza  
Elif UZUN  
01/08/2024

E-TİCARET SİTELERİNDE YER ALAN TÜKETİCİ YORUMLARININ METİN  
MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ  
(Yüksek Lisans Tezi)

Elif UZUN

İSKENDERUN TEKNİK ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

Ağustos 2024

ÖZET

Günümüzde teknolojiye yaşanan hızlı gelişmeler sonucunda tüketiciler, online alışveriş ortamını geleneksel alışveriş ortamına kıyasla daha fazla tercih etmektedir. Online alışveriş sürecinde müşterilerin bir ürünü satın alma aşamasında faydalandığı en önemli kaynaklardan biri tüketici yorumlarıdır. Ürünü satın alan müşterilerin paylaştığı yorumlar, başka müşterilerin satın alma kararını ve ürünün markası hakkındaki düşüncelerini etkilemektedir. Ancak ürün yorumlarında ürün ile ilgisi olmayan çok sayıda gereksiz yoruma rastlanılmakta ve bu durum ürün hakkında bilgi veren yorumlara ulaşmayı zorlaştırmaktadır. Öte yandan tüketicilerin paylaştığı yorumlarda kelimelerin konuşma dilinde yazılması, kısaltılması, noktalama işaretlerinin lüzumsuz veya fazla kullanılması gibi gürültü oluşturan durumlar da işlenecek metin verilerinin analizi zorlaşmakta ve bu zorlukların üstesinden gelebilmek için metin madenciliği tekniklerinden yararlanılmaktadır. Bu nedenle bu çalışmada metin madenciliği teknikleri kullanılarak online alışveriş sitelerinden daha önce alışveriş yapan müşterilerin, deneyimleri yönünde paylaştıkları çevrimiçi yorumların ürün ile alakalı olup olmadığının tespit edilmesi amaçlanmıştır. Bu amaçla öncelikle, Türkiye'nin önde gelen online pazar yerlerinden www.trendyol.com sitesinde yer alan farklı kategorilerdeki 21 ürüne ait 4795 adet yorum içeren bir veri seti oluşturulmuştur. Daha sonra bu veri seti üzerinde Terim Frekansı ve Ters Doküman Frekansı yöntemleri kullanılarak kelime vektörleri oluşturulmuştur. Son olarak, k-En Yakın Komşu (kNN), Naive Bayes, Karar Ağacı, Destek Vektör Makinesi (DVM) ve Derin Öğrenme sınıflandırma algoritmaları kullanılarak yorumlar sınıflandırılmış ve en yüksek başarı %81.61 doğruluk değeri ile DVM algoritması ile elde edilmiştir. Yapılan bu çalışma tüketicilere yalnızca ürün ile ilgili yorumları filtreleme imkânı sunacağından satın alma aşamasında karar vermelerini kolaylaştıracağı düşünülmektedir. Ayrıca online satış yapan firmalar ürün hakkında paylaşılan müşteri düşüncelerini dikkate alarak zayıf ve eksik yönlerini yeniden gözden geçirme fırsatı bulabileceklerdir.

Anahtar Kelimeler : Metin madenciliği, online alışveriş, tüketici yorumları, makine öğrenmesi

Sayfa Adedi : 65

Danışman : Dr. Öğr. Üyesi Murat UÇAR

# ANALYSING CONSUMER COMMENTS ON E-COMMERCE WEBSITES USING TEXT MINING METHODS

(M. Sc. Thesis)

Elif UZUN

ISKENDERUN TECHNICAL UNIVERSITY  
INSTITUTE OF GRADUATE STUDIES

August 2024

## ABSTRACT

Today, as a result of rapid developments in technology, consumers prefer the online shopping environment more than the traditional shopping environment. Consumer comments are one of the most essential resources that customers benefit from when purchasing a product online. The comments shared by the customers who bought the product affect the purchasing decision of other customers and their thoughts about the brand of the product. However, there are numerous needless remarks unrelated to the product in the product comments, making it difficult to find the comments that provide information on the goods. On the other hand, in the comments shared by consumers, in cases where the words are written in spoken language, shortened, or excessive use of punctuation marks, the analysis of the text data to be processed becomes difficult and text mining techniques are used to overcome these difficulties. Therefore, the purpose of this study is to figure out, using text mining techniques, if online comments posted by customers who have previously made purchases from online shopping sites are relevant to the product. For this purpose, at first, a data set of 4795 comments on 21 products from various categories was constructed on [www.trendyol.com](http://www.trendyol.com), one of Turkey's top online marketplaces. Then, word vectors were created on this data set by using Term Frequency and Inverse Document Frequency methods. Finally, the comments were classified using the k Nearest Neighbor, Naive Bayes, Decision Trees, Support Vector Machine (SVM) and Deep Learning classification algorithms, and the highest success was obtained with the SVM algorithm with an accuracy value of 81.61%. It is thought that this study will make it easier for consumers to decide at the purchasing stage, as it will provide the opportunity to filter only the comments about the product. Additionally, online retailers will have the chance to assess their flaws and shortcomings by considering the comments made by customers regarding the merchandise.

Key Words : Text mining, online shopping, consumer comments, machine learning

Page Number : 65

Supervisor : Asst. Prof. Dr. Murat UÇAR

## TEŞEKKÜR

Bu çalışmada her daim yanımda olan, zor zamanlarımda bana güç veren, hayatımın en değerli anılarını beraber geçirdiğim yol arkadaşım Sabahattin NUR'a göstermiş olduğu destek, sabır ve anlayış için sonsuz teşekkür ederim. Onunla bu yolculuğa çıkmak benim için büyük bir şans oldu ve onun desteği olmadan bu kadar ilerleyemezdim.

Yüksek lisans eğitimim boyunca bana rehberlik eden, bu çalışmanın hazırlanmasında desteğini esirgemeyen, bilgisi ve deneyimlerini benimle paylaşan saygı değer danışmanım Dr. Öğr. Üyesi Murat UÇAR ve değerli bilgilerini nezaketle aktaran Doç. Dr. Emine UÇAR'a gösterdikleri sabır ve anlayış için sonsuz teşekkürlerimi sunarım. Ayrıca bu eğitim süreci boyunca sağladığı maddi ve manevi desteklerinden dolayı Prof. Dr. Nuriye GÜNEBAKAN'a içten teşekkürlerimi sunarım.

**İÇİNDEKİLER****Sayfa**

|  |      |
|--|------|
| ÖZET .....                                     | v    |
| ABSTRACT .....                                 | vi   |
| TEŞEKKÜR .....                                 | vii  |
| İÇİNDEKİLER .....                              | viii |
| ÇİZELGELERİN LİSTESİ .....                     | x    |
| ŞEKİLLERİN LİSTESİ .....                       | xi   |
| SİMGELER VE KISALTMALAR .....                  | xii  |
| 1. GİRİŞ .....                                 | 1    |
| 2. VERİ MADENCİLİĞİ ve METİN MADENCİLİĞİ ..... | 4    |
| 2.1. Veri Madenciliği .....                    | 4    |
| 2.2. Metin Madenciliği .....                   | 9    |
| 3. LİTERATÜR .....                             | 13   |
| 3.1. Ulusal Literatür .....                    | 13   |
| 3.2. Uluslararası Literatür .....              | 19   |
| 4. MATERYAL VE METOD .....                     | 23   |
| 4.1. Araştırmada Kullanılan Veri Seti .....    | 23   |
| 4.2. Araştırmada Kullanılan Yazılım .....      | 24   |
| 4.3. Araştırmada Kullanılan Modeller .....     | 24   |
| 4.3.1. K-en yakın komşu .....                  | 25   |
| 4.3.2. Naive bayes .....                       | 26   |
| 4.3.3. Karar ağacı .....                       | 27   |
| 4.3.4. Destek vektör makinesi .....            | 29   |
| 4.3.5. Derin öğrenme .....                     | 30   |
| 4.4. Değerlendirme Metrikleri .....            | 31   |
| 4.4.1. Karmaşıklık matrisi .....               | 32   |

|   |           |
|---|-----------|
| 4.4.2. Doğruluk .....   | 33        |
| 4.4.3. Hata oranı .....   | 33        |
| 4.4.4. Kesinlik .....   | 33        |
| 4.4.5. Duyarlılık .....   | 34        |
| 4.4.6. Özgüllük .....   | 35        |
| 4.4.7. F- ölçütü .....  | 35        |
| 4.4.8. Eğri altında kalan alan (AUC).....                                       | 36        |
| <b>5. DENEYSEL ÇALIŞMALAR ve BULGULAR.....</b>                                  | <b>37</b> |
| 5.1. Araştırmada Kullanılan Veri Ön İşleme Adımları.....                        | 37        |
| 5.1.1. Tüm karakterlerin büyük harfe dönüştürülmesi.....                        | 37        |
| 5.1.2. Noktalama işaretlerinin kaldırılması .....                               | 38        |
| 5.1.3. Anlamsız kelimelerin filtrelenmesi .....                                 | 39        |
| 5.1.4. Kelimelerin köklerine ayrılması.....                                     | 40        |
| 5.1.5. Kelimelerin en düşük ve en yüksek uzunluk sınırlarının belirlenmesi..... | 41        |
| 5.1.6. Terim frekansı ve ters doküman frekansı .....                            | 42        |
| 5.2. Araştırmanın Bulguları ve Değerlendirme.....                               | 44        |
| 5.2.1. Naive bayes algoritması sonuçları .....                                  | 44        |
| 5.2.2. K-en yakın komşu algoritması sonuçları .....                             | 45        |
| 5.2.3. Karar ağacı algoritması sonuçları.....                                   | 46        |
| 5.2.4. Derin öğrenme algoritması sonuçları .....                                | 48        |
| 5.2.5. Destek vektör makinesi algoritması sonuçları .....                       | 50        |
| 5.3. Algoritmaların Performans Karşılaştırması .....                            | 51        |
| 5.4. Algoritmaların Sonuçlarının Tartışılması .....                             | 54        |
| <b>6. SONUÇ .....</b>   | <b>56</b> |
| <b>KAYNAKLAR .....</b>  | <b>58</b> |
| <b>DİZİN .....</b>  | <b>64</b> |

## ÇİZELGELERİN LİSTESİ

| <b>Çizelge</b>   | <b>Sayfa</b> |
|--|--------------|
| Çizelge 4.1. Veri seti içerisindeki kayıtlardan örnekler .....                       | 23           |
| Çizelge 4.2. Karmaşıklık matrisi.....  | 32           |
| Çizelge 5.1. Çalışmada kullanılan stopword listesinden örnekler .....                | 40           |
| Çizelge 5.2. Sınıflandırma algoritmalarının değerlendirme metrikleri sonuçları ..... | 52           |
| Çizelge 5.3. Sınıflandırma algoritmalarının karmaşıklık matrisi sonuçları.....       | 54           |



## ŞEKİLLERİN LİSTESİ

| Şekil  | Sayfa |
|--|-------|
| Şekil 2.1. Veri madenciliği ve disiplinler .....                         | 4     |
| Şekil 2.2. Bilgi keşfi sürecinde veri madenciliği .....                  | 6     |
| Şekil 2.3. Metin madenciliği süreci .....                                | 11    |
| Şekil 4.1. Veri setindeki yorumların kategorilere göre dağılımı .....    | 23    |
| Şekil 5.1. Veri ön işleme süreci .....                                   | 37    |
| Şekil 5.2. Transform cases operatörü.....                                | 38    |
| Şekil 5.3. Tokenize operatörü.....                                       | 39    |
| Şekil 5.4. Filter stopwords (dictionary) operatörü.....                  | 40    |
| Şekil 5.5. Stem (snowball) operatörü.....                                | 41    |
| Şekil 5.6. Filter tokens (by lenght) operatörü.....                      | 42    |
| Şekil 5.7. Generate n-grams (terms) operatörü.....                       | 43    |
| Şekil 5.8. Araştırmada kullanılan RapidMiner uygulaması .....            | 44    |
| Şekil 5.9. Naive bayes algoritmasının performans sonuçları .....         | 44    |
| Şekil 5.10. Naive bayes algoritmasının ROC eğrisi ve AUC değeri .....    | 45    |
| Şekil 5.11. KNN algoritmasının parametre değerleri.....                  | 45    |
| Şekil 5.12. KNN algoritmasının performans sonuçları.....                 | 46    |
| Şekil 5.13. KNN algoritmasının ROC eğrisi ve AUC değeri.....             | 46    |
| Şekil 5.14. Karar ağacı algoritmasının parametre değerleri .....         | 47    |
| Şekil 5.15. Karar ağacının görsel çıktısı.....                           | 47    |
| Şekil 5.16. Karar ağacı algoritmasının performans sonuçları.....         | 48    |
| Şekil 5.17. Karar ağacı algoritmasının ROC eğrisi ve AUC değeri .....    | 48    |
| Şekil 5.18. Derin öğrenme algoritmasının parametre değerleri.....        | 49    |
| Şekil 5.19. Derin öğrenme mimarisi .....                                 | 49    |
| Şekil 5.20. Derin öğrenme algoritmasının performans sonuçları .....      | 49    |
| Şekil 5.21. Derin öğrenme algoritmasının ROC eğrisi ve AUC değeri .....  | 50    |
| Şekil 5.22. DVM algoritmasının parametre değerleri.....                  | 50    |
| Şekil 5.23. DVM algoritmasının performans sonuçları .....                | 51    |
| Şekil 5.24. DVM algoritmasının ROC eğrisi ve AUC değeri .....            | 51    |
| Şekil 5.25. Uygulanan sınıflandırma algoritmalarının başarı grafiği..... | 52    |
| Şekil 5.26. Uygulanan sınıflandırma algoritmalarının AUC değerleri ..... | 53    |

## SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

### Simgeler

%

Log

### Açıklamalar

Yüzde

Logaritma

### Kısaltmalar

AUC

AVM

CNN

DVM

EDI

FN

FP

GRU

IDF

IMDB

KAÇD

KNIME

KNN

LDA

LSTM

NLP

NLTK

OneR

ROC

SMO

SMR

STPS

SQL

TF

TN

TP

TÜİK

WEKA

YALE

### Açıklamalar

Area Under the Curve

Alışveriş Merkezi

Convolutional Neural Network

Destek Vektör Makinesi

Elektronik Veri Değişimi

False Negative

False Positive

Gated Recurrent Unit

Inverse Document Frequency

Internet Movie Database

Kitlesel Açık Çevrimiçi Dersler

Konstanz Information Miner

K-Nearest Neighbor

Latent Dirichlet Allocation

Long-Short Term Memory

Neuro Linguistic Programming

Natural Language Toolkit

One Rule

Receiver Operating Characteristic

Sequential Minimal Optimization

Social Media Research

Sözlük Tabanlı Puanlama Sistemi

Structured Query Language

Term Frequency

True Negative

True Positive

Türkiye İstatistik Kurumu

Waikato Environment for Knowledge Analysis

Yet Another Learning Environment

## 1. GİRİŞ

İnternete erişimin kolay olması ile birlikte bilgiye ulaşmadaki zorluk büyük oranda azalma göstermiştir. İnternet aracılığıyla yapılan ve farklı bir alışveriş biçimi olan elektronik ticaretin özellikle pandeminin de etkisiyle önemi daha da artmıştır. Ülkemizde son yıllarda internet üzerinden yapılan alışveriş her alanda artış göstermeye devam etmektedir. Ayrıca elektronik ticaret; firmalara, tüketici pazarlarına, ulusal ve uluslararası pazarlara da bir değişim ve canlılık getirmiştir.

İnternet kullanıcı sayısının artması, online alışveriş yapma alışkanlığı olan kişilerin oranını da yükseltmektedir. Bu durum, hem tüketiciler hem de satıcılar için birçok avantaj sağlamaktadır. Tüketiciler, istedikleri ürün veya hizmete daha kolay, hızlı ve uygun fiyatlı bir şekilde ulaşabilmekte; satıcılar ise daha geniş bir kitleye ulaşarak daha fazla gelir elde edebilmektedir. Online alışverişin gelişmesi, aynı zamanda yeni iş alanları ve girişim fırsatları da yaratmaktadır. Örneğin, e-ticaret siteleri, lojistik firmaları, dijital pazarlama ajansları gibi sektörler online alışverişin büyümesinden doğrudan etkilenmektedir. Online alışverişin geleceği, internet kullanımının yaygınlaşması ve teknolojinin ilerlemesi ile birlikte daha da parlak görünmektedir (Çiçek ve Mürütsoy, 2014:293).

E-ticaret, internetin sağladığı platform sayesinde, bilgisayar ağları aracılığıyla ürün, hizmet ve bilgi gibi çeşitli kaynakların alım satımı, transferi ve değişimi gerçekleştirilebilen ileri bir ticaret modelidir. Bu model, geleneksel ticaret yöntemlerine göre daha hızlı, daha kolay ve daha verimli bir iş yapma imkanı sunmaktadır. İnternet ticareti, hem bireysel hem de kurumsal müşteriler için çeşitli avantajlar sağlamakta, rekabeti artırmakta ve ekonomik büyümeye katkıda bulunmaktadır. Bu nedenle, internet ticareti günümüzde hem bireysel hem de kurumsal düzeyde büyük bir önem taşımaktadır (Özkurt, 2023).

E-ticaret, 1990'ların ortalarında internetin yaygınlaşmasıyla birlikte hayatımıza giren ve ticaretin dijital ortamlarda yapılmasını sağlayan bir kavramdır. İlk olarak 1970'lerde Elektronik Veri Değişimi (EDI) ve tele-alışveriş gibi teknolojilerle başlayan e-ticaret, 1994 yılında Pizza Hut'ın internet üzerinden pizza satışı yapmasıyla modern anlamda ilk adımlarını atmıştır. Dünya genelinde e-ticaretin yaygınlaşması 2000'li yıllarda ticaretin serbestleşmesi ve internet kullanımının artmasıyla hız kazanmıştır. Amazon ve eBay gibi şirketlerin kurulması ve internet üzerinden satış yapmaya başlamalarıyla e-ticaret sektörü

büyük bir ivme kazanmış, bu platformlar dünya çapında e-ticaretin öncüleri haline gelmiştir (Gedik, 2021:190). Türkiye'de ise e-ticaretin tarihsel gelişimi 2007 yılından sonra hız kazanmış ve her geçen gün büyüyen bir sektör haline gelmiştir.

Ülkemizde son on yılın verileri incelendiğinde internete erişim hızlı bir artış göstermiştir. Hanehalkı bilişim teknolojileri kullanım araştırması sonuçlarına göre; 2012 yılında hanelerde internete erişim oranı %47,2 olarak belirlenirken bu oran 2023 yılında %95,5 oranına yükselmiştir. Bireylerde internet kullanımı ise 2012 yılında %47,4 iken 2023 yılına gelindiğinde %87,1 oranında gerçekleşmiştir. Dolayısıyla bireylerdeki internet kullanım oranı ile hanelerde internet erişim oranı birbirine paralellik göstermektedir. Türkiye İstatistik Kurumu verilerine göre internet kullanıcılarının internet üzerinden kişisel amaçla mal veya hizmet satın alması (e-ticaret) oranı, 2012 yılında %10,3 iken 2023 yılında %49,5 oranında gerçekleşmiştir (Türkiye İstatistik Kurumu [TÜİK], 2023).

Tüketiciler e-ticaret ortamında bir ürün satın alırken geleneksel alışveriş alışkanlıklarının yanı sıra o mal ve hizmeti önceden satın alan kişilerin yorumlarını da dikkate almaktadır (Cop ve Gümüş, 2009:180). Elektronik ortamda ürünü satın alan kişi ürünün marka, model ve fiyatı yanında ürün ve firma hakkında yapılmış değerlendirmeleri göz önünde bulundurarak karar vermektedir. Yorumlar, alıcıların ürün hakkında bilgi edinmelerine, karşılaştırma yapmalarına ve satın alma kararlarını verirken güven duymalarına yardımcı olur. Ayrıca satıcılar için ürünlerinin kalitesini ve müşteri memnuniyetini gösteren, potansiyel alıcıları çeken ve marka imajını güçlendiren bir araç konumundadır. Bu nedenle, e-ticaret sistemlerinde yorumların sayısı, niteliği ve olumlu ya da olumsuz olması, satış performansını doğrudan etkiler. Ancak paylaşılan kullanıcı yorumlarının çoğu ürün ile ilgili olmamaktadır. Çoğu yorum ürünün kalitesi, işlevi, boyutu, rengi, malzemesi gibi önemli özelliklerine değinmemektedir. Paylaşılan yorumlar daha çok “iyi”, “güzel”, “sorunsuz geldi”, “Tşk”, “hızlı kargo”, “kızım için aldım” gibi ürün hakkında gerçekçi ve yararlı bilgiler sunmak yerine, sadece olumlu veya olumsuz duyguları ifade eden gereksiz yorumlardan oluşmaktadır. Bu tür yorumların çok fazla olması, tüketicilerin ürünle ilgili bilgi veren ve satın alma kararını etkileyecek yorumlara ulaşmasını zorlaştırmaktadır. Ayrıca yorumlarda noktalama işaretlerinin gereksiz ve aşırı kullanımı, günlük konuşma dilinde ve kısaltılarak yazılması bu verilerin işlenmesini zorlaştırmaktadır.

Bu alıřmada tüketicide zaman kazandırmak ve satın alacağı ürün hakkında satın alma kararını etkileyecek yorumlara ulaşmasında kolaylık sağlanması konusu ele alınmıştır. Online alışveriş sitelerinden daha önce alışveriş yapan müşterilerin, deneyimleri yönünde paylaştıkları çevrimiçi yorumların ürün ile alakalı olup olmadığının metin madenciliği teknikleri kullanılarak tespit edilmesi amaçlanmıştır.



## 2. VERİ MADENCİLİĞİ ve METİN MADENCİLİĞİ

Veri madenciliği ve metin madenciliği, büyük veri kümelerinden bilgi çıkarma süreçleri olarak benzerlik gösterirken, ele aldıkları veri türleri açısından farklılık göstermektedir. Veri madenciliği genellikle sayısal verilerle çalışırken, metin madenciliği yapılandırılmamış metin verilerini analiz eder. Her iki disiplin de, verilerdeki gizli kalıpları ve ilişkileri keşfetmeye yöneliktir ve bu süreçte makine öğrenimi, istatistik ve algoritmalarından yararlanırlar.

### 2.1. Veri Madenciliği

Veri Madenciliği, büyük veri kümelerinin içerdiği bilgileri ortaya çıkarmak için bilgisayar programları kullanılarak yapılan bir veri analizi sürecidir. Büyük veri, geleneksel yöntemlerle işlenemeyecek kadar hacimli, çeşitli ve hızlı üretilen verilerdir. Bu verilerin içinde gizli olan bilgileri ortaya çıkarmak için Veri Madenciliği teknikleri kullanılır. Veri Madenciliği, verilerin içindeki kalıpları, eğilimleri, ilişkileri ve kuralları keşfederek gelecekteki olayları tahmin etmeye veya karar vermeye yardımcı olur. Veri Madenciliği, çok büyük boyutlardaki verilerin işlenmesini ve anlamlandırılmasını sağlayan bir veri bilimi alt dalıdır (Kalikov, 2006:7).

Veri Madenciliği, büyük hacimli veri kümelerinden anlamlı bilgiler elde etmek için verilerin analiz edilmesini sağlar. Veri Madenciliği, Şekil 2.1'de gösterildiği gibi, veri tabanı teknolojileri, enformasyon bilimi, görselleştirme, istatistik, Makine Öğrenmesi ve diğer disiplinlerden yararlanan bir alandır. Veri Madenciliği bu disiplinlerin kesişim noktasında yer alır (Özbay, 2015:263).



