

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

**METİN MADENCİLİĞİ İLE  
MÜŞTERİ GERİ BİLDİRİM ANALİZİ:  
MOBİLYA ŞİRKETLERİ ÖRNEĞİ**

**Hazırlayan  
Sevgi PEHLİVAN**

**Danışman  
Prof. Dr. Hülya TORUN**

**Yüksek Lisans Tezi**

**Ocak 2025  
KAYSERİ**



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

**METİN MADENCİLİĞİ İLE  
MÜŞTERİGERİ BİLDİRİM ANALİZİ:  
MOBİLYA ŞİRKETLERİ ÖRNEĞİ  
(Yüksek Lisans Tezi)**

**Hazırlayan  
Sevgi PEHLİVAN**

**Danışman  
Prof. Dr. Hülya TORUN**

**Ocak 2025  
KAYSERİ**

## **BİLİMSEL ETİĞE UYGUNLUK**

Bu alıřmadaki tm bilgilerin, akademik ve etik kurallara uygun bir řekilde elde edildiđini beyan ederim. Aynı zamanda bu kural ve davranıřların gerektirdiđi gibi, bu alıřmanın znde olmayan tm materyal ve sonuları tam olarak aktardıđımı ve referans gsterdiđimi belirtirim.

Sevgi PEHLİVAN

İmza

**“Metin Madenciliđi ile Müşteri Geri Bildirim Analizi: Mobilya Şirketleri Örneđi”**  
adlı Yüksek Lisans Tezi, Erciyes Üniversitesi Lisansüstü Tez Önerisi ve Tez Yazma  
Yönergesi’ ne uygun olarak hazırlanmıştır.

**Hazırlayan**

Sevgi PEHLİVAN

İmza

**Danışman**

Prof. Dr. Hülya TORUN

İmza

**Endüstri Mühendisliđi ABD Başkanı**

Prof. Dr. Lale ÖZBAKIR

İmza

## TEŐEKKÜR

Yüksek lisans ders dönemi ve tez yazma sürecim boyunca çalışmalarım adına yön gösterici olan, destekleyen ve emeklerini esirgemeyen değerli hocam Prof. Dr. Hülya Torun'a çok teşekkür ederim.

Ayrıca çalışma kapsamında hem kendimi geliştirmem adına hem de şirketin verilerini iyileştirmem amacıyla tarafıma paylaşım sağlayan ve tezimi destekleyen çalışma arkadaşlarımla aileme çok teşekkür ederim.

Sevgi PEHLİVAN

Ocak 2025, KAYSERİ

## METİN MADENCİLİĞİ İLE MÜŞTERİ GERİ BİLDİRİM ANALİZİ: MOBİLYA ŞİRKETLERİ ÖRNEĞİ

Sevgi PEHLİVAN

Erciyes Üniversitesi, Fen Bilimleri Enstitüsü  
Yüksek Lisans Tezi, Ocak 2025  
Danışman: Prof. Dr. Hülya TORUN

### ÖZET

Günümüz işletmeleri, müşteri memnuniyetini artırmak, müşteri sadakatini sağlamak ve müşteri portföyünü genişletmek için bir arayış içerisinde. Müşterilerini odak noktasına alarak hareket etme süreci, işletmelere rekabetçi avantajlar sağlamakta ve stratejilerinde kritik bir rol oynamaktadır. Bu bağlamda işletmeler, müşteri geri bildirimlerini etkili şekilde analiz etmeye odaklanmaktadır. Veri madenciliği ve metin madenciliği, büyük veri kümelerinin analizinde etkili yöntemlerdir. Bu yöntemler ile işletmenin, müşteri ile temas ettiği tüm noktalarda yer alan verilerin analiz edilmesi ve değerlendirilmesi mümkün hale gelmektedir. Bu araçların kullanımı ile müşteri geri bildirimleri analiz edilmekte, müşteri memnuniyetini etkileyen faktörler belirlenmekte, müşteri ilişkileri yönetim sürecinin etkinliği değerlendirilmekte ve çeşitli stratejiler geliştirilmektedir. Bu çalışmada, üç farklı mobilya şirketine ait müşteri geri bildirimleri, metin madenciliği ve veri madenciliği yöntemleriyle analiz edilmiştir. Analiz, müşteri hizmetleri biriminden elde edilen verilerin metin madenciliği araçlarıyla işlenmesi ve Gizli Dirichlet Ayırımı (Latent Dirichlet Allocation-LDA) yöntemiyle geri bildirimlerin konu dağılımları ile çağrı tipleri arasındaki uyumunun incelenmesi üzerine odaklanmıştır. Elde edilen detaylı veri kümesi, sınıflandırma algoritmalarıyla işlenmiş ve yüksek doğrulukla çalışan bir çağrı atama modeli geliştirilmiştir. Önerilen model, literatürdeki benzer çalışmalardan farklı olarak müşteri hizmetleri yönetim sürecine yönelik sistematik iyileştirmeler sunmuş ve farklı sınıflandırıcılarla test edilerek yüksek performans sonuçları elde etmiştir.

**Anahtar Kelimeler:** Konu Modelleme, Metin Madenciliği, Veri Madenciliği, Sınıflandırma.

# ANALYSIS OF CUSTOMER FEEDBACK USING TEXT MINING: A CASE STUDY OF FURNITURE COMPANIES

Sevgi PEHLIVAN

Erciyes University, Graduate School of Natural and Applied Sciences  
Master Thesis, January 2025  
Supervisor: Prof. Dr. Hülya TORUN

## ABSTRACT

Today's businesses are in pursuit of increasing customer satisfaction, ensuring customer loyalty, and expanding their customer portfolios. The process of acting with a customer-centric approach provides competitive advantages to businesses and plays a critical role in their strategies. In this context, businesses focus on effectively analyzing customer feedback. Data mining and text mining are effective methods for analyzing large datasets. Through these methods, it becomes possible to analyze and evaluate data collected at all touchpoints where the business interacts with its customers. By utilizing these tools, customer feedback can be analyzed, factors affecting customer satisfaction can be identified, the effectiveness of customer relationship management processes can be evaluated, and various strategies can be developed. In this study, customer feedback from three different furniture companies was analyzed using text mining and data mining methods. The analysis focused on processing the data obtained from the customer service department with text mining tools and examining the alignment between the topic distributions of feedback and call types through the Latent Dirichlet Allocation method. The detailed dataset obtained was processed with classification algorithms to develop a call assignment model that operates with high accuracy. The proposed model differs from similar studies in literature by providing systematic improvements for customer service management processes and has achieved high performance results when tested with different classifiers.

**Keywords:** Topic Modeling, Text Mining, Data Mining, Classification.

## İÇİNDEKİLER

### METİN MADENCİLİĞİ İLE MÜŞTERİ GERİ BİLDİRİM ANALİZİ: MOBİLYA ŞİRKETLERİ ÖRNEĞİ

BİLİMSEL ETİĞE UYGUNLUK.....	ii
YÖNERGEYE UYGUNLUK.....	iii
KABUL VE ONAY .....	iv
TEŞEKKÜR.....	v
ÖZET.....	vi
ABSTRACT.....	vii
İÇİNDEKİLER .....	viii
KISALTMALAR .....	xi
TABLolar LİSTESİ.....	xii
ŞEKİLLER LİSTESİ .....	xiii
GİRİŞ .....	1

#### 1. BÖLÜM GENEL BİLGİLER ve LİTERATÜR ÇALIŞMASI

1.1. Problem Durumu .....	8
1.2. Araştırmanın Amacı .....	9
1.3. Araştırmanın Önemi.....	10

#### 2. BÖLÜM YÖNTEM VE MATERYAL

2.1. Veri Madenciliği.....	11
2.2. Veri Madenciliği Türleri.....	14
2.2.1. Tanımlayıcı Metotlar.....	14
2.2.1.1. Kümeleme.....	14

2.2.1.2. Birliktelik Kuralı Analizi.....	15
2.2.2. Tahminleyici Metotlar.....	15
2.2.2.1. Regresyon.....	15
2.2.2.2. Sınıflandırma.....	16
2.3. Metin Madenciliği.....	22
2.3.1. Metin Madenciliği Süreci.....	24
2.3.2. Doğal Dil İşleme.....	27
2.3.3. Bilgi Çıkarımı.....	28
2.3.4. Duygu Analizi .....	29
2.3.5. Metin Kümeleme.....	30
2.3.6. Metin Görselleştirme.....	30
2.3.7. Konu Modelleme.....	31
2.3.7.1. LDA Algoritması.....	33

### 3. BÖLÜM

#### METİN MADENCİLİĞİ İLE MÜŞTERİ GERİ BİLDİRİM ANALİZİ: MOBİLYA ŞİRKETLERİ ÖRNEĞİ

3.1. Uygulamanın Konusu Amacı ve Önemi .....	36
3.2. Uygulamanın Kapsamı.....	37
3.3. Uygulamanın Evreni ve Örneklemi.....	40
3.4. Metin Madenciliği Süreci.....	46
3.4.1. Metin Ön İşleme.....	47
3.4.2. Özellik Çıkarımı.....	49
3.4.2. Konu Modelleme.....	55
3.5. Sınıflandırma.....	62
3.5.1. Naive Bayes.....	63
3.5.2. Yapay Sinir Ağları.....	64
3.5.3. Destek Vektör Makinesi.....	65
3.5.4. Karar Ağaçları.....	66
3.5.5. LightGBM Model.....	67
3.5.6. AdaBoost Model.....	68
3.5.7. Rastgele Orman.....	69

## 4. BÖLÜM SONUÇ ve ÖNERİLER

KAYNAKÇA .....	74
ÖZGEÇMİŞ.....	78



## KISALTMALAR

LDA: Latent Dirichlet Allocation (Gizli Dirichlet Ayırımı)

TF-IDF: Term Frequency-Inverse Document Frequency (Terim Frekansı-Ters Doküman Frekansı)

NMF: Non-Negative Matrix Factorization

LSI: Latent Semantic Indexing (Gizli Anlamsal İndeksleme)

NLP: Natural Language Processing (Doğal Dil İşleme)

SVM: Support Vector Machine (Destek Vektör Makinesi)

## TABLolar LİSTESİ

Tablo 1. Karmaşıklık Matrisi .....	20
Tablo 2. Birinci Örnek Veri Seti .....	41
Tablo 3. İkinci Örnek Veri Seti.....	45
Tablo 4. Metin Ön İşleme Adımları .....	48
Tablo 5. Birinci Kelime Listesi Örneği.....	49
Tablo 6. İkinci Kelime Listesi Örneği.....	50
Tablo 7. Marka Bazlı Kelime Listeleri .....	53
Tablo 8. Anahtar Kelimeler.....	56
Tablo 9. LDA ile Olasılıklı Atama Örneği.....	57
Tablo 10. Çağrı Dağılım Olasılıkları .....	60
Tablo 11. Çağrı Tipi Eşleşmeleri .....	61
Tablo 12. Naive Bayes Sonuçları.....	63
Tablo 13. Yapay Sinir Ağları Sonuçları.....	64
Tablo 14. Destek Vektör Makinesi Sonuçları .....	65
Tablo 15. Karar Ağaçları Sonuçları .....	66
Tablo 16. LightGBM Sonuçları .....	67
Tablo 17. AdaBoost Model Sonuçları.....	68
Tablo 18. Rastgele Orman Sonuçları .....	69
Tablo 19. Karşılaştırma Tablosu .....	70

## ŞEKİLLER LİSTESİ

Şekil 1. Veri Madenciliği Süreci.....	13
Şekil 2. Veri Madenciliği Türleri.....	14
Şekil 3. Metin Madenciliği Süreci.....	24
Şekil 4. LDA Modeli İşleyiş Süreci.....	34
Şekil 5. Uygulama Süreç Akış Şeması.....	38
Şekil 6. Veri Kaynak Dağılımı.....	43
Şekil 7. Marka Dağılımı.....	43
Şekil 8. Çağrı Tipi Dağılımı.....	44
Şekil 9. Metin Ön İşleme Süreci Akış Şeması.....	47
Şekil 10. Mevcut Sınıflandırma Veri Seti.....	62
Şekil 11. Yeni Sınıflandırma Veri Seti.....	62
Şekil 12. Çağrı Atama Modeli Örneği.....	71

## GİRİŞ

Artan küresel rekabet koşulları altında, müşteri odaklılık, işletmeler için vazgeçilmez bir strateji haline gelmiştir. Müşteri memnuniyetini sürekli izlemek ve müşterilerin ihtiyaçlarına hızlı ve etkili çözümler sunabilmek, işletmelerin pazar konumlarını güçlendirmekte ve rekabet avantajı sağlamalarına katkıda bulunmaktadır. Dijital dönüşümün hızlanması ve Endüstri 4.0'ın etkisiyle, müşteri geri bildirimlerinin etkin bir şekilde analiz edilmesi ve bu verilere dayalı stratejik kararların alınması, işletmelerin uzun vadeli başarısı için temel bir unsur olarak öne çıkmaktadır. Bu doğrultuda, müşteri verilerinin etkili bir şekilde yönetilmesi için veri odaklı stratejilere her zamankinden daha fazla ihtiyaç duyulmaktadır.

Müşteri geri bildirimleri, genellikle işletmelerin çağrı merkezleri veya web siteleri üzerinden; talep, şikâyet, öneri ve soru gibi çeşitli amaçlarla iletilir. Çoğunlukla yapısal olmayan ve doğal dilde ifade edilen bu geri bildirimler, ürün veya hizmet eksiklikleri, hatalar, rakiplerle karşılaştırmalar, risk faktörleri ve iyileştirme fırsatları hakkında değerli ve gizli bilgiler içermektedir. Büyük veri analitiği, veri madenciliği ve metin madenciliği araçları, işletmelerin bu yapısal olmayan geniş veri yığınlarını analiz ederek anlamlı içgörüler elde etmesine olanak sağlamaktadır. Bu analizler, müşteri şikâyetleri ve taleplerinden stratejik karar alma süreçlerine katkı sunacak önemli sonuçlar çıkarmaktadır.

Günümüzde sosyal medya kullanımının yaygınlaşmasıyla bireyler, herhangi bir konudaki duygu ve düşüncelerini milyonlarca kişiyle anlık olarak paylaşabilmektedir. Bu paylaşımlar arasında müşterilerin alışveriş deneyimleri de sıklıkla yer almakta ve bu deneyimler, diğer kullanıcıların satın alma kararlarını etkileyebilmektedir. Ayrıca, kazanılmış müşterilerin kaybına da yol açabilmektedir. Özellikle Twitter, Facebook, alışveriş ve şikâyet siteleri gibi platformlar, müşterilerin ürün veya hizmetlerle ilgili şikâyetlerini dile getirmesi için etkin bir ortam sunmaktadır. Sosyal medya, yalnızca bir

Pazarlama platformu değil, aynı zamanda müşterilerle derin bağlar kurarak davranışlarını etkileyen ve marka sadakatini artıran güçlü bir araçtır [1]. Bu durum, işletmelerin müşteri geri bildirimlerini etkili bir şekilde işleyebilmesi ve yönetebilmesi amacıyla müşteri hizmetleri departmanlarının kurulması, çağrı merkezi sistemlerinin oluşturulması ve servis hizmetlerine yoğunlaşılması gibi uygulamalarla sonuçlanmıştır.

Bu araştırma, çevrimiçi platform olan Şikayetvar.com ve çağrı merkezi kayıtlarından alınan mobilya alışveriş deneyimlerine ait müşteri geri bildirimlerini metin madenciliği ve veri madenciliği yöntemleri ile analiz etmeyi amaçlamıştır. Çalışma, müşteri hizmetleri süreçlerinin etkinliğinin değerlendirilmesi ve marka bazlı iyileştirmeye açık alanların tespiti amacıyla dört bölümde ele alınmıştır.

Birinci bölümde genel bilgilere ve literatür taramasına yer verilmiştir. Literatür araştırmasıyla birlikte mevcut çalışmalara dair bir bakış açısı kazandırılmıştır. Problemin tanımlanması, araştırmanın amacı ve araştırmanın önemi, müşteri hizmetleri yönetim süreciyle birlikte ele alınarak bütünsel bir yaklaşım benimsenmiştir.

İkinci bölümde yöntem ve materyal açıklanmıştır. Veri madenciliği kavramı ve veri madenciliğinin bir alt dalı olan metin madenciliği ele alınmıştır. Veri madenciliğinin türleri, teknikleri ve uygulama alanları detaylandırılmış; metin madenciliği ile arasındaki ilişkiler ve farklar açıklanmıştır.

Üçüncü bölümde ise araştırma konusu olan müşteri geri bildirim analizi için elde edilen 9.930 adet müşteri yorumu, Python programlama dili kullanılarak metin madenciliği yöntemleriyle işlenmiş ve LDA yöntemiyle konu modellemesi uygulanmıştır. Konu modellemesi sonucunda elde edilen detaylı verilerle müşteri hizmetleri çağrı atama süreci değerlendirilmiş; sürecin etkinliği ve atanmış çağrıların yorumlarla uyumu incelenmiştir. Bu aşamadan sonra metin madenciliği araçları ile ulaşılan belge-kelime matrisi özellikleri, konu modellemesi sonucunda ulaşılan yeni çağrı tipleriyle birleştirilerek elde edilen veriye sınıflandırma modelleri uygulanmıştır. Sınıflandırma performansı en yüksek algoritma ile yeni gelen çağrılar için çağrı atama modeli önerilmiş; bu sayede gelen müşteri geri bildirimlerinin çağrı tipleri arasındaki dağılımı ortaya çıkarılmış ve sürecin iyileştirilmesi amaçlanmıştır.

Son bölümde ise, çalışma sonuçları ve gelecek çalışmalara dair öneriler sunulmuştur.

## 1. BÖLÜM

### GENEL BİLGİLER VE LİTERATÜR ÇALIŞMASI

Müşteri hizmetleri yönetimi, bir işletmenin mevcut ve potansiyel müşterileriyle olan ilişkilerini geliştirmek, müşteri ihtiyaçlarını karşılamak, müşteri memnuniyetini artırmak, müşteri bağlılığını sağlamak, olumlu marka imajını korumak ve temelde operasyon verimliliğini sağlamak amacıyla uyguladığı stratejik bir süreçtir. Bu süreç, müşterinin işletme ile ilk temas anından başlayarak satış sonrası hizmetleri ve uzun vadeli ilişki yönetimini kapsar. Modern işletmeler için müşteri hizmetleri yönetimi, yalnızca bir destek fonksiyonu değil, aynı zamanda rekabet avantajı sağlayan stratejik bir araç olarak konumlanmıştır.

Literatürde, müşteri hizmetleri yönetim süreci alanında yapılan çalışmalarda, müşteri hizmetleri yönetiminin sadece bir destek fonksiyonu olmaktan ziyade, işletmelerin stratejik karar alma süreçlerini yönlendiren bir yapı taşına dönüştüğü gözlemlenmiştir. Parasuraman ve arkadaşları tarafından geliştirilen "SERVQUAL" modeli, hizmet kalitesinin müşteri beklentilerine uygunluğunu ölçmek için kullanılan bir çerçeve sunmaktadır. Hizmet kalitesinin, müşteri beklentileri ile algılanan performans arasındaki fark üzerinden beş boyut (güvenilirlik, yanıt verebilirlik, empati, fiziksel görünüm, güven) ile değerlendirilmesine odaklanır ve işletmeler için hizmet kalitesini ölçmek adına kapsamlı bir çerçeve sunar. SERVQUAL modeli, farklı sektörlerde geniş bir uygulama alanı bulmuştur. Yücel, [2] tarafından yapılan çalışmada, SERVQUAL modeli kullanılarak bankacılık sektöründe hizmet kalitesi analiz edilmiştir. Araştırma, Malatya ilindeki bir özel bankanın şubesinde rastgele seçilen müşterilerle yapılan anketlere dayanmaktadır. SERVQUAL ölçeği kapsamında bankanın sunduğu hizmet kalitesi beş boyutta değerlendirilmiş ve hizmet kalitesinin müşteri beklentilerini karşılamadığı tespit edilmiştir. Araştırmada, güvenilirlik boyutunun müşteri memnuniyeti açısından en kritik unsur olduğu bulunmuş, diğer

Boyutlarda da eksiklikler gözlemlenmiştir. Benzer şekilde, Chen vd. [3] çalışmalarında müşteri geri bildirimlerinin işletmeler için stratejik bir değer kaynağı olarak kullanılabileceğini vurgulamışlardır. Çalışma, çevrimiçi tüketici incelemelerinin pazarlama iletişimde yeni bir bileşen olarak değerini ve stratejik etkisini araştırmaktadır. Çalışma, tüketici geri bildirimlerinin işletmelerin karar alma süreçlerine nasıl entegre edilebileceğine ve müşteri memnuniyeti ile bağlılığını artırmak için bu geri bildirimlerden nasıl yararlanılabileceğine odaklanmıştır. Ayrıca, müşteri memnuniyetine etki eden faktörlerin belirlenmesi ve işletmelerin bu faktörlere yönelik iyileştirme stratejileri geliştirmesi açısından veri analitiğinin önemini ortaya koymuştur.

Özellikle dijitalleşme ve büyük veri analitiğinin gelişmesiyle birlikte, müşterilerin davranışlarını ve ihtiyaçlarını daha derinlemesine anlamak mümkün hale gelmiştir. Bu bağlamda literatürde, müşteri hizmetleri yönetimi süreçlerinin artan dijitalleşme ile daha fazla araştırma konusu haline gelmesi ve dijitalleşmenin katkısı ile uygulanan örnek çalışmalar ön plana çıkmaktadır. Costa vd. [4] dijitalleşmenin küçük ve orta ölçekli işletmelerde müşteri hizmetlerini geliştirme potansiyelini araştırmıştır. Büyük veri analitiği, talep tahmini ve süreç iyileştirme için stratejik bir araç olarak vurgulanmıştır. Chernova vd. [5] çalışmalarında, dijitalleşmenin perakende hizmetlerinde müşteri değerini nasıl artırdığını analiz etmektedir. Büyük veri analitiği, müşteri davranışlarını daha iyi anlamak ve müşteri deneyimini optimize etmek için kullanılmaktadır. Çalışma, dijitalleşmenin tüketici memnuniyetini artırmada ve kişiselleştirilmiş hizmetler sunmada oynadığı rolü detaylandırmakta, ayrıca bu süreçte işletmelerin operasyonel verimliliğini artıracak yöntemler önermektedir. Varol vd. [6] çalışmalarında, dijitalleşmenin müşteri deneyimi ve müşteri sadakati üzerindeki etkilerini incelemektedir. Çalışmada, bankacılık sektöründe dijital araçların ve büyük veri analitiğinin müşterilerin hizmet algısı ve işletmelere olan bağlılığı üzerindeki etkileri araştırılmıştır. Çalışma, dijitalleşmenin müşteri deneyimi ve sadakat üzerindeki olumlu etkilerini net bir şekilde ortaya koymakta ve bu süreçte büyük veri analitiğinin önemine dikkat çekmektedir. Araştırmacılar, işletmelerin dijitalleşme sürecinde müşteri ihtiyaçlarını ve memnuniyetini ön planda tutmaları gerektiğini önermektedir. Kobayashi vd. [7] çalışmalarında, yapay zekâ destekli şikâyet yönetimi ile metin madenciliği tekniklerinin bir araya gelerek müşteri deneyimini iyileştirme potansiyeline sahip olduğunu göstermiştir. Yapay zekâ ve doğal dil işleme tekniklerini kullanarak 50 yıl boyunca

sağlık alanında yapılan arařtırmaları analiz etmiřtir. alıřma, zellikle metin madencilięi yntemlerini kullanarak mřteri deneyimini ve saęlık hizmetleri sektrndeki geri bildirimlerin deęerlendirilmesini ele almıřtır. Generatif yapay zek araları ile elde edilen verilerin, hizmet iyileřtirme ve mřteri ihtiyalarının belirlenmesi gibi alanlarda nasıl kullanılabilceęini vurgulamaktadır.

Dijitalleřme ve byk veri analitięi, mřteri hizmetleri ynetiminde yeni fırsatlar sunsa da literatr incelendięinde dijitalleřmenin mobilya sektrne zel etkilerinin yeterince incelenmedięi grlmektedir. Bu durum, dijitalleřmenin sektrel baęlamda ne lde rekabet avantajı saęladıęı ve mřteri memnuniyetini nasıl etkiledięi sorularını aık bırakmaktadır. Metin madencilięi, veri madencilięi ve veri analitięi yaklařımları mřteri hizmetleri ynetimini daha etkin ve stratejik hale getirmek iin kullanılan gl aralardır. Literatrde bu alanda uygulanan alıřmalar, metin madencilięi yaklařımlarından; belge iřleme, konu modelleme, bilgi ıkarımı ve duygu analizi alanlarında yoęunlařmıřtır.

Blei vd. [8] tarafından geliřtirilen LDA yntemi, metinlerden anlamlı temaların ıkarılmasında yaygın olarak kullanılan bir model haline gelmiřtir. Bu yntem, zellikle byk veri setlerinde kullanılmak zere her bir belgeyi belirli bir konu daęılımına gre daęıtacak ve her konuyu kelime daęılımı ile tanımlayacak řekilde modellenmiřtir. Arařtırmacılar bu yntemin, metinlerden anlamlı temaların ıkarılmasında yaygın bir ara olarak kullanılmasını amalamıřlardır. Hong vd. [9] alıřmalarında LDA modeli ile kısa metinlerden (tweetlerden) kendilięinden anlamlı konuları ıkarmak ve metin yıęınları iindeki gizli konu yapılarını belirleyerek bu konuları kullanıcılara veya verilere daęıtmayı amalamıřlardır. Ayrıca Twitter gibi sosyal medya platformlarında konu modellemesi alıřmalarının yapılabilirlięi zerine katkı saęlamıřlardır. Budak vd. [10] alıřmalarında otel hizmetlerine ynelik evrimii mřteri yorumlarını analiz etmek iin LDA yntemini kullanmıřtır. LDA ile yorumlardan elde edilen temaların aęırlıklı olarak hizmet kalitesi, temizlik, konfor ve alıřan davranıřlarıyla ilgili olduęunu ortaya koymuřlardır. Kaynar vd. [11] alıřmalarında evrimii platformlarda otellerle ilgili mřteri yorumlarını analiz etmiř ve bu geri bildirimlerden anlamlı igrler ıkarmıřtır. alıřmada, duygu analizi ve metin madencilięi yntemleri kullanarak mřteri deneyimini anlamaya ynelik detaylı bir yaklařım sunulmaktadır. Yorumların yaklařık %75'inin olumlu olduęu tespit edilmiř, en ok memnuniyet

belirten konular arasında otelin konumu ve personelin ilgisi yer alırken, yemek kalitesi ve oda temizliği gibi alanlarda eksiklikler öne çıkmıştır. LDA ve Terim Frekansı-Ters Doküman Frekansı (Term Frequency-Inverse Document Frequency – TF-IDF) gibi metin madenciliği yöntemleri kullanılarak yorumlar yapılandırılmış, makine öğrenimi algoritmaları; Destek Vektör Makinesi (Support Vector Machine – SVM), Naive Bayes, Lojistik Regresyon ile duygu analizi gerçekleştirilmiş ve SVM en yüksek doğruluğu sağlamıştır. Çalışma, otel sektöründe müşteri deneyimlerini optimize etmek için veri odaklı bir yaklaşım sunmaktadır. Sutherland vd. [12] çalışmalarında çevrimiçi konaklama platformlarından toplanan müşteri yorumlarını analiz ederek, müşteri memnuniyetine yönelik anahtar temaları belirlemek için LDA yöntemini kullanmıştır. Analiz sonucunda, otel konumları, oda temizliği, müşteri hizmetleri ve yemek kalitesi gibi temel temalar ortaya çıkarılmıştır. Müşteri yorumlarının büyük çoğunluğu olumlu olsa da yemek kalitesi ve hijyen gibi alanlarda olumsuz geri bildirimler dikkat çekmiştir. Çalışma, otel yöneticilerinin müşteri geri bildirimlerinden anlamlı içgörüler çıkararak hizmet kalitesini artırmalarına yardımcı olacak veri odaklı bir yaklaşım sunmaktadır.

Göker vd. [13] çalışmalarında FATİH projesine ilişkin internet ortamında bulunan görüşlerin metin madenciliği yöntemleri kullanılarak otomatik olarak tespit edilmesini amaçlamışlardır. Çalışma iki temel kısımdan oluşmaktadır. İlk aşamada, internet ortamındaki yapısal olmayan veri kümeleri yapısal veri haline dönüştürülmektedir. Bu amaçla, metin madenciliği veri önileme yazılımı geliştirilmiştir. İkinci aşamada, geliştirilen metin madenciliği veri önileme yazılımıyla yapısal veri kümesine dönüştürülen veriler üzerinde makine öğrenmesi algoritmaları uygulanarak yorumlar otomatik olarak sınıflandırılmıştır. FATİH projesine yönelik 444 görüş içeren metin dosyasındaki metinler, TF-IDF ağırlıklandırma yöntemi ile vektörel olarak temsil edilerek sınıflandırma algoritmalarının model başarımleri ölçütleri karşılaştırılmıştır. Sonuç olarak, performansı karşılaştırılan algoritmalar arasında en yüksek başarımın Ardışık Minimal Optimizasyon Algoritmasına ait olduğu (%88,73) gözlemlenmiştir. Büyükeke vd. [14] yaptıkları çalışmada otel işletmelerinin müşteri yorumlarını analiz etmeyi ve anlamlandırmayı amaçlamaktadır. Tripadvisor platformundan toplanan 212.435 yorumun kullanıldığı çalışmada, metin madenciliği yöntemleriyle bu verilerden rekabetçi bir zekâ oluşturulmaya çalışılmıştır. Bu amaçla, duygu analizi için Lojistik

Regresyon, SVM ve Naive Bayes gibi yöntemler kullanmışlardır. Analiz sonuçlarına göre, yorumların %80'inin olumlu, %20'sinin ise olumsuz olduğu tespit edilmiştir. Ayrıca, LDA ile konu analizi yapılarak yorumların büyük çoğunluğunun deneyim, değer ve eğlence ile ilgili olduğu belirlenmiştir. Yorumların sınıflandırılması için Lojistik Regresyon ve SVM algoritmaları daha iyi sonuçlar vermiştir.

Literatürde yer alan çalışmalar, müşteri hizmetleri yönetim sürecinin ve müşteri memnuniyetinin iyileştirilmesi adına incelemeler yaparken metin madenciliği ve veri madenciliği yaklaşımlarını çoğunlukla birlikte ele almıştır. Çalışmalarda gerekli analizlerin uygulanması için Python programlama dili ve kütüphaneleri, Weka programı, Statistica programı ve RapidMiner yazılımı yaygın olarak kullanılmaktadır. Melek, [15] yaptığı çalışmada veri madenciliği ve metin madenciliği tekniklerinin kullanılmasıyla şirket vizyon ifadelerinin analizini gerçekleştirmiştir. Çalışmanın temel amacı, Capital dergisinin En Beğenilen Şirketler araştırmasında 2010 yılında yer alan şirketlerin vizyon ifadelerini itibar boyutları ile ilişkilendirmektir. Statistica programı kullanılarak vizyon ifadelerine metin madenciliği adımları uygulanmış ve oluşturulan kelime belge listesine göre bağımlı değişkenler için histogram grafikleri oluşturulmuştur. Son olarak veriye faktör analizi uygulanarak kelimeler faktörler bazında gruplandırılmıştır. Sözen vd. [16] yaptıkları çalışmada e-ticaret web sitelerindeki mobilya ürünleri hakkında yapılan yorumları RapidMiner yazılımını kullanarak değerlendirmişlerdir. Ahşap ve metal malzemedен üretilmiş mobilyaların yorumları incelenerek web madenciliği yöntemleriyle değerlendirilmiştir. Çalışmada, günlük hayatta sıkça kullanılan çalışma masası, kütüphane ve sandalye gibi mobilya türleri üzerine yoğunlaşmıştır. Her mobilya türü için ayrı ayrı web ve metin madenciliği teknikleri kullanılarak kelime bulutları oluşturulmuştur. Bu şekilde, mobilya ürünleriyle ilgili yapılan yorumlardaki anahtar kelimeler görselleştirilmiştir. Çalışmanın sonuçlarına göre, ahşap sandalyelerde sağlamlık, metal sandalyelerde ise oturma ile ilgili yorumlar öne çıkmıştır. Ahşap ve metal çalışma masalarında güzel kelimesi öne çıkarılmıştır. Ahşap kütüphanelerde kolay, metal kütüphanelerde ise ofis ve paketleme anahtar kelimeleri belirgin olarak ortaya çıkmıştır. Yapay zekâ destekli metin madenciliği yöntemleri, müşteri hizmetleri yönetiminde önemli bir araç haline gelmiştir.

Bununla birlikte, sektörel uygulamalar da literatürde dikkat çekmektedir. Özellikle sosyal medya yönetiminin ön plana çıktığı sektörlerde daha yaygın olarak araştırma alanı oluşturulmuştur. Erduran, [17] yaptığı çalışmada Türkiye’de faaliyet gösteren bankalara yönelik müşteri şikayetlerinin veri madenciliği yöntemleri ile analizini yapmıştır. Toplam 100.000 şikâyetten oluşan iki farklı veri seti ile bir veri tabanı oluşturulmuş ve bu veri tabanı üzerinde RapidMiner programı ile veri madenciliği tekniklerinden kümeleme ve birliktelik analizleri uygulanmıştır. Yapılan analizler sonucunda, müşteri şikayetlerinde ortak ve en sık kullanılan kelimelerin ve kelime gruplarının belirlendiği ve şikayetlerin anlamlı bir şekilde (k-means, Silhouette skoru > 0,5) gruplandırıldığı ifade edilmiştir.

### 1.1. Problem Durumu

Mobilya sektöründe müşteri memnuniyeti, satış sonrası hizmetlerin etkinliği ile doğrudan ilişkilidir. Bu süreç, teslimat, kurulum ve ürünün kullanım aşamalarında ortaya çıkan müşteri geri bildirimlerini kapsamaktadır. Ancak, müşteri geri bildirimlerinin büyük bir kısmı yapısal olmayan verilerden oluştuğu için analiz edilmesi zor ve zaman alıcıdır. Şikayetlerin doğru sınıflandırılmaması ve bu doğrultuda çağruların uygun birimlere yönlendirilememesi, müşteri hizmetleri yönetiminde aksaklıklara yol açmakta ve müşteri memnuniyetsizliğine neden olmaktadır. Bu durum, işletmelerin müşteri beklentilerini karşılamakta yetersiz kalmalarına, marka bağlılığının azalmasına ve rekabet avantajının kaybedilmesine yol açabilmektedir. Özellikle mobilya sektöründe şikayet yönetim süreçlerindeki bu sorunlar, müşteri hizmetleri departmanlarının operasyonel verimliliğini düşürmekte ve işletmelerin stratejik karar alma süreçlerini olumsuz etkilemektedir. Bu bağlamda aşağıdaki temel sorular, araştırmanın probleminin konusunu şekillendirmektedir:

- Müşteri geri bildirimlerinden elde edilen veriler, mobilya sektöründe şikâyet yönetimini nasıl iyileştirebilir?
- Metin madenciliği teknikleri, müşteri şikayetlerinin analiz edilmesi ve çağrı atama süreçlerinin iyileştirilmesinde ne derece etkilidir?
- Mobilya sektörüne özgü şikayetler için hangi temalar ön plana çıkmaktadır ve bu temalar müşteri hizmetleri süreçlerini iyileştirmek için nasıl kullanılabilir?

## 1.2. Araştırmanın Amacı

Günümüz teknolojisi, işletmelere müşteri geri bildirimlerini hızlı ve etkin bir şekilde analiz etme fırsatı sunmaktadır. Müşteri memnuniyetinin sağlanmasında, satış sonrası hizmetlerin etkinliği, ürün kalitesi kadar kritik bir rol oynamaktadır. Bu bağlamda, müşteri geri bildirimlerinin analizi, işletmelerin müşteri beklentilerini anlaması, sıkça şikayet edilen alanları tespit etmesi ve bu alanlarda iyileştirmeler yapması açısından stratejik bir öneme sahiptir.

Satış sonrası hizmetlerde ortaya çıkan teslimat gecikmeleri, kurulum problemleri ve eksik ürün gibi sorunlar, müşteri memnuniyetini olumsuz etkileyen en yaygın problemler arasında yer almaktadır. Bu tür sorunların işletme düzeyinde hızlı bir şekilde çözülebilmesi için müşteri geri bildirimlerinin sistematik bir şekilde takip edilmesi ve analiz edilmesi gerekmektedir. Sistematik analiz süreçleri, yalnızca sorunları çözme hızını artırmakla kalmaz, aynı zamanda sıkça karşılaşılan problemleri önceden tespit ederek süreçlerin iyileştirilmesine olanak tanır. Bu durum, müşteri memnuniyetini artırırken işletmelerin rekabet avantajını güçlendirmesine ve marka bağlılığını sağlamasına da katkı sunar.