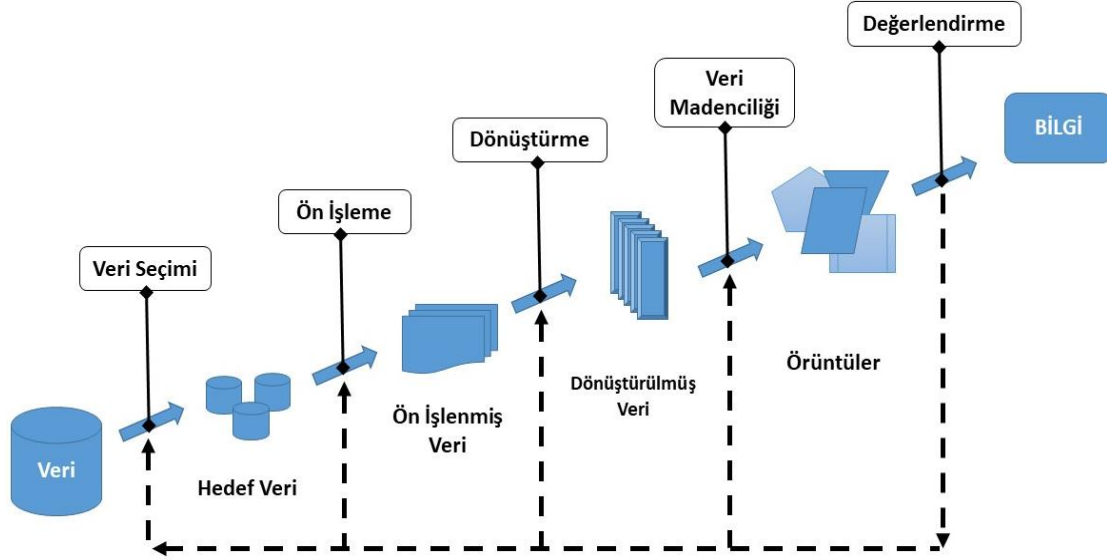
Şekil 2.1. Veri madenciliği ve disiplinler (Özbay, 2015:264)

Veri Madenciliği, farklı disiplinlerden gelen yöntemleri birleştirerek, verilerin özelliklerini, ilişkilerini ve kalıplarını keşfetmeyi amaçlar. Veri Madenciliği, üç ana alandan beslenerek ortaya çıkmıştır (Ergün, 2012:16). Bunlar:

- **Klasik istatistik:** İstatistik, veri madenciliğinin temel taşlarından biri olarak kabul edilir ve bilimin en eski dallarından biridir. Veri toplama, analiz etme, yorumlama ve sunma süreçleriyle yakından ilgilenir. İstatistik, veri setlerinin dağılımını, merkezi eğilim ve dağılım ölçülerini, varyansını, korelasyonunu ve test edilebilirliğini detaylı bir şekilde inceleyerek, verilerin anlamlandırılmasında kritik bir rol oynar. Ayrıca, çeşitli hipotezleri formüle etmek ve bu hipotezleri matematiksel modeller aracılığıyla test etmek için kullanılır. Bu süreçler, veri bilimindeki karmaşık problemleri çözmek ve bilgiye dayalı kararlar almak için temel oluşturur.
- **Yapay zeka:** Veri madenciliğinin ilerlemesine önemli katkılar sağlayan ve insan zekasını modellemeye yönelik çalışmalar yapan bir teknoloji alanıdır. Bu alan, verileri analiz etmek, öğrenme süreçlerini gerçekleştirmek, mantıksal sonuçlar çıkarmak ve kararlar almak için gelişmiş algoritmalar kullanır. Ayrıca, yapay zeka, toplanan verilerden elde edilen bilgilerle yeni sorunları çözme veya daha önce karşılaşılmamış durumlar karşısında uygun davranışlar geliştirme yeteneğine sahiptir.
- **Makine öğrenmesi:** Makine Öğrenimi, İstatistik ve yapay zekanın temel ilkeleri üzerine kurulu bir bilim alanı olarak tanımlanabilir. Bu disiplin, bilgisayar sistemlerinin topladığı verileri analiz etme, bu veriler arasından değerli bilgiler elde etme ve elde edilen bu bilgilerle akıllı kararlar alma yeteneğini geliştirmek için matematiksel modeller ve algoritmalar kullanır. Günümüzde, Makine Öğrenimi teknolojisi, sağlık hizmetlerinden finansa, eğitimden otomasyon endüstrisine kadar geniş bir yelpazede kullanılarak, insan yaşamını daha verimli ve gelişmiş hale getirmekte büyük rol oynamaktadır.

Veri Madenciliği, veri yığınlarından bilgi çıkarmak için kullanılan bir tekniktir. Veri Madenciliği, veriyi anlamak, özetlemek, sınıflandırmak, kümelemek, ilişkilendirmek ve tahmin etmek için çeşitli yöntemler ve algoritmalar uygular. Veri Madenciliği, bilgi keşfi sürecinin önemli bir adımıdır. Bilgi keşfi süreci, Şekil 2.2'de gösterildiği gibi, veriyi

temizleme, bütünleştirme, dönüştürme, seçme, modelleme ve değerlendirme gibi aşamalardan oluşur (Savaş, Topaloğlu ve Yılmaz, 2012:8).



Şekil 2.2. Bilgi keşfi sürecinde veri madenciliği (Savaş ve diğerleri, 2012:8)

Veri Madenciliği süreci, aşağıdaki adımları içerir (Savaş ve diğerleri, 2012:8);

- **Veri seçimi:** Bu adımda, Veri Madenciliği uygulanacak veri kaynakları belirlenir ve toplanır. Veri kaynakları, veri tabanları, web siteleri, sosyal medya, sensörler veya başka türde veriler olabilir. Veri seçimi, Veri Madenciliği sürecinin başarısını etkileyen önemli bir adımdır. Çünkü uygun ve kaliteli veriler, daha iyi bilgiler üretmeye yardımcı olur.
- **Veri ön işleme:** Toplanan verilerin, Veri Madenciliği algoritmalarına uygun hale getirilmesi işlemidir. Veri ön işleme, verilerin temizlenmesi, bütünleştirilmesi, dönüştürülmesi ve indirgenmesi gibi alt adımları içerir. Verilerin temizlenmesi, eksik, gürültülü veya tutarsız verilerin düzeltilmesi veya çıkarılmasıdır. Verilerin bütünleştirilmesi, farklı kaynaklardan gelen verilerin bir araya getirilmesidir. Verilerin dönüştürülmesi, verilerin standartlaştırılması, normalleştirilmesi veya öznitelik çıkarılmasıdır. Verilerin indirgenmesi, verilerin boyutunun azaltılmasıdır. Örneğin, veriler; örnekleme, kümeleme veya boyut indirgeme gibi tekniklerle indirgenebilir.

- **Veri madenciliği:** Ön işlenmiş veriler üzerinde, istatistiksel, Makine Öğrenmesi veya yapay zeka gibi yöntemlerle bilgi çıkarılması işlemidir. Veri Madenciliği, sınıflandırma, kümeleme, ilişki kuralları, derin öğrenme veya metin madenciliği gibi farklı teknikleri kapsar. Veri Madenciliği sonucunda elde edilen bilgiler, modeller, desenler veya kurallar şeklinde ifade edilebilir.
- **Değerlendirme:** Veri Madenciliği sonucunda elde edilen bilgilerin doğruluk, güvenilirlik ve anlamlılık açısından değerlendirilmesi işlemidir. Değerlendirme, farklı ölçütler kullanılarak yapılabilir. Örneğin, sınıflandırma modellerinin doğruluğu, karmaşıklık matrisi veya ROC eğrisi gibi metriklerle ölçülebilir. Kümeleme modellerinin kalitesi, siluet katsayısı veya Davies-Bouldin indeksi gibi metriklerle değerlendirilebilir. İlişki kurallarının anlamlılığı ise destek, güven veya kaldıraç gibi metriklerle belirlenebilir.
- **Bilgi sunumu:** Değerlendirilen bilgilerin ilgili kullanıcılara sunulması işlemidir. Bilgi sunumu, bilgilerin anlaşılır ve etkili bir şekilde iletilmesini sağlar. Sunum, grafikler, tablolar, raporlar veya görselleştirme gibi tekniklerle yapılabilir.

Veri Madenciliği sürecinin bu beş adımını takip ederek, büyük miktarda veriden anlamlı bilgiler çıkarmak mümkündür. Veri Madenciliği sürecinin her adımında farklı teknikler, araçlar ve yöntemler kullanılabilir. Veri Madenciliği sürecinin sonucunda, verilerden elde edilen bilgiler, karar verme, tahmin etme, keşfetme veya öğrenme gibi farklı amaçlar için kullanılabilir.

Veri Madenciliği, farklı sektörlerde ve alanlarda çeşitli amaçlarla kullanılmaktadır (Özbay, 2015:265). Bu alanlardan bazıları şunlardır:

- **Sağlık:** Veri madenciliği teknikleri, sağlık sektöründe ilaç geliştirme süreçlerinden hastalık tanı ve tedavi stratejilerine, ilaçların etkilerinin ve yan etkilerinin değerlendirilmesinden hasta test sonuçlarının analizine kadar geniş bir yelpazede önemli faydalar sağlamaktadır. Bu yöntemler, sağlık risklerinin belirlenmesinde ve yönetilmesinde de kritik rol oynamaktadır.
- **Endüstri ve mühendislik:** Veri Madenciliği, bilgisayarlarla üretilen ya da toplanan verilerin analizi, üretim süreçlerinin takibi, geliştirilmesi, kalite kontrol, güvence ve

sistem performansının optimizasyonu gibi çeşitli işlevlerde önemli katkılar sağlamaktadır. Bu süreçler, verilerin daha iyi anlaşılmasını ve işlenmesini sağlayarak, sektörlerin daha verimli ve etkin çalışmasına olanak tanımaktadır.

- **Kamu:** Veri Madenciliği, kamu sektörünün kaynaklarını daha etkin ve verimli bir şekilde kullanma, kamu güvenliği ve emniyetini sağlama, suçla mücadele etme, terörizm ve siber saldırılara karşı koyma gibi konularda kamu sektörüne destek olmaktadır.
- **Pazarlama:** Veri Madenciliği, pazarlama sektöründe çok geniş bir uygulama alanına sahiptir. Pazarlama sektörü, satış tahminlerinden müşteri segmentasyonuna, müşteri sadakatinden memnuniyet ölçümlmelerine, kârlılık ve rekabet analizlerine kadar birçok alanda veri madenciliğinden yararlanmaktadır. Bu sayede, işletmeler müşteri davranışlarını daha iyi anlayabilir, pazar trendlerini öngörebilir ve stratejik kararlar alabilirler. Veri madenciliği, pazarlamacılara derinlemesine bilgiler sunmakta ve sektördeki rekabet avantajını artırmalarına olanak tanımaktadır.
- **Bankacılık, finans ve borsa:** Veri Madenciliği, bankacılık, finans ve borsa sektöründe en yaygın kullanılan yöntemlerden biridir. Kredi kartı ve kredi değerlendirme, risk analizi ve yönetimi, hisse senedi fiyat tahmini, yatırım kararı verme gibi konularda bankacılık, finans ve borsa sektörüne katkıda bulunmaktadır.
- **İnternet:** Veri Madenciliği, internet ortamında büyük miktarda verinin üretilmesi ve toplanması nedeniyle internet sektöründe önemli bir rol oynamaktadır. Kullanıcı davranışları analizi, siber güvenlik, web arama motoru optimizasyonu ve e-ticaret gibi konularda internet sektörüne hizmet etmektedir. Özellikle e-ticaret platformları, müşteri davranışlarını analiz ederek kişiselleştirilmiş öneriler sunmak ve stok yönetimini optimize etmek için bu yöntemi kullanır.
- **Eğitim:** Veri Madenciliği, eğitim sektöründe de giderek artan bir öneme sahiptir. Öğrenci verilerinin analizi, öğrenci başarı ve başarısızlık faktörleri, öğrenci performansı ve gelişimi, eğitim-öğretim ortamlarının iyileştirilmesi gibi konularda eğitim sektörüne yararlı olmaktadır. Bu teknolojiler, eğitimde kişiselleştirilmiş öğrenme yaklaşımlarını destekleyerek, öğrencilerin ihtiyaçlarına daha iyi yanıt verilmesini sağlamaktadır.

## 2.2. Metin Madenciliği

Metin madenciliği, metin türündeki verilerin işlenerek analiz edilmesini sağlayan bir tekniktir. Metin madenciliği, Veri Madenciliğinin bir alt alanı olarak kabul edilir. Veri Madenciliğinde olduğu gibi, metin madenciliği de anlamsız veriden anlamlı bilgi çıkarımı yapmayı amaçlar. Metin madenciliği, metin verilerinin içerdiği bilgileri ortaya çıkarmak için çeşitli yöntemler kullanır. Bu yöntemler arasında metin sınıflandırma, kümeleme, özetleme, Duygu Analizi, konu modelleme ve ilişki çıkarımı sayılabilir (Yaman, 2022:18).

Metin madenciliği, doğal dil işleme, Makine Öğrenimi ve istatistik gibi alanları kullanan, metin verilerinden anlamlı bilgiler elde etmeyi amaçlayan bir Veri Madenciliği tekniğidir. Metin madenciliği, metin verilerinin ön işleme, analiz ve görselleştirme gibi aşamalardan geçirilerek, metinlerde gizli olan kalıpların, eğilimlerin, ilişkilerin ve bilgilerin keşfedilmesini sağlar. Metin madenciliği, 2000'li yıllarda internetin yaygınlaşması, sosyal medyanın ortaya çıkması ve büyük verinin öneminin artmasıyla birlikte daha fazla popülerlik kazanmıştır (Yıldız ve Ağdeniz, 2018:291).

Metin madenciliği, büyük miktarda metin verisinden bilgi çıkarmak için kullanılan bir tekniktir. Metin madenciliğinin çalışma alanları çok geniştir ve birçok uygulama alanı vardır (Şeker, 2015:32). Bazı çalışma alanları şunlardır:

- **Enformasyon getirimi (information retrieval):** Kullanıcının ihtiyacına uygun olan metin belgelerini bulmak için kullanılan bir süreçtir. Enformasyon Getirimi, arama motorları, web madenciliği, metin sınıflandırma ve özetleme gibi uygulamalarda kullanılır. Enformasyon Getirimi, metin belgelerini temsil etmek, benzerlik ölçmek ve sorgu ile eşleşen belgeleri sıralamak gibi adımları içerir.
- **Doğal dil işleme (natural language processing):** İnsanların kullandığı doğal dilleri anlamak ve işlemek için kullanılan bir yöntemdir. Doğal Dil İşleme, metin madenciliğinin önemli bir aşamasıdır. Doğal Dil İşleme, metinleri dilbilgisine göre analiz etmek, anlam çıkarmak, çevirmek ve üretmek gibi işlemleri gerçekleştirir. Doğal Dil İşleme, ses tanıma, makine çevirisi, sohbet robotları ve doğal dil üretimi gibi uygulamalarda kullanılır.

- **Adlandırılmış varlık tanıma (named entity recognition):** Metin içindeki isimlendirilmiş varlıkları (kişi, yer, kurum, tarih, saat vb.) tanımak ve sınıflandırmak için kullanılan bir süreçtir. Adlandırılmış Varlık Tanıma, metinlerden bilgi çıkarmak için önemli bir adımdır. Adlandırılmış Varlık Tanıma, istatistiksel yöntemler, Makine Öğrenimi algoritmaları veya derin öğrenme modelleri gibi çeşitli teknikler kullanarak gerçekleştirilebilir. Metinlerdeki bilgileri zenginleştirmek, ilişkileri bulmak ve bilgi tabanları oluşturmak için kullanılır. Örneğin, bir haber metnindeki "Cumhurbaşkanı", "Ankara", "Türkiye" gibi ifadeler Adlandırılmış Varlık olarak tanınabilir ve sırasıyla "Kişi", "Yer" ve "Ülke" olarak sınıflandırılabilir. Adlandırılmış Varlık Tanıma, bilgi getirme, bilgi çıkarma, soru-cevap sistemleri ve biyomedikal metin madenciliği gibi uygulamalarda kullanılır.
- **Örüntüsü tanımlı varlıkların bulunması (pattern identified entities):** Metin içindeki belirli bir örüntüye sahip varlıkları (telefon numarası, e-posta adresi, plaka vb.) tanımak için kullanılan bir süreçtir. Örüntüsü Tanımlı Varlıkların Bulunması, düzenli ifadeler (regular expressions) gibi basit yöntemler veya daha gelişmiş yöntemler kullanarak gerçekleştirilebilir. Metinlerdeki kişisel veya hassas bilgileri tespit etmek, filtrelemek veya anonimleştirmek, veri güvenliği, veri temizliği ve veri entegrasyonu gibi uygulamalarda kullanılır.
- **Eş atıf (coreference):** Metin içindeki farklı ifadelerin aynı varlığı veya kavramı gösterip göstermediğini belirlemek için kullanılan bir yöntemdir. Eş Atıf, metin madenciliğinde anlam çözümlemesinin önemli bir parçasıdır. Metinlerdeki bağlamları anlamak, tutarlılık sağlamak, metin üretimi, metin özetleme ve metin zenginleştirme gibi uygulamalarda kullanılır.
- **İlişki, kural ve olay çıkarımı:** Metin madenciliğinde ilişki, kural ve olay çıkarımı, metinlerdeki anlamsal bilgileri otomatik olarak tanımlama ve çıkarma işlemidir. Bu işlem, metinlerdeki özneler, nesnelere, eylemler, özellikler ve ilişkiler gibi unsurları belirlemek ve bunların arasındaki mantıksal bağlantıları kurmak için doğal dil işleme (NLP) teknikleri kullanır. İlişki çıkarımı, metinlerdeki iki veya daha fazla unsur arasındaki anlamsal ilişkileri tanımlar. Örneğin, "Ali kitabı okudu" cümlesinde Ali ile kitap arasında bir sahiplik ilişkisi olduğu çıkarılabilir. Kural çıkarımı, metinlerdeki unsurların genel veya özel kurallara uygunluğunu veya uymazlığını belirler. Örneğin, "Ali kitabı okudu" cümlesinde Ali'nin okuma eyleminin bir kurala

uygun olduğu varsayılabilir. Olay çıkarımı, metinlerdeki unsurların bir olayın parçası olup olmadığını veya bir olayın sonucu olup olmadığını saptar. Örneğin, "Ali kitabı okudu" cümlesinde Ali'nin okuma eyleminin bir olayın parçası olduğu veya bir olayın sonucu olduğu belirlenebilir. Metin madenciliğinde ilişki, kural ve olay çıkarımı, metinlerden bilgi elde etmek ve anlamak için önemli bir yöntemdir.

- **Duygu analizi (sentiment analysis):** Metnin yazarının veya konuşmacının duygusunu veya tutumunu tanımak ve sınıflandırmak için kullanılan bir süreçtir. Duygu Analizi, metinleri genellikle olumlu, olumsuz veya nötr olarak sınıflandırmak için kullanılır. Ancak, daha karmaşık duygu ölçekleri de mevcuttur. Örneğin, metinlerin mutluluk, üzüntü, kızgınlık, korku, nefret veya sevgi gibi belirli duygu türlerini içerip içermediğini belirlemek mümkündür. Duygu Analizi, sosyal medya, müşteri hizmetleri, pazarlama, ürün incelemeleri, eğitim ve sağlık gibi çeşitli uygulama alanlarında önemli bir rol oynamaktadır.

Metin madenciliği, çeşitli kaynaklardan elde edilen belgeler üzerinden bilgi çıkarma işlemidir. Metin madenciliği, her bir belgeyi önceden işlemek için alır ve onun biçimini ve karakter kümesini doğrular. Ardından belge, metinden anlamlı bilgi elde etmek için anlamsal analiz yapan bir metin analizi aşamasına girer. Metin analizi için farklı teknikler kullanılabilir. Bazı durumlarda metin analiz teknikleri, istenen bilgiye ulaşana kadar tekrar edilir. Elde edilen bilgi, bir yönetim bilgi sistemine aktarılır ve bu sistemden kullanıcılar için değerli bilgiler sağlanır.

Metin madenciliği; veri toplama, veri ön işleme, metin madenciliği teknikleri, metin analizi ve bilginin keşfi süreçlerinden oluşmaktadır. Bu süreç Şekil 2.3'de gösterilmiştir (Yaman, 2022:17).



**Şekil 2.3.** Metin madenciliği süreci (Yaman, 2022:18).

Metin Madenciliği süreci, aşağıdaki adımları içerir (Yaman, 2022:18);

- **Veri toplama:** Bu aşamada, metin madenciliği yapılacak veri kaynakları belirlenir ve veriler toplanır. Veri kaynakları, web sayfaları, sosyal medya, haberler, raporlar, kitaplar, makaleler gibi farklı türlerde olabilir. Veriler, dosya formatlarına, boyutlarına ve yapısına göre farklı yöntemlerle toplanabilir.
- **Veri ön işleme:** Bu aşamada, toplanan veriler metin madenciliği için uygun hale getirilir. Veri ön işleme işlemleri arasında, metin temizleme, dil tanıma, sözcük köküne indirgeme, sözcük çantası oluşturma, terim ağırlıkları hesaplama gibi adımlar bulunur. Bu işlemler, verilerin gürültüden arındırılmasını, standartlaştırılmasını ve vektörel bir forma dönüştürülmesini sağlar.
- **Metin madenciliği teknikleri:** Bu aşamada, ön işlenmiş veriler üzerinde metin madenciliği teknikleri uygulanır. Metin madenciliği teknikleri, verilerden bilgi çıkarmak için kullanılan algoritmalarıdır. Metin madenciliği teknikleri arasında, sınıflandırma, kümeleme, öbekleme, ilişki analizi, Duygu Analizi, konu modelleme gibi yöntemler bulunur. Bu yöntemler, verileri farklı kriterlere göre gruplandırmak, benzerlikleri ve farklılıkları bulmak, ilişkileri ve desenleri ortaya çıkarmak, duyguları ve konuları belirlemek gibi amaçlarla kullanılır.
- **Metin analizi:** Bu aşamada, metin madenciliği teknikleri ile elde edilen sonuçlar analiz edilir. Metin analizi işlemleri arasında, sonuçların yorumlanması, değerlendirilmesi, görselleştirilmesi ve raporlanması gibi adımlar bulunur. Bu işlemler, sonuçların anlamlandırılmasını, doğruluğunun ve güvenilirliğinin ölçülmesini, görsel olarak sunulmasını ve paylaşılmasını sağlar.
- **Bilginin keşfi:** Bu aşamada, metin analizi ile elde edilen bilgiler kullanılır. Bilginin keşfi işlemleri arasında, bilgilerin karar verme süreçlerine katkısı, bilgilerin yeni sorulara cevap olması, bilgilerin yeni fikirler üretmesi gibi adımlar bulunur. Bu işlemler, bilgilerin değerinin artırılmasını ve yeni bilgilerin üretilmesini sağlar.

### 3. LİTERATÜR

İnternet ortamında paylaşılan kullanıcı yorumlarının analizi için metin madenciliği tekniklerinin kullanıldığı ulusal ve uluslararası birçok çalışma bulunmaktadır.

#### 3.1. Ulusal Literatür

Ergün (2012), tüketicilerin ürün seçiminde karar vermesine yardımcı olmak amacıyla, www.hepsiburada.com web sitesinden alınan ürün yorumlarından özet bilgiler üretmeyi amaçlamıştır. Bu amaçla, yorumlardaki kelimeleri morfolojik analiz ile işleyerek sözcük türlerini ve eklerini belirlemiştir. Ürün özelliklerini ve bunların değerlendirmelerini içeren kelimeleri önceden tanımlamış, yorumlardaki anlam bilgisini çıkarmak için Türkçe dilbilgisi kurallarına göre cümle Ağacı oluşturmuştur. Bu ağaçlarda Derinlik Öncelikli Arama algoritmasını kullanarak istenilen özellikleri tespit etmiştir. Bu işlemi, Java programlama dili ile geliştirmiştir. Zemberek doğal dil işleme kütüphanesini de bu yazılımda kullanmıştır. Yazılımın geliştirilmesinde NetBeans ortamını tercih etmiştir. Çıkarılan bilgileri SQL veri tabanında saklamıştır. Veri tabanındaki bilgileri sorgulayarak tüketicilerin ürünler hakkındaki memnuniyet oranlarını sayısal olarak elde etmiştir (Ergün, 2012).

Çeliksi (2017), seçilen yabancı dizilerin Türkçe altyazılarını ve twitter yorumlarını, açık kaynak R programını kullanarak metin madenciliği yöntemleriyle analiz etmeyi amaçlamıştır. Çalışmasında, dizi türleri ve senaryoları arasındaki ilişkiyi incelemiş ve dizi türlerinin senaryo içeriklerine nasıl yansıdığını ortaya koymuştur. Araştırmanın bulgularına göre, aksiyon ve komedi dizileri türlerine uygun senaryolarla hazırlanırken, cinsel içerikli dizilerin saptırıcı ve kışkırtıcı senaryolarla desteklendiği tespit edilmiştir (Çeliksi, 2017).

Sönmez (2017), İstanbul'daki alışveriş merkezlerinin popülerliklerini ve konuşulan konuları belirlemek için Foursquare adlı sosyal medya platformundan yararlanmıştır. Foursquare aracılığıyla alışveriş merkezlerine bırakılan yorumları, check-in sayıları, kullanıcı sayıları, puanlar gibi verileri toplamıştır. Bu verilerin analizi için metin madenciliği yöntemlerinden Gizli Dirichlet Tahsisi (LDA) kullanmıştır. LDA ile yorumlardan 6 farklı konu çıkarmış ve bu konulara etiketler vermiştir. Ayrıca her bir alışveriş merkezinin hangi konuya ne kadar yakın olduğunu hesaplamıştır. Sonuç olarak, alışveriş merkezlerinin en çok Genel AVM

Atmosferi, Otopark, Kitle-Yoğunluk, Yiyecek-İçecek Hizmetleri, AVM Konumu ve Eğlence-Ek Hizmetler konularında değerlendirildiğini bildirmiştir (Sönmez, 2017).

Işık (2019), sosyal medya yorumlarının Duygu Analizini yapmak için denetimli öğrenme yöntemlerini kullanmıştır. Denetimli öğrenme yöntemleri arasında Naive Bayes, Sıralı Minimal Optimizasyon (SMO) ve k-En Yakın Komşu algoritmalarını seçmiştir. Veri kümesi olarak, Twitter üzerinden bazı e-ticaret firmalarına ait ürün veya hizmet yorumlarını toplamıştır. Yorumları olumlu, olumsuz ve nötr olmak üzere üç sınıfa manuel olarak etiketlemiştir. Çalışmasında, veri kümesinin sınıflara dağılımının ve öznitelik seçiminin sınıflandırma performansına etkisini araştırmıştır. Bu amaçla, Weka 3.8 yazılımını kullanarak 16 farklı model oluşturmuş ve karşılaştırmıştır. Deneysel sonuçlar, dengesiz veri kümesinin dengeli veri kümesinden daha yüksek doğruluk oranına sahip olduğunu göstermiştir. Ayrıca, öznitelik seçimi yapılan veri kümelerinin de daha iyi sonuçlar verdiğini tespit etmiştir. En yüksek sınıflandırma doğruluğunu %93,52 ile dengesiz veri kümesinde öznitelik seçimi yapılan kNN algoritması ile elde etmiştir (Işık, 2019).