Bu araştırmanın temel amacı, Türkiye’de faaliyet gösteren üç büyük mobilya şirketinin müşteri geri bildirimlerini metin madenciliği teknikleriyle analiz ederek, çağrı merkezi süreçlerini iyileştirmeye yönelik stratejik bir model geliştirmektir. Araştırma kapsamında, müşteri şikayetlerinden yola çıkarak çağrı atama süreçlerini optimize edecek bir sistemin tasarlanması hedeflenmektedir. Ayrıca, mobilya sektörüne özgü satış sonrası hizmetlerin geliştirilmesi için somut öneriler sunulacak ve müşteri şikayetlerinden elde edilen verilerin sektörel iyileştirmelere nasıl dönüştürülebileceği ortaya konacaktır.

Bu çalışmanın sonuçları, müşteri hizmetleri süreçlerinde operasyonel verimliliği artırarak, müşteri memnuniyetini sağlama noktasında sektöre rehberlik edecek nitelikte bir strateji sunmayı hedeflemektedir.

### 1.3. Araştırmanın Önemi

Mobilya sektörü, müşteri hizmetleri yönetiminde kendine has dinamiklere sahip bir alan olup, literatürde bu sektörün müşteri geri bildirimleri ve şikayet yönetimi üzerine detaylı incelemeleri sınırlıdır. Duygun vd. [18] çalışmalarında mobilya sektöründeki ürünlerin fiziksel doğası ve satış sonrası süreçlerin karmaşıklığı nedeniyle kendine özgü müşteri beklentilerine sahip olduğunu vurgulamışlardır. Şikayetlerin genellikle teslimat, kurulum ve ürün kalitesi üzerine yoğunlaştığını, bu alanların müşteri memnuniyeti üzerinde kritik bir etkiye sahip olduğunu vurgulamışlardır.

Bu çalışma, mobilya sektörüne özgü müşteri şikayetleri ve satış sonrası hizmet süreçlerini kapsamlı bir biçimde ele alarak, literatürdeki teorik ve uygulamalı eksiklikleri doldurmayı hedeflemektedir. Metin madenciliği uygulamasının, aynı sistemden beslenen ama rakip rollerde yer alan işletmeler açısından ele alınması yaygın olmayan bir çalışmaya örnek niteliğindedir. Ayrıca, metin madenciliği ve büyük veri analitiği gibi yenilikçi yaklaşımları mobilya sektörü bağlamında uygulayarak, müşteri memnuniyetini artırmaya ve sektörel rekabet avantajı sağlamaya yönelik stratejik öneriler geliştirmektedir. Bu araştırma, hem akademik dünyaya hem de sektöre rehberlik edecek bir kaynak niteliği taşımaktadır.

## 2. BÖLÜM YÖNTEM VE MATERYAL

Bu bölümde, veri madenciliği kavramı ve bir alt çalışma alanı olan metin madenciliği tanımlanarak kavramsal çerçeve oluşturulmuştur. Veri madenciliği türleri, teknikleri ve uygulama alanları detaylandırılarak metin madenciliği ile arasındaki ilişkiler ve farklar açıklanmıştır.

### 2.1. Veri Madenciliği

Veri madenciliği, büyük ve karmaşık veri kümelerindeki gizli desenleri, ilişkileri ve eğilimleri keşfetme sürecidir. Bu süreç, istatistik, yapay zekâ, makine öğrenimi ve veri tabanı yönetimi gibi çeşitli disiplinlerin tekniklerinden faydalanarak verilerin derinlemesine analiz edilmesini sağlar [19]. Özellikle büyük veri setleri üzerinde yapılan analizlerde bu teknikler, anlamlı ve kullanışlı bilgilerin otomatik olarak keşfedilmesine olanak tanır.

Veri madenciliğinin temelleri, 1960'larda istatistiksel analiz ve veri tabanı yönetim sistemlerinin gelişimi ile atılmıştır. Bu dönemde, verilerin organize edilmesi ve temel analizlerin gerçekleştirilmesi sağlanmıştır. 1970'lerde ilişkisel veri tabanlarının geliştirilmesiyle veri yönetimi kolaylaşmış ve veriler arasında ilişkilerin daha etkili bir şekilde analiz edilmesi mümkün hale gelmiştir. 1980'lerde veri ambarları ve çevrimiçi analitik işleme sistemlerinin ortaya çıkışı, farklı kaynaklardan gelen verilerin birleştirilmesini ve çok boyutlu analizlerin yapılmasını sağlamıştır. Bu yenilikler, büyük miktardaki verinin daha sistematik bir şekilde işlenmesine olanak tanımıştır. 1990'lara gelindiğinde, veri madenciliği, gelişmiş algoritmalar ve tekniklerin yardımıyla bağımsız

bir disiplin haline gelmiştir. Bu dönemde, veri madenciliği iş dünyasında karar destek sistemleri içinde yaygın olarak kullanılmaya başlanmıştır.

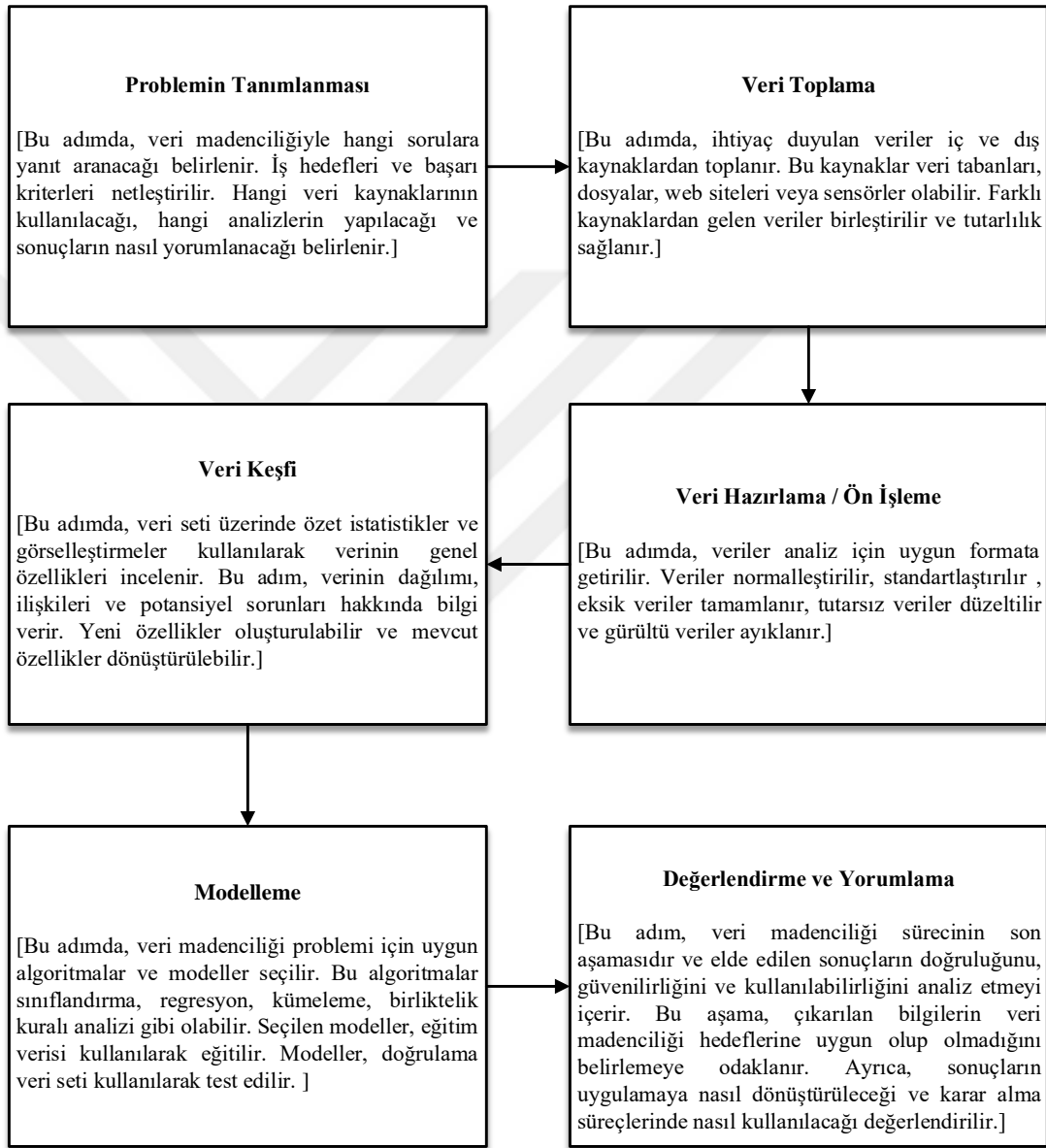
Özellikle müşteri ilişkileri yönetimi, finansal analiz ve pazar tahmini gibi uygulama alanlarında etkin bir araç haline gelmiştir. 2000'li yıllardan itibaren büyük veri teknolojilerindeki hızlı gelişmeler ve yapay zekâ uygulamalarının yaygınlaşması, veri madenciliğinin daha büyük ve karmaşık veri setlerini analiz etmesini mümkün kılmıştır. Bu gelişmeler, işletmelerin ve araştırmacıların daha geniş kapsamlı ve derinlemesine analizler yaparak stratejik kararlar almalarına katkıda bulunmuştur. Günümüzde veri madenciliği, sadece mevcut verilerin analizini değil, aynı zamanda gelecekteki eğilimlerin tahmini ve stratejik kararlar için rehberlik sağlama işleviyle de kritik bir araç olarak kullanılmaktadır.

Veri madenciliğinin günümüzdeki önemini artıran unsurlardan biri de yapay zekâ, makine öğrenimi ve büyük veri teknolojilerindeki ilerlemelerdir. Bu teknolojiler, veri madenciliğinin doğruluğunu ve hızını artırarak daha karmaşık veri kümeleri üzerinde daha derin analizler yapılmasını mümkün kılar. Böylece, işletmeler ve araştırmacılar, daha geniş kapsamlı ve daha güvenilir bilgilere daha hızlı bir şekilde ulaşabilirler. Bu bağlamda, veri madenciliği sadece mevcut durumun analiziyle kalmayıp, aynı zamanda gelecekteki eğilimleri tahmin etme ve stratejik kararlar için bir rehber sunma işlevi de görür.

Veri madenciliği, ham verilerin anlamlı bilgilere dönüştürülmesi için kullanılan çok aşamalı bir süreçtir. Bu süreç, genellikle aşağıdaki temel adımlardan oluşmaktadır: veri temizleme, veri entegrasyonu, veri seçimi, veri dönüştürme, veri madenciliği, sonuçların değerlendirilmesi ve sonuçların sunulması [19].

Veri madenciliği, iş dünyasından sağlık ve eğitim sektörüne, pazarlamadan sosyal medyaya kadar geniş bir uygulama alanına sahiptir. Müşteri ilişkileri yönetiminde müşteri segmentasyonu ve davranış analiziyle, finans sektöründe kredi riski analizi ve dolandırıcılık tespiti için kullanılır. Sağlık sektöründe, hastalıkların erken teşhisi ve tedavi süreçlerinin iyileştirilmesi amacıyla genetik veriler ve hasta kayıtları analiz edilir. Eğitim alanında, öğrenci performansı ve bireysel öğrenme güçlükleri veri madenciliği ile tespit edilerek daha etkili eğitim programları oluşturulabilir. Perakende sektöründe

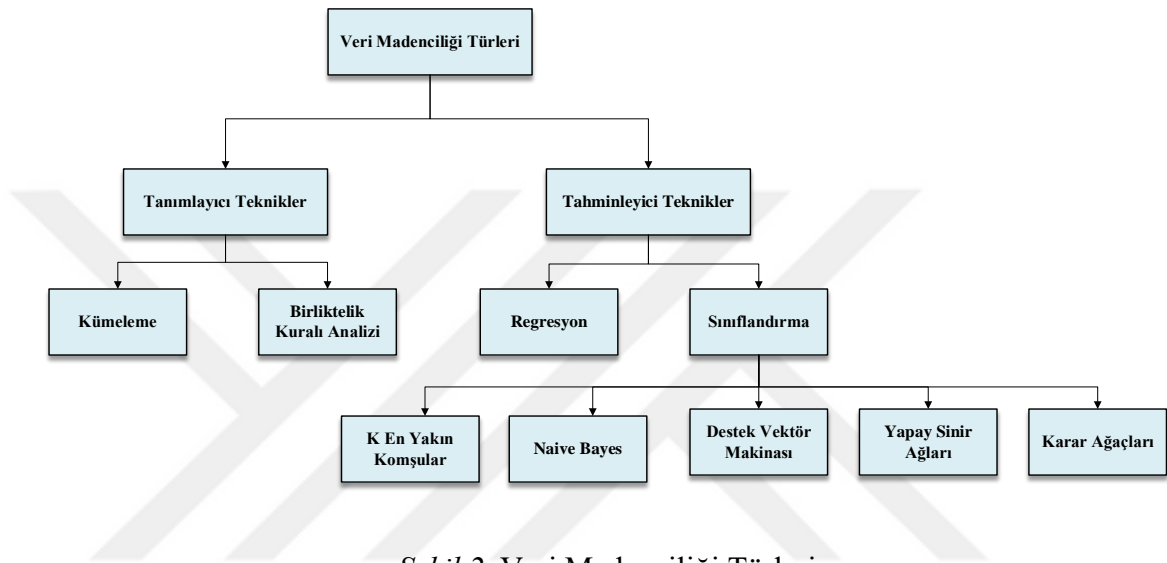
müşteri alışkanlıkları analiz edilerek çapraz satış ve yukarı satış stratejileri geliştirilirken, sosyal medya ve web analitiği sayesinde kullanıcı davranışları ve trendler belirlenebilir. Üretim ve lojistikte ise süreçlerin optimize edilmesi ve maliyetlerin düşürülmesi sağlanır. Şekil 1’de veri madenciliği sürecinin detayları yer almaktadır.



Şekil 1. Veri Madenciliği Süreci

## 2.2. Veri Madenciliği Türleri

Veri madenciliği, iki temel amaç doğrultusunda gerçekleştirilebilir: tahmin ve tanımlama. Her iki amaç da farklı veri madenciliği teknikleri ve yöntemleri kullanılarak gerçekleştirilir ve verilerden anlamlı içgörüler elde edilmesini sağlar. Veri madenciliği türleri Şekil 2'deki gibi sınıflandırılabilir.



Şekil 2. Veri Madenciliği Türleri

### 2.2.1. Tanımlayıcı Metotlar

Tanımlayıcı metotlar, mevcut verilerden anlamlı ve yorumlanabilir örüntüler ile ilişkiler bulmayı amaçlar. Bu metotlar, veriyi özetlemek, önemli özellikleri belirlemek ve verinin genel davranışını anlamak için kullanılır. Tanımlama metotları, veri kümesinin kısa ve öz bir tanımını sunar ve veriden açıklayıcı örüntüler çıkarır. Tanımlama metotları genellikle kümeleme, birliktelik kuralı analizi ve anomali tespiti olarak karşımıza çıkar.

#### 2.2.1.1. Kümeleme

Kümeleme, veri madenciliğinde denetimsiz öğrenme yöntemlerinden biridir ve veri setini benzer özelliklere sahip gruplara ayırma işlemidir. Bu yöntem, veri içindeki doğal yapıları ve kalıpları ortaya çıkararak benzer verileri aynı grupta toplar. Kümeleme, müşteri segmentasyonu, pazar araştırmaları, biyolojik veri analizi, görüntü işleme ve

birçok diğ er alanda yaygın olarak kullanılır. Kümeleme algoritmalarının temel amacı, içsel benzerliklere dayanarak veri noktalarını mantıklı gruplara ayırmaktır. Kümeleme analizinde kullanılan yöntemler arasında bölme, hiyerarşik, yoğunluk tabanlı, ızgara tabanlı ve model tabanlı yöntemler bulunur. En yaygın kümeleme yöntemleri arasında K-means, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), hiyerarşik ve model tabanlı algoritmalar yer alır. Veri miktarının artmasıyla birlikte bu analizin önemi ve kullanım alanları da genişlemektedir.

### **2.2.1.2. Birliktelik Kuralı Analizi**

Birliktelik kuralı analizi, veri madenciliği ve pazarlama arařtırmalarında yaygın olarak kullanılan etkili bir tekniktir. Bu yöntem, veri setinde birlikte gözlemlenen özellikler arasındaki ilişkileri tespit ederek anlamlı kurallar oluşturmayı amaçlar. Özellikle "pazar sepeti analizi" olarak bilinen uygulamasıyla, müşteri alışveriş verilerinden hangi ürünlerin sıklıkla birlikte satın alındığını belirlemek için kullanılır. Bu analiz, işletmelerin ürün önerileri geliřtirmesi, çapraz satış stratejileri oluşturması ve stok yönetimi gibi operasyonel süreçlerini optimize etmesi açısından kritik bir rol oynar. Birliktelik kuralı analizinde en yaygın kullanılan algoritmalar arasında Apriori Algoritması, FP-Growth Algoritması ve Apriori-TID Algoritması yer alır. Bu algoritmalar, veri setindeki ilişki ve kalıpları ortaya çıkarmak için etkili bir şekilde çalışarak işletmelere stratejik karar alma süreçlerinde rehberlik eder.

### **2.2.2. Tahminleyici Metotlar**

Tahminleyici metotlar, veri kümesinin mevcut bilgilerini kullanarak gelecekteki olayları veya bilinmeyen deęişkenleri tahmin etmeyi amaçlar. Bu metotlar, özellikle iş zekâsı, finans, saęlık ve pazarlama gibi alanlarda yaygın olarak kullanılır. Tahmin metotları, genellikle iki ana kategoriye ayrılır: sınıflandırma ve regresyon.

#### **2.2.2.1. Regresyon**

Regresyon, sürekli bir hedef deęişkenin gelecekteki deęerlerini tahmin etmek için kullanılır. Regresyon metotları, bağımlı ve bağımsız deęişkenler arasındaki ilişkiyi modelleyerek, bağımlı deęişkenin deęerlerini tahmin eder. Regresyon analizinde,

bağımsız değişkenler (girdi) bağımlı değişkeni (çıktı) etkileyen faktörler olarak değerlendirilir. Modelleme sürecinde, bağımsız değişkenlerin bağımlı değişken üzerindeki etkisi matematiksel bir fonksiyonla ifade edilir. Doğrusal regresyon, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi düz bir doğru ile açıklarken, lojistik regresyon sınıflandırma problemlerinde kullanılır. Çoklu regresyon, birden fazla bağımsız değişkenin etkisini analiz ederken, polinom regresyon doğrusal olmayan ilişkileri modellemek için kullanılır. Zaman serisi regresyonu ise zamanla değişen verilerde gelecekteki eğilimleri öngörür. Bu metotlar, ev fiyatlarının tahmini, satış tahmini ve risk analizi gibi uygulamalarda yaygın olarak kullanılır.

#### **2.2.2.2. Sınıflandırma**

Sınıflandırma, veri madenciliğinde, veriyi önceden tanımlanmış kategorilere ayırmayı amaçlayan yaygın bir tekniktir. Bu yöntem, veri setindeki örüntüleri belirleyerek yeni gelen verilerin sınıflandırılmasına olanak tanır [20]. Sınıflandırma, verileri belirli kategorilere ayırma işlemidir. Etiketlenmiş veri setleriyle çalışır ve yeni verilerin hangi kategoriye ait olduğunu tahmin eder. Kullanım alanları arasında kredi riski analizi, hastalık teşhisi ve spam tespiti gibi çeşitli uygulamalar bulunur.

Sınıflandırma, belirli adımları takip eden bir süreçtir. İlk olarak, veri ön işleme aşamasında eksik veya hatalı veriler temizlenir, veriler uygun bir forma dönüştürülür ve sınıflandırma için önemli özellikler seçilir. Bu adım, modelin performansını artırmak için kritik öneme sahiptir. Sonraki aşamada, veriler eğitim ve test verisi olmak üzere ayrıştırılır. Eğitim verisi, modelin öğrenmesi için kullanılan veri setidir. Bu veri seti, modelin örüntüleri tanımasını ve kategorileri öğrenmesini sağlar. Test verisi ise modelin performansını değerlendirmek için kullanılan veri setidir. Bu veri, modelin yeni veriler üzerindeki tahmin başarısını ölçer. Verilerin ayrıştırılmasında, veri setinin büyüklüğüne göre farklı teknikler uygulanarak model performansı artırılabilir. En yaygın kullanılan ayrıştırma yöntemleri, rastgele bölme ve k katlı çapraz doğrulamadır.

Sınıflandırma, makine öğreniminin temel tekniklerinden biri olup, çeşitli modeller ve algoritmalar kullanılarak gerçekleştirilir. Bu modeller, farklı veri türlerine ve problemlerine uygun çözümler sunar. Aşağıda, sınıflandırmada en çok kullanılan ve bu

arařtırmada uygulanan algoritmaların genel özellikleri, avantajları ve dezavantajları açıklanmaktadır.

**Naive Bayes:** Naive Bayes, olasılık teorisine dayanan ve Bayes teoremini kullanan bir sınıflandırma algoritmasıdır. Bu algoritma, özelliklerin birbirinden bağımsız olduğu varsayımına dayanır, bu nedenle "naive" (saf) olarak adlandırılır. Hızlı ve verimlidir. Küçük veri setlerinde bile iyi performans gösterir. Özelliklerin sayısı arttıkça performansı düşmez. Çok hızlı ve verimli çalışır. Özellikle metin madenciliği, spam tespiti gibi uygulamalarda güçlüdür. Özelliklerin birbirinden bağımsız olduğunu varsayar, bu genellikle gerçek dünyadaki verilerde geçerli değildir. Bu nedenle bazı veri setlerinde performansı düşük olabilir. Naive Bayes algoritmasının temelinde Bayes teoremi yer alır. Bayes teoremi, bir olayın olasılığını, diğer bir olayın gerçekleştiği bilgisine dayanarak hesaplar. Bayes teoreminin matematiksel ifadesi şu şekildedir:

$$P(C | X) = P(C) * P(X | C) / P(X) \quad (1)$$

- $P(C|X)$ : X özellik vektörü verildiğinde C sınıfının posterior olasılığı.
- $P(X|C)$ : C sınıfı verildiğinde X özellik vektörünün olasılığı.
- $P(C)$ : C sınıfının ön
- $P(X)$ : X özellik vektörünün ön olasılığı

Naive Bayes, Bayes teoremini kullanarak bir veri noktasının belirli bir sınıfa ait olma olasılığını hesaplar. Özelliklerin bağımsız olduğu varsayımını kullanarak,  $P(X|C)$  olasılığı şu şekilde ayrıştırılır:

$$P(X | C) = P(x_1 | C) * P(x_2 | C) \dots P(x_n | C) \quad (2)$$

Burada  $x_1, x_2, \dots, x_n$ , x özellik vektöründeki bireysel özellikleri temsil eder. Naive Bayes algoritması, her sınıf için posterior olasılıkları hesaplar ve en yüksek posterior olasılığı olan sınıfı seçer.

Şahinaslan vd. [21] çalışmalarında Naive Bayes algoritmasını kullanarak çok dilli sosyal medya yorumlarının duygu analizini yapmayı amaçlamışlardır. WEKA programı ile farklı dillerde (Türkçe, İngilizce, Almanca, Fransızca, Arapça, Rusça, Korece) YouTube yorumlarını analiz ederek pozitif, negatif ve nötr duyguları

sınıflandırmışlardır. Elde ettikleri doğruluk oranı, çok dilli veri analizi için Naive Bayes algoritmasının etkili bir araç olduğunu göstermektedir.

**Destek Vektör Makineleri:** SVM, sınıflandırma ve regresyon analizlerinde kullanılan güçlü bir makine öğrenimi algoritmasıdır. SVM, veriyi iki sınıfa ayıran en iyi hiper düzlemi bulmayı amaçlar. Yüksek boyutlu veri setlerinde ve doğrusal olmayan verilerde de etkili sonuçlar verebilir. SVM, sınıflar arasındaki ayrımı maksimize eden hiper düzlemi bulur. Bu hiper düzlem, sınıflar arasındaki marjini en büyük yapacak şekilde seçilir. Sınıflar arasındaki marjin, destek vektörleri adı verilen kritik veri noktaları tarafından belirlenir. Yüksek boyutlu veri setlerinde iyi çalışır. Doğrusal olmayan sınıflandırma problemleri için çekirdek fonksiyonları (kernel functions) kullanarak sınıflar arasındaki en iyi ayrımı yapar. Aşırı öğrenmeye daha az yatkındır. Büyük veri setlerinde eğitim süreci yavaş olabilir. Parametre ayarı (örneğin, C ve gamma değerleri) performans üzerinde büyük etkiye sahiptir.

**LightGBM Algoritması:** Karar ağaçlarına dayalı hızlı ve güçlü bir makine öğrenimi algoritmasıdır. Büyük veri kümelerinde yüksek performans ve düşük bellek kullanımı sunar. Özellikle, histogram tabanlı bölme ve yaprak odaklı büyüme gibi yenilikçi teknikler sayesinde eğitim süresini hızlandırır ve doğruluğu artırır. Kategorik veri desteği ve paralel işleme yeteneği ile esnek ve etkili bir algoritmadır. LightGBM, sınıflandırma, regresyon ve sıralama problemleri için geniş bir uygulama alanına sahiptir ve doğru hiperparametrelerle aşırı öğrenme riski düşük, yüksek doğruluk sağlayan bir model sunar.

**Yapay Sinir Ağları:** Yapay sinir ağları, insan beyninin çalışma prensiplerinden esinlenerek geliştirilmiş makine öğrenimi modelleridir. Sinir ağları, büyük veri setlerinde karmaşık ilişkileri ve kalıpları öğrenebilir. Özellikle derin öğrenme teknikleri, çok katmanlı sinir ağları kullanarak yüksek doğruluklu tahminler ve sınıflandırmalar yapabilir. Karmaşık veri yapıları ve ilişkiler için uygundur. Büyük veri setlerinde karmaşık ilişkileri ve kalıpları öğrenebilir. Derin öğrenme ile, çok katmanlı sinir ağları yüksek doğruluklu tahminler yapabilir. Yüksek hesaplama maliyeti ve uzun eğitim süresi gerektirir. Aşırı öğrenme riski vardır ve eğitim için büyük miktarda veri gereklidir.

**Karar Ağaçları:** Karar ağaçları, sınıflandırma ve regresyon analizlerinde kullanılan popüler bir tekniktir. Karar ağaçları, veriyi dallara ayırarak karar kuralları oluşturur ve bu kurallara dayanarak veriyi sınıflandırır veya tahmin eder. Bu teknik, veriyi anlaşılır bir şekilde görselleştirir ve karmaşık karar süreçlerini basitleştirir. Kolay yorumlanabilir, görselleştirilebilir ve anlaşılırdır. Hem kategorik hem de sürekli veriyle çalışabilir. Özellikle küçük veri setlerinde iyi performans gösterir. Aşırı öğrenmeye eğilimlidir. Derin ağaçlar, veri setini gereğinden fazla öğrenerek, yeni verilerde düşük performans gösterebilir. Gelişmiş Versiyonlar; Random Forest ve Gradient Boosting Trees gibi yöntemler, birçok karar ağacını bir araya getirerek aşırı öğrenmeyi önleyip model performansını artırır.

**AdaBoost Model:** AdaBoost, sınıflandırma ve regresyon analizlerinde kullanılan etkili bir makine öğrenimi algoritmasıdır. Temel olarak, bir dizi zayıf öğreniciyi (örneğin, karar ağaçlarını) bir araya getirerek güçlü bir model oluşturur. AdaBoost, her bir zayıf modelin hatalarını dikkate alarak bir sonraki modeli eğitir ve bu süreçte hataları daha iyi sınıflandırmaya çalışır. Bu, modelin doğruluğunu artıran iteratif bir süreçtir. AdaBoost, Boosting yönteminin temel bir versiyonudur ve daha gelişmiş sürümleri bulunmaktadır. Gradient Boosting ve XGBoost gibi yöntemler, AdaBoost'un prensiplerini daha karmaşık optimizasyon teknikleriyle genişleterek model performansını artırır. Bu yöntemler, daha iyi genelleme yeteneği ve daha yüksek doğruluk sunarak, AdaBoost'un sınırlamalarını aşmaya çalışır.

**Rastgele Ormanlar:** Rastgele Ormanlar, birden fazla karar ağacını birleştirerek sınıflandırma ve regresyon problemlerinde yüksek doğruluk sağlayan güçlü bir makine öğrenimi algoritmasıdır. Her ağaç, rastgele bir alt veri kümesi ve öznitelikler kullanılarak eğitilir ve sonuçlar çoğunluk oyu (sınıflandırma) veya ortalama (regresyon) ile birleştirilir. Aşırı öğrenmeye dayanıklıdır ve öznitelik önemini belirleme gibi avantajlar sunar. Ancak, hesaplama maliyeti yüksek olabilir ve yorumlanabilirliği düşüktür. Finans, sağlık, pazarlama ve e-ticaret gibi çeşitli alanlarda yaygın olarak uygulanır.

Sınıflandırma modellerinin performansını değerlendirmek için kullanılan metrikler, modelin doğruluğunu, hassasiyetini, hatırlama oranını ve genel dengeyi anlamak açısından kritik öneme sahiptir. Bu metrikler, özellikle gerçek dünyada sınıflandırma

problemlerine uygulanan modellerin başarısını analiz etmek ve iyileştirme yapmak için kullanılır. Sınıflandırma modellerinin değerlendirilmesinde en yaygın olarak kullanılan metrikler; karmaşıklık matrisi, doğruluk değeri, kesinlik değeri, hatırlama değeri, F1 skoru, kappa istatistiği ve ROC eğrisidir.

**Karmaşıklık Matrisi:** Karmaşıklık matrisi bir sınıflandırma modelinin performansını görsel olarak değerlendirmek için kullanılır. Bu matris, modelin doğru ve yanlış tahminlerini ayrıntılı olarak gösterir ve modelin başarısız olduğu alanları net bir şekilde görmeyi sağlar. Matris, dört farklı bileşenden oluşur (Tablo 1).

Tablo 1. Karmaşıklık Matrisi

	<b>Gerçek Pozitif</b>	<b>Gerçek Negatif</b>
<b>Tahmin Pozitif</b>	Doğru Pozitif (TP)	Yanlış Pozitif (FP)
<b>Tahmin Negatif</b>	Yanlış Negatif (FN)	Doğru Negatif (TN)

- Doğru Pozitif (TP): Gerçekte pozitif olan ve model tarafından doğru tahmin edilen örnek sayısı.
- Yanlış Pozitif (FP): Modelin pozitif olarak tahmin ettiği ancak gerçekte negatif olan örnek sayısı.
- Yanlış Negatif (FN): Modelin negatif olarak tahmin ettiği ancak gerçekte pozitif olan örnek sayısı.
- Doğru Negatif (TN): Gerçekte negatif olan ve model tarafından doğru tahmin edilen örnek sayısı.

**Doğruluk:** Modelin tüm veri setindeki doğru tahmin oranını ifade eder. Genel başarıyı gösterir. Yüksek doğruluk değeri model performansının yüksek olduğunu gösterir. Doğruluk değeri, sınıflar arasında dengesizlik olduğunda yanıltıcı olabilir. Örneğin, nadir görülen bir sınıfın tahminlerinde doğruluk yüksek olsa bile, modelin başarısı düşük olabilir. Doğruluk, ek metrikler ile daha kapsamlı bir şekilde değerlendirilmelidir. Doğruluk metriğinin formülü aşağıdaki gibidir:

$$\text{Doğruluk} = \frac{TP + TN}{TP + FP + FN + TN} \quad (3)$$

**Kesinlik (Precision):** Pozitif olarak tahmin edilen verilerin gerçekten pozitif olma oranını ölçer. Yanlış pozitifleri azaltmada önemlidir. Yüksek kesinlik değeri model performansının yüksek olduğunu gösterir. Kesinlik metriğinin formülü aşağıdaki gibidir:

$$\text{Kesinlik} = \frac{TP}{(TP + FP)} \quad (4)$$

**Hatırlama (Recall):** Pozitif sınıfa ait gerçek verilerin ne kadarının doğru tahmin edildiğini ölçer. Yanlış negatifleri azaltmada önemli bir metriğe sahiptir. Yüksek hatırlama değeri, model performansının yüksek olduğunu gösterir. Hatırlama metriğinin formülü aşağıdaki gibidir:

$$\text{Hatırlama} = \frac{TP}{TP + FN} \quad (5)$$

**F1 Skoru:** Kesinlik ve hatırlama arasında bir denge sağlar. Dengesiz veri setlerinde kullanışlıdır. Yüksek F1 skoru kesinlik ve hatırlama arasında iyi bir denge olduğunu gösterir. F1 skorunun formülü aşağıdaki gibidir:

$$\text{F1 Skoru} = 2 \times \frac{\text{Kesinlik} \times \text{Hatırlama}}{\text{Kesinlik} + \text{Hatırlama}} \quad (6)$$

**ROC Eğrisi:** Farklı eşik değerleri için modelin hatırlama oranı ile yanlış pozitif oranını karşılaştırır. Eğrinin altında kalan alan (AUC), modelin genel performansını değerlendirir. Değerin yüksek olması, modelin farklı eşik değerleri arasında iyi bir performans sergilediğini ve pozitif sınıfları negatiflerden ayırmada başarılı olduğunu gösterir.

**Kappa İstatistiği:** Rastgele tahminlere göre modelin ne kadar iyi performans gösterdiğini ölçer. Değerin yüksek olması modelin rastgele tahminlerden çok daha iyi performans gösterdiğini ifade eder. Kappa istatistiğinin formülü aşağıdaki gibidir:

$$\text{Kappa} = \frac{\text{Gözlenen Doğruluk} - \text{Beklenen Doğruluk}}{1 - \text{Beklenen Doğruluk}} \quad (7)$$

### 2.3. Metin Madenciliği

Metin madenciliği, geniş metin koleksiyonları üzerinde bilgi keşfetme sürecidir. Bu süreç, yapılandırılmamış metin verilerini işleyerek anlamlı bilgiler, kalıplar ve ilişkiler elde etmeyi amaçlar. Metin madenciliği, doğal dil işleme (Natural Language Processing-NLP), bilgi çıkarımı, veri madenciliği ve makine öğrenimi gibi çeşitli disiplinlerin tekniklerini bir araya getirir.

Metin madenciliği; metinlerden bilgi çıkarma, metin sınıflandırma, kümeleme, duygu analizi, konu modelleme ve daha birçok analitik görevi içerir. Veri tabanlarından bilgi çıkarma yöntemlerinin metin verilerine uygulanmış halidir. Yapılandırılmamış veya yarı yapılandırılmış metinlerden anlamlı ve kullanışlı bilgiler çıkarılmasını sağlar. Hearst, [22] metin madenciliğini "metin koleksiyonları üzerinde bilgi keşfetme süreci" olarak tanımlar ve metin madenciliğinin doğal dil işleme, bilgi çıkarımı ve veri madenciliği gibi disiplinlerin yöntemlerini bir araya getirdiğini vurgular. Mooney vd. [23] metin madenciliğini "veri tabanlarından bilgi çıkarma yöntemlerinin yapılandırılmamış veya yarı yapılandırılmış metinlere uygulanması" olarak tanımlamışlardır. Metin madenciliğinin bilgi çıkarma, doğal dil işleme ve makine öğrenimi gibi disiplinlerin yöntemlerini bir araya getirdiğini belirtmişlerdir.

Metin madenciliği, 20. yüzyılın sonlarından itibaren bilgi keşfi ve veri madenciliği alanlarının bir alt disiplini olarak gelişmiştir. İlk dönemlerde doğal dil işleme ve bilgi çıkarma odaklı çalışmalarla başlayan metin madenciliği, 1990'larda TF-IDF ve Gizli Anlamsal İndeksleme (Latent Semantic Indexing-LSI) gibi tekniklerle yapılandırılmamış metinlerden bilgi çıkarma alanında önemli adımlar atmıştır. 2000'lerde dijital verilerin hızla artmasıyla büyük veri analitiği ve makine öğrenimi süreçlerine entegre edilmiş, 2010'lardan itibaren ise derin öğrenme, sözcük gömme yöntemleri ve Transformer tabanlı modeller (BERT, GPT) ile büyük ilerlemeler kaydedilmiştir. Bu gelişmeler, metin madenciliğinin performansını ve uygulama alanlarını önemli ölçüde genişletmiştir.