Tuna (2019), bir çevrimiçi rezervasyon sitesinde yer alan Antalya bölgesindeki otellere ilişkin şikâyet ve yorumları metin madenciliği tekniklerinden Duygu Analizi ile incelemiştir. Çalışmasında, açık kaynak kodlu Python yazılımı ile Spyder ve Anaconda platformlarından faydalanmış, 7 farklı algoritma kullanarak yapılan yorumların ve beğeni derecelerinin uyumluluğu arasındaki doğruluğu ortaya koymuştur. Bunların içerisinde ise en başarılı yöntemin Lojistik Regresyon (%87,9) olduğunu tespit etmiştir. Okuyuculara metinsel ifadelerin konaklama işletmeleri açısından başarılı bir şekilde çözümlenebileceğini kanıtlamıştır (Tuna, 2019).

Yılmaz (2019), spam (sahte yorum) kullanıcı yorumlarını tespit etmek için DocRep, NodeRep ve SPR2EP olarak üç farklı yaklaşımı karşılaştırmıştır. Bu yaklaşımlar, yorum metinlerinin içeriği ve sosyal ağlardaki kullanıcı-ürün ilişkileri arasındaki bağlantıları dikkate almıştır. Bu bilgileri birleştirerek, ürün, kullanıcı ve her yorum için öznitelik vektörleri oluşturmuştur. Bu vektörleri, Makine Öğrenmesi algoritmalarıyla sınıflandırarak, yorumların spam olup olmadığını belirlemiştir. Çalışmasında, hem yorum metinlerinden hem de sosyal ağ bağlantı bilgilerinden faydalanmıştır. Sonuç olarak sosyal ağ bilgisine daha fazla ağırlık veren SPR2EP yaklaşımının diğer iki yaklaşımdan daha yüksek doğruluk oranına sahip olduğunu göstermiştir (Yılmaz, 2019).

Atabay (2020), seyahat yorum sitelerindeki kullanıcı yorumlarını içerik analizi ve Veri Madenciliği teknikleriyle değerlendirmiştir. Çalışmasında, her şey dahil hizmet sunan Antalya, Mayorka ve Şarm El Şeyh otellerini ziyaret eden turistlerin otellerin hizmet kalitesi, özellikleri ve duygu durumları hakkındaki görüşlerini ortaya çıkarmayı amaçlamıştır. Çalışmada, bu bölgelerdeki en popüler üç otelin İngilizce dilinde yazılmış 3588 kullanıcı yorumunu analiz etmiştir. Verilerin analizi için R programlama dilini ve ilgili kütüphaneleri kullanmıştır. Çalışmanın sonucunda, her üç bölge ve otel için de benzer tema ve duyguları tespit etmiş, ancak Antalya ve Mayorka arasında daha yüksek bir korelasyon bulmuştur. Şarm El Şeyh otelinin ise diğer iki bölgeden daha fazla olumlu duygu içeren yorumlara sahip olduğunu görmüştür (Atabay, 2020).

Tığa Öztürk (2020), kruvaziyer seyahat deneyimlerinin kruvaziyer firmalarının değerlendirilmesinde nasıl bir rol oynadığını araştırmıştır. CruiseCritic web sitesinde paylaşılan Royal Caribbean International firmasının gemileriyle yapılan seyahatlerin yorumlarını analiz ederek, yorumların genel yıldız derecelendirme değerlerini denetimli Makine Öğrenmesi yöntemleriyle tahmin etmeye çalışmıştır. Tahmin sonuçları ile gerçek değerlendirmeler karşılaştırıldığında, seyahatçilerin firmaya verdikleri puanların, firmanın sunduğu hizmet kalitesiyle ilgili duydukları memnuniyet veya memnuniyetsizliğe bağlı olarak değiştiğini görmüştür. Ayrıca, seyahatçilerin firmayı değerlendirirken en çok dikkate aldıkları özelliklerin "kıyı gezileri", "gemi deneyimleri", "hizmet", "yiyecekler", "eğlence", "kabin", "ortak kullanım alanları", "fiyat-değer" ve "gemiye binme" olduğunu, ancak "zenginleştirilmiş aktiviteler" ve "fitness ve rekreasyon" özelliklerinin puanlamada etkili olmadığını tespit etmiştir (Tığa Öztürk, 2020).

Aktaş, Coskuner ve Soner (2021), Yemek Sepeti platformundan elde edilen verileri ve bu verilerin farklı şekillerde işlenmesini kullanarak, çeşitli Makine Öğrenmesi algoritmalarının performansını karşılaştırmışlardır. Çalışmalarında, veriler üzerinde doğal dil işleme teknikleri uygulanmış ve farklı algoritmaların doğruluk oranları hesaplanmıştır. Ayrıca, algoritmaların parametre ayarları da en iyi sonucu elde etmek için optimize edilmiştir. Bu çalışmanın sonucunda, etiketli verilerle eğitilen modellerin etiketsiz verilerde de kullanılabileceği ve müşteri memnuniyeti analizi için yararlı olabileceği sonucuna varılmıştır. Doğal dil işleme tekniklerinin uygulanmasının, modellerin doğruluk oranlarını yaklaşık %5 artırdığı görülmüştür (Aktaş ve diğerleri, 2021).

Budak (2021), rekabete dayalı pazarda hava yolu firmalarının güçlü ve zayıf yönlerini belirlemek için veri ve metin madenciliği yöntemlerini kullanmıştır. Araştırma verisi olarak, TripAdvisor adlı sosyal medya sitesinde yer alan çevrimiçi seyahat yorumlarını seçmiştir. Star Alliance üyesi 26 hava yolu firmasını bu veriler ışığında değerlendirmiştir. Değerlendirme kriterlerini, kullanıcıların yaptığı yorum ve verdikleri skorlara dayanarak oluşturmuştur. Kullanıcı yorumlarından duygu polaritesini ve polarite değerlerini Duygu Analizi ile tespit etmiş, Destek Vektör Makineleri, Naive Bayes ve Derin Öğrenme algoritmaları ile sınıflandırma ve tahminleme yapmıştır. Elde ettiği sonuçlar arasından Derin Öğrenme algoritmasının daha başarılı olduğunu görmüştür. Ayrıca, Karar Ağacı ile hava yolu firmalarının güzergâh, uçuş noktası, kabin sınıfı ve duygu polaritesine göre kriterlerdeki performanslarını ölçmüş ve yorumlamıştır. Sonuç olarak, müşteri yorum ve puanları bakımından en iyi hava yolunu EVA olarak saptamıştır (Budak, 2021).

Öztürk (2021), sosyal medya yorumlarının duygu durumunu belirlemek için metin madenciliği tekniklerini kullanan bir denetimli öğrenme modeli sunmuştur. Model, Naive Bayes, SMO ve J48 gibi farklı sınıflandırma algoritmalarını karşılaştırmıştır. Veri seti olarak, Knime programı ile Twitter'dan toplanan etkileşimli tahta kullanımına ilişkin yorumları kullanmıştır. Bu yorumları Knime programı ile toplamış, manuel olarak olumlu, olumsuz ve nötr olarak etiketlemiştir. Öznitelik seçimi yapılan ve yapılmayan verileri, Weka 3.8 programı ile analiz etmiştir. Dengeli ve dengesiz veri kümeleri için 12 farklı model oluşturmuş ve performanslarını değerlendirmiştir (Öztürk, 2021).

Yılmaz (2021) çalışmasında, e-ticaret sistemlerinde satılan ürünler hakkında yapılan yorumların metin madenciliği teknikleri ile analiz ederek türlerine göre ayrılmasını amaçlamıştır. E-ticaret sistemi olarak OpenCart altyapısından yararlanmış ve "ecommercecomments.store" alan adı üzerinden bir web sitesi oluşturmuştur. Web sitesinin kurulum aşamalarını detaylı bir şekilde anlatmıştır. Çalışmasında, ürün değerlendirmelerinin metin madenciliği ile nasıl sınıflandırılabileceğini göstermiştir. Çalışmanın sonucunda, metin madenciliği ile işlenen verilerin beklenen sonuçlara ulaştığını görmüştür. Sınıflandırma işlemi sonunda, güvenilirlik değeri yüksek olan metinleri doğru kategoriler altında listelemiştir (Yılmaz, 2021).

Akgöz ve Tengilimođlu (2022), booking.com'da yer alan 5 yıldızlı Bakü otellerinin müşteri memnuniyetini metin madenciliđi teknikleri ile deđerlendirmeyi amaçlamışlardır. Bu amaçla, 3.275 adet yorum veri ön işleme aşamalarından geçirek kelime sıklık ve ađırlıklarını belirlemişlerdir. Ayrıca, birliktelik analizleri ile yorumlarda birlikte anılan terimleri ortaya çıkarmışlardır. Çalışmalarının bulgularına göre, yorumlarda en çok personel, oda ve otel terimleri kullanılmıştır. Bu terimleri, müşterilerin otelleri deđerlendirirken en çok dikkat ettikleri unsurlar olarak yorumlamışlardır (Akgöz ve Tengilimođlu, 2022).

Çakmak (2022), platform ekonomilerinin oluşturacağı yeni ekonomik çevreyi anlamak için sosyal medyada platform ekonomisi kullanıcı yorumlarının metin madenciliđi tekniđi ile analiz edilmesini ve bu analizlerin yeni ekonomik çevrenin etkilerini ortaya koymasını hedeflemiştir. Çalışmanın yöntemini, araştırma amacına uygun olarak seçtiđi on adet konu etiketi üzerine kurmuştur. Araştırmada kullandığı veriler, herkesin erişebileceđi sosyal medya platformu Twitter'dan Twitter Developer aracılıđıyla Social Media Research (SMR) Foundation ve "TAGS" programından izin alarak elde etmiş ve sınıflandırmıştır. Araştırmasında belirlediđi e-ticaret, indirim, online alışveriş, fırsat, blackfriday, novemberdeals, karacuma, muhteşemcuma, efsanekasım, kasımindirim konu etiketlerini içeren tweetler, 14 Kasım - 5 Aralık 2021 ve 01-20 Ekim 2022 tarihleri arasındaki dönemi kapsamıştır. Yaptığı analizler sonucunda, tweetlerde belli başlı firmaların adlarının sıkça geçtiđi ve kullanıcıların platformlarla ilgili birtakım sorunlar yaşadıklarını görmüştür (Çakmak, 2022).

Yaman (2022), Türkçe müşteri yorumlarının Duygu Analizi için yeni bir yöntem sunmuştur. Yöntem, metin madenciliđi ve dođal dil işleme tekniklerini kullanarak yorumları olumlu, olumsuz ve tarafsız olarak etiketlemiştir. Yöntemin temelini, Türkçe için özel olarak geliştirilen sözlük tabanlı bir puanlama sistemi oluşturmuştur. Sözlük Tabanlı Puanlama Sistemi (STPS) algoritmasını bir e-ticaret sitesinden topladıđı müşteri yorumları üzerinde uygulamıştır. Algoritmanın sonuçlarını, beş farklı Makine Öğrenmesi yöntemiyle çeşitli performans ölçütleri açısından deđerlendirmiştir. STPS algoritmasının Duygu Analizinde yüksek başarı oranlarına ulaştığını görmüştür (Yaman, 2022).

Çabuk, Yücalar ve Toçođlu (2023), e-ticaret platformlarında yer alan ürün yorumlarının Duygu Analizine odaklanmışlardır. Çalışmalarında, Türkiye'deki e-ticaret sitelerinden 15170 adet ürün yorumu toplanarak olumlu ve olumsuz olarak iki kez etiketlenmiştir. Bu

veri seti kullanılarak Uzun-Kısa Süreli Bellek algoritması ile bir model oluşturulmuş ve bu model ile 203274 adet yeni yorum otomatik olarak sınıflandırılmıştır. Çalışmanın devamında, otomatik olarak sınıflandırılan veri setinin doğruluğu dört farklı derin öğrenme algoritması ile test edilmiştir. Son olarak, otomatik olarak sınıflandırılan veri seti ile dört geleneksel Makine Öğrenmesi ve dört derin öğrenme algoritmasının performansları karşılaştırılmıştır. Elde edilen sonuçlar, derin öğrenme algoritmalarının Makine Öğrenmesi algoritmalarından daha üstün olduğunu ve otomatik olarak sınıflandırılan veri setinin yüksek doğruluk sağladığını göstermiştir (Çabuk ve diğerleri, 2023).

Daşgın (2023), Kitlesele Açık Çevrimiçi Dersler (KAÇD) platformlarından biri olan Udeemy sitesindeki kurslara ilişkin yorumların olumlu ya da olumsuz olarak sınıflandırılmasını amaçlamıştır. Böylece kullanıcılar, yorumları okumadan kursların kalitesi hakkında fikir sahibi olabileceklerdir. Bu amaçla, klasik Makine Öğrenmesi ve derin öğrenme yöntemlerini kullanarak yorumlara Duygu Analizini uygulamıştır. Klasik Makine Öğrenmesi yöntemleri arasında BayesNet, J48 ve One Rule (OneR) algoritmalarını test etmiş ve BayesNet algoritması %91.576 doğruluk oranıyla en iyi performansı göstermiştir. Derin öğrenme yöntemleri arasında ise, veri setine Random, GloVe ve Word2Vec kelime gömme tekniklerini uygulayarak Kapılı Tekrarlayan Birim (GRU) ve Evrişimsel Sinir Ağları (CNN) ve Uzun-Kısa Süreli Hafıza (LSTM) hibrit mimarilerini denemiştir. Bu yöntemlerden en yüksek doğruluk oranını %95.67 ile GloVe kelime gömme tekniği kullanılan GRU mimarisinden elde etmiştir (Daşgın, 2023).

Kapukaya (2023), Internet Movie Database (IMDb) "En iyi 250 Filmler" listesindeki ilk üç film olan Esaretin Bedeli, Baba ve Kara Şövalye filmlerinin Twitter gerçek kişi kullanıcılarının yorumlarını R programlama dili ile analiz etmeyi amaçlamıştır. Bu filmlerin IMDb puanlamasına göre yüksek derecelendirilmesinin, Twitter kullanıcılarının duygu değerleri ile uyumlu olup olmadığını araştırmıştır. Bu amaçla, Twitter platformundan yapılandırılmamış verileri çekmiş, R programı ile metin ön temizleme işlemlerini uygulamış, meta veriye dönüştürmüş ve metin madenciliği yöntemlerinden Terim Ağırlığı (TF-IDF), N-gram modelleme ve Duygu Analizini kullanmıştır. Her bir film için kelime bulutları oluşturarak kullanıcı yorumlarının Duygu Analizini yapmıştır. Sonuç olarak, ilgili filmler hakkında atılan tweet metinlerinde pozitif duygu kelimelerinin baskın olduğunu görmüştür (Kapukaya, 2023).

Sayın Okatan (2023), banka müşterilerinin dijital bankacılık hizmetleri ile ilgili sosyal medya platformlarında yaptıkları yorumları metin madenciliği yöntemleri ile incelemeyi amaçlamıştır. Bu sayede, Türkiye'deki bankaların dijital bankacılıkta müşteri beklentilerini ne derecede karşıladıklarını ve bu alanda nasıl iyileştirmeler yapabileceklerini ortaya koymuştur. Çalışmasında, Türkiye Bankalar Birliği verilerine göre en çok kullanılan 3 devlet bankası ve 10 özel bankanın 2020 Ocak - 2022 Ağustos dönemindeki dijital bankacılık verilerini kullanmıştır. Bankaların kelime yoğunluklarını, wordcloud görsellerini ve Duygu Analizi sonuçlarını incelemiş ve her banka için ayrıntılı yorumlar yapmıştır. Çalışmasındaki bulgularda, banka müşterileri dijital bankacılıkta en çok; kullanım kolaylığı, hizmet kalitesi, ücretlendirme, mobil uygulamanın erişilebilirliği ve hızı gibi konulara değinmiştir. Dijital bankacılıkta özel bankalar ile kamu bankaları arasında büyük farklar olmamakla birlikte, farklılıkların daha çok kullanıcı deneyimine dayalı olduğunu görmüştür (Sayın Okatan, 2023).

### **3.2. Uluslararası Literatür**

Go, Bhayani ve Huang (2009), Twitter mesajlarının Duygu Analizini otomatik olarak yapmak için bir yöntem geliştirmişlerdir. Twitter mesajlarını pozitif ve negatif olmak üzere iki kategoriye ayırmışlar ve uzaktan denetimli öğrenme tekniğini kullanarak bu kategorileri tahmin etmeye çalışmışlardır. Yazarlar, tweetleri unigram, bigram ve unigram ile bigramın birleşimi şeklinde farklı şekillerde temsil etmişler ve bu temsilleri DVM, Naive Bayes ve Maximum Entropi gibi farklı sınıflandırma algoritmalarıyla test etmişlerdir. Elde ettikleri sonuçlara göre, unigram ile bigramın birleşimi temsiliyle %83 doğruluk oranına ulaşan Maximum Entropi algoritması en iyi performansı sergilemiştir (Go ve diğerleri, 2009).

Zhang, Ye, Zhang ve Li (2011), Naive Bayes ve DVM gibi geleneksel Makine Öğrenimi yöntemlerini, internet üzerinden Kantonca yazılmış restoran yorumlarının otomatik olarak pozitif veya negatif olarak etiketlemek için uygulamıştır. Çalışmalarında, farklı özellik gösterimleri ve özellik boyutlarının sınıflandırma başarımı üzerindeki etkisi incelenmiştir. Sınıflandırma modelleri ile özellik seçimleri arasındaki etkileşimin doğruluğu belirlediği görülmüştür. Naive Bayes sınıflandırıcısı, DVM ile karşılaştırıldığında daha yüksek veya eşit doğruluk sağlamıştır. Karakter bazlı bigramlar, Kantonca Duygu Analizinde unigram ve trigramlardan daha iyi özellikler olduğunu göstermiştir (Zhang ve diğerleri, 2011).

Lei ve Law (2015), Macau bölgesindeki restoranlarda yemek deneyimini değerlendiren bir web sitesini incelemiştir. Çalışmada, TripAdvisor'dan alınan çevrimiçi tüketici yorumları, içerik analizi ve sınıflandırma yöntemleriyle değerlendirilmiştir. Olumlu, olumsuz ve nötr yorumların frekans dağılımları hesaplanmıştır. Araştırmanın bulguları, Macau'daki yemek deneyimlerinin genel olarak müşteriler tarafından beğenildiğini ortaya koymuştur. Ancak lüks restoranlar, daha ekonomik restoranlara göre daha fazla olumsuz yorum almıştır. İçerik analizi, tüketicilerin yemek deneyimlerini etkileyen faktörleri belirlemiş ve hizmet sektörüne yönelik öneriler sunmuştur (Lei ve Law, 2015).

Felbermayr ve Nanopoulos (2016), çevrimiçi yorumlardan duygu içeriğini çıkarmak için bir yöntem geliştirmişlerdir. Çalışmalarında, duygu sözlüğü kullanarak yorumların kalitesine göre duygu boyutlarının önemini belirlemişlerdir. Çalışmalarında, güven, neşe ve beklenti duygu boyutlarının en önemli kategoriler olduğunu bulmuşlardır. Ayrıca, kalabalık kaynaklı duygu sözlüğünün sınıflandırma doğruluğu açısından daha iyi performans gösterdiğini tespit etmişlerdir. Bu çalışma ile pazarlamacılara çevrimiçi yorum analizinde duygusal bir bakış açısı sunmayı amaçlamışlardır (Felbermayr ve Nanopoulos, 2016).

Sparks, So ve Bradley (2016), çevrimiçi bir olumsuz yorum ve otel tarafından verilen yanıtın potansiyel müşterilerin işletme hakkındaki algılarını nasıl etkilediğini araştırmışlardır. Çalışma, olumsuz yorumlara verilen yanıtların varlığının (veya yokluğunun) ve yanıtların niteliğinin (yanıtlayanın kimliği, sesi, hızı ve eylem planı) müşterilerin işletmeye duyduğu güven ve endişe düzeylerini nasıl değiştirdiğini incelemektedir. Araştırma sonuçları, çevrimiçi bir yanıtın sunulmasının (yanıtın olmamasına göre) işletmenin güvenilirliği ve müşteri odaklılığı hakkında olumlu izlenimler bıraktığını göstermiştir. Çalışma, olumsuz çevrimiçi yorumlara etkili bir şekilde nasıl yanıt verilebileceği konusunda öneriler sunmuştur (Sparks ve diğerleri, 2016).

Guo, Barnes ve Jia (2017), otel ziyaretçilerinin müşteri hizmetleriyle ilgili beklenti ve deneyimlerini analiz etmek için gizli dirichlet analizi (LDA) adlı Veri Madenciliği yöntemini kullanmışlardır. Çalışmalarında, 16 ülkede yer alan 25.670 otel hakkında 266.544 çevrimiçi yorum incelenmiştir. LDA sonucunda, müşteri hizmetleriyle ilgili 19 farklı boyut belirlenmiştir. Çalışmada ayrıca, demografik özelliklere göre boyutların önem derecelerinde farklılıklar olduğu gözlemlenmiştir. Algısal haritalama tekniğiyle, otellerin yıldız sayısına göre hangi boyutların daha öne çıktığı ortaya konmuştur. Çalışmalarında, otel yöneticilerine

ve arařtırmacılara müşteri hizmetlerinin iyileřtirilmesi için öneriler sunmuřlardır (Guo ve diđerleri, 2017).

Lohman (2017), biliřim teknolojileri ve moda sektöründe faaliyet gösteren beřer firmaya ait hisse senedi yatırımcılarının Twitter üzerindeki duygularını incelemiřtir. Twitter aracılıđıyla elde edilen tweetleri, Naive Bayes algoritması kullanılarak Duygu Analizine tabi tutmuřtur. Duygu Analizi sonuçları, piyasa kapitalizasyonu esas alınarak oluřturulan bir portföy stratejisine uyarlanmış ve bu stratejinin performansı, Sharpe oranı temel alınarak oluřturulan bařka bir portföy stratejisiyle karřılařtırılmıřtır. Çalışmanın bulguları, Duygu Analizinin biliřim teknolojileri firmalarına yatırım yaparken daha iyi bir performans sağladığını göstermiřtir (Lohman, 2017).

Barushka ve Hajek (2019), spam tespitine yönelik bir yaklařım sunmuřlardır. Yaklařım, Bag of Words ve n-grams gibi metin madenciliđi tekniklerini birleřtirmektedir. Yaklařımın etkinliđini göstermek için iki farklı otel hakkında yapılan olumlu ve olumsuz yorumlar bir veri seti olarak kullanılmıřtır. Yaklařımın sonuçlarını deđerlendirmek için ileri beslemeli bir yapay sinir ađı eđitilmiřtir. Yapay sinir ađı, yorumların polarite deđerlerini tahmin etmek için kullanılmıřtır. Yaklařımın bařarısı, dođruluk, hassasiyet, geri çağırma ve F1 ölçütleri ile ölçülmüřtür (Barushka ve Hajek, 2019).

Chakravarthy, Kameswari, Mydeen ve Seenivasan (2021), Coursera platformundaki Makine Öğrenmesi dersine iliřkin yorumları veri seti olarak kullanmıřlardır. Bu veriler üzerinde metin madenciliđi tekniklerini uygulayarak, yorumların olumlu, olumsuz veya nötr olma durumlarını belirlemiřlerdir. Çalışmalarında, metin madenciliđi için dođal dil işleme yöntemleri kullanılarak, metinler ön işlemden geçirilmifitir. Metin madenciliđi performansının deđerlendirilmesinde ise kNN, Lojistik Regresyon gibi Makine Öğrenimi algoritmaları kullanılmıřtır. Sonuç olarak, kNN algoritması özgüllük ve duyarlılık açısından daha iyi performans gösterirken, Lojistik Regresyon algoritması %94 gibi yüksek bir dođruluk oranına ulařmıřtır (Chakravarthy ve diđerleri, 2021).

Khan, Khan, Shamim, Gupta ve Sherwani (2023), çevrimiçi yemek teslimatları için müşterilerin yorumlarını ve derecelendirmelerini analiz etmek amacıyla bir metin madenciliđi yaklařımını kullanmıřlardır. Arařtırmacılar, çeřitli metin madenciliđi teknikleri ve algoritmalarını kullanarak müşteri geri bildirimlerini incelemiřlerdir. Çalışmalarının

bulguları, müşteri memnuniyeti ve hizmet kalitesi gibi faktörlerin, çevrimiçi yemek teslimat hizmetlerinin başarısını nasıl etkilediğini ortaya koymuştur. Ayrıca, bu analizler sayesinde işletmelerin müşteri ihtiyaçlarını daha iyi anlamaları ve hizmetlerini iyileştirmeleri için öneriler sunmuşlardır (Khan ve diğerleri, 2023).

Zhang, Lopez, Lam, Yenumula ve Vu (2023), çevrimiçi tüketici yorumlarını analiz etmek için kullanılan farklı metin madenciliği yöntemlerini karşılaştırmışlardır. Araştırmacılar, duygu analizi, konu modelleme ve sınıflandırma gibi çeşitli tekniklerin etkinliğini değerlendirmişlerdir. Elde ettikleri bulgular, duygu analizinin tüketici memnuniyetini anlamada en etkili yöntem olduğunu, konu modellemenin ise yorumların ana temalarını belirlemede başarılı olduğunu göstermiştir. Ayrıca, sınıflandırma tekniklerinin doğru ürün kategorilerini belirlemede yüksek doğruluk oranına sahip olduğunu tespit etmişlerdir (Zhang ve diğerleri, 2023).

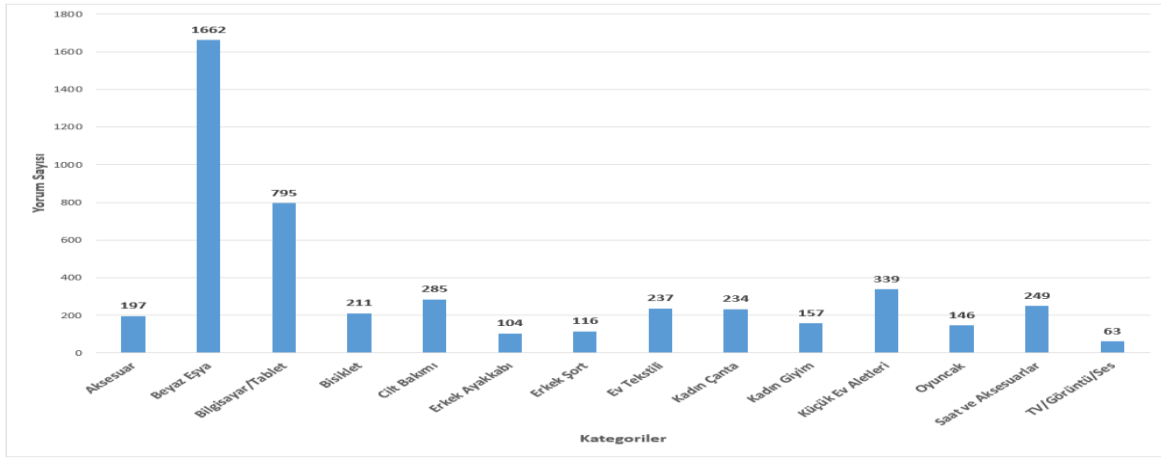
Li, Park ve Kim (2024) çalışmalarında, e-ticarette bilgi tutarlılığının algılanan yararlılığı nasıl etkilediğini araştırmak için metin madenciliği yaklaşımı kullanarak hipotezler geliştirip test etmişlerdir. Amazon.com'dan altı kategoride 600 ürünü içeren 61.900 çevrimiçi yorumu analiz etmişlerdir. Çalışmalarında, derecelendirme tutarsızlığı ve metin benzerliğinin algılanan yararlılığı olumsuz etkilediğini, inceleme tutarlılığının ise olumlu etkilediğini bulmuşlardır. Ayrıca, derecelendirme tutarsızlığının deneyim ürünlerinin incelemelerinde arama ürünlerine göre algılanan yararlılığı olumlu etkilediğini tespit etmişlerdir (Li ve diğerleri, 2024).

Rahmawati, Fuadi ve Afrillia'nın (2024), TikTok Shop'ta ürün değerlendirmeleri için duygu analizi teknikleri kullanmışlardır. Araştırmalarında, TikTok Shop kullanıcılarından 500 yorum toplanmış ve bu yorumlar eğitim ve test verisi olarak ayrılmıştır. Yorumlar, Python'un Selenium modülü ile otomatik olarak toplanmış ve Destek Vektör Makinesi kullanılarak analiz edilmiştir. Test sonuçlarına göre, 121 yorum olumlu, 29 yorum ise olumsuz olarak sınıflandırılmıştır. Sistemin doğruluğu %78, yanlışlık oranı %22 olarak ölçülmüştür. Elde edilen sonuçlar, sistemin e-ticaret uygulamalarında önemli bir potansiyele sahip olduğunu göstermiştir (Rahmawati ve diğerleri, 2024).

## 4. MATERYAL VE METOD

### 4.1. Araştırmada Kullanılan Veri Seti

Online alışveriş sitesi [www.trendyol.com](http://www.trendyol.com) adresinden Şekil 4.1’de görüldüğü gibi 14 farklı kategorideki ürünlere ait tüketici yorumları, Foxpro diliyle hazırlanan yazılım ile toplanarak veri seti oluşturulmuştur. Sonrasında yine Foxpro yazılımı ile tekrar eden yorumlar kaldırılmıştır. Veri seti; kategori, tür, ürün adı, yorum, derece ve ilgi olarak 6 sütun ve 4795 satırdan oluşmaktadır.



Şekil 4.1. Veri setindeki yorumların kategorilere göre dağılımı.

İlk aşamada kullanıcı yorumları ilgili ve ilgisiz olarak etiketlenmiştir. Veri setinde yer alan kayıtların 2683 adedi ilgili, 2112 adedi ilgisiz olarak belirlenmiştir. Veri setinde yer alan kayıtlardan örnekler Çizelge 4.1’de yer almaktadır.

Çizelge 4.1. Veri seti içerisindeki kayıtlardan örnekler

| Yorum Sınıfı     | Yorum Metni  |
|------------------|--|
| İlgili Yorumlar  | Çok kaliteli ekranı var kanallar arası geçiş biraz yavaş sadece. Android özellikli olması telefon gibi kullanmanızı sağlıyor çok iyi.                |
|                  | Güzel nemlendiriyor ve yapış yapış bir his bırakmıyor hemen emiliyor. Kokusu çok hoş kesinlikle tavsiye ederim.                                      |
|                  | Son derece hoş ve şık. Yumuşak ve duruşu gayet güzel. Aşırı soğuk günler dışında rahatlıkla giyilebilir. Karlı, donlu günlerde biraz ince kalabilir. |
|                  | Wi-fi çekim performansı çok kötü ses ekran pil iyi ancak hızlı şarj olmaması kötü ram ve hafıza yeterli performans orta üstü.                        |
| İlgisiz yorumlar | Henüz kullanmadım ama paketleme güzeldi.   |
|                  | Yanlış ürün geldi, iade ile uğraşmak istemedim.  |
|                  | Arkadaşımla doğum günü hediyesi aldım.   |
|                  | Sorunsuz teslimat, kapıya kadar geldi, ertesi gün servis kurulum için geldi.   |

## 4.2. Arařtırmada Kullanılan Yazılım

Bu alıřmada veri n iřleme, makine ğrenimi, metin madencilięi, tahmin modelleri ve optimizasyon gibi birok alanda gl zmler sunan, veri madencilięi ve makine ğrenimi srelerini kolaylařtırmak iin tasarlanmış RapidMiner programı tercih edilmiřtir.

RapidMiner, Java tabanlı bir yazılım olup kullanıcıların Java kodu eklemelerine de olanak tanımakta ve aynı zamanda Python, Weka ve R gibi dięer popler programlama dilleri ve ortamlarıyla da entegre alıřabilmektedir. Bu ok ynllk, kullanıcıların farklı beceri seviyelerine ve ihtiyalarına uyum saęlayarak, veri analizi projelerinde esneklik ve geniř bir yelpazede fonksiyonellik sunmaktadır (Uysal ve Bilen, 2017:45).

RapidMiner, 2001 yılında YALE (Yet Another Learning Environment) adı altında yapay zeka birimleri tarafından geliřtirilmeye bařlanmış ve zamanla endstri standartlarını karřılayacak řekilde geliřtirilmiřtir. 2006 yılında Rapid-I řirketi tarafından daha da geliřtirilen bu platform, 2013 yılında RapidMiner olarak yeniden markalařmıřtır. RapidMiner, kod yazma ihtiyaını ortadan kaldıran grsel bir programlama ortamı sunmaktadır. Bu zellięi, zellikle programlama bilgisi olmayan veya kod yazmaya zaman ayırmak istemeyen kullanıcılar iin byk bir avantaj saęlamıřtır. Kullanıcılar, srkle bırak mantıęıyla alıřan bu platformda, veri iřleme, modelleme ve analiz iřlemlerini kolaylıkla gerekleřtirebilmektedirler. Ayrıca, RapidMiner'in geniř ktphane ve eklenti desteęi sayesinde, kullanıcılar ihtiya duydukları zellikleri kolayca entegre edebilmekte ve zelleřtirebilmektedirler (zkul, 2021:49).

## 4.3. Arařtırmada Kullanılan Modeller

Bu alıřmada, k-En Yakın Komřu, Naive Bayes, Karar Aęacı, Destek Vektr Makinesi ve Derin ğrenme sınıflandırma algoritmaları kullanılmıřtır. Bu algoritmalar veri setindeki rneklerin hangi sınıfa ait olduęunu belirlemek iin kullanılmıřtır. Her bir algoritmanın kendine has avantajları ve kullanım senaryoları vardır, bu nedenle belirli bir problem iin en uygun olanın seilmesi nemlidir.

### 4.3.1. K-en yakın komşu

Makine Öğrenmesinde k-En Yakın Komşu (kNN) algoritması hem sınıflandırma hem de regresyon yöntemlerinde kullanılan çok yönlü bir algoritmadır. Bu algoritma sınıfı bilinmeyen bir veriyi, veri setindeki diğer veriler ile karşılaştırarak hesaplanan uzaklığa göre veriyi en uygun sınıfa yerleştirir (Dilki, 2020:227). Verilerin boyut ve yapısına bağlı olarak önceden belirlenen k değerine göre sınıflandırma gerçekleştirir (Khan, Ding ve Perrizo, 2002:518).

KNN algoritması, basitliği ve sağlamlığı nedeniyle sınıflandırma problemlerinde sıklıkla tercih edilen bir Makine Öğrenme algoritmasıdır. Eğitime ihtiyaç duymaması, kolay uygulanabilir olması, analitik olarak takip edilebilirliği, yerel bilgilere uyarlanabilirliği, paralel işleme uygunluğu ve gürültülü eğitim verilerine karşı dirençliliği gibi birçok avantajı bulunmaktadır (Bhatia, 2010:302).

KNN algoritmasının birçok avantajı olmasına rağmen, bazı dezavantajları da göz önünde bulundurulmalıdır (Taşçı ve Onan, 2017). Bu dezavantajlar şunlardır:

- **Yüksek bellek ihtiyacı:** KNN algoritması, tüm eğitim verilerini bellekte tutmak zorundadır. Bu durum, büyük veri kümeleri için önemli bir bellek ihtiyacı doğurur.
- **İşlem yükü ve maliyet:** Veri seti ve öznitelik boyutu arttıkça, en yakın komşuları bulmak için gereken işlem yükü ve hesaplama maliyeti de artar. Bu durum, algoritmanın performansını ve maliyetini olumsuz etkileyebilir.
- **Parametrelere bağımlılık:** KNN algoritmasının performansı, k komşu sayısı, uzaklık ölçütü ve öznitelik sayısı gibi parametrelere bağlıdır. Bu parametrelerin optimal değerlerini belirlemek zor olabilir ve performansı önemli ölçüde etkileyebilir.
- **Gürültüye duyarlılık:** KNN algoritması, gürültülü verilerden etkilenebilir. Gürültülü veriler, algoritmanın yanlış sınıflandırmalar yapmasına neden olabilir.

- **Ölçeklendirme sorunu:** KNN algoritması, büyük veri kümeleri için ölçeklendirme sorunu yaşayabilir. Büyük veri kümelerinde, en yakın komşuları bulmak için gereken işlem yükü ve hesaplama maliyeti artar.
- **Yorumlanabilirlik:** KNN algoritması, karmaşık karar verme süreçlerine sahip olabilir. Bu durum, algoritmanın kararlarının yorumlanmasını zorlaştırabilir.

KNN algoritmasında kullanılan uzaklık ölçütü, algoritmanın performansını önemli ölçüde etkileyebilir. Bu nedenle, veri kümesinin özelliklerine ve algoritmanın gereksinimlerine göre en uygun uzaklık ölçütü seçilmelidir. Sayısal veriler için Öklid (Euclidean) uzaklığı veya Minkowski uzaklığı, kategorik veriler için Manhattan uzaklığı veya Hamming uzaklığı, metinsel veriler için Cosine uzaklığı veya Jaccard uzaklığı daha uygun olabilir. Euclidean mesafe hesaplamada en çok kullanılan uzaklık ölçüsüdür (Hu, Huang, Ke ve Tsai, 2016:3). Euclidean uzaklığı, iki nokta arasında,

$$x_1 = (x_{11}, x_{12} \dots x_{1n}) \text{ ve } x_2 = (x_{21}, x_{22} \dots x_{2n}) \quad 1$$

olmak üzere,

$$Dist_{euclidean}(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2} \quad 2$$

formülü ile hesaplanmaktadır.

#### 4.3.2. Naive bayes

Naive Bayes sınıflandırıcı algoritması, metin dokümanlarını sınıflandırmak için kullanılan en başarılı algoritmalarından biridir. Naive Bayes modeli, basitliği, hızı ve yüksek boyutlu verilerle çalışma kabiliyeti nedeniyle popüler bir sınıflandırma yöntemidir (Pilavcılar, 2007:23). Ancak, bağımsızlık varsayımının ihlal edildiği durumlarda performansı düşebilir.

Bayes teoremine dayanan Naive Bayes sınıflandırıcısı, ilgili kayıtların hangi sınıfa hangi olasılıkla ilişkili olduklarını belirlemektedir. Veri setindeki özellikleri birbirinden bağımsız ve tüm özelliklerin aynı düzeyde önemli olduğunu kabul etmektedir. Bayes teoremine göre bir C sınıfının bilinmesi koşulunda x özneliğinin var olma olasılığı ile x özneliğinin bilinmesi koşulunda C sınıfının var olma olasılığı mevcuttur (Sarikoç, 2012:72).

$$P(C).P(x | C) P(C | x) = P(x)$$

Naive Bayes sınıflandırıcısı, temel olarak iki farklı model versiyonu ile çalışır: binary bağımsızlık modeli ve multinominal model. Her iki model de kelimelerin sınıftan bağımsız olduğunu varsayar (Pilavcılar, 2007:23).

- **Binary bağımsızlık modeli:** Bu model, metin dokümanlarını iki farklı sınıfa ayırmak için kullanılır (örneğin, spam ve spam değil). Her kelime için sınıfın olasılığını hesaplar ve bu olasılıkları birbirleriyle karşılaştırarak en yüksek olasılığa sahip sınıfı tahmin eder. Örneğin, bir e-posta içerisinde geçen belirli kelimelerin spam veya spam değil sınıfına ait olma olasılıkları hesaplanır.
- **Multinominal model:** Bu model, metin dokümanlarını birden fazla sınıfa ayırmak için kullanılır. Her kelimenin her sınıfa ait olma olasılığını hesaplar ve bu olasılıkları kullanarak en uygun sınıfı tahmin eder. Örneğin, bir haber makalesinde geçen kelimelerin spor, ekonomi veya siyaset gibi farklı kategorilere ait olma olasılıkları hesaplanır.

Bu algoritma, sınıflandırma işleminde değerleri birbirinden bağımsız olarak ele alır. Sınıfların ve örnek verilerin hangi sınıflara ait olduğu bellidir. Algoritma bir eleman için her durumun olasılığını hesaplar ve olasılık değeri en yüksek olana göre sınıflandırma yapar. Ancak, test kümesindeki bir verinin eğitim kümesinde bir karşılığı mevcut değilse, bu veri için olasılık değeri olarak 0 atanır ve tahmin yapılamaz. Yani, eğitim verilerinde hiç gözlenmemiş bir durumla karşılaşıldığında, bu durumun olasılığı sıfır kabul edilir (Çelik, Dal ve Aydın, 2021:884). Bu çalışmada bu sıfır değerinin yanıltıcı sonuçlara yol açmaması için Naive Bayes modeli Laplace düzeltmesi (laplace correction) ile birlikte kullanılmıştır. Laplace düzeltmesi, bu sorunu önlemek için basit bir yöntemdir ve sıfır değerlerinin oluşmasını önlemek için her sayıya bir eklemektedir.

### 4.3.3. Karar ağacı

Karar Ağacı, sade ve anlaşılır bir yapıya sahip olan, verileri böl ve yönet yöntemi ile sınıflandırma ve regresyon gerçekleştiren bir algoritmadır (Kalaycı, 2018:873). Bu yöntem,

veri setini ağaç yapısı içinde bölerek kararlar alır. Her düğüm, bir özellik üzerinde bir eşik değeri belirler ve veriyi bu eşik değerine göre iki alt küme halinde böler. Bu işlem, ağacın yaprak düğümlerine kadar devam eder. Yaprak düğümlerinde sınıflandırma veya regresyon işlemi gerçekleştirilir.

Karar Ağacı, bölme ve budama olarak iki temel işlem gerçekleştirir ve algoritma durdurma kriterine göre sonlandırılır (Bounsaythip ve Rinta-Runsala, 2001:12). Bölme işlemi, veri kümesini alt kümelerine böler ve bu alt kümeler üzerinde işlem yapmayı sağlar. Budama ise gereksiz dalları keserek ağacın boyutunu azaltır ve daha hızlı sonuçlar elde edilmesini sağlar. Algoritma durdurma kriteri ise genellikle bir hedef değere ulaşıldığında veya belirli bir süre geçtiğinde algoritmayı sonlandırır. Bu sayede verimli ve etkili sonuçlar elde edilir.

Karar Ağacı oluşturmada düğüm ayırma ölçütleri önemli bir rol oynar ve bilgi kazancı, ki-kare istatistiği, kazanç oranı ve GINI indeksi en sık kullanılan ölçütlerdendir (Kothari ve Dong., 2001:171).

- **Bilgi kazancı (information\_gain):** Bulunulan durumun entropi değerini (rastgelelik derecesini) düşüren seçimler yaparak bilgi kazancını maksimize etmeye çalışır. Her soruda (düğüm/node), hata fonksiyonu tekrar hesaplanır ve en düşük hataya sahip soru/durum seçilir. Bilgi kazancı, bir özelliğin bir veri setindeki belirsizliği ne kadar azalttığını ölçer. Daha yüksek bilgi kazancı, daha az belirsizlik ve daha iyi bir bölme anlamına gelir.
- **Ki-kare istatistiği:** Ki-kare istatistiği, bir özelliğin sınıf etiketleri arasındaki dağılım üzerindeki etkisini değerlendirir. Yüksek bir ki-kare değeri, özelliğin ve sınıf etiketinin dağılımı arasında güçlü bir ilişki olduğunu gösterir ve bu da daha etkili bir sınıflandırma performansına işaret eder. Bu istatistiksel ölçüm, özellik seçimi ve model eğitimi sırasında önemli bir rol oynar ve makine öğrenmesi uygulamalarında sıkça kullanılır.
- **Kazanç oranı (gain\_ratio):** Bilgi kazancını düğümün bölünmesi için kullanılan özniteliklerin sayısına göre düzelten bir oran olarak tanımlar. Bilgi kazancının bölünme sayısına göre düzeltilmiş halidir ve aşırı dallanmayı engellemeye yardımcı olabilir.

- **GINI indeksi (gini\_index):** Bir veri setindeki sınıfların eşit dağılımını ölçer. Düğümdeki sınıf dağılımının ne kadar homojen olduğunu ifade eder. Düğümdeki sınıfların eşit dağıldığı durumda GINI indeksi minimum değere sahiptir. Daha düşük Gini indeksi, daha eşit bir dağılım ve daha iyi bir bölme anlamına gelir. Bu indeks, sınıf homojenliği sağlamak için kullanılır.

Karar Ağacı oluştururken hangi ölçütün kullanılacağına karar verirken, veri setinin türü (sınıflandırma veya regresyon), özelliklerin türü (sayısal veya kategorik), veri setinin büyüklüğü, hesaplama karmaşıklığı gibi faktörlerin göz önünde bulundurulması gerekir. En iyi ölçüt, belirli bir veri seti ve problem için deneysel olarak belirlenmelidir. Farklı ölçütleri denemek ve sonuçları karşılaştırmak önemlidir.

1986 yılında J. Ross Quinlan tarafından geliştirilen ID3 algoritması, Karar Ağacı alanında önemli bir dönüm noktası olmuştur. ID3, bilgi kazancı ölçütünü kullanarak bir veri setindeki en iyi bölme özelliğini seçmiştir. Daha sonra Quinlan, ID3 algoritmasını güncelleyerek C4.5 algoritmasını geliştirmiştir. C4.5, bilgi kazancı yerine gain ratio ölçütünü kullanarak daha iyi bölme kararları verebilmiştir. Ayrıca, eksik veri değerleri ile başa çıkabilme ve sürekli değerlere sahip özellikler için eşik değerleri belirleyebilme gibi yenilikler de getirmiştir. C4.5 algoritması, halen geleneksel istatistiksel yöntemlerle elde edilen Karar Ağacına temel teşkil etmektedir. Bu algoritma, basitliği, yorumlanabilirliği ve yüksek performansı sayesinde birçok farklı alanda kullanılmaktadır (Aksu ve Doğan, 2019:1186).

Çalışmada en çok tercih edilen C4.5 algoritması (Quinlan, 1993:37) ve bilgi kazancı düğüm ayırma ölçütü kullanılmıştır. Bu algoritma Karar Ağacı oluşturma sürecinde veya sonrasında bazı düğüm ve alt ağacı silerek aşırı uygunluk problemini ortadan kaldırmakta ve veri setindeki gürültülü kayıtların kaldırılmasını sağlamaktadır (Niuniu ve Yuxun, 2010:106).

#### 4.3.4. Destek vektör makinesi

Destek Vektör Makinesi (DVM), sınıflandırma ve regresyon problemlerinde değişkenler arası ilişkilerin bilinmediği durumlarda kullanılan bir Makine Öğrenmesi tekniğidir. DVM, eğitim verilerini analiz ederek yeni veriler için doğru tahminler yapmayı ve genelleyci bir model oluşturmayı hedefler. DVM, parametrik olmayan bir sınıflandırıcıdır, yani verilerin

dağılımına ilişkin önceden bir varsayım yapmaz. Eğitim kümelerinde girişler ve çıktılar birbirine bağlanır (Cüvitoğlu, 2017:17).

DVM, veri noktalarını iki farklı sınıfa ayırmak amacıyla kullanılan bir makine öğrenimi modelidir. Bu model, sınıflar arasındaki en geniş marjı sağlayacak şekilde bir ayırım hiperdüzlemi oluşturmayı hedefler. Hiperdüzlem, her iki sınıfın da en yakın veri noktalarına olan uzaklığını en üst düzeye çıkaracak biçimde konumlandırılır. Bu sayede, model yeni bir veri noktasıyla karşılaştığında, bu noktanın hangi sınıfa ait olduğunu etkin bir şekilde tahmin edebilir. (Cortes ve Vapnik, 1995:274).

DVM, veri kümesinin doğrusallığına göre doğrusal (Linear) ve Doğrusal olmayan (non-linear) olarak ikiye ayrılır. Veriler doğrusal olarak ayrılabilirse, buna doğrusal DVM denir. Doğrusal DVM, veri kümesindeki noktaları bir doğru ile ayırarak, iki sınıf arasındaki en geniş marjı bulur. Veriler doğrusal olarak ayrılamazsa, buna da doğrusal olmayan DVM denir. Doğrusal olmayan DVM, veri kümesindeki noktaları bir eğri ile ayırır (Cüvitoğlu, 2017:18).

Verilerin doğrusal olarak ayrılamadığı durumlarda, doğrusal DVM yetersiz kalabilir. Bu gibi durumlarda, doğrusal olmayan DVM daha iyi sonuçlar verebilir. Doğrusal olmayan DVM, verileri orijinal giriş uzayından yüksek boyutlu bir nitelik uzayına haritalayarak, bu uzayda doğrusal olarak ayrılabilir hale getirir. Bu haritalama işlemi, çekirdek fonksiyonları kullanılarak yapılır. Çekirdek fonksiyonları, verilerin nitelik uzayındaki iç çarpımlarını doğrudan hesaplamaya yarar. Böylece, çok yüksek boyutlu bir nitelik uzayıyla uğraşmak zorunda kalmadan, verileri doğrusal olarak sınıflandırmak mümkün olur (Yakut, Elmas ve Yavuz, 2014:143).

#### **4.3.5. Derin öğrenme**

Derin öğrenme, yapay zeka araştırmalarında son yıllarda önemli bir ilerleme kaydeden bir teknolojidir. Bu yöntem, insan beyninin bilgi işleme şeklini taklit eden yapay sinir ağlarından esinlenmiştir. Derin öğrenme modelleri, görüntü ve ses tanıma, doğal dil işleme ve karmaşık desen tanıma gibi görevlerde olağanüstü başarılar göstermiştir. Özellikle, derin öğrenme, büyük veri kümelerinden karmaşık özellikleri otomatik olarak çıkarma yeteneği sayesinde, Makine Öğrenmesinin diğer yöntemlerine göre bazı alanlarda daha üstün

sonular elde etmektedir. Denetimli ve denetimsiz ğrenme arasındaki fark, algoritmaların eđitim sırasında etiketli verilere ihtiya duyup duymamasına dayanır. Denetimli ğrenme, etiketli veri setleri kullanarak modelin dođru tahminler yapmasını sađlarken, denetimsiz ğrenme, veri iindeki gizli yapıları veya desenleri keřfetmeye alıřır (Őeker, Diri ve Balık, 2017:48).