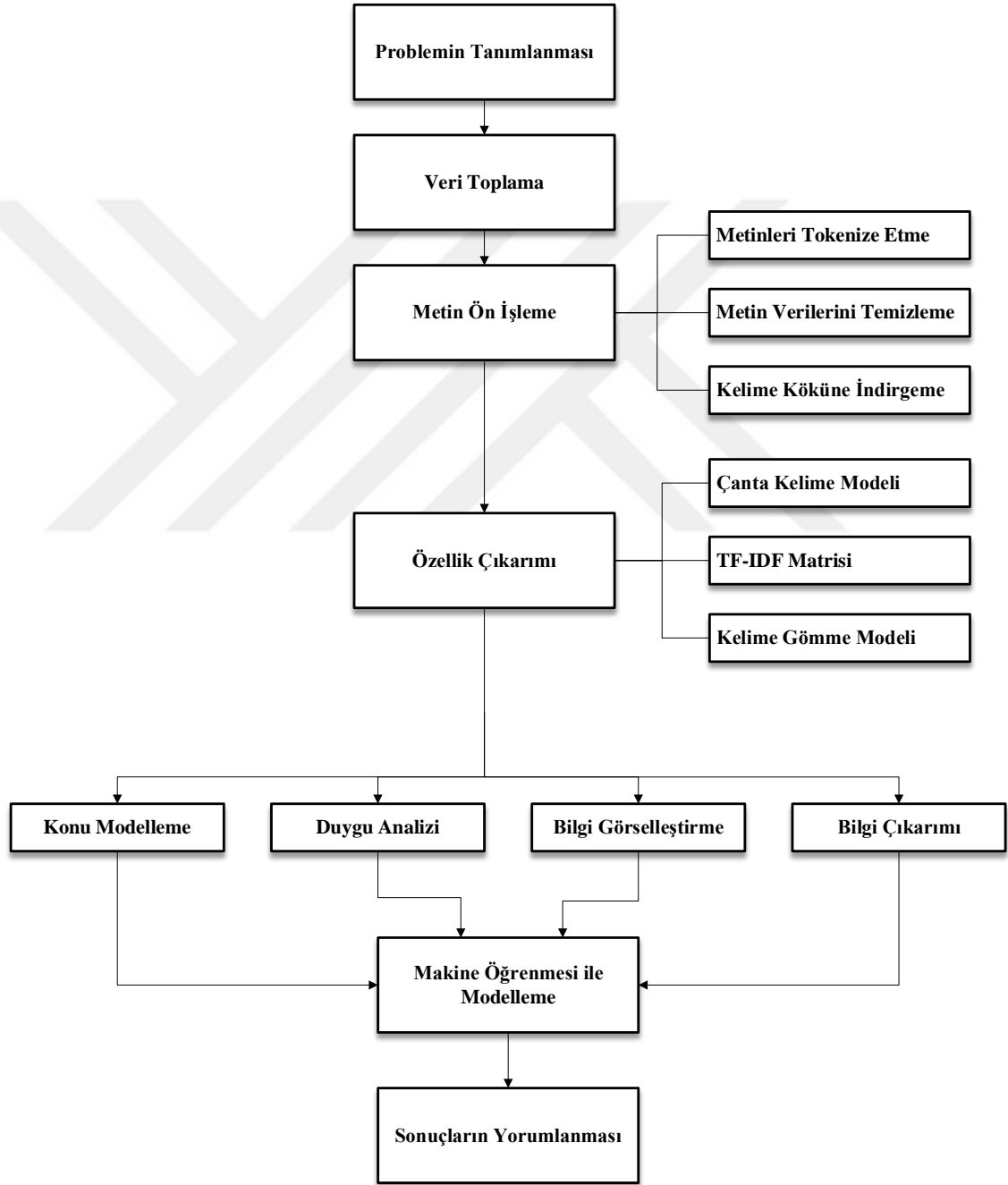
Günümüzde metin madenciliği; iş zekâsı ve pazarlama alanında müşteri geri bildirimlerinin ve sosyal medya verilerinin analiziyle müşteri memnuniyetini artırırken, pazarlama kampanyalarını optimize eder. Sağlık sektöründe elektronik sağlık kayıtları

ve tıp literatürünün analiziyle hastalık trendleri ve tedavi yöntemleri hakkında değerli bilgiler sağlar. Finans sektöründe duygu analizi ve risk yönetimi, metin madenciliği teknikleriyle etkinleştirilir. Hukuk alanında yasal belgelerin analizi ve dolandırıcılık tespiti gerçekleştirilirken, eğitim sektöründe öğrenci geri bildirimleri ve eğitim materyalleri üzerinde çalışmalar yapılabilir. Medya ve eğlence sektöründe içerik öneri sistemleri ve trend analizi öne çıkarken, sosyal bilimlerde sosyal medya analizi ve kültürel çalışmalarda kullanılır. Bilimsel araştırmalarda ise literatür taraması ve atıf analizleriyle araştırma trendleri ve etkili çalışmalar belirlenir. Metin madenciliği, farklı sektörlerdeki geniş uygulama yelpazesıyla büyük bir potansiyele sahiptir.



### 2.3.1. Metin Madenciliği Süreci

Metin madenciliği, yapılandırılmamış metin verilerini analiz ederek anlamlı bilgiler elde etme sürecidir. Bu süreçte kullanılan temel adımlar Şekil 3'te gösterilmiş ve metin madenciliği sürecindeki temel adımlar aşağıda detaylı olarak açıklanmıştır:



Şekil 3. Metin Madenciliği Süreci

**1. Problemin Tanımlanması:** Bu aşamada, metin madenciliği çalışmasının amacı ve kapsamı net bir şekilde belirlenir. Örneğin, müşteri yorumlarından duygu analizi gerçekleştirmek, belgeleri sınıflandırmak ya da metinlerden konu modelleme yapmak gibi spesifik hedefler tanımlanır. Ayrıca, kullanılacak metin kaynakları ve uygulanacak analiz yöntemleri detaylandırılarak çalışmanın çerçevesi çizilir.

**2. Veri Toplama:** Bu adımda, metin verileri sosyal medya paylaşımları, müşteri geri bildirimleri, haber makaleleri gibi farklı kaynaklardan toplanır. Verilerin bir araya getirilmesi sırasında, farklı formatlarda ve yapılarında bulunan metinlerin tutarlı bir veri seti haline getirilmesi sağlanır. Böylece, analiz sürecine uygun bir veri altyapısı oluşturulur.

**3. Metin Ön İşleme:** Metin ön işleme aşaması, ham metni analiz için uygun hale getirmek amacıyla gerçekleştirilen kritik bir adımdır. Bu süreç, metin verilerinin işlenebilir bir formata dönüştürülmesini ve analizlerin doğruluğunu artırmayı hedefler. Metin ön işlemede uygulanan temel adımlar aşağıdaki şekilde sıralanabilir:

- Gereksiz Karakterlerin Temizlenmesi: Metin içerisinde yer alan gereksiz karakterler (noktalama işaretleri, sayılar) ve HTML etiketleri temizlenerek, analiz için önemli olmayan unsurlar çıkarılır.
- Durak Kelimelerin Kaldırılması: Metin içerisinde anlam taşımayan "durak kelimeler" (örneğin "ve", "bir", "bu") temizlenir.
- Küçük Harfe Dönüştürme: Metinde yer alan tüm harfler küçük harfe dönüştürülerek tutarlılık sağlanır.
- Tokenizasyon: Tokenizasyon işlemi ile metin, kelime, cümle veya karakter gibi anlamlı parçalara ayrılır.
- Kök Bulma: Kök Bulma işlemi ile kelimelerin son ekleri çıkarılarak kök forma indirgenir.

**4. Özellik Çıkarımı:** Özellik çıkarımı, ham metin verilerinin anlamlı ve analiz edilebilir bir forma dönüştürülmesidir. Bu aşama, metindeki bilgileri sayısal bir temsil haline

getirerek, makine öğrenimi modellerinin ve diğer analiz araçlarının bu verilerle çalışmasını sağlar. En yaygın olarak kullanılan yöntemler; Çanta Kelime Modeli, TF-IDF Matrisi ve Kelime Gömme modelidir.

- **Çanta Kelime Modeli:** Çanta Kelime Modeli, metindeki her kelimenin frekansını hesaplayarak bir özellik vektörü oluşturur. Bu yöntem, kelimelerin sırasını dikkate almaz, yalnızca metinde hangi kelimelerin geçtiğine ve ne sıklıkta kullanıldığına odaklanır. Kolay uygulanabilir olması ve düşük maliyetle çalışması avantaj sağlarken, bir kelimenin diğer kelimelerle ilişkisini göz ardı etmesi özellikle NLP uygulamalarında modelin performansını olumsuz etkileyebilir.

Örnek Metin: "Kediler köpekleri sever."

Sözlük: {"kediler", "köpekleri", "sever"}

Vektör: [1, 1, 1] (her kelime metinde bir kez geçtiği için)

- **TF-IDF Matrisi:** TF-IDF, metindeki önemli kelimeleri belirlemek için hem kelime frekansını (TF) hem de bu kelimenin kaç farklı belgede geçtiğini (IDF) dikkate alır. Bu yöntem, sık geçen genel kelimelerin önemini azaltırken, nadir fakat kritik öneme sahip kelimelere daha fazla ağırlık verir.

**Terim Frekansı (TF):** Bir kelimenin belirli bir belgede kaç kez geçtiğini ifade eder.

$$TF = \frac{\text{Kelimenin Belgedeki Frekansı}}{\text{Toplam Kelime Sayısı}} \quad (8)$$

**Ters Doküman Frekansı (IDF):** Kelimenin toplam kaç belgede geçtiğini ölçer ve ters logaritmasını alır.

$$IDF = \log \left( \frac{\text{Toplam Belge Sayısı}}{\text{Kelimenin Geçtiği Belge Sayısı}} \right) \quad (9)$$

**TF-IDF Hesaplama:** Her kelimenin TF ve IDF değerlerinin çarpımıdır.

$$TF - IDF = TF * IDF \quad (10)$$

- **Kelime Gömme (Word Embeddings):** Kelime gömme yöntemleri, kelimeleri çok boyutlu vektörler olarak temsil eder ve kelimeler arasındaki anlamsal ilişkileri yakalar. Word2Vec, GloVe ve FastText gibi popüler modeller bu yöntemi uygular.

**Word2Vec:** Kelimeleri, belirli bir bağlamda görülen diğer kelimelerle ilişkisine dayanarak temsil eder.

**GloVe:** Küresel kelime-ortam matrislerinden anlamsal bilgi çıkarır.

**FastText:** Kelimelerin alt birimlerini de dikkate alarak daha zengin bir temsil sağlar.

### 2.3.2. Doğal Dil İşleme

Doğal dil işleme, bilgisayarların insan dilini anlamasını, işlemesini ve üretmesini mümkün kılan bir yapay zekâ alanıdır. NLP, dilbilim, bilgisayar bilimi ve makine öğrenimi tekniklerini bir araya getirerek, metin veya konuşma tabanlı veriyle çalışır. İnsan dili son derece karmaşık bir yapıya sahiptir; konuşma kuralları, kelime anlamları, bağlam ve kültürel referanslar gibi unsurları içerir. NLP'nin temel amacı, insan dilini bilgisayarların anlayabileceği bir forma dönüştürerek, doğal dilde iletişimi kolaylaştırmaktır. Bu amaçla NLP;

- Bilgisayarların insan dilini anlamasını ve bu dilde yanıt verebilmesini sağlar. Örnek olarak; Sesli komutlarla çalışan cihazlar (Siri, Alexa) ve müşteri hizmetlerinde kullanılan chatbotlar verilebilir.
- Yapılandırılmamış metin veya konuşma verilerinden anlamlı bilgiler elde eder. Örnek olarak; müşteri geri bildirimlerinden duygu analizi ve büyük metin koleksiyonlarından bilgi çıkarımı.
- Metin sınıflandırma, dil çevirisi, özetleme gibi uygulamaları mümkün kılar.

1950'lerde Alan Turing'in "Turing Testi" ile temelleri atılan bu alan, 1960'larda kural tabanlı çeviri sistemleriyle başlamıştır. Turing, bilgisayarların doğal dilde insanlarla iletişim kurabilme yeteneğini sorgulamıştır. 1970'lerde anlambilim odaklı çalışmalara yönelmiştir. 1980'lerde istatistiksel yöntemler 1990'larda makine öğrenimi (örneğin, TF-IDF, LSI) ile daha güçlü hale gelen NLP, 2000'lerde büyük veri analitiği ve otomatik çeviri gibi uygulamalarla genişlemiştir. 2010'larda derin öğrenme teknikleri ve Transformer mimarisıyla (BERT, GPT) devrim yaşayan NLP, günümüzde büyük dil modelleriyle tıp, hukuk, pazarlama ve sosyal bilimler gibi çeşitli alanlarda etkili çözümler sunmaktadır.

NLP, müşteri geri bildirimlerinden sağlık kayıtlarına, yasal belgelerden sosyal medya analizine kadar geniş bir uygulama yelpazesine sahiptir. İş zekâsı ve pazarlamada müşteri yorumları analiz edilerek stratejiler geliştirilebilirken, sağlık sektöründe hastalık tahmini ve tıbbi araştırmalar için veriler işlenir. Hukuk alanında mahkeme kararları ve sözleşmeler analiz edilirken, eğitimde öğrenci geri bildirimleri ve materyal geliştirme süreçleri desteklenir. Sosyal bilimlerde toplumsal eğilimlerin ve kültürel değişimlerin analizi yapılırken, bilimsel araştırmalarda literatür taraması ve atıf analizleriyle araştırma trendleri belirlenir. Bu uygulamalar, NLP'nin yapılandırılmamış verileri anlamlı bilgilere dönüştürme kapasitesini ortaya koyar.

### 2.3.3. Bilgi Çıkarımı

Bilgi çıkarımı, yapılandırılmamış veya yarı yapılandırılmış metinlerden belirli bilgiler veya veriler elde etme sürecidir. Bu tür, genellikle isimlendirilmiş varlık tanıma, ilişki çıkarımı ve olay çıkarımı gibi alt bileşenleri içerir.

- İsimlendirilmiş Varlık Tanıma (NER): Metinlerdeki kişi isimleri, yer adları, organizasyon isimleri gibi belirli varlıkları tanımlama ve etiketleme süreci.
- İlişki Çıkarımı: Metinler arasındaki ilişki ve bağlantıları belirleme. Örneğin, bir şirket ile CEO'su arasındaki ilişki.
- Olay Çıkarımı: Metinlerde belirtilen olayları ve bu olaylarla ilgili önemli bilgileri tespit etme şeklinde karşımıza çıkabilir.

### 2.3.4. Duygu Analizi

Duygu analizi, metin tabanlı verilerdeki duygusal tonları belirlemeye yönelik bir doğal dil işleme tekniğidir. Bu analiz, bir metinde ifade edilen duygu ve düşünceleri algılayarak, bunları pozitif, negatif veya nötr gibi kategorilere ayırmayı amaçlar. Duygu analizi, metin tabanlı verilerin duygusal tonlarını anlamak ve sınıflandırmak için kullanılan bir doğal dil işleme tekniği olarak, müşteri geri bildirimlerinden sosyal medya yorumlarına kadar çeşitli kaynaklardan anlamlı bilgiler elde edilmesine olanak tanır. İş dünyasında, duygu analizi müşteri memnuniyetini artırmak, ürün geliştirme süreçlerini iyileştirmek ve pazarlama stratejilerini optimize etmek için kritik bir araçtır. Örneğin, bir markaya yönelik pozitif veya negatif algılar analiz edilerek, marka imajı güçlendirilebilir veya olası krizler önceden tespit edilebilir. Sosyal medyada kamuoyu görüşleri ve toplumsal eğilimler analiz edilerek sosyal, ekonomik veya politik kararlar şekillendirilebilir. Sağlık sektöründe, hasta yorumlarının analiziyle tedavi süreçleri optimize edilebilirken, finans dünyasında piyasa trendleri ve yatırımcı duyarlılığı duygu analizi ile öngörülebilir. Duygu analizi, bireylerin metinlerdeki duygusal ifadelerini algılayarak, işletmelerin, araştırmacıların ve politikacıların daha iyi stratejik kararlar almasını sağlar. Bu bağlamda, duygu analizi, yalnızca bir analiz yöntemi değil, aynı zamanda insan davranışını anlama ve tahmin etmede vazgeçilmez bir araçtır.

Doğal dil işleme sürecinde kullanılan yöntemler, problemin doğasına ve verinin yapısına göre farklılık gösterir. Bu yöntemlere dair bilgiler aşağıdaki şekildedir;

- **Kural tabanlı yöntemler:** Duygu analizi sürecinde genellikle önceden tanımlanmış duygu sözlüklerine dayanır. Kelimelerin veya ifadelerin duygu puanları kullanılarak metnin genel duygusal tonu belirlenir. Örneğin, müşteri geri bildirimlerinde sıkça geçen "harika" gibi pozitif ifadeler veya "korkunç" gibi negatif ifadeler değerlendirilerek ürün veya hizmete yönelik genel memnuniyet düzeyi analiz edilebilir.
- **Makine öğrenimi yöntemleri:** Duygu analizi için yaygın bir şekilde kullanılan başka bir tekniktir. SVM, Naive Bayes ve Lojistik Regresyon gibi algoritmalar, eğitilmiş modeller aracılığıyla duygu sınıflandırması yapar. Örneğin, bir sosyal

medya platformundaki kullanıcı yorumlarının pozitif, negatif veya nötr olarak sınıflandırılması, bu algoritmalarla etkin bir şekilde gerçekleştirilebilir.

- **Derin öğrenme yöntemleri:** Duygu analizinde transformer tabanlı modeller (örneğin, BERT ve GPT), dilin bağlamsal özelliklerini anlamak ve daha karmaşık metin yapılarında bile yüksek doğrulukla analiz yapmak için kullanılır. Bu yöntemler, özellikle ironi veya sarkazm gibi daha karmaşık duygusal ifadelerin çözümünde etkilidir. Bu yöntemler, yalnızca metnin duygusal tonunu belirlemekle kalmaz, aynı zamanda insan davranışını daha iyi anlamak, pazarlama stratejilerini optimize etmek, toplumsal eğilimleri analiz etmek ve karar alma süreçlerini desteklemek için işletmelere, araştırmacılara ve politika yapıcılara güçlü araçlar sunar. Duygu analizi, kullanılan yöntemlerin doğrudan problemin çözümüne adapte edilmesiyle, modern veri analitiği süreçlerinde stratejik bir rol oynamaktadır.

### 2.3.5. Metin Kümeleme

Metin kümeleme, benzer metin belgelerinin gruplandırılması sürecidir. Bu tür, yapılandırılmamış metin verilerinde benzerlik ve farklılıkları tespit ederek, belgeleri anlamlı gruplara ayırır. Metin kümeleme sürecinde kullanılan yöntemler, problemin doğasına ve verinin yapısına göre farklılık gösterir. Bu yöntemlere dair bilgiler aşağıdaki şekildedir;

- **Hiyerarşik Kümeleme:** Belgelerin hiyerarşik bir ağaç yapısında gruplanmasıdır.
- **K-Ortalamalar Kümeleme:** Belgelerin önceden belirlenmiş k sayıda küme içinde gruplanmasıdır.