Derin ğrenme, zellikle byk veri setleri zerinde karmařık rntleri tanıma ve ğrenme yeteneđi ile bilinir. Bu yetenekler, eřitli niversiteler ve řirketler tarafından geliřtirilen ve srekli olarak iyileřtirilen ktphaneler ve API'lar sayesinde mmkn olmaktadır. Bu ktphaneler, arařtırmacıların ve geliřtiricilerin derin ğrenme modellerini daha hızlı ve etkili bir řekilde oluřturmasına olanak tanır. TensorFlow, Keras, PyTorch ve Theano, bu alanda en ok kullanılan ktphanelerden bazılarıdır. TensorFlow, zellikle byk veri kmeleri ve karmařık hesaplamalar iin uygun olan geniř lekli dađıtık sistemlerde mkemmел performans sergiler. Keras ise, TensorFlow'un bir st katmanı olarak dřnlebilir ve daha basit ve hızlı prototipleme iin kullanıcı dostu bir arayz sađlar. PyTorch, dinamik hesaplama grafikleri ve hızlı deney yapma imkanı ile bilim insanları ve arařtırmacılar arasında poplerdir. Theano ise, matematiksel ifadelerin derinlemesine optimize edilmesi ve hızlı hesaplama iin kullanılır. Her bir ktphane, belirli senaryolarda avantajlar sunar ve kullanıcıların ihtiyalarına gre en uygun olanı semeleri nemlidir. rneđin, endstriyel uygulamalar ve retim iin TensorFlow tercih edilirken, akademik arařtırmalar ve prototip geliřtirme iin PyTorch veya Keras daha uygun olabilir. Theano, zellikle matematiksel hesaplamaların yođun olduđu projelerde tercih edilebilir. Bu ktphanelerin her biri, Makine ğrenimi modellerinin geliřtirilmesi ve eđitilmesinde nemli bir rol oynar ve yapay zekanın geleceđini řekillendirmede kritik neme sahiptir (Gndz ve Cedimođlu, 2019:11).

#### **4. 4. Deđerlendirme Metrikleri**

Deđerlendirme metrikleri, Makine ğrenimi modellerinin performansını lmek ve bu modellerin gerek dnya verileri zerinde ne kadar iyi alıřtıđını belirlemek iin kullanılır. Bu metrikler, modelin dođruluđunu, duyarlılıđını, kesinliđini ve diđer nemli zelliklerini deđerlendirerek, modelin tahminlerinin ne kadar gvenilir olduđunu anlamamıza yardımcı olur. Deđerlendirme metrikleri, modelin eđitim srecindeki performansını deđerlendirmenin yanı sıra, modelin yeni ve grlmemiř verilere ne kadar iyi genelleme yaptđını anlamak

için de kullanılır. Ayrıca, farklı modeller arasında karşılaştırma yaparken veya modelin parametrelerini ayarlarken de bu metriklerden yararlanır.

Makine Öğrenmesi projelerinde doğru metriklerin seçilmesi, modelin başarısını doğru bir şekilde ölçebilmek ve modeli sürekli iyileştirebilmek için kritik öneme sahiptir. Çalışmada sınıflandırma modellerinin performansını ölçmek için doğruluk, hata oranı, kesinlik, duyarlılık, özgüllük ve F ölçütü gibi değerlendirme metrikleri kullanılmıştır. Ayrıca bu metriklerden yararlanabilmek için karmaşıklık matrisinin doğru bir şekilde hesaplanması gerekir (Yaman, 2022:30).

#### 4.4.1. Karmaşıklık matrisi (confusion matrix)

Karmaşıklık matrisi (Confusion Matrix), veri setindeki gerçek değerler ile sınıflandırma algoritması tarafından oluşturulan tahmin değerlerini gösterir. Bu matris, modelin tahminlerini gerçek sınıflarla karşılaştırarak, doğru pozitif (TP), yanlış pozitif (FP), doğru negatif (TN) ve yanlış negatif (FN) gibi temel bileşenleri içerir. Makine Öğrenmesi algoritmalarının değerlendirilmesinde ve karşılaştırılmasında gerek duyulan ölçütlerin çoğu aşağıdaki Çizelge 4.2’de gösterilen karmaşıklık matrisine göre hesaplanır (Yaman, 2022:31).

**Çizelge 4.2.** Karmaşıklık matrisi (Yaman, 2022:31).

|                 |       | Tahmin Değerleri |       |
|-----------------|-------|------------------|-------|
|                 |       | True             | False |
| Gerçek Değerler | True  | TP               | FN    |
|                 | False | FP               | TN    |

TP (Doğru Pozitif) : Doğru tahmin edilen pozitif kayıt sayısı

TN (Doğru Negatif) : Doğru tahmin edilen negatif kayıt sayısı

FP (Yanlış Pozitif) : Yanlışlıkla pozitif olarak tahmin edilen kayıt sayısı

FN (Yanlış Negatif) : Yanlışlıkla negatif olarak tahmin edilen kayıt sayısı

#### 4.4.2. Doğruluk (accuracy)

Doğruluk, bir sınıflandırma modelinin performansını ölçmek için kullanılan bir metriktir. Doğru tahmin edilen kayıtların toplam kayıt sayısına olan oranıdır. Örneğin, 100 kayıttan 80'ini doğru tahmin eden bir modelin doğruluk değeri 0.8'dir. Bütün kayıtlar doğru bir şekilde tahmin edildiği ölçüde bu değer 1 (en yüksek), doğru tahmin edilmediği takdirde bu değer 0 (en düşük) olarak sonuçlanır. Doğruluk değeri, modelin ne kadar güvenilir olduğunu gösterir.

$$\text{Doğruluk} = (TN+TP) / N \quad 4$$

Doğruluk, basit ve anlaşılır bir metrik olmasına rağmen, her durumda uygun bir seçim olmayabilir. Özellikle dengesiz veri setleri için, doğruluk yanıltıcı olabilir. Bu nedenle, doğruluk metriğini kullanırken veri setinin dağılımını ve problem türünü göz önünde bulundurmak gerekir.

#### 4.4.3. Hata oranı (classification error)

Bir sınıflandırma modelinin performansını ölçmek için kullanılan bir metriktir. Hata oranı, yanlış tahmin edilen kayıtların toplam kayıt sayısına bölünmesiyle hesaplanır.

$$\text{Hata oranı} = (FN+FP) / N \quad 5$$

$$\text{Hata oranı} = 1 - (TN+TP) / N \quad 6$$

Hata oranı ne kadar yüksekse, modelin doğruluğu o kadar düşüktür. Hata oranı ne kadar düşükse, modelin doğruluğu o kadar yüksektir. Bütün kayıtlar yanlış bir şekilde tahmin edildiği ölçüde bu değer 1 (en yüksek), yanlış tahmin edildiği takdirde bu değer 0 (en düşük) olarak sonuçlanır.

#### 4.4.4. Kesinlik (precision)

Bir sınıflandırma modelinin performansını ölçmek için kullanılan bir metriktir. Kesinlik, doğru tahmin edilen pozitif değerlerin, tahmin edilen toplam pozitif değerlere oranıdır.

Örneğin, bir e-posta filtresi spam olarak işaretlediği 100 e-postadan 90'nın gerçekten spam olduğunu varsayalım. Bu durumda, filtre kesinliği %90'dır.

$$\text{Kesinlik} = TP / (TP+FP) \quad 7$$

Kesinlik, modelin yanlış pozitifleri ne kadar az ürettiğini gösterir. Kesinlik, gerçek değerlerle karşılaştırıldığında, modelin pozitif olarak tahmin ettiği değerlerin ne kadarının doğru olduğunu belirtir. Kesinlik, duyarlılık veya geri çağırma ile birlikte kullanıldığında, modelin hem doğru pozitifleri hem de yanlış negatifleri nasıl yönettiğini anlamak için yararlıdır.

#### 4.4.5. Duyarlılık (sensitivity)

Bir sınıflandırma modelinin gerçekten pozitif olan durumları ne kadar iyi tespit ettiğini ölçen önemli bir metriktir. Doğru tahmin edilen pozitif değerlerin, doğru tahmin edilen pozitif değerler ile yanlışlıkla negatif tahmin edilen değerlerin toplamına olan oranıdır.

$$\text{Duyarlılık} = TP / (TP+FN) \quad 8$$

Duyarlılık, 0 ile 1 arasında bir değer alır. Pozitiflerin tamamı doğru tahmin edilmiş olursa bu değer 1 (en yüksek) olur. Bu, ideal bir durumdur ve modelin mükemmel performans gösterdiği anlamına gelir. Pozitiflerin tamamı doğru tahmin edilmemiş olursa bu değer 0 (en düşük) olur. Bu, kötü bir durumdur ve modelin hiç güvenilir olmadığı anlamına gelir.

Duyarlılık, modelin pozitif sınıfı kaçırmaması gereken durumlarda çok önemlidir. Örneğin, hastalık teşhisi gibi hayati bir konuda, modelin hastalığı olan kişileri doğru olarak tespit etmesi gerekir. Aksi takdirde, hastalığı olan kişilere yanlışlıkla negatif sonuç verilirse, bu ciddi sağlık sorunlarına veya ölüme yol açabilir. Bu nedenle, duyarlılığı yüksek olan bir model tercih edilir.

#### 4.4.6. Özgüllük (specificity)

Bir sınıflandırma modelinin negatif sınıfı ne kadar iyi ayırt ettiğini gösteren önemli bir metriktir. Özgüllük, modelin gerçekten negatif olan örnekleri doğru bir şekilde negatif

olarak tahmin etme olasılığını ifade eder. Doğru tahmin edilen negatif değerlerin, doğru tahmin edilen negatif değerler ile yanlışlıkla pozitif tahmin edilen değerlerin toplamına olan oranıdır.

$$\text{Özgüllük} = \text{TN} / (\text{TN} + \text{FP}) \quad 9$$

Özgüllük, 0 ile 1 arasında bir değer alır. Özgüllük ne kadar yüksekse, modelin negatif sınıfı ayırt etme yeteneği o kadar iyidir. Özgüllük ne kadar düşükse, modelin negatif sınıfı yanlış tanıma olasılığı o kadar yüksektir.

Özgüllük, modelin negatif sınıfı kaçırmaması gereken durumlarda çok önemlidir. Örneğin, kredi verme gibi bir konuda, modelin kredi alamayacak kişileri doğru olarak reddetmesi gerekir. Aksi takdirde, kredi alamayacak kişilere yanlışlıkla pozitif sonuç verilirse, bu ciddi finansal zararlara yol açabilir. Bu nedenle, özgüllüğü yüksek olan bir model tercih edilir.

Negatif sınıfın maliyeti veya riski yüksek olduğu durumlarda özgüllüğe dikkat etmek gerekir. Örneğin, bir hastalığı teşhis etmek için kullanılan bir testin özgüllüğü düşükse, sağlıklı kişiler yanlışlıkla hastalıklı olarak tanımlanabilir ve gereksiz tedaviye maruz kalabilirler. Bu nedenle, özgüllük hesaplamak ve optimize etmek için uygun yöntemleri kullanmak önemlidir.

Özgüllük, duyarlılık ile birlikte kullanıldığında, modelin hem pozitif hem de negatif sınıfları ne kadar iyi tanımladığını gösterir. Duyarlılık ve özgüllük arasında genellikle bir denge vardır. Yani, duyarlılığı arttırmak özgüllüğü azaltabilir ve tam tersi de geçerlidir.

#### **4.4.7. F ölçütü (f measure)**

F-ölçütü, bir modelin duyarlılık ve kesinlik değerlerinin harmonik ortalamasını alarak performansını ölçer. Bu ölçüt, her iki metriğin dengesini sağlamak için kullanılır. Yüksek duyarlılığa sahip bir model, tüm pozitif durumları doğru bir şekilde saptasa da, birçok negatif durumu da yanlışlıkla pozitif olarak işaretleyebilir, bu da kesinliğin düşmesine neden olur. Öte yandan, yüksek kesinlik oranına sahip bir model, sadece gerçekten pozitif olan durumları doğru tahmin ederken, bazı pozitif durumları gözden kaçırabilir, bu da duyarlılığın azalmasına yol açar. Dolayısıyla, F-ölçütü bu iki önemli metriği dengeler ve

modelin genel performansını daha iyi bir şekilde değerlendirmemizi sağlar. F-ölçütünün formülü şöyledir:

$$F - \text{Ölçütü} = 2 \times (\text{Kesinlik} \times \text{Duyarlılık}) / (\text{Kesinlik} + \text{Duyarlılık}) \quad 10$$

Örneğin, bir modelin duyarlılığının 0.8 ve kesinliğinin 0.6 olması durumunda; F-ölçütü değeri,  $F = 2 * (0.8 * 0.6) / (0.8 + 0.6) = 0.68$  olarak hesaplanır.

F-ölçütü, sınıflandırma problemlerinde sıkça kullanılan bir metriktir. Özellikle dengesiz veri setleri için uygun bir seçimdir. Dengesiz veri setleri, bir sınıfın diğerlerine göre çok daha fazla veya az sayıda örneğe sahip olduğu veri setleridir. Bu durumda, basit doğruluk metriği yanıltıcı olabilir. Örneğin, %90'ı A sınıfına, %10'u B sınıfına ait olan bir veri setinde, model tüm örnekleri A sınıfına atarsa, doğruluk %90 olur. Ancak bu model B sınıfını hiç tanımamıştır ve bu nedenle iyi bir performans göstermemiştir. F-ölçütü, bu tür problemlerde daha adil bir değerlendirme sağlar. F-ölçütü, her sınıf için ayrı ayrı hesaplanabilir ve daha sonra ortalaması alınabilir. Böylece, modelin tüm sınıfları ne kadar iyi ayırt ettiği görülebilir.

F-ölçütü, sadece ikili sınıflandırma problemleri için değil, çoklu sınıflandırma problemleri için de kullanılabilir. Çoklu sınıflandırma problemlerinde, her sınıf için duyarlılık ve kesinlik hesaplanır ve daha sonra F-ölçütü formülü uygulanır. F-ölçütü ile modelin hem doğru hem de dengeli bir şekilde tahmin yapma yeteneği değerlendirilebilir.

#### **4.4.8. Eğri altında kalan alan (AUC)**

AUC değeri, sınıflandırma modelinin performansını değerlendirmek için kullanılan bir metriktir. ROC (Alıcı İşletim Karakteristiği) eğrisinin bir parçası olan AUC, modelin farklı sınıflandırma eşiklerindeki performansını ölçer. AUC değeri 0 ile 1 arasında değişir; 1'e yakın bir değer, modelin pozitif ve negatif sınıfları mükemmel bir şekilde ayırt ettiğini gösterirken, 0.5 değeri ise modelin rastgele tahminler yaptığını ve ayırt edici bir kapasitesinin olmadığını gösterir. Bu metrik, modelin ne kadar iyi genelleme yaptığını ve özellikle dengesiz veri setlerinde ne kadar etkili olduğunun anlaşılmasına yardımcı olur (Tomak ve Bek, 2009:60).

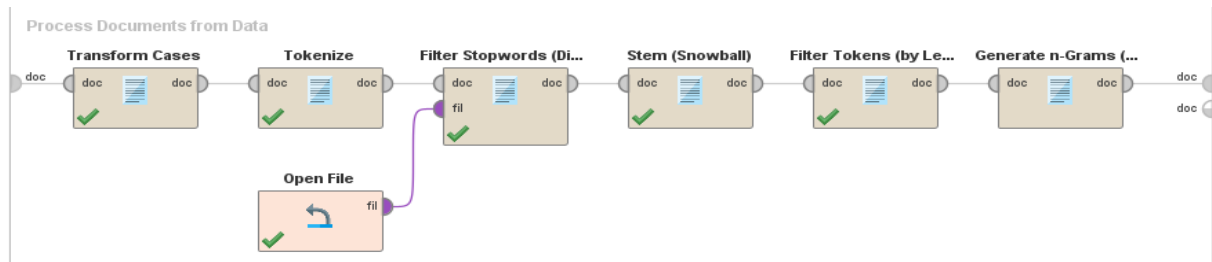
## 5. DENEYSEL ÇALIŞMALAR ve BULGULAR

### 5.1. Araştırmada Kullanılan Veri Ön İşleme Adımları

Veri ön işleme, metin madenciliği ve doğal dil işleme gibi Makine Öğrenimi uygulamalarında verilerin analize hazır hale getirilmesi için yapılan bir dizi işlemdir. Veri ön işleme, eksik veya hatalı verileri düzeltmek, gürültüyü azaltmak, boyut indirgenmesi ve dönüştürülmesi gibi işlemleri içerir. Bu işlemler, verilerin kalitesini artırmak ve model performansını iyileştirmek için gereklidir.

Bu çalışmanın ön işleme sürecinde tüm karakterlerin büyük harfe dönüştürülmesi, noktalama işaretlerinin kaldırılması, anlamsız kelimelerin filtrelenmesi, kelimelerin köklerine ayrılması, kelimelerin en düşük ve en yüksek uzunluk sınırlarının belirlenmesi yöntemleri sırasıyla gerçekleştirilmiştir.

Kullanıcı yorumları RapidMiner Studio programındaki metin madenciliğinin veri ön işleme operatörleri kullanılarak ön işleme sürecine tabi tutulmuştur. Veri ön işlemede sırasıyla Transform Cases, Tokenize, Filter Stopwords (Dictionary), Stem (Snowball), Filter Tokens (by Length), Generate n-Grams (Terms) operatörleri Şekil 5.1'de gösterildiği gibi kullanılmıştır. Metin madenciliği ile ilgili çalışmalarda ön işleme sürecinde çoğunlukla bu yöntemler kullanılmaktadır (Hofmann ve Klinkenberg, 2016:402).



Şekil 5.1. Veri ön işleme süreci (Hofmann ve Klinkenberg, 2016:402).

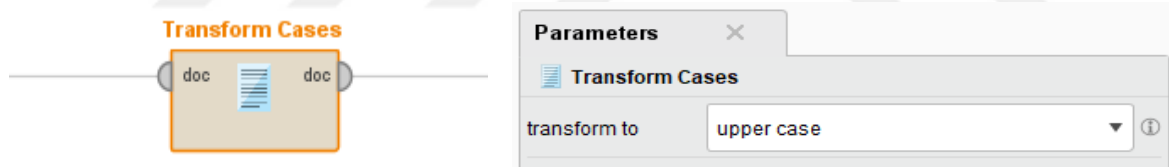
#### 5.1.1. Tüm karakterlerin büyük harfe dönüştürülmesi

Veri ön işleme yöntemlerinden biri tüm karakterlerin büyük harfe dönüştürülmesidir. Bu yöntem, metin verilerindeki büyük-küçük harf farklılıklarını ortadan kaldırmak için kullanılır. Örneğin, "İskenderun" ve "iskenderun" kelimeleri aynı anlama gelir, ancak farklı

şekilde yazılır. Bu durumda, tüm karakterler büyük harfe dönüştürülerek "İSKENDERUN" olarak standartlaştırılabilir. Böylece, metin verilerindeki tutarsızlıklar azaltılabilir ve veri analizi veya Makine Öğrenimi modellerinin performansı artırılabilir.

Tüm karakterlerin büyük harfe dönüştürülmesi, Python gibi programlama dillerinde kolayca yapılabilmektedir. Örneğin, bir metin değişkeninin adı "text" ise, `text.upper()` komutu ile bu değişkenin tüm karakterleri büyük harfe dönüştürülebilir. Bu komut, metin değişkeninin orijinal değerini değiştirmez, sadece yeni bir değer döndürür. Eğer orijinal değer de değiştirilmek istenirse, `text = text.upper()` şeklinde atama yapılabilir.

Bu çalışmada tüm karakterlerin büyük harfe dönüştürülmesi için RapidMiner uygulamasında Transform Cases operatörü kullanılmıştır. Bu operatör, bir belgedeki tüm karakterleri sırasıyla küçük harfe veya büyük harfe dönüştürmektedir. Çalışmada kullanılan sınıflandırma algoritmalarının performansını arttırmak amacıyla, veri setindeki tüm karakterlerin büyük harfe dönüştürülmesi tercih edilmiş ve bunun için Şekil 5.2’de görüldüğü gibi “transform to” parametresi “upper case” olarak seçilmiştir.

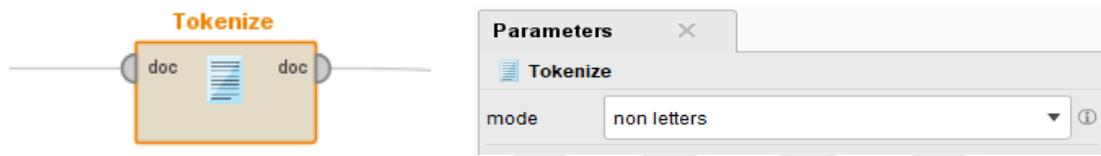


**Şekil 5.2.** Transform cases operatörü (Hofmann ve Klinkenberg, 2016:245).

### 5.1.2. Noktalama işaretlerinin kaldırılması

Veri ön işlemede kullanılan yöntemlerden biri de noktalama işaretlerinin kaldırılmasıdır. Bu işlemde, metinden nokta, virgül, soru işareti, ünlem işareti, tırnak işareti, parantez gibi semboller çıkarılır. Noktalama işaretleri, metnin anlamını ve yapısını belirlemek için gereklidir, ancak bazı durumlarda veri analizini zorlaştırmakta ve yanlış sonuçlara yol açabilmektedir. Noktalama işaretlerinin kaldırılması metin verilerinin boyutunu azaltarak metin sınıflandırma, Duygu Analizi, metin özetleme gibi işlemlerde performansı artırmaktadır (Hofmann ve Klinkenberg, 2016:245).

Noktalama işaretlerinin kaldırılması için çeşitli yöntemler vardır. Bunlardan en basit olanı, metindeki noktalama işaretlerini boşluk veya başka bir karakter ile değiştirmektir. Bu çalışmada noktalama işaretlerinin kaldırılması için RapidMiner uygulamasında Tokenize operatörü kullanılmıştır. Tokenize operatörü, metin verilerini küçük parçalara ayırmak için kullanılan bir işlemdir (Hofmann ve Klinkenberg, 2016:228). Yapılan çalışmada Tokenize operatörünün “mode” parametresinin değeri Şekil 5.3’de görüldüğü gibi “non letters” seçilerek veri setindeki harf olmayan tüm karakterler kaldırılmıştır.



Şekil 5.3. Tokenize operatörü (Hofmann ve Klinkenberg, 2016:228).

### 5.1.3. Anlamsız kelimelerin filtrelenmesi

Veri ön işlemede, anlamsız kelimelerin filtrelenmesi, verilerdeki gürültüyü azaltmak ve işlem süresini kısaltmak için önemli bir adımdır. Anlamsız kelimeler, metnin anlamına katkıda bulunmayan veya analiz için gereksiz olan kelimelerdir. Örneğin, "ve", "ama", "bir", "de" gibi bağlaçlar, edatlar ve belirteçler anlamsız kelimeler olarak kabul edilebilir. Anlamsız kelimelerin filtrelenmesi için genellikle şu yöntemler kullanılır (Hofmann ve Klinkenberg, 2016:207):

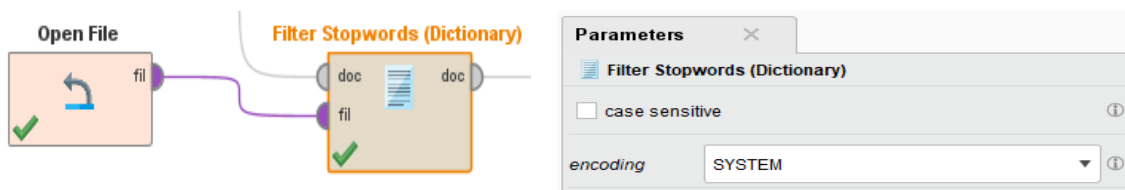
- **Sözlük tabanlı yöntem:** Bu yöntemde, önceden tanımlanmış bir anlamsız kelime listesi kullanılır. Metindeki her kelime bu listeye karşı kontrol edilir ve listede varsa metinden çıkarılır. Bu yöntemin avantajı basit ve hızlı olmasıdır. Ancak, bu yöntemin dezavantajı, anlamsız kelime listesinin dil, alan ve uygulamaya göre değişebileceği ve her duruma uygun bir liste bulmanın zor olabileceğidir.
- **İstatistiksel yöntem:** Bu yöntemde, metindeki kelimelerin sıklığına veya belgelere dağılımına göre anlamsız kelimeler belirlenir. Örneğin, çok sık veya çok seyrek görünen kelimeler anlamsız kabul edilebilir. Bu yöntemin avantajı, anlamsız kelime listesine ihtiyaç duymaması ve verilere uyum sağlayabilmesidir. Ancak, bu yöntemin dezavantajı, sıklık veya dağılım ölçütlerinin belirlenmesinin zor olması ve bazı anlamlı kelimelerin yanlışlıkla filtrelenmesi riskidir.

Bu çalışmada anlamsız kelimelerin filtrelenmesi için RapidMiner uygulamasında Filter Stopwords (Dictionary) operatörü kullanılmış ve metin analizi için önemli olmayan kelimeler veri setinden çıkarılmıştır. Filter Stopwords (Dictionary) operatörü, bir metin veri setindeki gereksiz veya anlamsız kelimeleri kaldırmak için kullanılır. Bu operatör, belirli bir dil için önceden tanımlanmış bir sözlük dosyası alır ve veri setindeki her metin belgesini bu sözlüğe göre filtreler. Çalışmada Çizelge 5.1’de görüldüğü gibi önceden tanımlanmış bir stopword listesi oluşturulmuştur.