### 2.3.6. Metin Görselleştirme

Metin görselleştirme, metin madenciliği sonuçlarını grafiksel ve görsel formatlarda sunma sürecidir. Bu tür, verilerin daha anlaşılır ve analiz edilebilir hale gelmesini sağlar. Metin görselleştirme sürecinde kullanılan yöntemler, problemin doğasına ve verinin yapısına göre farklılık gösterir. Bu yöntemlere dair bilgiler aşağıdaki şekildedir;

- **Kelime Bulutları:** Metinlerdeki en sık geçen kelimelerin, sıklığına göre daha baskın olacak şekilde yazılar ile görselleştirilmesidir.
- **Grafik Tabanlı Görselleştirmeler:** Konu modelleme ve kümeleme sonuçlarının grafikler ve diyagramlar aracılığıyla gösterilmesidir.

### 2.3.7. Konu Modelleme

Konu modelleme, yapılandırılmamış metin verilerinde gizli temaları (konuları) keşfetmek için kullanılan bir doğal dil işleme ve makine öğrenimi tekniğidir. Bu yöntem, büyük miktarda metin verisini analiz ederek, belgelerin içeriğinde hangi konuların baskın olduğunu otomatik olarak belirlemeye yardımcı olur. Konu modelleme, herhangi bir insan müdahalesi olmadan metin verilerinin anlamını yakalamayı amaçlayan denetimsiz öğrenme yöntemlerinden biridir.

Konu modelleme, 20. yüzyılın sonlarında bilgi geri getirme ve istatistiksel dil modelleme çalışmalarından doğmuş, özellikle 1990'larda LSI gibi tekniklerle ilk adımlarını atmıştır. Blei vd. [8] tarafından geliştirilen LDA, konu modellemede bir dönüm noktası olmuş ve belgelerdeki gizli temaları keşfetmek için probabilistik bir yaklaşım sunmuştur. 2010'lu yıllarda Non-Negative Matrix Factorization (NMF) gibi yöntemlerle çeşitlenen konu modelleme, derin öğrenmenin yükselişiyle sinirsel konu modelleri gibi bağlama duyarlı modellerle daha sofistike bir hale gelmiştir. Günümüzde, Transformer tabanlı modeller (BERT, GPT) konu modellemenin doğruluğunu artırırken, hibrit yaklaşımlar ve bulut teknolojileri daha büyük veri setlerinde uygulanabilirliği kolaylaştırmıştır. Bu evrim, konu modellemeyi hem akademik hem de ticari uygulamalarda vazgeçilmez bir araç haline getirmiştir.

Konu modelleme, geniş bir uygulama alanına sahiptir ve iş dünyasından sağlığa, eğitimden hukuka kadar çeşitli sektörlerde kullanılır. İş dünyasında müşteri geri bildirimleri ve pazarlama stratejilerinde ortak temaları belirlemek için kullanılırken, sağlık alanında hastalık belirtileri ve tedavi yöntemleri gibi verilerin analizinde kritik bir rol oynar. Eğitim sektöründe öğrenci geri bildirimlerini ve akademik eğilimleri analiz ederken, hukukta yasal belgeler ve mahkeme kararlarındaki eğilimlerin incelenmesini sağlar. Ayrıca, sosyal medya ve haber analizinde trendleri ve halkın ilgi alanlarını

anlamak için kullanılır. Bu kapsamlı kullanım, konu modellemeyi yapılandırılmamış metin verilerini anlamlandırmada güçlü bir araç haline getirmiştir.

Konu modelleme, bir belge koleksiyonundaki her bir belgenin belirli bir oranda birden fazla konuyla ilişkilendirilebileceğini varsayar. Bu işlem, kelimelerin belgelerdeki ortaklıklarına dayanır. Konu modelleme, veri setinin yapısına, boyutuna ve hedefe bağlı olarak farklı yöntemlerle uygulanabilir. LDA, probabilistik yapı nedeniyle geniş veri setlerinde tercih edilirken, NMF yorumlanabilirliğiyle daha küçük veri setlerinde öne çıkar. Sinirsel konu modelleme gibi yeni yöntemler, bağlam bilgisiyle daha derin analizler yapmayı mümkün kılmaktadır. Bu çeşitlilik, konu modelleme tekniklerinin her türlü metin analizi ihtiyacına uyarlanabilirliğini sağlamaktadır. Konu modellemede yaygın olarak kullanılan yöntemler şunlardır:

- **Latent Dirichlet Allocation (LDA):** Probabilistik bir model olup, belgelerdeki gizli konuları keşfetmek için yaygın olarak kullanılır.
- **Non-Negative Matrix Factorization (NMF):** Pozitif faktörizasyon yöntemine dayanır ve kolay yorumlanabilir çıktılar sunar.
- **Probabilistic Latent Semantic Analysis (pLSA):** Konu modellemede probabilistik bir yaklaşım sağlayarak kelime ve belgeler arasındaki ilişkileri analiz eder.
- **Structural Topic Modeling (STM):** Konularla bağlamsal bilgiler arasındaki ilişkiyi anlamak için meta-verilerden faydalanır.
- **Neural Topic Models (NTM):** Derin öğrenme algoritmalarıyla konuları bağlama duyarlı bir şekilde analiz eder.

Python, konu modelleme çalışması için oldukça yaygın olarak kullanılmaktadır. Bunun sebebi, Python dilinin metin madenciliği süreci için çeşitli kütüphaneleri ve araçlarının olmasından kaynaklanmaktadır. Python'da konu modelleme için yaygın olarak kullanılan kütüphaneler; Gensim, Scikit-Learn, NLTK, spaCy, TextBlob, BERTopic ve Mallet olarak öne çıkmaktadır.

### 2.3.7.1. LDA Algoritması

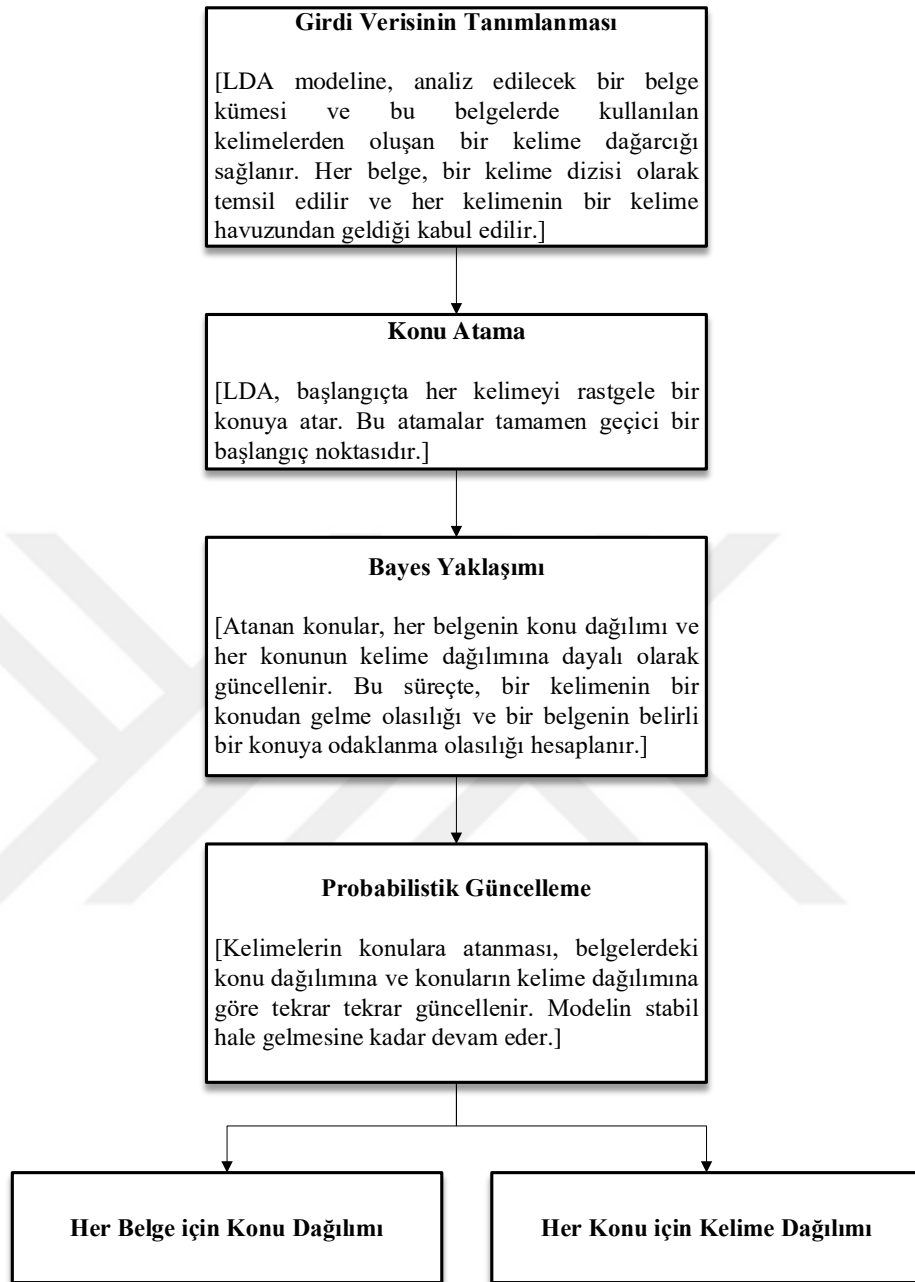
Latent Dirichlet Allocation (LDA), 2003 yılında Blei, Ng ve Jordan tarafından geliştirilen ve metinlerdeki gizli konuları keşfetmek için kullanılan bir probabilistik konu modelleme yöntemidir. LDA, yapılandırılmamış metin verilerinde gizli temaların belirlenmesi için en yaygın kullanılan modellerden biridir.

LDA, yapılandırılmamış metin verilerinden anlamlı konular çıkarmak için kullanılan probabilistik bir modeldir. Her belgeyi belirli konulara, her konuyu ise kelime dağılımlarına dayandırarak analiz eder. Geniş veri setlerinde esneklik ve ölçeklenebilirlik sunar, müşteri geri bildirim analizi ve sosyal medya araştırmaları gibi çeşitli alanlarda kullanılır. Ancak bağlama duyarsızlık, hesaplama karmaşıklığı ve ideal konu sayısının belirlenmesindeki zorluklar gibi sınırlamaları vardır. Alternatiflerinden farklı olarak, konuların probabilistik yapısını modelleyerek belgeler arası ilişkileri daha etkili bir şekilde ortaya çıkarır.

LDA'nın temel çalışma ilkeleri aşağıdaki gibidir;

- **Belge ve Kelime Dağılımı:** LDA, her belgenin farklı konulardan oluştuğunu ve her konunun farklı kelimelerle temsil edildiğini varsayar. Her belge, belirli bir konu dağılımına sahiptir. Her konu, kelimelerden oluşan bir dağılıma sahiptir.
- **Probabilistik Yapı:** LDA, bir belgeyi gözlemlenen kelimelerin ardışık bir dizisi olarak ele alır ve bu kelimelerin altında yatan gizli konuları modellemek için Bayes teoremini kullanır. Her kelimenin bir konudan geldiği varsayılır ve bu süreç probabilistik bir şekilde işler.
- **Varsayımlar:** Konular Dirichlet dağılımı kullanılarak modellenir. Belgelerdeki kelimelerin belirli bir konudan gelme olasılığı, o konunun kelime dağılımına dayanır.

LDA modeli çalışma prensibinde iki farklı çıktı üretmektedir. LDA Modeli ve İşleyişi Şekil 4'te yer almaktadır.



Şekil 4. LDA Modeli İşleyiş Süreci

### 3. BÖLÜM

## METİN MADENCİLİĞİ İLE MÜŞTERİ GERİ BİLDİRİM ANALİZİ: MOBİLYA ŞİRKETLERİ ÖRNEĞİ

Bu çalışmanın uygulama aşamasında, Türkiye’de faaliyet gösteren üç büyük rakip mobilya şirketinin satış sonrası hizmetler birimlerinden alınan toplam 9.930 müşteri şikâyeti, metin madenciliği ve veri madenciliği teknikleri kullanılarak analiz edilmiştir. Bu analizle, şirketlerin çağrı merkezi operasyonlarında karşılaşılan müşteri sorunlarının detaylı bir şekilde anlaşılması ve bu süreçlerde iyileştirmeye ihtiyaç duyulan alanların belirlenmesi hedeflenmiştir. Çalışma, ilk aşamada müşteri geri bildirimlerinin yapısal olmayan doğasından yola çıkarak markalar ve konular bazında yoğunlaştığı alanları tespit etmeye odaklanmıştır. Ayrıca, buradan hareketle verilerin müşteri hizmetleri süreç yönetimi kapsamında ne kadar etkili işlendiğini ve süreç yönetiminin verimliliğini tespit etmeye odaklanılmıştır.

Araştırma kapsamında incelenen süreçlerde, müşteri geri bildirimlerinin kaydedilmesi, analiz edilmesi ve çözüm süreçlerine yönlendirilmesi için kullanılan mevcut ERP sistemleri temel alınmıştır. Her üç şirketin de çağrı merkezlerine gelen müşteri bildirimleri, ortak bir işletim sistemi aracılığıyla dört ana çağrı tipinden birine atanarak kayıt altına alınmaktadır. Bu süreçte müşteri temsilcilerinin bireysel yorumları doğrultusunda bir çağrı tipi belirlenmekte ya da sistem üzerinde mevcut kayıtlar manuel olarak incelenip güncellenmektedir. Ancak bu yaklaşım, insan faktörüne dayalı olduğu için hatalara açık bir yapı sergilemekte ve doğru çağrı tipi ile müşteri geri bildiriminin uyumunu tam olarak sağlayamama riski taşımaktadır. Çalışmada metin madenciliği işleme adımlarından sonra ilk olarak mevcut çağrı tiplerinin müşteri geri bildirimlerini ne ölçüde doğru temsil ettiği ve uyumluluğu değerlendirilmiştir.

Çağrı tipleri ile uyumluluğu değerlendirilen geri bildirimlerde gizlenen; talep, şikâyet, istek vb. ana temaların ortaya çıkarılması amacıyla konu modelleme uygulanmış ve uyumluluğu değerlendirilen geri bildirimlere göre sınıflandırma modeli geliştirilerek müşteri geri bildirimleri için en doğru çağrı yönlendirmesi sağlayan algoritma belirlenmeye çalışılmıştır.

Araştırmanın nihai amacı, mevcut çağrı merkezi süreçlerinin doğruluğunu ve etkinliğini artırarak müşteri memnuniyetini en üst düzeye çıkarmak için iyileştirme önerileri sunmaktır. Bu doğrultuda çalışmada, araştırma konusu, uygulanan yöntemler, sınırlılıklar ve elde edilen bulgular ayrıntılı bir şekilde incelenmiştir. Elde edilen sonuçlar, mobilya sektöründeki müşteri şikâyet yönetiminin daha verimli hale getirilmesi ve müşteri odaklı stratejiler geliştirilmesine katkı sunmayı hedeflemektedir.

### **3.1. Uygulamanın Konusu Amacı ve Önemi**

Günümüz teknolojisi, işletmelere müşteri geri bildirimlerini hızlı ve etkin bir şekilde analiz etme fırsatı sunmaktadır. Özellikle satış sonrası hizmetlerin etkinliği, müşteri memnuniyetinin sağlanmasında ürün kalitesi kadar kritik bir rol oynamaktadır. Bu bağlamda, müşteri geri bildirimlerinin analizi, işletmelerin müşteri beklentilerini anlaması, sıkça şikâyet edilen alanları tespit etmesi ve bu alanlarda iyileştirmeler yapması için stratejik bir öneme sahiptir.

Bu çalışmanın amacı, Türkiye’de faaliyet gösteren üç büyük mobilya şirketine ait müşteri geri bildirimlerini metin madenciliği yöntemleri ile analiz ederek çağrı merkezi süreçlerini iyileştirmektir. Araştırma kapsamında, müşteri geri bildirimleri doğal dil işleme teknikleriyle analiz edilmiş, bu sayede şikâyet yönetiminde karşılaşılan sorunlar belirlenmiş ve doğru çağrı yönlendirmesi için çağrı atama modeli geliştirilmiştir.

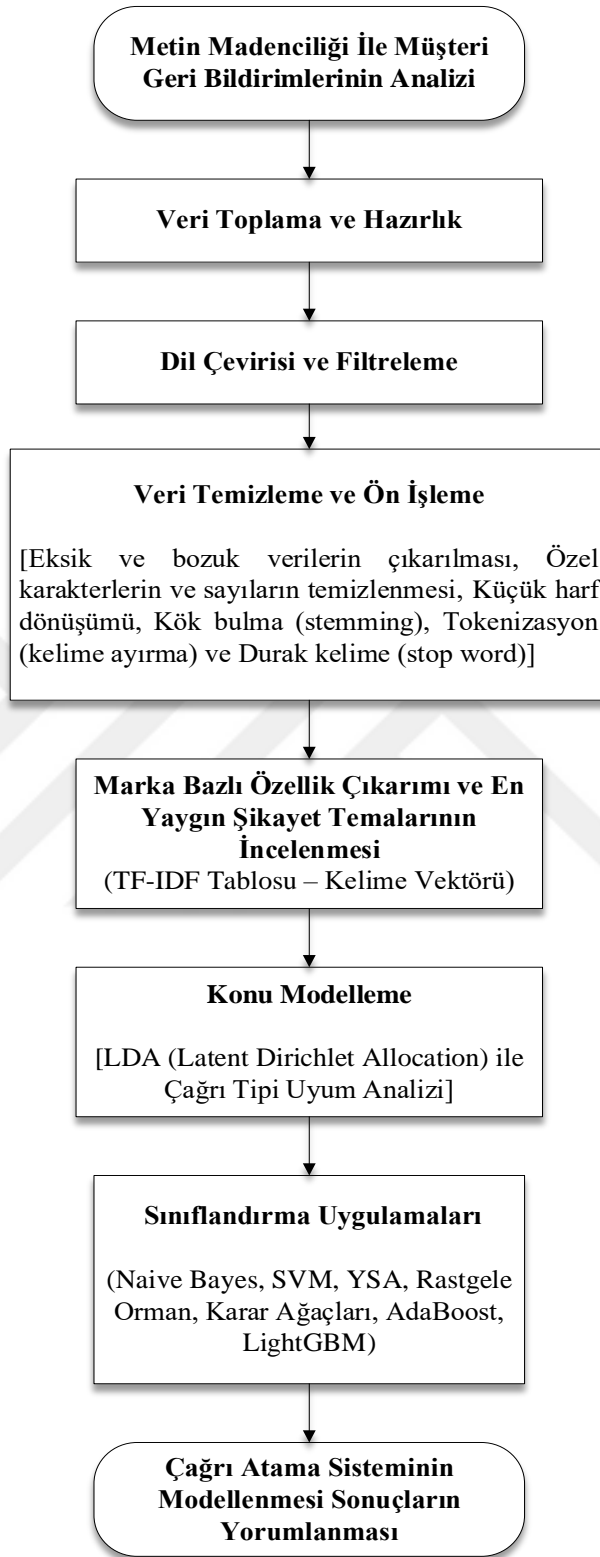
Müşteri geri bildirimlerinin doğru bir şekilde analiz edilmesi, çağrı yönlendirme sürecinin hızlanmasını sağlarken müşteri memnuniyetini artırır ve işletmelerin stratejik karar alma süreçlerine katkı sunar. Bu çalışma, mobilya sektörüne özgü şikâyetlerin analizini derinlemesine ele alarak sektörün satış sonrası hizmetlerde iyileştirme yapmasına ve rekabet avantajı elde etmesine rehberlik etmektedir. Ayrıca, metin

madenciliği metodolojisinin uygulama alanına yönelik yenilikçi bir perspektif sunmaktadır.