**Çizelge 5.1.** Çalışmada kullanılan stopword listesinden örnekler.

|        |         |        |        |         |         |          |        |        |             |
|--------|---------|--------|--------|---------|---------|----------|--------|--------|-------------|
| acaba  | aslında | benden | bin    | birşey  | bizim   | bundan   | burada | de     | dokuz       |
| altmış | ayrıca  | beni   | bir    | birşeyi | böyle   | bunlar   | çok    | defa   | dolayı      |
| altı   | bana    | benim  | birçok | biz     | böylece | bunları  | çünkü  | değil  | dolayısıyla |
| ama    | bazı    | beri   | biri   | bize    | bu      | bunların | da     | diğer  | dört        |
| ancak  | belki   | beş    | birkaç | bizden  | buna    | bunu     | daha   | diye   | edecek      |
| arada  | ben     | bile   | birkez | bizi    | bunda   | bunun    | dahi   | doksan | eden        |

Oluşturulan stopword listesi, Filter Stopwords (Dictionary) operatörüne, Open File operatörü yardımıyla sözlük dosyası olarak eklenmiştir. Sözlük dosyasındaki kelimelerin doğru şekilde okunmasını ve işlenmesini sağlamak için Şekil 5.4’de görüldüğü gibi “*Encoding*” parametresinin değeri “*SYSTEM*” seçilmiştir.



**Şekil 5.4.** Filter stopwords (dictionary) operatörü (Hofmann ve Klinkenberg, 2016:245).

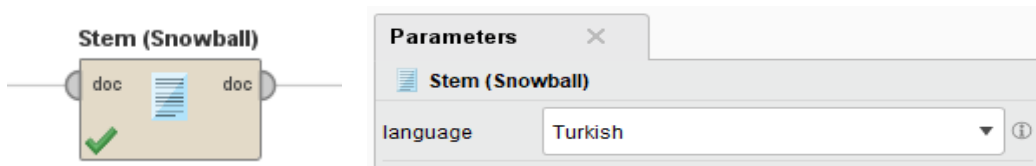
#### 5.1.4. Kelimelerin köklerine ayrılması

Veri ön işlemede kullanılan bir diğer yöntem ise kelimelerin köklerine ayrılmasıdır. Bu yöntem, metin içindeki kelimelerin anlamsal olarak en temel biçimlerine indirgenmesini sağlar. Örneğin, "geliyor", "geldi" ve "gelecek" kelimelerinin hepsinin kökü "gel"dir. Kelimelerin köklerine ayrılması, metin analizi ve sınıflandırma gibi doğal dil işleme uygulamalarında veri boyutunu azaltmak, veri kalitesini artırmak ve model performansını

iyileştirmek için kullanılmaktadır (Hofmann ve Klinkenberg, 2016:255). Kelimelerin köklerine ayrılması için çeşitli algoritmalar ve kütüphaneler mevcuttur. Bunlardan bazıları şunlardır:

- **Zemberek:** Türkçe doğal dil işleme uygulamaları geliştirmek için kullanılan açık kaynaklı bir Java kütüphanesidir. Türkçe metinleri analiz etmek, düzeltmek, sınıflandırmak ve işlemek için çeşitli araçlar ve algoritmalar sunmaktadır (Akın ve Akın, 2007).
- **NLTK:** Python programlama dilinde yazılmış doğal dil işleme kütüphanesidir. Kelimelerin köklerine ayrılması için Porter, Snowball, Lancaster gibi algoritmaları içerir (Aydın ve Kantarcı, 2024:11).
- **Pandas:** Veri analizi için tasarlanmış, Python programlama dilinde yazılmış açık kaynaklı bir kütüphanedir. Zaman serisi verileri, tablolar ve diğer çeşitli veri yapılarını işlemek için geniş bir fonksiyon yelpazesi sunar (Demirhan ve Demiralay, 2023:1740).

Bu çalışmada kelimelerin köklerine ayrılması için RapidMiner uygulamasında Stem (Snowball) operatörü uygulanmıştır. Stem (Snowball) operatörü, kelimelerin sonundaki ekleri atarak, aynı anlama gelen farklı biçimlerde yazılmış kelimeleri tek bir köke indirger. Bu operatör 15 farklı dili desteklemektedir. Çalışmada kullanılan veri setinin Türkçe olmasından dolayı Şekil 5.5’de gösterildiği gibi Stem (Snowball) operatörünün “*language*” parametre değeri “*Turkish*” olarak seçilmiştir.



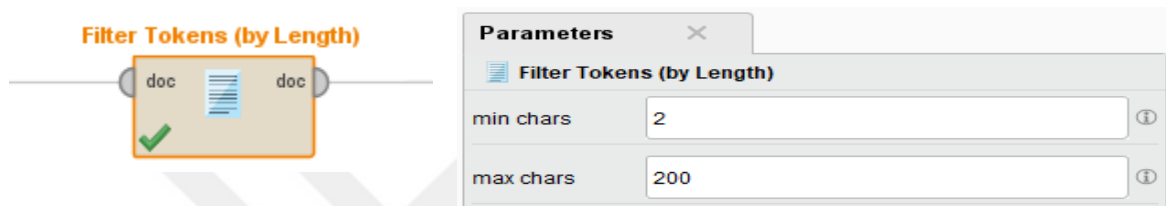
Şekil 5.5. Stem (snowball) operatörü (Hofmann ve Klinkenberg, 2016:255).

### 5.1.5. Kelimelerin en düşük ve en yüksek uzunluk sınırlarının belirlenmesi

Veri ön işleme aşamasında yapılması gereken önemli bir adım da kelimelerin en düşük ve en yüksek uzunluk sınırlarının belirlenmesidir. Bu sınırlar, veri setinin dağılımına, kullanılan

modelin mimarisine ve uygulamanın amacına göre değişebilir. Bu işlem, veri setindeki gürültüyü azaltmak ve veri boyutunu küçülterek model performansını artırmak için yapılmaktadır (Hofmann ve Klinkenberg, 2016:229).

Çalışmada kelimelerin en düşük ve en yüksek uzunluk sınırlarını belirlemek için RapidMiner uygulamasında Filter Tokens (by Length) operatörü kullanılmıştır. Bu işlem için Şekil 5.6'da görüldüğü gibi “*min chars*” parametresinin değeri 2 ve “*max chars*” parametresinin değeri 200 olarak belirlenmiştir.



Şekil 5.6. Filter tokens (by length) operatörü (Hofmann ve Klinkenberg, 2016:229).

### 5.1.6. Terim frekansı ve ters doküman frekansı

Terim frekansı (TF) ve Ters doküman frekansı (IDF), metin madenciliği ve bilgi erişimi alanlarında yaygın olarak kullanılan istatistiksel ölçülerdir. Terim Frekansı, bir metin içerisinde bir terimin ne kadar geçtiğini, metindeki frekans değerini gösterir. Ters Doküman Frekansı ise birden fazla metinde bir kelimenin tekrar etme sayısına göre kelimenin terim olup olmadığının anlaşılmasını sağlar (Kuş, 2023: 26).

Terim Frekansı, matematiksel olarak  $t$  teriminin bir belgede kaç kez geçtiğinin, belgedeki toplam kelime sayısına bölünmesiyle hesaplanır (Kuş, 2023:27).

$$TF(t) = \frac{t \text{ teriminin belgede geçtiği toplam sayı}}{\text{belgedeki toplam terim sayısı}}$$

11

TF değeri, bir terimin bir dokümanda kaç kez geçtiğini sayarak hesaplanabilir. Örneğin, "ürün" terimi bir dokümanda 3 kez geçiyorsa, TF değeri 3'tür. Ancak, bu değer dokümanın uzunluğuna bağlıdır. Uzun bir dokümanda aynı terim daha fazla geçebilir. Bu nedenle, TF değerini dokümanın toplam kelime sayısına bölerek normalize etmek gerekir. Böylece, TF

değeri bir frekansa dönüşür. Örneğin, 100 kelimelik bir dokümanda "ürün" terimi 3 kez geçiyorsa, normalize edilmiş TF değeri 0.03'tür.

IDF değeri ise bir terimin kaç belgede geçtiğini gösterir ve belgeler arasındaki belge sıklığının tersini hesaplar (Kuş, 2023:27).

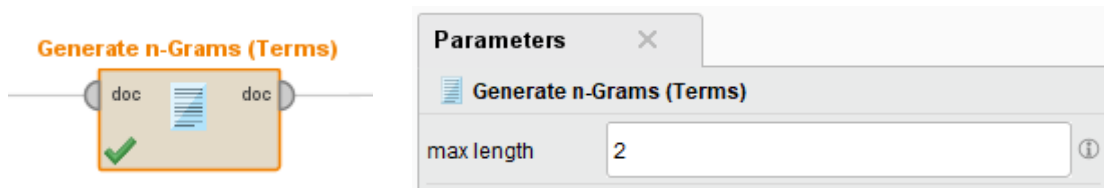
$$IDF(t) = \log \frac{\text{tüm belgelerin sayısı}}{t \text{ teriminin geçtiği belge sayısı}} \quad 12$$

Bir terim çok fazla dokümanda geçiyorsa, o terim sorguyla alakalılığı düşük olan genel bir terimdir. Örneğin, "ve", "bir", "ile" gibi terimler hemen hemen her dokümanda geçer ve sorguyla alakalı bilgi taşımazlar. Bu nedenle, bu tür terimlere düşük bir ağırlık vermek gerekir. IDF değeri, tüm doküman sayısını içeren bir terimin doküman sayısına bölerek hesaplanabilir. Örneğin, 1000 dokümanlık bir kümede "ürün" terimi 10 dokümanda geçiyorsa, IDF değeri  $1000/10 = 100$ 'dür. Ancak, bu değer çok büyük olabilir ve logaritmik olarak ölçeklendirmek gerekir. Böylece, IDF değeri  $\log(1000/10) = 2$ 'dir.

Terim frekansı ve ters doküman frekansı, birlikte kullanılarak terim ağırlığı (TF-IDF) adı verilen bir değer elde edilir. Terim ağırlığı, bir kelimenin hem bir metinde hem de genel olarak ne kadar önemli olduğunu gösterir. TF ve IDF değerlerinin hesaplanmasının ardından elde edilen sonuçların çarpımları ile TF-IDF değeri hesaplanır (Kuş, 2023:27).

$$TF-IDF = TF \times IDF \quad 13$$

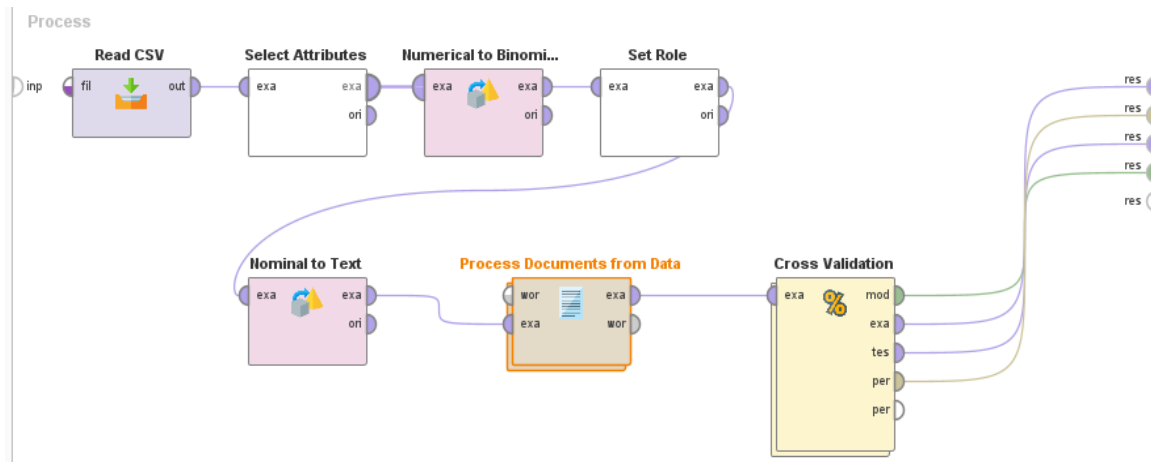
Çalışmada kelime vektörlerinin oluşturulması ve yorumların kelime kelime analiz edilmesi için n-gram yöntemi uygulanmıştır. Bunun için RapidMiner uygulamasında Şekil 5.7'de görülen Generate n-Grams (Terms) operatörü kullanılmış ve "max length" değeri 2 verilerek kelimelerin ikili birleşim vektörleri oluşturulmuştur.



Şekil 5.7. Generate n-grams (terms) operatörü (Hofmann ve Klinkenberg, 2016:249).

## 5.2. Araştırmanın Bulguları ve Değerlendirme

Çalışmada, kNN, Naive Bayes, Karar Ağacı, DVM ve Derin Öğrenme sınıflandırma algoritmaları kullanılarak sınıflandırma yapılmıştır. Çalışmada kullanılan algoritmaların mümkün olduğunca objektif ve doğru bir şekilde performans tahmini elde edebilmesi için Şekil 5.8’de görüldüğü gibi 10 kat çapraz doğrulama tekniği (Cross Validation) kullanılmıştır.



Şekil 5.8. Araştırmada kullanılan RapidMiner uygulaması.

### 5.2.1. Naive bayes algoritması sonuçları

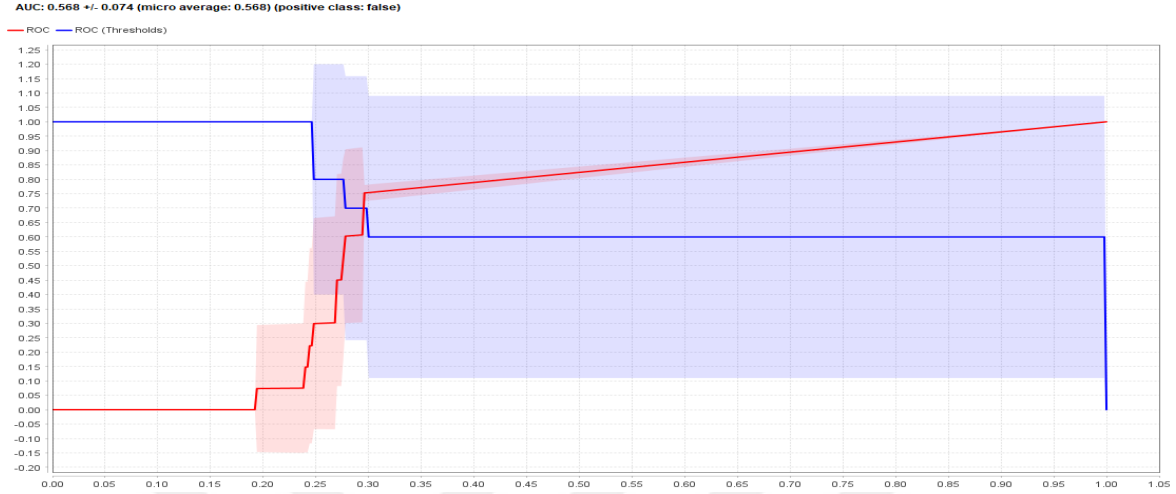
Naive Bayes algoritmasının performans sonuçları incelendiğinde Şekil 5.9’da görüldüğü gibi %73.99 oranında başarı elde edilmiştir.

The screenshot shows the PerformanceVector (Performance) window in RapidMiner. The window displays the performance metrics for the Naive Bayes algorithm. The accuracy is 73.99% +/- 2.12% (micro average: 73.99%). The window also shows a confusion matrix and class precision.

|              | true true | true false | class precision |
|--------------|-----------|------------|-----------------|
| pred. true   | 1982      | 550        | 78.28%          |
| pred. false  | 697       | 1566       | 69.20%          |
| class recall | 73.98%    | 74.01%     |                 |

Şekil 5.9. Naive bayes algoritmasının performans sonuçları.

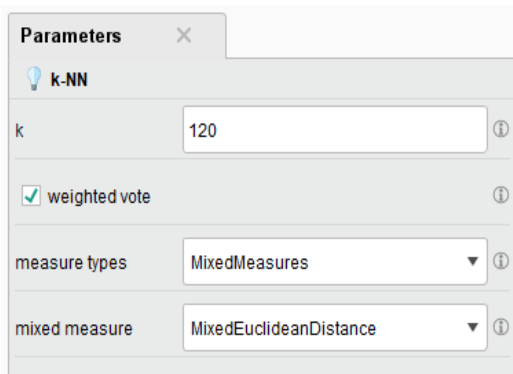
Şekil 5.10'daki ROC eğrisi ve AUC değeri incelendiğinde Naive Bayes algoritmasının 0.568 AUC değeri ile pozitif ve negatif sınıfları iyi bir şekilde ayırt edemediği görülmektedir.



Şekil 5.10. Naive bayes algoritmasının ROC eğrisi ve AUC değeri.

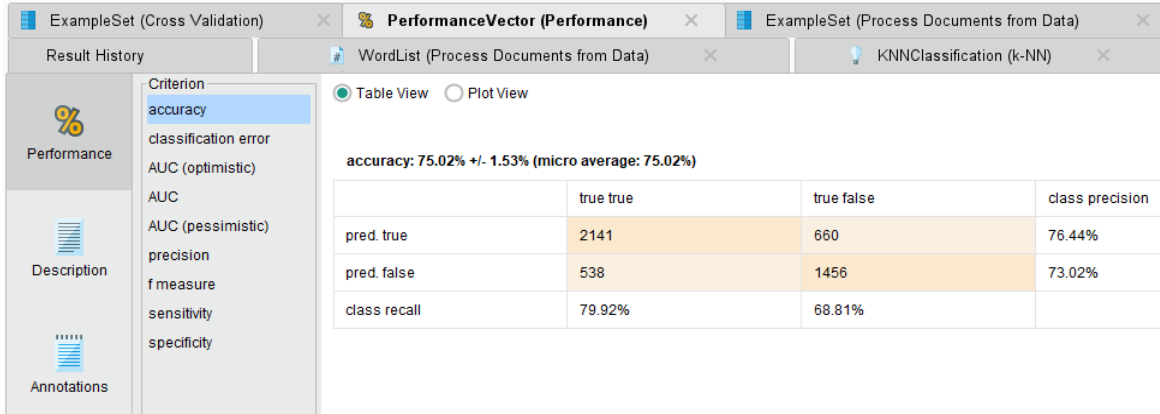
### 5.2.2. K-en yakın komşu algoritması sonuçları

KNN algoritmasında optimal k sınıf değerini belirlemek için Optimize Parameters (Grid) operatörü kullanılmış ve en uygun k değeri 120 olarak bulunmuştur. Şekil 5.11'de görüldüğü gibi uzaklık ölçümü için ise “Euclidean” mesafe ölçütü kullanılmıştır.



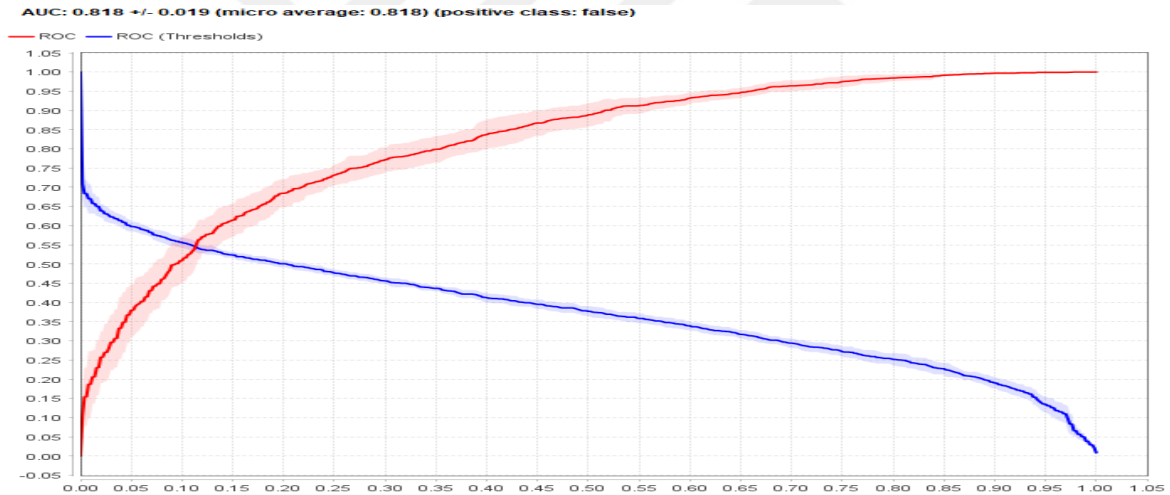
Şekil 5.11. KNN algoritmasının parametre değerleri.

KNN algoritmasının performans sonuçları incelendiğinde Şekil 5.12'de görüldüğü gibi %75.02 oranında başarı elde edilmiştir.



Şekil 5.12. KNN algoritmasının performans sonuçları.

Şekil 5.13'deki ROC eğrisi ve AUC değeri incelendiğinde kNN algoritmasının 0.818 AUC değeri ile pozitif ve negatif sınıfları iyi bir şekilde ayırt ettiği görülmektedir.

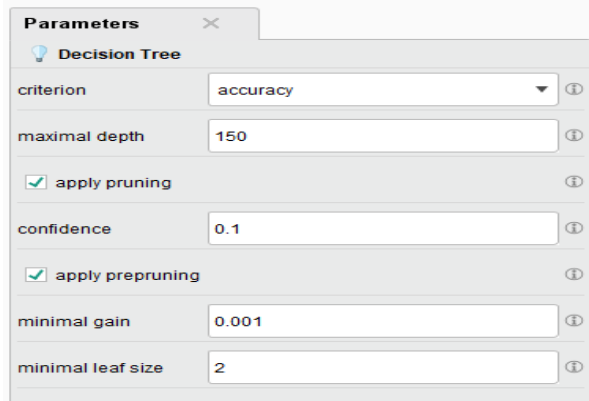


Şekil 5.13. KNN algoritmasının ROC eğrisi ve AUC değeri.

### 5.2.3. Karar ağacı algoritması sonuçları

Karar Ağacı algoritmasında en iyi performans Şekil 5.14'de gösterilen parametrelerle elde edilmiştir. En iyi bölünmeyi belirlemek için dallanma kriteri (criterion) parametresi “accuracy” seçilmiştir. Karar ağacının ne kadar derin oluşturulacağını belirlemek için “maximal depth” parametre değeri 150 olarak belirlenmiştir. Ağacın dallanma sürecinde karar verme güvenilirliğini artırmak için “confidence” parametre değeri 0.1 olarak verilmiştir. Ağacın aşırı dallanmasını önlemek ve modelin genelleme yeteneğini artırmak için ise “minimal gain” parametre değeri 0.001 olarak belirlenmiştir. Modelin aşırı

öğrenmesini (overfitting) önlemeye yardımcı olması için ağacın yaprak düğümlerinde bulunabilecek minimum örnek sayısı “*minimal leaf size*” parametresi ile 2 olarak belirlenmiştir. Ayrıca ağacın aşırı büyümesini önlemek için “*apply pruning*” ve ağacın oluşturulma sürecini belli bir noktada durdurarak modelin aşırı uyuma karşı direncini arttırmak için “*apply prepruning*” parametresi aktif hale getirilmiştir.



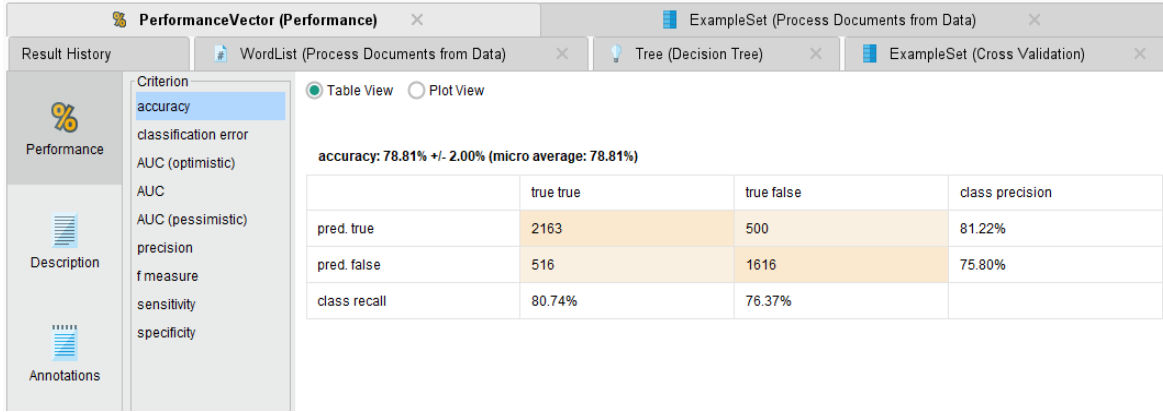
Şekil 5.14. Karar ağacı algoritmasının parametre değerleri.

Karar ağacı algoritması için belirlenen parametreler sonucunda Şekil 5.15’de görülen görsel çıktı elde edilmiştir.