### **3.2. Uygulamanın Kapsamı**

Bu araştırma, Türkiye’de faaliyet gösteren ve mobilya üretim sektöründe lider konumda bulunan üç farklı markanın satış sonrası hizmet birimlerinden elde edilen müşteri geri bildirimlerini kapsamaktadır. Araştırma, 2022 yılının ikinci yarısında Şikayetvar.com ve sosyal medya platformları aracılığıyla çağrı merkezine ulaşan ve Türkçe yazı diliyle ifade edilmiş toplam 9.930 müşteri geri bildirimini analizini içermektedir. Analiz sürecinde, müşterilerin sorunlarını, memnuniyetsizliklerini, taleplerini ve önerilerini içeren yazılı yorumlar değerlendirilmiştir.

Araştırmada kullanılan veri seti, çağrı merkezlerinde tutulan veri tabanlarından elde edilen, yapılandırılmamış metin formatındaki müşteri geri bildirimlerinden oluşmaktadır. Araştırmanın uygulama sürecinde, Python programlama dili kullanılarak geri bildirimlerin detaylı analizi gerçekleştirilmiştir. Veri işleme adımlarında metin madenciliği teknikleri, NLP yöntemleri ve konu modelleme gibi araçlardan yararlanılmıştır. Bu tekniklerin uygulanması, hem çağrı merkezi süreçlerinin daha etkin yönetilmesi hem de müşteri memnuniyetinin artırılmasına yönelik stratejik iyileştirmelere ışık tutmaktadır. Çalışmanın uygulama sürecini görselleştiren akış şeması Şekil 5’te yer almaktadır. Bu şema, veri toplama, işleme, analiz ve sonuçların değerlendirilmesi aşamalarını detaylı bir şekilde açıklamaktadır.



Şekil 5. Uygulama Süreç Akış Şeması

Şekil 5'te gösterilen uygulama aşamaları aşağıda detaylandırılmıştır:

**Veri Toplama ve Hazırlık:** Bu aşamada, çağrı merkezlerinden elde edilen toplam 9.930 adet müşteri geri bildirimleri toplanmıştır. Çalışmanın temel veri seti oluşturulmuş ve analize uygun formata dönüştürülmüştür. Müşteri ve markalara ait kişisel bilgiler, gizlilik politikaları doğrultusunda filtrelenmiştir. Metinsel veriler, analiz kolaylığı sağlamak amacıyla İngilizceye çevrilmiştir.

**Dil Çevirisi ve Filtreleme:** Bu aşamada, elde edilen verilerin dili Türkçe olduğu için, Python kütüphanelerinde işlenebilmesi amacıyla geri bildirim metinleri NLP teknikleri kullanılarak İngilizceye çevrilmiştir. Çeviri işlemi, geri bildirimlerin anlam bütünlüğünü koruyacak şekilde otomatik çeviri araçlarıyla yapılmış ve doğruluğu manuel olarak kontrol edilmiştir. Marka ve müşteri bilgileri anonim şekilde filtrelenerek veriye dahil edilmiştir.

**Veri Temizleme ve Metin Ön İşleme:** Bu adımda, veri seti içindeki gürültüyü ve gereksiz unsurları ortadan kaldırmak için veri temizleme işlemleri gerçekleştirilmiştir. Eksik ve bozuk verilerin çıkarılması, özel karakterlerin ve sayıların temizlenmesi işlemleri Python programı kullanılarak uygulanmıştır. Metinler üzerinde küçük harf dönüşümü, Tokenizasyon, kök bulma ve durak kelime çıkarımı işlemleri gerçekleştirilmiştir.

**Özellik Çıkarımı:** Bu aşamada, ön işleme tabi tutulan veriler kullanılarak TF-IDF matrisi ve kelime vektörü çıkarılmıştır. Bu işlem, marka bazında özelleştirilerek müşteri geri bildirimleri için önemli terimler belirlenmiştir.

**Konu Modelleme:** Geri bildirimlerin mevcut çağrı tipleri ile uyumlarının belirlenmesi amacıyla, literatürde yaygın olarak kullanılan LDA konu modelleme algoritması uygulanarak yorumların çağrılara olasılıklı olarak dağıtılması sağlanmıştır. Ayrıca, bu adımda geri bildirimlerin içerdiği ana temalar LDA'ya göre modellenmiş ve mevcut çağrı tipleriyle birlikte değerlendirilerek marka bazlı olarak incelenmiştir.

**Sınıflandırma Uygulamaları:** İlk aşamada mevcut çağrı tipleri hedef değişken olarak alınmış ve metin madenciliği süreçleri uygulanarak oluşturulan TF-IDF özellikleri kullanılarak sınıflandırılmıştır. İkinci aşamada, LDA olasılık dağılımı dikkate alınarak

belirlenen (en yüksek olasılıklı çağrı tipi) yeni çağrı tipini hedef değişken olarak SVM, Naive Bayes, Yapay Sinir Ağları, Karar Ağaçları, Rastgele Orman, Adaboost ve LightGBM yöntemleri ile sınıflandırılmıştır. Bu aşamada, en iyi performansı gösteren algoritmalar kıyaslanarak geri bildirimlerin doğru çağrı tipine yönlendirilmesi için kullanılmıştır.

**Sonuçların Yorumlanması:** Son adımda, elde edilen analiz sonuçları yorumlanmış ve müşteri geri bildirimlerinin işletme için hangi alanlarda iyileştirme yapılabileceği konusunda içgörüler elde edilmiştir. Sonuçların, marka bazında müşteri memnuniyetinin artırılmasına ve satış sonrası hizmet süreçlerinin optimize edilmesine yönelik stratejiler geliştirilmesinde kullanılması hedeflenmiştir.

### 3.3. Uygulamanın Evreni ve Örneklemi

Bu araştırmanın evreni, Türkiye'de mobilya sektöründe faaliyet gösteren ve satış sonrası hizmetler birimleri aracılığıyla müşteri geri bildirimleri toplayan üç büyük mobilya şirketinin müşterilerinden oluşmaktadır. Bu şirketler, sektörlerinde lider konumda olup geniş bir müşteri tabanına sahiptir. Dolayısıyla, araştırma kapsamında değerlendirilen müşteri geri bildirimleri, Türkiye'deki mobilya sektöründe yaygın olarak karşılaşılan şikâyetleri ve geri bildirimleri temsil etmektedir.

Araştırmada kullanılan örneklem, bu üç mobilya şirketinin çağrı merkezine 2022 Haziran-Aralık döneminde gelen 9.930 müşteri geri bildiriminden oluşmaktadır. Örneklem seçiminde, yılın ikinci yarısında yazılı olarak farklı kaynaklardan kayıt altına alınan müşteri geri bildirimleri dikkate alınmıştır. Bu seçim, geri bildirimlerin temsil gücünü artırmış ve analiz edilen konu modelleme sonuçlarının geçerliliğini sağlamıştır. Örneklem, müşteri memnuniyetsizliği ve beklentilerine ilişkin geniş bir perspektif sunmuş, mobilya sektöründe karşılaşılan sorun alanlarının tespit edilmesine imkân vermiştir.

Metinler toplanırken birçok farklı formatta depolanabilmektedir. Örneğin; xml, csv, txt, pdf, xls, rtf, img vb. Elde edilen veriler xlsx formatındadır. Burada önemli nokta, kullanılacak metin madenciliği programının verileri işleyebileceği formatın belirlenmesidir. Çalışmamda Python'un tercih edilme sebebi ise, bu çeşitli formatları

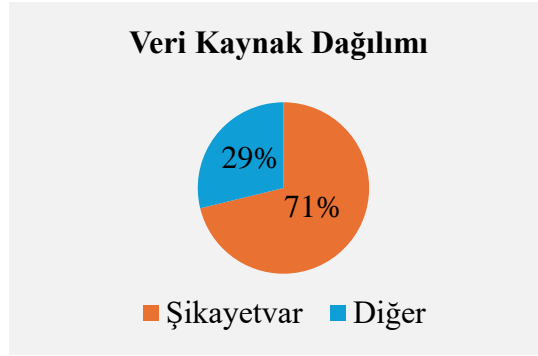
işleyebilme kabiliyetine sahip geniş kütüphane desteği ve metin madenciliği süreçlerinde hızlı, etkili veri işleme olanakları sunmasıdır. Ayrıca, Python kütüphanelerinin büyük çoğunluğu İngilizce metinler üzerinde optimize edilmiş olduğundan, metinlerin İngilizceye çevrilmesi analizlerin daha doğru ve verimli olmasını sağlamaktadır. Tablo 2’de veri setinde yer alan değişken örnekleri gösterilmiştir.

Tablo 2. Birinci Örnek Veri Seti

Kayıt Yeri	Marka	Çağrı Tipi	Yorum
Şikayetvar	A	Servis Talebi	Yay sesi
Diğer	B	SSH Sorunları	A'dan aldığımız koltuk takımından hiç memnun değiliz A bizi gerçekten hayal kırıklığına uğrattı kalitesi kesinlikle eskidenmiş maalesef koltuğa oturduğu an her yerinden ses geliyor çok rahatsız edici onca para verip alıyorsun markasız koltuklardan hiçbir farkı yok kesinlikle tavsiye etmem kimseye #Markadan beklentileriniz nedir?# #Tamir-Düzeltilme#
Şikayetvar	B	SSH Sorunları	2018 yılında almış olduğunuz B four seasons çift kişilik yatakta çökmeler meydana geldi. Her gün bel ağrısı ile uyanıyoruz. 3 kez müşteri hizmetleri geldi inceleme yaptı çökmesi normal dediler sünger değişimi yaptılar ama 2 ay geçmeden yeniden çökme başladı. Hiç memnun kalmadık B markasına güvenerek aldık ancak mağdur olduk. #Fatura ve diğer kayıtlarında geçen müşteri adı-soyadı nedir?# #E. P.# #Markadan beklentileriniz nedir?# #Ücret İade, Değişim#

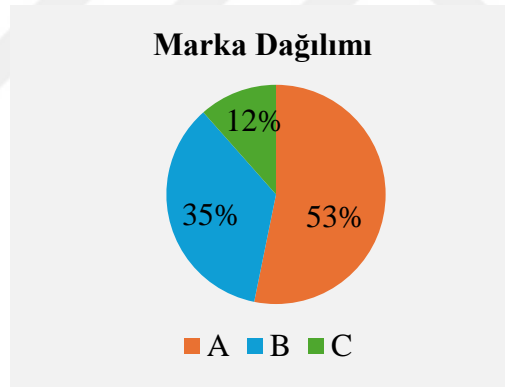
Diğer	A	SSH Sorunları	Benim sizden aldığım iki koltuk takımını var ikisinde de sorun yaşıyorum daha 1 buçuk senelik koltuklar hatta şu an da O. koltuk takımım tamire gitti M. koltuk takımında da aynı sorunlar yaşıyorum koltuk çöktü gıcırıyor yazık günak ben alırken 3 takım aldım 2 si bozuk çıktı acil ürünümün değişimini istiyorum #Fatura ve diğer kayıtlarında geçen müşteri adı-soyadı nedir? #Markadan beklentileriniz nedir?# #Özür, Ücret İade, Değişim# #Şikâyet konu bayii/şube/mağaza/servis adı ve adresi (cadde-ilçe-il veya açıklayıcı konum) nedir?# #H. mobilya Elazığ aldığım 2 oturma odasına ayıplı ürün çıktı#
-------	---	---------------	---

**Kayıt Yeri:** Yorumların hangi sistemden elde edildiğini gösteren kategorik bir değişkendir. Kayıt yeri değişkeni; “Şikayetvar ve Diğer” olmak üzere iki farklı değerden oluşmaktadır. Şikayetvar, Türkiye'deki tüketicilerin ürün ve hizmetlerle ilgili yaşadıkları sorunları paylaşarak çözüm aradığı, popüler bir şikâyet platformudur. Bu platform, kullanıcıların sorunlarını dile getirmelerine ve firmalarla iletişim kurmalarına olanak tanır. Bu nedenle, araştırmada Şikayetvar'dan elde edilen geri bildirimler, müşteri memnuniyetsizliği ve beklentileri hakkında kapsamlı ve gerçekçi bir bakış açısı sunmaktadır. Diğer türünde elde edilen yorumlar çağrı merkezine sosyal medya aracılığı ile ulaşan yorumları göstermektedir. Kayıt yeri değişkenine ait dağılım oransal olarak Şekil 6'da yer almaktadır.



Şekil 6. Veri Kaynak Dağılımı

**Marka:** Çağrı merkezine farklı kayıt yerleri ile gelen yorumların hangi mobilya markasına ait olduğunu gösteren kategorik bir değişkendir. İlk aşamada verilerin elde edilmesinde markaların gizliliğini korumak için her markaya bir ID tanımlaması yapılmıştır. Kayıt yeri değişkenine ait dağılım Şekil 7’de yer almaktadır.



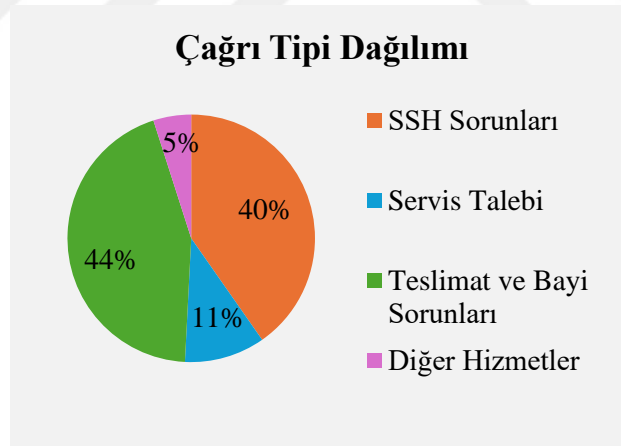
Şekil 7. Marka Dağılımı

Toplam veri seti içerisindeki yorum sayısının A markasında yoğunlaşmış olmasının temel sebebi; A markasının diğer markalara göre daha geniş bir satış ağının bulunması ve marka değerinin daha büyük olmasıdır. Satış ağı gelişmiş olan A markası için satılan ürün sayısı ve verilen hizmet sayısı diğerlerinden daha fazladır. Bu sebeple çağrı merkezine geri dönüş oranının da daha fazla olması beklenen bir durumdur. Aynı durum B ve C markaları arasında yer alan dağılım içinde geçerlidir.

**Çağrı Tipi:** Çağrı Tipi, müşteri yorumlarının hangi tür sorun veya talep ile ilgili olduğunu gösteren kategorik bir değişkendir. Müşterilerin çağrı merkezine ilettiği oldukları talepler, şikayetler ve öneriler, çağrı merkezi müşteri temsilcileri veya Şikayetvar platformundaki kullanıcılar tarafından manuel olarak dört ana kategoriye atanmıştır:

1. SSH Sorunları: Satış Sonrası Hizmetlerle ilgili sorunlar,
2. Servis Talebi: Teknik destek ve bakım talepleri,
3. Teslimat ve Bayi Sorunları: Ürün teslimatı ve mağazalarla ilgili problemler,
4. Diğer Hizmetler: Yukarıdaki kategorilere girmeyen hizmet sorunları.

Şekil 8’de çağrı tipi değişkeninin dağılımı gösterilmiştir. Bu dağılım incelendiğinde, müşteri geri bildirimlerinin özellikle SSH Sorunları ve Teslimat ve Bayi Sorunları kategorilerinde yoğunlaştığı görülmektedir. Bu iki kategori, müşteri memnuniyetsizliğinin büyük ölçüde satış sonrası hizmetler ve teslimat süreçlerinde yaşandığını ortaya koymaktadır.



Şekil 8. Çağrı Tipi Dağılımı

Çağrı tiplerinin manuel olarak belirlenmesi, bazı durumlarda subjektif etkiler yaratabilir. Bu durum, aynı tür sorunların farklı müşteri temsilcileri tarafından farklı kategorilere atanmasına yol açabilir. Örneğin, bazı karmaşık ya da kapsamlı sorunlar bir temsilci tarafından "SSH Sorunları" olarak atanırken, başka bir temsilci tarafından "Teslimat ve Bayi Sorunları" ya da "Servis Talebi" olarak sınıflandırılabilir.

Tablo 3’te yer alan örnekler, aynı marka için alınan farklı çağruların benzer şikayetlere rağmen farklı çağrı tiplerine atanabileceğini göstermektedir. Örneğin:

- Teslimat ve Bayi Sorunları kategorisine atanan bir geri bildirimde, koltukların yaylarında ses problemi olduğu belirtilmiştir.
- Aynı tür şikâyet içeren bir başka geri bildirim, SSH Sorunları kategorisine atanmıştır ve bu atama farkı, manuel sınıflandırmanın etkisini ortaya koymaktadır.