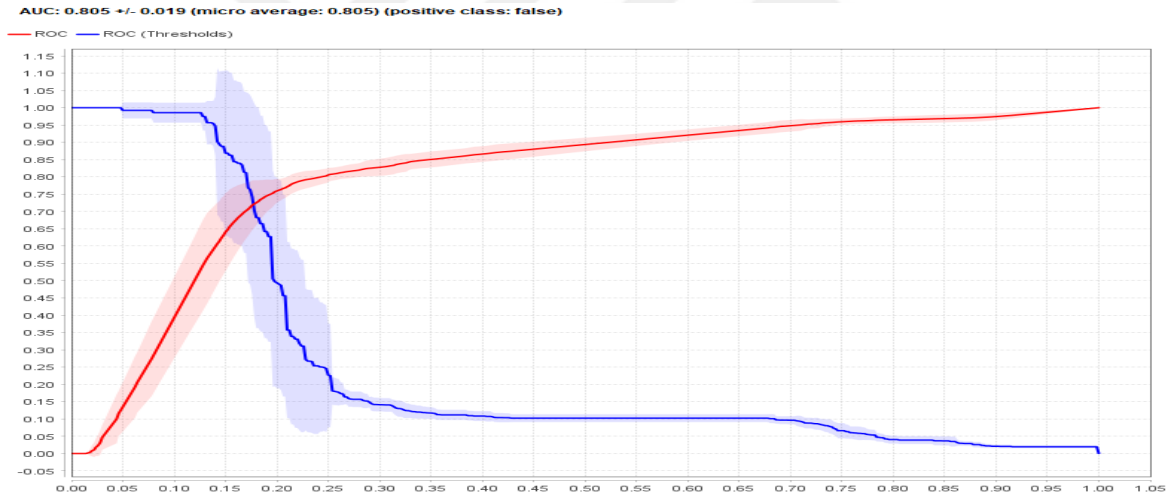
Şekil 5.15. Karar ağacının görsel çıktısı.

Karar Ağacı algoritmasının performans sonuçları incelendiğinde Şekil 5.16’da görüldüğü gibi %78.81 oranında başarı elde edilmiştir.



Şekil 5.16. Karar ağacı algoritmasının performans sonuçları.

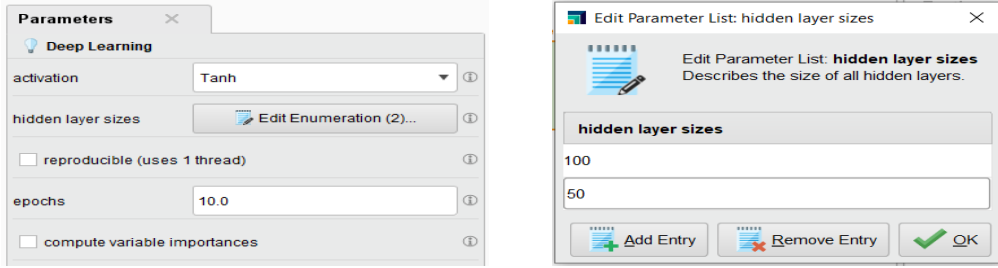
Şekil 5.17’deki ROC eğrisi ve AUC değeri incelendiğinde Karar Ağacı algoritmasının 0.805 AUC değeri ile pozitif ve negatif sınıfları iyi bir şekilde ayırt ettiği görülmektedir.



Şekil 5.17. Karar ağacı algoritmasının ROC eğrisi ve AUC değeri.

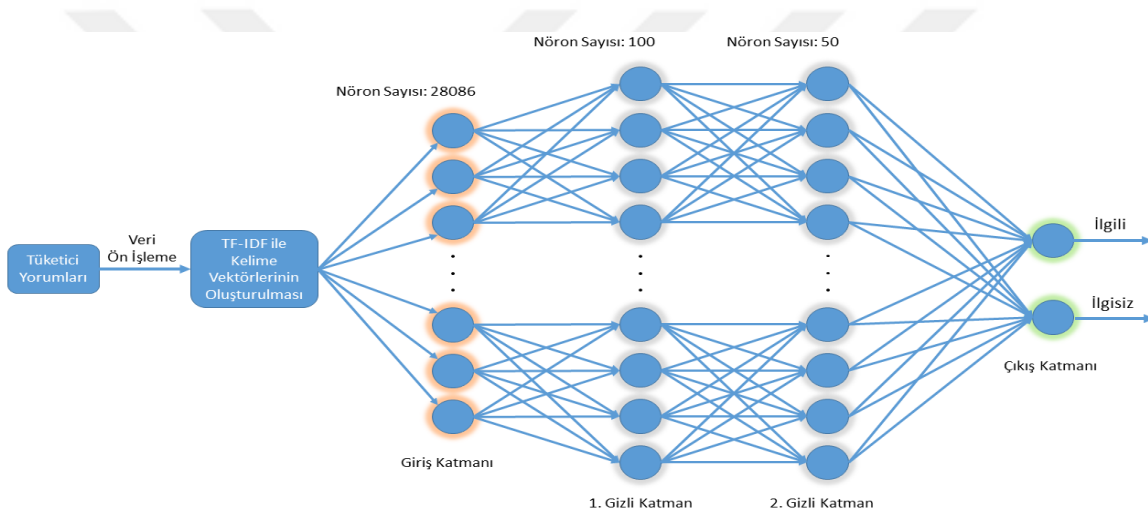
#### 5.2.4. Derin öğrenme algoritması sonuçları

Derin Öğrenme algoritmasında en iyi performans Şekil 5.18’de gösterilen parametrelerle elde edilmiştir. Bir nöronun çıktısını belirleyerek, ağın doğru tahminler yapmasını sağlamak için “activation” parametre değeri “Tanh” olarak seçilmiştir. Modelin eğitim veri seti üzerinde kaç kez çalıştırılacağını belirleyen “epochs” parametre değeri 10.0 olarak ve yapay sinir ağındaki gizli katmanlarının nöron sayısı 100, 50 olarak belirlenmiştir.



Şekil 5.18. Derin öğrenme algoritmasının parametre değerleri.

Derin öğrenme algoritması için belirlenen parametreler sonucunda Şekil 5.19'da görülen derin öğrenme mimarisi elde edilmiştir.



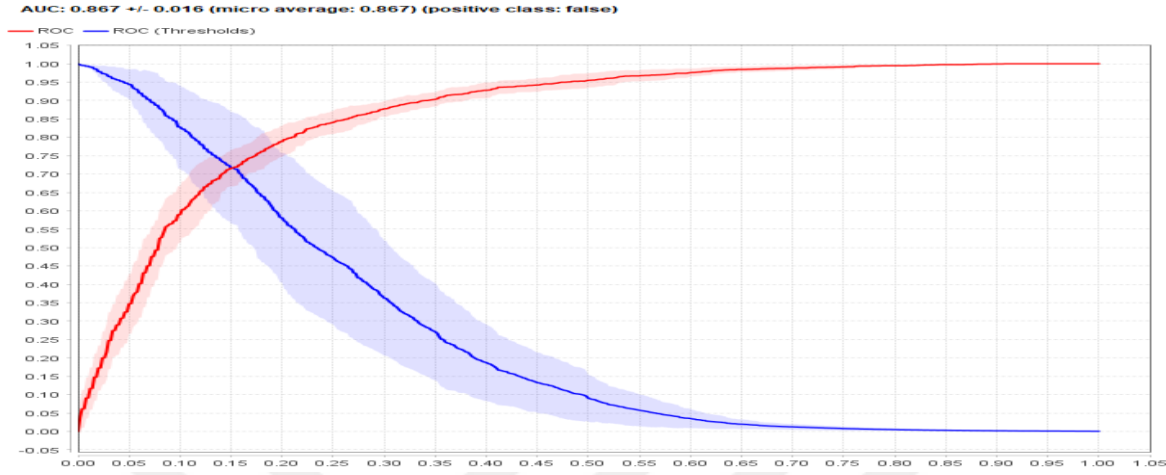
Şekil 5.19. Derin öğrenme mimarisi.

Derin Öğrenme algoritmasının performans sonuçları incelendiğinde Şekil 5.20'de görüldüğü gibi %78.62 oranında başarı elde edilmiştir.

|              | true true | true false | class precision |
|--------------|-----------|------------|-----------------|
| pred. true   | 2009      | 355        | 84.98%          |
| pred. false  | 670       | 1761       | 72.44%          |
| class recall | 74.99%    | 83.22%     |                 |

Şekil 5.20. Derin öğrenme algoritmasının performans sonuçları.

Şekil 5.21'deki ROC eğrisi ve AUC değeri incelendiğinde Derin Öğrenme algoritmasının 0.867 AUC değeri ile pozitif ve negatif sınıfları iyi bir şekilde ayırt ettiği görülmektedir.



Şekil 5.21. Derin öğrenme algoritmasının ROC eğrisi ve AUC değeri.

### 5.2.5. Destek vektör makinesi algoritması sonuçları

DVM algoritmasında en iyi performans RapidMiner programında yer alan Support Vector Machine (LibSVM) operatörü ve Şekil 5.22'de gösterilen parametrelerle elde edilmiştir. DVM algoritmasının veri setindeki örüntüleri nasıl işleyeceğini belirleyen “*type*” parametre değeri “*nu-SVC*” olarak seçilmiştir. DVM'nin veri setindeki ilişkileri modelleyebilmesini sağlamak için “*kernel type*” parametre değeri “*linear*” seçilmiştir. Modelin hata toleransını kontrol eden “*nu*” parametre değeri 0.5 ve “*epsilon*” parametre değeri 0.1 olarak verilmiştir.

Parameters

SVM (Support Vector Machine (LibSVM))

svm type: nu-SVC

kernel type: linear

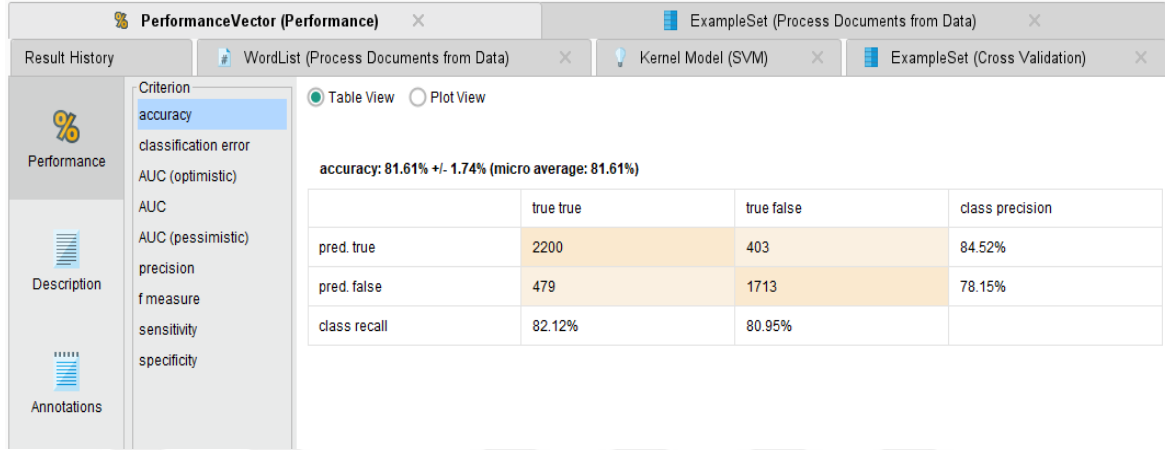
nu: 0.5

epsilon: 0.1

calculate confidences

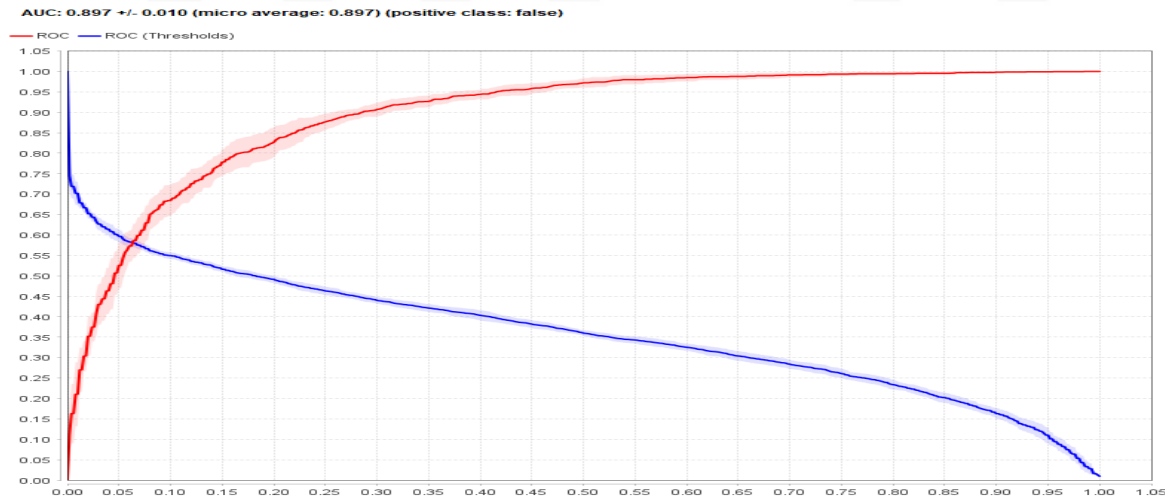
Şekil 5.22. DVM algoritmasının parametre değerleri.

DVM algoritmasının performans sonuçları incelendiğinde Şekil 5.23’de görüldüğü gibi %81.61 oranında başarı elde edilmiştir.



Şekil 5.23. DVM algoritmasının performans sonuçları.

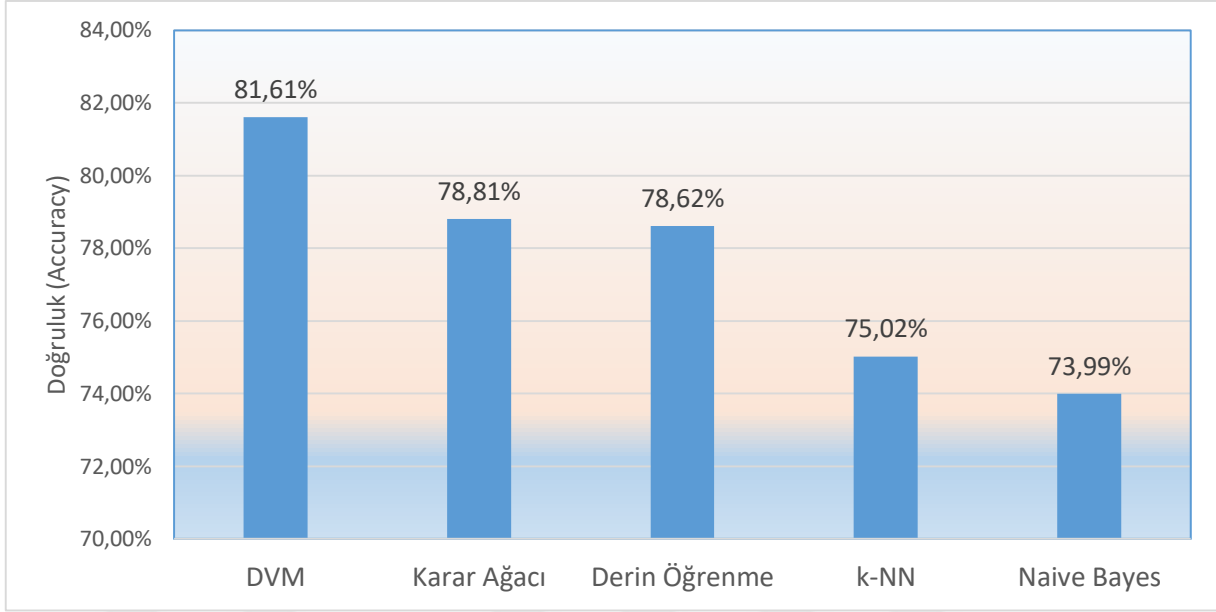
Şekil 5.24’teki ROC eğrisi ve AUC değeri incelendiğinde DVM algoritmasının 0.897 AUC değeri ile pozitif ve negatif sınıfları en iyi şekilde ayırt ettiği görülmektedir.



Şekil 5.24. DVM algoritmasının ROC eğrisi ve AUC değeri.

### 5.3. Algoritmaların Performans Karşılaştırması

Çalışmada uygulanan sınıflandırma algoritmalarından Şekil 5.25’te görüldüğü gibi %81.61 doğruluk değeri ile en yüksek başarı DVM algoritması ile elde edilmiştir.



**Şekil 5.25.** Uygulanan sınıflandırma algoritmalarının başarı grafiği.

Çizelge 5.2’de çalışmada uygulanan sınıflandırma algoritmalarının değerlendirme metrikleri sonuçları yer almaktadır. Bu sonuçlara göre de en iyi performansın DVM algoritmasına ait olduğu görülmektedir.

**Çizelge 5.2.** Sınıflandırma algoritmalarının değerlendirme metrikleri sonuçları.

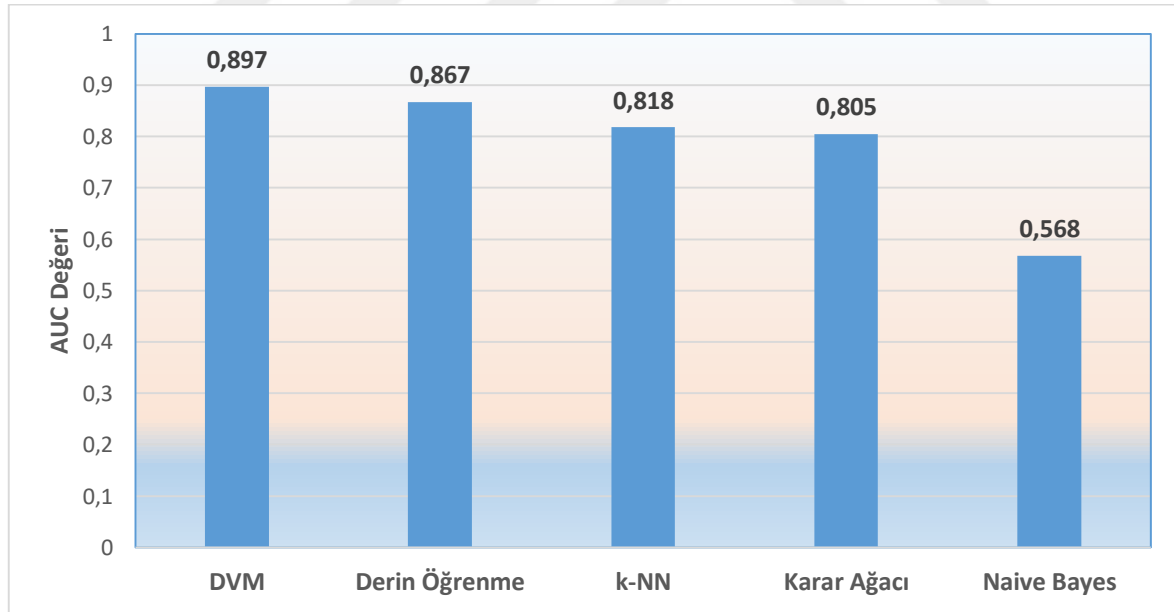
| Sınıflandırma Algoritmaları | Doğruluk (%) | Hata Oranı (%) | Kesinlik (%) | Duyarlılık (%) | Özgüllük (%) | F-Ölçütü (%) |
|-----------------------------|--------------|----------------|--------------|----------------|--------------|--------------|
| <b>DVM</b>                  | 81,61        | 18,39          | 78,16        | 80,96          | 82,12        | 79,51        |
| <b>Karar Ağacı</b>          | 78,81        | 21,19          | 75,89        | 76,37          | 80,74        | 76,10        |
| <b>Derin Öğrenme</b>        | 78,62        | 21,38          | 72,92        | 83,22          | 74,99        | 77,40        |
| <b>KNN</b>                  | 75,02        | 24,98          | 73,06        | 68,81          | 79,92        | 70,82        |
| <b>Naive Bayes</b>          | 73,99        | 26,01          | 69,27        | 74,01          | 73,98        | 71,53        |

DVM algoritmasında özellikle LibSVM operatörünün “nu-SVC” ve “Linear” çekirdek türleri ile yüksek başarı elde edilmiştir. DVM algoritmasının Linear, Evolutionary, PSO operatörleri ve diğer parametreleri ile elde edilen sonuçlarda doğruluk değeri %75 ile %80

arasında değişmektedir. Bu sonuçlar, DVM algoritmasının tüketici yorumlarının metin madenciliği yöntemleri ile analizi konusunda potansiyel olarak güçlü bir potansiyele sahip olduğunu göstermektedir.

Karar Ağacı ve Derin Öğrenme algoritmaları kabul edilebilir bir doğruluk değerine sahiptir, ancak başarıları DVM algoritmasının biraz gerisinde kalmıştır. KNN ve Naive Bayes algoritmaları ise en düşük performansa sahip algoritmalar. Bu sonuçlar özellikle Naive Bayes algoritmasının metin madenciliği alanındaki potansiyelinin ve tüketici yorumları üzerinde analiz yapma kapasitesinin diğer modellere göre daha sınırlı performansa sahip olduğunu göstermektedir.

Çalışmada uygulanan sınıflandırma algoritmalarının AUC değerleri karşılaştırıldığında, Şekil 5.26'da görüldüğü gibi en yüksek performansı 0.897 AUC değeri ile DVM algoritması sergilemiştir. Bu sonuç, DVM algoritmasının, veri setindeki pozitif ve negatif sınıfları ayırt etme performansının diğer algoritmalara kıyasla daha üstün olduğunu göstermektedir.



**Şekil 5.26.** Uygulanan sınıflandırma algoritmalarının AUC değerleri.

Algoritmaların AUC değerleri, özellikle Naive Bayes algoritmasının diğer algoritmalara göre oldukça düşük bir AUC değerine sahip olduğunu göstermektedir. AUC değerinin 0.5'e yakın olması, Naive Bayes algoritmasının rastgele tahmin yapma olasılığı ile tüketici

yorumlarının analizi konusunda diğer modellere kıyasla daha kısıtlı bir performans sergilediğini ortaya koymaktadır.

Sonuç olarak başta DVM algoritması olmak üzere Karar Ağacı ve Derin Öğrenme algoritmalarının, Çizelge 5.2’de görülen değerlendirme metrikleri sonuçlarına göre başarılı sonuçlar verdiği görülmektedir. Bu algoritmalar, tüketici yorumlarının metin madenciliği ile analizi konusunda yüksek başarı elde edebilmek için dikkate alınabilir.

#### 5.4. Algoritmaların Sonuçlarının Tartışılması

Veri setine uygulanan sınıflandırma algoritmalarının karmaşıklık matrisi sonuçları Çizelge 5.3’de verilmiştir. Bu sonuçlara göre modellerin performansı gerçek değerler, tahmin değerleri ve doğruluk değerleri üzerinden tartışılmıştır.

**Çizelge 5.3** Sınıflandırma algoritmalarının karmaşıklık matrisi sonuçları.

|                  |               |         | Gerçek Değerler |         | Doğruluk (%) |
|------------------|---------------|---------|-----------------|---------|--------------|
|                  |               |         | İlgili          | İlgisiz |              |
| Tahmin Değerleri | DVM           | İlgili  | 2200            | 403     | 81,61        |
|                  |               | İlgisiz | 479             | 1713    |              |
|                  | Karar Ağacı   | İlgili  | 2163            | 500     | 78,81        |
|                  |               | İlgisiz | 516             | 1616    |              |
|                  | Derin Öğrenme | İlgili  | 2009            | 355     | 78,62        |
|                  |               | İlgisiz | 670             | 1761    |              |
|                  | kNN           | İlgili  | 2141            | 660     | 75,02        |
|                  |               | İlgisiz | 538             | 1456    |              |
|                  | Naive Bayes   | İlgili  | 1982            | 550     | 73,99        |
|                  |               | İlgisiz | 697             | 1566    |              |

Tüketici yorumlarının ürün ile ilgili olup olmadığının tespit edilmesi için kullanılan sınıflandırma algoritmaları ile elde edilen sonuçlar farklılık göstermektedir. Özellikle DVM sınıflandırma algoritmasından elde edilen doğruluk değerleri, tüketici yorumlarını tahmin etmede bu modelin güvenilir ve etkili bir şekilde kullanılabileceğini göstermektedir.

Derin Öğrenme ve Karar Ağacı sınıflandırma algoritmaları da iyi bir performans göstermiş olup tüketici yorumlarının analizi ve sınıflandırılması için tercih edilebilir. Özellikle Derin

Öğrenme algoritmasının çok katmanlı yapısı, tüketici yorumları gibi metin tabanlı verilerdeki ince nüansları ve karmaşık desenleri tanıma yeteneğini artırmaktadır. KNN ve Naive Bayes Algoritmaları ise, diğer algoritmalara göre daha düşük bir performans göstermiştir. Bu algoritmaların tahmin performansı, diğerlerine göre daha az güvenilir olabilir, tüketici yorumlarının analizi için daha az tercih edilmelidir.

Tüketici yorumlarının analizinde, DVM, Karar Ağacı ve Derin Öğrenme sınıflandırma algoritmaları öne çıkmaktadır. Bu algoritmaların yüksek tahmin performansı, hem işletmeler için stratejik kararlar alırken hem de müşteri memnuniyetini artırmada yardımcı olabilir. Her modelin kendine özgü güçlü ve zayıf yönleri vardır, bu nedenle model seçimi yapılırken, veri setinin niteliği, modelin karmaşıklığı ve eğitim süresi gibi faktörler göz önünde bulundurulmalıdır. Ayrıca, modelin performansını değerlendirmek için çapraz doğrulama ve diğer istatistiksel testler kullanılabilir.

## 6. SONUÇ

Günümüzde internet üzerinden yapılan alışveriş başta ülkemiz olmak üzere tüm dünyada hızla artmaktadır. Tüketicilere zaman ve mekandan bağımsız olarak alışveriş yapma imkanı sunan e-ticaret, geniş ürün yelpazesi ve rekabetçi fiyatlarla cazip hale gelmiştir. Tüketiciler, e-ticareti geleneksel alışverişe kıyasla daha fazla tercih etmektedir. Ayrıca, e-ticaret, zaman tasarrufu ve ulaşım maliyetlerini azaltma gibi ek faydalar sağlayarak, tüketicilerin hayatını kolaylaştırmaktadır.

E-ticarete ürünü satın alan kişiler ürünün markası ve fiyatı yanında ürün hakkında yapılmış yorumları dikkate alarak karar vermektedirler. Ancak paylaşılan kullanıcı yorumlarının çoğunun ürün ile ilgili olmaması ve ilgisiz yorumların aşırı olması müşterilerin sipariş aşamasında karar vermelerini olumsuz yönde etkilemektedir. Bu çalışmada metin madenciliği teknikleri kullanılarak, e ticaret sitelerinden bir ürünü satın alan müşterinin, satın aldığı ürün deneyimi sonrasında paylaştığı çevrimiçi yorumların ürün ile ilgili olup olmadığı tespit edilmesi amaçlanmıştır.

Çalışmada, Türkiye'nin önde gelen online pazar yerlerinden biri olan trendyol.com'da yer alan farklı kategorilerde 21 ürüne ait 4795 yorum metin madenciliği teknikleri kullanılarak analiz edilmiştir. Terim Frekansı ve Ters Doküman Frekansı (TF-IDF) yöntemleri kullanılarak kelime vektörleri oluşturulmuş ve kNN, Naive Bayes, Karar Ağacı, DVM ve Derin Öğrenme algoritmaları ile yorumlar sınıflandırılmıştır. En yüksek başarı, %81.61 doğruluk değeri ile DVM algoritması ile elde edilmiştir.

Pazarlama stratejilerinin geliştirilmesinde, tüketici yorumlarının analizi önemli bir yer tutmaktadır. Bu çalışma, müşteri memnuniyetini artırmak, ürün ve hizmetlerin iyileştirilmesi için geri bildirim sağlamak ve potansiyel müşterilere güven vermek için kullanılabilir. Çalışma, e-ticaret sitelerindeki tüketici yorumlarının, pazarlama alanında ağızdan ağıza iletişimin modern bir yansıması olarak değerlendirilebileceğini ve bu yorumların stratejik bir şekilde nasıl kullanılabileceğini ortaya koymaktadır.

Yapılan bu çalışma, ürün hakkında bilgi veren yorumları filtreleme olanağı sunarak müşterilerin sipariş sürecinde daha bilinçli ve hızlı karar vermelerini sağlayacaktır. Bu sayede, kullanıcılar gereksiz ve yanıltıcı yorumlarla vakit kaybetmeden, doğrudan ürünün

kalitesi ve performansı hakkında net bilgilere ulaşabileceklerdir. Müşteriler aradıkları ürünle ilgili en doğru ve güvenilir bilgilere ulaşarak, ihtiyaçlarına en uygun ürünü seçme konusunda büyük bir avantaj elde edeceklerdir. Doğru ürün seçimi sayesinde müşteri memnuniyeti artacak ve olası iade süreçleri de azalacaktır. Ayrıca e-ticaret firmaları ürün hakkında bilgi veren müşteri yorumlarını dikkate alarak ürünlerinin zayıf ve eksik yönlerini görebileceklerdir. Böylece ürün geliştirme sürecine de büyük ölçüde katkı sağlayacaktır. Farklı sektörlerdeki firmalar ürünleri üzerinde daha iyi stratejiler geliştirebileceklerdir.

Sonuç olarak bu çalışma, e-ticaret platformları, pazarlamacılar ve akademisyenler için değerli bir kaynak olarak hizmet edebilir ve sektördeki yenilikçi yaklaşımların geliştirilmesine yardımcı olabilir. Ayrıca, bu alanda yapılan literatürün daha da zenginleştirilmesine katkıda bulunarak, dijital pazarlama uygulamalarının geleceğine ışık tutabilir.

## KAYNAKLAR

- Akgöz, E. ve Tengilimoğlu, E. (2022). Online Otel Yorumlarının Metin Madenciliği Teknikleri ile İncelenmesi: Bakü Otelleri Örneği. *Selçuk Turizm ve Bilişim Araştırmaları Dergisi*, 1(1), 1-8.
- Akın, A. A. Ve Akın, M. D. (2007). Zemberek, An Open Source NLP Framework For Turkic Languages. *Structure*, 10(2007), 1-5.
- Aksu, G. ve Doğan, N. (2019). Comparison of Decision Trees Used İn Data Mining. *Pegem Eğitim ve Öğretim Dergisi*, 9(4), 1183-1208.
- Aktaş, Ö., Coskuner, B. ve Soner, İ. 2021. Turkish Sentiment Analysis Using Machine Learning Methods: Application on Online Food Order Site Reviews. *Journal of Artificial Intelligence and Data Science*, (1) 1-10.
- Atabay, L. (2020). Otel Yorumlarının Metin Madenciliği Yöntemleri ile Karşılaştırmalı Analizi: Mayorka, Antalya, Şarm-El Şeyh Örneği. Yüksek Lisans Tezi, *Akdeniz Üniversitesi, Sosyal Bilimler Enstitüsü, Antalya*.
- Aydın, Ö. Ve Kantarcı, H. (2024). Türkçe Anahtar Sözcük Çıkarımında LSTM ve BERT Tabanlı Modellerin Karşılaştırılması. *Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 17(1), 9-18.
- Barushka, A. ve Hajek, P. (2019). Review Spam Detection Using Word Embeddings and Deep Neural Networks. In *Artificial Intelligence Applications and Innovations: 15th IFIP WG 12.5 International Conference, AIAI 2019, Hersonissos, Crete, Greece, May 24-26, 2019, Proceedings 15* (Pp. 340-350). Springer International Publishing.
- Bhatia, N. (2010). Survey of Nearest Neighbor Techniques. Arxiv Preprint Arxiv: 1007.0085. *IJCSIS Vol. 8, No. 2, 2010 Pp. 302-305*.
- Bounsaythip, C. ve Rinta-Runsala, E. (2001). Overview of Data Mining For Customer Behavior Modeling. VTT Information Technology Research Report, Version, 1, 1-53.
- Budak, İ. (2021). Veri ve Metin Madenciliği ile Hava Yolu İşletmelerinin Sosyal Medya Yorum ve Skorlarının Değerlendirilmesi. Doktora Tezi, *Pamukkale Üniversitesi, Sosyal Bilimler Enstitüsü, Denizli*.
- Chakravarthy, V. J., Kameswari, M., Mydeen, H. D. ve Seenivasan, M. (2021). Opinion Mining From Student Text Review For Choosing Better Online Courses. In *IOP Conference Series: Materials Science and Engineering (Vol. 1070, No. 1, P. 012067)*. IOP Publishing.
- Cop, R. ve Gümüş, N. (2009). Pazarlamada Ağızdan Ağıza İletişimin Tüketici Davranışlarındaki Rolü ve Bir Araştırma. *Ticaret ve Turizm Eğitim Fakültesi Dergisi*, 2, 179-202.
- Cortes, C. ve Vapnik, V. (1995). Support-Vector Networks. *Machine Learning*, 20, 273-297.

- Cüvitoğlu, A. (2017). Prediction of Synergistic Drug Combinations By Using MachineLearning Methods. Master's Thesis. *Dokuz Eylül University, Graduate School of Natural and Applied Sciences*, 17-21.
- Çabuk, M., Yücalar, F. ve Toçoğlu, M. A. (2023). Makine Öğrenmesi ile E-Ticaret Ürün Yorumlarının Otomatik Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, (52), 110-121.
- Çakmak, C. (2022). Platform Ekonomilerinin Gelecek Perspektifi ve Kullanıcı Yorumlarının Metin Madenciliği ile Analizi. Yüksek Lisans Tezi, *Sivas Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Sivas*.
- Çelik, E., Dal, D. ve Aydın, T. (2021). Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması. *Avrupa Bilim ve Teknoloji Dergisi*, (27), 880-889.
- Çeliksü, Z. (2017). Yabancı Dizilerin Alt Yazı ve Twitter Yorumlarının Metin Madenciliği ile İncelenmesi. Yüksek Lisans Tezi, *Mimar Sinan Güzel Sanatlar Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul*.
- Çiçek, R. ve Mürütsoy, M. (2014). İnternet Tüketicisinin Satın Alma Davranışlarının İncelenmesi Üzerine Bir Araştırma: Niğde Üniversitesi Akademik ve İdari Personeli Örneği. *C.Ü. İktisadi ve İdari Bilimler Dergisi*, Cilt 15, Sayı 2.
- Daşgın, R. (2023). Kitleleşen Çevrimiçi Ders Platformlarında Kurslara Yapılan Yorumların Metin Madenciliği Kullanılarak Duygu Analizi. Yüksek Lisans Tezi, *Aksaray Üniversitesi, Sosyal Bilimler Enstitüsü, Aksaray*.
- Demirhan, T. ve Demiralay, İ. (2023). Öğrencilerin Öğrenme Yönetim Sistemi Kullanım Alışkanlıklarının Pandas ile Analizi. *Trakya Eğitim Dergisi*, 13(3), 1736-1751.
- Dilki, G., ve Deniz Başar, Ö. (2020). Comparison Study of Distance Measures Using K-Nearest Neighbor Algorithm on Bankruptcy Prediction. *İstanbul Ticaret Üniversitesi, Fen Bilimleri Dergisi*, 19(38), 224–233.
- Ergün, K. (2012). Metin Madenciliği Yöntemleri ile Ürün Yorumlarının Otomatik Değerlendirilmesi. Doktora Tezi, *Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Sakarya*.
- Felbermayr, A. ve Nanopoulos, A. (2016). The Role of Emotions For The Perceived Usefulness In Online Customer Reviews. *Journal of Interactive Marketing*, 36(1), 60-76.
- Gedik, Y. (2021). E-Ticaret: Teorik Bir Çerçeve. *Ankara Üniversitesi Sosyal Bilimler Dergisi* / 12(1) / 2021 / 184 – 198.
- Go, A., R., Bhayani, L., Huang. (2009). Twitter Sentiment Classification Using Distant Supervision. *CS224N Project Report, Stanford(2009): S.1-12*.
- Guo, Y., Barnes, S. J. ve Jia, Q. (2017). Mining Meaning From Online Ratings and Reviews: Tourist Satisfaction Analysis Using Latent Dirichlet Allocation. *Tourism Management*, 59, 467-483.

- Gündüz, G. ve Cedimoğlu, H. (2019). Derin Öğrenme Algoritmalarını Kullanarak Görüntüden Cinsiyet Tahmini. *Sakarya University Journal of Computer and Information Sciences Vol. 2, No. 1, April 2019*.
- Hofmann, M. ve Klinkenberg, R. (2016). Rapidminer: Data Mining Use Cases and Business Analytics Applications. *CRC Press. Chapman & Hall/CRC. Data Mining and Knowledge Discovery Series*.
- Hu, L.Y., Huang, M.W., Ke, S.W., Tsai, C.F. (2016). The Distance Function Effect on kNearest Neighbor Classification for Medical Datasets. *Springer Plus, 5(1), 1-9*.
- Işık, N. (2019). Metin Madenciliği Yöntemleri ile E-Ticaret Markalarına Yönelik Sosyal Medya Yorumlarının Analizi. Yüksek Lisans Tezi, *Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, İstanbul*.
- İnternet: Özkurt, S. (2023). E-ticaret Nedir? Nasıl Yapılır? (9 Adımda Başlayın). URL: <https://ikas.com.tr/blog/e-ticaret-nasil-yapilir>, Son Erişim Tarihi: 20.12.2023.
- İnternet: Taşçı, E. ve Onan, A. (2017). K-En Yakın Komşu Algoritması Parametrelerinin Sınıflandırma Performansı Üzerine Etkisinin İncelenmesi. Erişim Adresi: <https://ab.org.tr/ab16/bildiri/102.pdf>, Son Erişim Tarihi: 21.02.2024.
- İnternet: TÜİK, (2023). Hanehalkı Bilişim Teknolojileri (BT) Kullanım Araştırması, 2023. URL: [https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-\(BT\)-Kullanim-Arastirmasi-2023-49407](https://data.tuik.gov.tr/Bulten/Index?p=Hanehalki-Bilisim-Teknolojileri-(BT)-Kullanim-Arastirmasi-2023-49407), Son Erişim Tarihi: 26.03.2024.
- Kalaycı, T. (2018). Kimlik Hırsızları Web Sitelerinin Sınıflandırılması İçin Makine Öğrenmesi Yöntemlerinin Karşılaştırılması. *Pamukkale Üniversitesi, Mühendislik Bilimleri Dergisi, 24 (5) , 870-878*.
- Kalikov, A. (2006). Veri Madenciliği ve Bir E-Ticaret Uygulaması. Yüksek Lisans Tezi, *Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara*.
- Kapukaya, N. (2023). Metin Madenciliği ve Duygu Analizi: IMDB En İyi Üç Filmin Twitter Yorumlarının Analizi. Yüksek Lisans Tezi, *Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul*.
- Khan, F. M., Khan, S. A., Shamim, K., Gupta, Y. ve Sherwani, S. I. (2023). Analysing Customers' Reviews and Ratings For Online Food Deliveries: A Text Mining Approach. *International Journal of Consumer Studies, 47(3), 953-976*.
- Khan, M., Ding, Q. ve Perrizo, W. (2002). K-Nearest Neighbor Classification on Spatial Data Streams Using P-Trees. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (Pp. 517-528)*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Kothari, R. ve Dong, M. I. N. G. (2001). Decision Trees For Classification: A Review and Some New Results. *Pattern Recognition: From Classical To Modern Approaches, 169-184*.

- Kuş, İ. (2023). Akademik Makale Öneri Sistemi. Yüksek Lisans Tezi, *Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Eskişehir*.
- Lei, S. ve Law, R. (2015). Content Analysis of Tripadvisor Reviews on Restaurants: A Case Study of Macau. *Journal of Tourism, 16(1)*.  
[https://ertr.tamu.edu/files/2015/02/SP03\\_ReviewSession\\_Lei.pdf](https://ertr.tamu.edu/files/2015/02/SP03_ReviewSession_Lei.pdf)
- Li, Q., Park, J. ve Kim, J. (2024). Impact of Information Consistency in Online Reviews On Consumer Behavior in The E-Commerce Industry: A Text Mining Approach. *Data Technologies and Applications, 58(1), 132-149*.
- Lohman, P. (2017). Portfolio Investment Strategy Based on Twitter Sentiment. Independent Thesis, *Umeå University, Faculty of Science and Technology*. id: *Diva2:1113099*.
- Niuniu, X. ve Yuxun, L. (2010). Notice of Retraction: Review of Decision Trees. In *2010 3rd International Conference on Computer Science and Information Technology (Vol. 5, Pp. 105-109)*. IEEE.
- Özbay, Ö. (2015). Veri Madenciliği Kavramı ve Eğitimde Veri Madenciliği Uygulamaları. *Inesjournal Uluslararası Eğitim Bilimleri Dergisi / The Journal of International Education Science, (5), 262-272*.
- Özkul, M. (2019). Metin Madenciliği Yöntemi ile Akıllı Ulaşım Sistemleri Algısının Duygu Analizi İle Ölçülmesi. Yüksek Lisans Tezi, *Bandırma Onyedli Eylül Üniversitesi Sosyal Bilimler Enstitüsü, Samsun*
- Öztürk, Y. (2021). Etkileşimli Tahta Kullanımına Yönelik Sosyal Medya Yorumlarının Metin Madenciliği Yöntemleri ile Analizi. Yüksek Lisans Tezi, *Sivas Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Sivas*.
- Pilavcılar, İ. F. (2007). Metin Madenciliği ile Metin Sınıflandırma. Yüksek Lisans Tezi. *Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul, 23-30*.
- Rahmawati, R., Fuadi, W. ve Afrillia, Y. (2024). Cosmetic Shop Sentiment Analysis on TikTok Shop Using the Support Vector Machine Method. *International Journal of Engineering, Science and Information Technology, 4(2), 31-37*.
- Quinlan, J. R. (1993). C4.5: Programs for Machine Learning, Morgan Kauffman Publishers, *San Mateo, CA, USA*.
- Sarıkoç, F. (2012). Meme Kanseri Biyopsi Örneklerinde Östrojen Hormonu Alıcı Varlığının Zeki Yöntemlerle Tespiti. Doktora Tezi, *Erciyes Üniversitesi, Fen Bilimleri Enstitüsü, Kayseri, 148s*.
- Savaş, S., Topaloğlu, N. ve Yılmaz, M. (2012). Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi, Fen Bilimleri Dergisi, Yıl:11 Sayı: 21 Bahar 2012 s. 1-23*.

- Sayın Okatan, B. (2023). Dijital Bankacılık Uygulamalarına İlişkin Müşteri Yorumlarının Metin Madenciliği Yöntemleri ile İncelenmesi. Doktora Tezi, *Gümüşhane Üniversitesi, Lisansüstü Eğitim Enstitüsü, Gümüşhane*.
- Sönmez, N. (2017). Çevrimiçi Yorumların Metin Madenciliği ile Analizi: İstanbul'daki Alışveriş Merkezleri Üzerine Bir Çalışma. Yüksek Lisans Tezi, *İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul*.
- Sparks, B. A., So, K. K. F., ve Bradley, G. L. (2016). Responding to Negative Online Reviews: The Effects of Hotel Responses on Customer Inferences of Trust and Concern. *Tourism Management*, 53, 74-85.
- Şeker, A., Diri, B. ve Balık, H. H. (2017). Derin Öğrenme Yöntemleri ve Uygulamaları Hakkında Bir İnceleme. *Gazi Mühendislik Bilimleri Dergisi* 3 (3). (2017) 47-64.
- Şeker, S. E. (2015). Metin Madenciliği (Text Mining). *Yönetim Bilişim Sistemleri Ansiklopedisi*, 2(3), 30-32.
- Tığa Öztürk, G. (2020). Kruvaziyer Turistlere Ait Çevrimiçi Seyahat Yorumlarının Metin Madenciliği Yöntemi ile Değerlendirilmesi. Doktora Tezi, *Aydın Adnan Menderes Üniversitesi, Sosyal Bilimler Enstitüsü, Aydın*.
- Tomak, L. ve Bek, Y. (2009). İşlem Karakteristik Eğrisi Analizi ve Eğri Altında Kalan Alanların Karşılaştırılması. *Journal of Experimental and Clinical Medicine*, 27(2).
- Tuna, M. F. (2019). Çevrimiçi Yorum ve Şikâyetlerin Otel İşletmeleri Üzerinden Duygu Analizi ile İncelenmesi. Doktora Tezi, *Erciyes Üniversitesi, Sosyal Bilimler Enstitüsü, Kayseri*.
- Uysal, İ. ve Bilen, M. (2017). Rapidminer ile Biyolojik Verilerin Sezgisel Algoritmalar Kullanılarak Sınıflandırılması. *1st International Turkish World Engineering and Science Congress in Antalya, December 7-10, 2017, Turkey*.
- Yakut, E., Elmas, B. ve Yavuz, S. (2014). Yapay Sinir Ağları ve Destek Vektör Makineleri Yöntemleriyle Borsa Endeksi Tahmini. *Süleyman Demirel Üniversitesi, İktisadi ve İdari Bilimler Fakültesi Dergisi*, Y.2014, C.19, S.1, s.139-157.
- Yaman, U. C. (2022). Metin Madenciliği Teknikleri ile Türkçe Müşteri Yorumlarının Analizi. Yüksek Lisans Tezi, *Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Eskişehir*.
- Yıldız, B. ve Ağdeniz, Ş. (2018). Muhasebede Analiz Yöntemi Olarak Metin Madenciliği. *Muhasebe Bilim Dünyası Dergisi*, Haziran 2018; 20(2); 286-315.
- Yılmaz, C. M. (2019). Sosyal Ağ Analizi ve Metin Madenciliği Kullanarak Yanıltıcı Yorumları Tespit Etme. Yüksek Lisans Tezi, *Orta Doğu Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Ankara*.

- Yılmaz G. (2021). E-Ticaret Sistemlerinde Yapılan Ürün Yorumlarının Metin Madenciliği Uygulaması ile İncelenmesi. Yüksek Lisans Tezi, *İstanbul Aydın Üniversitesi, Lisansüstü Eğitim Enstitüsü, İstanbul*.
- Zhang, S., Lopez, D., Lam, L., Yenumula, A. ve Vu, T. (2023). A Comparative Analysis of Text Mining Methodologies for Online Consumer Reviews. *Journal of Applied Business & Economics*, 25(7).
- Zhang, Z., Ye, Q., Zhang, Z. ve Li, Y. (2011). Sentiment Classification of Internet Restaurant Reviews Written in Cantonese. *Expert Systems With Applications*, 38(6), 7674-7682. <https://doi.org/10.1016/j.eswa.2010.12.147>.



## DİZİN

---

### A

Abstract · vi, viii

Algoritma · v, ix, x, xi, 4, 5, 6, 10, 12, 13, 14, 15, 16, 17, 18, 19, 21, 24, 25, 26, 27, 28, 29, 31, 32, 38, 41, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 59, 60, 62

Alışveriş · v, xii, 1, 2, 3, 13, 17, 23, 56, 62

Analiz · v, 4, 5, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 29, 37, 38, 39, 40, 41, 43, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62

---

### B

Bilgi keşfi süreci · xi, 5, 6

---

### Ç

Çevrimiçi · v, xii, 3, 14, 16, 18, 20, 21, 22, 56, 59, 62

Çizelge · x, 23, 32, 40, 52, 54

---

### D

Derin öğrenme · v, viii, ix, xi, 7, 10, 16, 18, 24, 30, 31, 44, 48, 49, 50, 52, 53, 54, 55, 56, 60, 62

Destek vektör makinesi · v, viii, ix, xii, 22, 24, 29, 50, 62

---

### E

E-ticaret · 1, 2, 8, 14, 16, 17, 22, 56, 57, 59, 60, 63

---

### F

Formül · 5, 26, 36

---

### İ

İlgili · v, vii, 2, 7, 15, 17, 18, 19, 20, 23, 26, 37, 54, 56, 57

İlgisiz · 23, 54, 56

---

### K

Karar ağacı · v, viii, ix, xi, 16, 24, 27, 28, 29, 44, 46, 47, 48, 52, 53, 54, 55, 56

K-en yakın komşu · v, viii, ix, 14, 24, 25, 45, 60

---

### M

Makine öğrenmesi · v, 4, 5, 7, 14, 15, 17, 18, 21, 25, 28, 29, 30, 32, 59, 60

Metin madenciliği · v, viii, xi, 3, 4, 7, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 21, 22, 24, 37, 42, 53, 54, 56, 58, 59, 60, 61, 62, 63

Müşteri · v, 1, 2, 3, 8, 11, 15, 16, 17, 19, 20, 21, 22, 55, 56, 57, 62

---

**N**

Naive Bayes · v, vi, viii, ix, xi, 14, 16, 19, 21, 24, 26, 27, 44, 45, 52, 53, 54, 55, 56

---

**O**

Online · v, vi, 1, 3, 17, 23, 56, 58, 59, 60, 61, 62, 63

---

**Ö**

Ön işleme · ix, 6, 9, 11, 12, 17, 24, 37, 38, 39, 40, 41

---

**P**

Parametre · xi, 15, 25, 32, 38, 39, 40, 41, 42, 45, 46, 47, 48, 49, 50, 52, 60  
Performans · ix, xi, 2, 8, 14, 15, 16, 17, 18, 19, 20, 21, 23, 25, 26, 28, 29, 31, 32, 33, 34, 35, 36, 37, 38, 40, 42, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 57, 60

---

**R**

RapidMiner · xi, 24, 37, 38, 39, 40, 41, 42, 43, 44, 50, 60, 62

---

**S**

Sembol · 38  
Sınıflandırma · v, x, xi, 5, 7, 9, 10, 11, 12, 14, 16, 19, 20, 22, 24, 25, 26, 27, 28, 29, 30, 32, 33, 34, 36, 38, 40, 41, 44, 51, 52, 53, 54, 55, 59, 60, 61

---

**T**

Teknoloji · v, 1, 2, 4, 5, 8, 21, 30, 59, 60  
Terim frekansı · v, ix, 42, 43, 56  
Ters doküman frekansı · v, ix, 42, 43, 56  
Trendyol · v, vi, 23, 56  
Tüketici · v, 1, 2, 3, 13, 20, 22, 23, 53, 54, 55, 56, 58, 59

---

**Ü**

Ürün · v, 1, 2, 3, 9, 11, 13, 14, 16, 17, 22, 23, 33, 42, 43, 54, 56, 57, 59, 63

---

**V**

Veri madenciliği · viii, xi, 4, 5, 6, 7, 8, 9, 15, 20, 24, 59, 60, 61

---

**Y**

Yapay zeka · 5, 7, 24, 30, 31  
Yorum · v, xi, 2, 3, 5, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 26, 29, 37, 43, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63



**TEKNOVERSİTE**



teknoversite **AYRICALIĞINDASINIZ**

**İSTE**