Tablo 3. İkinci Örnek Veri Seti

Kayıt Yeri	Marka	Çağrı Tipi	Yorum
Diğer	A	Teslimat ve Bayi Sorunları	Van İ. T. A mağazasından ürünleri alalı yaklaşık 2-3 yıl oldu. Koltukların yaylarında ses geliyor. Dış görünümünde herhangi bir olumsuzluk olmamasına rağmen rahatsız edici bir şekilde ses var. Bunun giderilmesini talep etmekteyim. #Fatura ve diğer kayıtlarında geçen müşteri adı-soyadı nedir?# #H. Y.# #Şikâyete konu bayii/şube/mağaza/servis adı ve adresi (cadde-ilçe-il veya açıklayıcı konum) nedir?# #VAN İ. T. A MAĞAZASI# #Markadan beklentileriniz nedir?# #Tamir-Düzeltilme#
Diğer	A	SSH sorunları	Benim sizden aldığım iki koltuk takımını var ikisinde de sorun yaşıyorum daha 1 buçuk senelik koltuklar hatta şu an da O. koltuk takımım tamire gitti M. koltuk takımında da aynı sorunlar yaşıyorum koltuk çöktü gıcırıyor yazık günak ben alırken 3 takım aldım 2 si bozuk çıktı acil ürünümün değişimini istiyorum #Fatura ve diğer kayıtlarında geçen müşteri adı-soyadı

			<p>nedir?# #Y. E. Ç.# #Markadan          beklentileriniz nedir?# #Özür, Ücret İade,          Değişim# #Şikâyete konu          bayii/şube/mağaza/servis adı ve adresi          (cadde-ilçe-il veya açıklayıcı konum)          nedir?# #H. mobilya Elazığ aldığım 2          oturma odasına ayıplı ürün çıktı#</p>
--	--	--	--

Manuel sınıflandırma süreçlerinde, müşteri temsilcilerinin bireysel yorumlarına dayalı olarak yapılan atamalar, geri bildirimlerin kategorilere doğru bir şekilde dağıtılmasında subjektif bir etki yaratmaktadır. Bu durum, aynı tür sorunların farklı çağrı tiplerine atanmasına ve dolayısıyla veri analiz süreçlerinde hatalara yol açabilmektedir. Bu nedenle, bu sınıflandırma problemini çözmek ve analiz sürecini objektif bir şekilde gerçekleştirmek amacıyla metin madenciliği ve NLP tekniklerinin uygulanması hedeflenmiştir. Özellikle, konu modelleme algoritmaları ve makine öğrenimi tabanlı sınıflandırma yöntemlerinin kullanılmasıyla, müşteri geri bildirimleri daha doğru bir şekilde kategorize edilebilecektir.

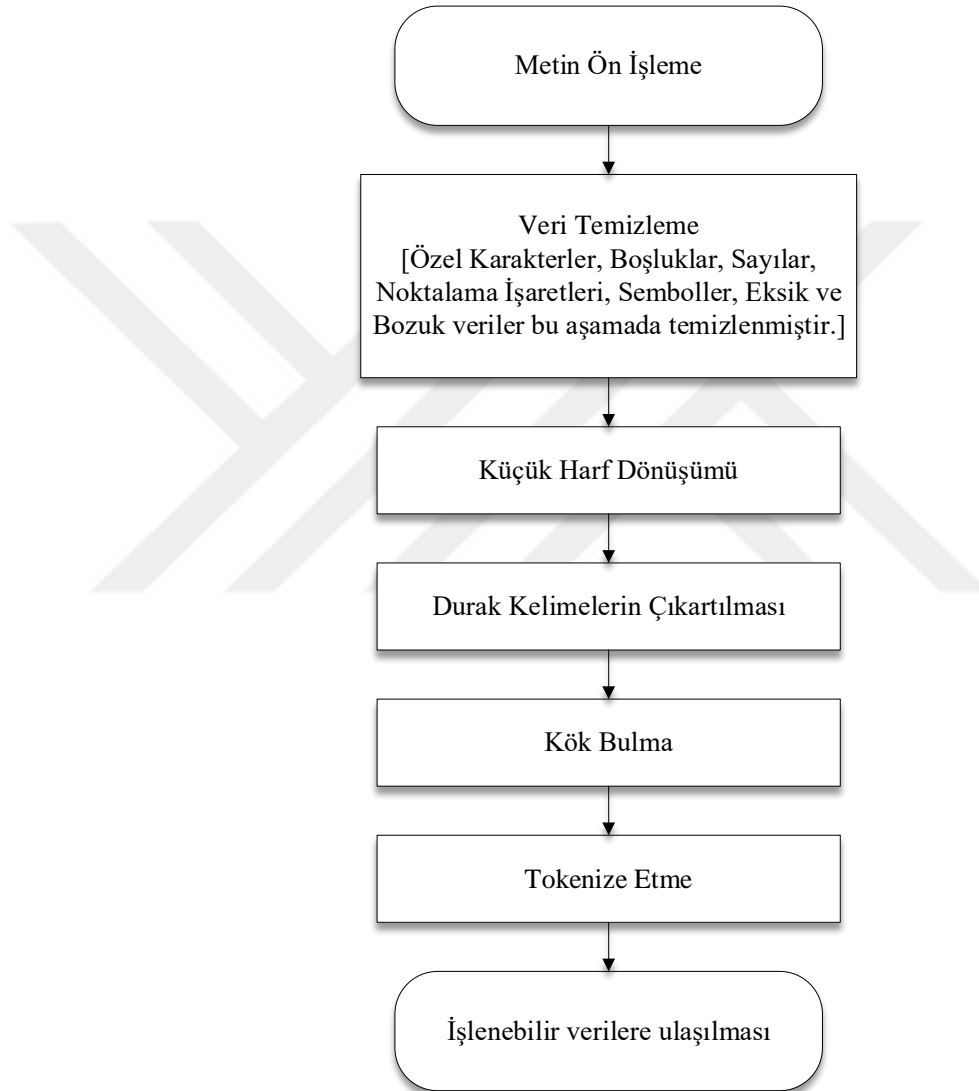
**Yorum:** Müşterilerin ürün ve hizmetlerle ilgili şikayetlerini ve taleplerini doğal dilde ifade ettiği metinlerdir. Bu yorumlar, analiz için metin madenciliği süreçlerinde kullanılan temel veriyi oluşturmuştur.

### 3.4. Metin Madenciliği Süreci

Bu bölümde, müşteri geri bildirimlerinin metin madenciliği teknikleri kullanılarak nasıl analiz edildiği detaylandırılmıştır. Bu bölüm, analiz verilerinin Python programı ile; metin ön işleme, konu modelleme ve sınıflandırma süreci olarak üç ana başlık altında ele alınmıştır.

### 3.4.1. Metin Ön İşleme

Metin ön işleme adımı, doğal dilde elde edilen müşteri geri bildirimlerinin analiz edilebilir hale getirilmesi için Python Pandas, NLTK ve re kütüphaneleri kullanılmıştır. Bu süreç Şekil 9'da yer alan adımları içermektedir.



Şekil 9. Metin Ön İşleme Süreci Akış Şeması

Şekil 9'da gösterilen metin ön işleme süreci, müşteri geri bildirimlerini analiz edilebilir hale getirmek için çeşitli adımlardan oluşmaktadır. Bu süreçte, ham metinler üzerinde temizleme, dönüşüm ve yapılandırma işlemleri gerçekleştirilmiştir.

Aşağıda yer alan tabloda (Tablo 4), metin ön işleme sürecindeki her bir adımın amacı ve uygulanma süreci özetlenmiştir:

Tablo 4. Metin Ön İşleme Adımları

İşlem Adımı	Veri	Açıklama
Ham Veri	"A marka baza koku yaptı iade kabul edilmesi için servisin ürünü göndermesi gerekiyormuş yaklaşık 2 aya yakın bazasız kalacağım."	
Temizlenmiş Veri	"A marka baza koku yaptı iade kabul edilmesi servisin ürünü göndermesi gerekiyormuş aya yakın bazasız kalacağım"	Bu aşamada noktalama işaretleri ve "2" sayısı temizlenmiştir. Büyük harfler küçük harfe dönüştürüldü
Küçük Harf Dönüşümü	"a marka baza koku yaptı iade kabul edilmesi servisin ürünü göndermesi gerekiyormuş aya yakın bazasız kalacağım"	Bu aşamada metindeki tüm büyük harfler küçük harflere çevrilerek veri tutarlılığı sağlanmıştır.
Durak Kelimelerin Çıkartılması	"A marka baza koku yaptı iade kabul servis ürün gönder gerek aya yakın bazasız kalacağım"	Bu aşamada "için", "ve", "gibi", "bu", "yaklaşık" gibi durak kelimeler çıkarılmıştır.
Kök Bulma	"marka baza koku yap iade kabul servis ürün gönder gerek ay yakın baza kal"	Bu aşamada kelimenin farklı biçimlerinin tek bir biçime indirgenmesi sağlanmıştır.
Tokenizasyon	"marka", "baza", "koku", "yap", "iade", "kabul", "servis", "ürün", "gönder", "gerek", "ay", "yakın", "bazasız", "kal"	Bu aşamada her bir kelime bağımsız olarak ele alınmıştır.

Metin ön işleme sürecinin tamamlanmasıyla birlikte, müşteri geri bildirimleri analiz edilebilir bir formata getirilmiştir. Bu aşamada, metinlerin daha anlamlı özelliklerle temsil edilmesi sağlanmıştır.

### 3.4.2. Özellik Çıkarımı

Özellik çıkarımı, metin madenciliğinde metin ön işleme ile hazırlanan verileri sayısallaştırarak anlamlı özellikler elde etme sürecidir. Bu aşamada, TF-IDF yöntemi kullanılarak müşteri geri bildirimleri için bir kelime listesi ve kelime-belge matrisi oluşturulmuştur. Kelime listesi, belgelerde öne çıkan ve farklılık gösteren terimleri içerirken, kelime-belge matrisi bu terimlerin her bir belgede ne kadar önemli olduğunu gösteren sayısal değerleri temsil etmektedir.

Tablo 5'te yer alan kelime tablosundan yararlanarak sadece veri ön işleme sonucunda bile yorumlanabilecek pek çok sonuç elde edilmiştir. Kelimeler incelediğinde elde edilen sonuçlar, veri tabanında yer alan şikayetlerden bazı çıkarımlar yapılmasını sağlamaktadır. Tablo 5 müşteri şikayetlerinde sıklıkla tekrar eden anahtar kelimeleri ve bu kelimelerin dokümanlar içerisindeki önemini göstermektedir. Özellikle "ürün", satın alma ve "hizmet" gibi kelimelerin yüksek frekansları, bu alanlarda müşterilerin yaşadığı problemlerin yoğun olduğunu işaret etmektedir. Ürün kelimesi en yüksek toplam frekansa (11.381) sahip olan kelime olması bu konunun şikayetlerde oldukça yoğun olarak yer aldığını gösterir. Bu, müşterilerin en çok ürünlerle ilgili sorunlar yaşadığını ya da ürünlerin şikayetlerde sıkça bahsedildiğini ortaya koymaktadır.

Tablo 5. Birinci Kelime Listesi Örneği

Kelime	Toplam Kelime Sayısı	Dokümanda Bulunma Sayısı
satın alındı	9.063	6.390
marka	7.197	5.593
beklemek	5.460	5.177
ürün	11.381	4.907
özel	6.935	4.904
isim	7.885	4.521
kayıt	4.419	4.092
hizmet	8.359	4.064
fatura	4.243	4.027
takım	5.730	3.955

şikayet	4.740	3.917
soyadı	3.888	3.884
dedi	6.989	3.834
çağrı	6.614	3.817
bahsetmek	3.858	3.767
açıklama	3.621	3.604
adres	3.680	3.582
konum	3.598	3.537
olurdu	5.413	3.418
kanepe	5.710	3.333

Kanepe kelimesinin hem toplam frekansta hem de doküman sayısında belirgin bir şekilde yer alması, kanepe ürün grubunun diğer mobilyalara göre müşteriler arasında sıkça dile getirilen bir konu olduğunu gösterir. Kanepe, genellikle evlerde en çok kullanılan mobilya parçalarından biridir. Bu nedenle, daha fazla sayıda müşteri bu ürün grubunu satın almış ve kullanıyor olabilir. Satışların yüksek olması, doğal olarak şikâyet sayısını da artırabilir, çünkü daha fazla kullanıcı deneyimi geri bildirimini doğurur. Ayrıca kanepe kelimesinin yüksek frekansı, bu ürün grubunda belirli sorunların tekrarlandığını gösterebilir. Bu da kanepe ürün grubunun kalite kontrol, tasarım veya hizmet sonrası süreçlerinde iyileştirme gerektirdiğine işaret eder.

Tablo 6: İkinci Kelime Listesi Örneği

<b>Kelime</b>	<b>Toplam Kelime Sayısı</b>	<b>Dokümanda Bulunma Sayısı</b>
ay	4.904	3.319
konu	3.272	3.183
sorun	5.090	3.111
değişim	4.356	3.098
bayi/şube/mağaza/servis	3.083	3.083
gün	4.644	2.931
Marka A	4.245	2.763
bile	3.653	2.760
bir	3.815	2.699
istemek	3.776	2.692

zaman	3.393	2.518
Marka B	5.057	2.517
ulaşmak	3.881	2.516
demek	3.445	2.436
mağaza	4.136	2.365
geldi	2.996	2.323
sipariş	3.078	2.213
koltuk	4.064	2.206
gelmek	3.032	2.141
beğenmek	2.643	2.102
sokak/ilçe/il	2.042	2.042
teslimat	3.024	1.963
talep	2.360	1.885
yatak	4.014	1.875

Tablo 6’da yer alan "değişim" kelimesinin yüksek sıklığı (4.356), müşterilerin ürün değişimi talep ettiklerine işaret etmektedir. Bu durum, ürünlerin kalitesi veya beklentileri karşılamaması nedeniyle yaşanan memnuniyetsizliği işaret edebilir. Özellikle mobilya sektöründe, müşterilerin ürünleri uzun süre kullanma beklentisi göz önüne alındığında, ürün değişimi taleplerinin bu kadar yaygın olması, kalite kontrol ve ürün geliştirme alanlarında iyileştirme fırsatlarını ortaya koyar.

Tablo 6’da yer alan “bayi/şube/mağaza/servis” kelimesinin yüksek sıklığı (3.083), müşterilerin büyük bir bölümünün sorunlarının bayiler, mağazalar veya servislerle ilgili olduğunu göstermektedir. Özellikle mobilya sektörü gibi mağazacılık sektörünün büyük önem taşıdığı sektörlerde marka algısını üst düzeye taşımak için müşterilerin mağazalarda yaşadığı sorunlar ile yakından ilgilenilmesi gerekmektedir. Kelimenin toplam geri bildirimlerden %31’inde bulunması ile önem taşıyan bir geri bildirimdir.

Tablo 6’da yer alan "Teslimat" kelimesinin sıkça tekrarlanması, teslimat sürecinin müşteri memnuniyetini etkileyen kritik bir faktör olduğunu gösterir. Müşteriler, teslimatın zamanında yapılmaması, hasarlı ürünlerin gönderilmesi veya teslimat sırasında yaşanan diğer sorunlar nedeniyle şikâyetle bulunabilirler. Bu durum

markaların teslimat süreçleri ve bu süreçlerde yaşanan problemlerin üzerine odaklanılması gerektiğine işaret eder.

Markaların isimlerinin yüksek frekansı, müşterilerin şikâyet ve geri bildirimlerinde bu markaları sıkça dile getirdiklerini gösterir. Bu durum, markaların müşteri memnuniyetini artırmak için belirli sorun alanlarına odaklanmaları gerektiğine işaret eder. Tabloya göre A markası 2.763 farklı yorumda 4.245 kez tekrar ederek toplam en yoğun geri bildirim verilen marka haline gelmiştir. Şikâyet alma yoğunluğuna göre A markasını B markası takip etmektedir. B markası tüm yorumlarda toplamda 5.057 kez geçerek A markasından daha çok tekrar etmiştir.

Marka A:  $4.245 / 2.763 \approx 1.54$  - Her bir dokümanda yaklaşık 1.54 kez geçmektedir.

Marka B:  $5.057 / 2.517 \approx 2.01$  - Her bir dokümanda yaklaşık 2.01 kez geçmektedir.

Marka B, geri bildirim verilen dokümanlarda daha sık tekrar eden bir marka olarak öne çıkmaktadır. Bu, her bir müşteri geri bildiriminde Marka B'nin daha fazla vurgulandığını veya markaya dair sorunların daha yoğun ifade edildiğini gösterir. Diğer bir deyişle, Marka B ile ilgili geri bildirimler, belirli sorunlarda daha fazla odaklanmayı işaret edebilir.

“Koltuk” ve “yatak” gibi ürün isimlerinin yüksek frekansı, özellikle bu ürün gruplarında yaşanan sorunların sıklıkla dile getirildiğini göstermektedir. Özellikle kanepe ve yatak gibi mobilyalar, günlük kullanıma uygun ve dayanıklı olmalıdır. Koltuk kelimesinin tüm yorumların %22’sini, yatak kelimesinin tüm yorumların %18’ini oluşturması bu iki ürün grubunun hem çok satan ürünler olduğuna hem de bu ürün gruplarında çeşitli sebeplerden sorunlar ile karşılaşıldığına işaret eder. Bu nedenle, bu ürün grupları için üretim süreçlerinin gözden geçirilmesi ve müşteri geri bildirimlerine dayalı olarak iyileştirmeler yapılması gerekmektedir.

Tablo 7’de marka bazlı kelime listesi sunulmuştur.

Tablo 7. Marka Bazlı Kelime Listesi

A MARKASI KELİME LİSTESİ			B MARKASI KELİME LİSTESİ			C MARKASI KELİME LİSTESİ		
Kelime	Toplam Kelime Sayısı	Dokümanda Bulunma Sayısı	Kelime	Toplam Kelime Sayısı	Dokümanda Bulunma Sayısı	Kelime	Toplam Kelime Sayısı	Dokümanda Bulunma Sayısı
ürün	6.064	2.608	ürün	3.961	1.739	ürün	1.356	560
satın aldı	4.889	3.416	satın aldı	3.139	2.238	satın aldı	1.035	736
hizmet	4.472	2.202	hizmet	3.011	1.438	hizmet	876	424
isim	4.320	2.468	isim	2.702	1.556	isim	863	497
marka	3.857	3.023	marka	2.547	1.952	marka	793	618
dedi	3.749	2.035	dedi	2.475	1.380	dedi	765	419
müşteri	3.727	2.669	müşteri	2.461	1.710	çağrı	759	438
çağrı	3.531	2.049	çağrı	2.324	1.330	müşteri	747	525
takım	3.005	2.101	kanepa	2.095	1.188	takım	645	461
kanepa	2.974	1.759	takım	2.080	1.393	kanepa	641	386
beklenti	2.945	2.807	olurdu	1.956	1.240	olurdu	608	394
olurdu	2.849	1.784	beklenti	1.925	1.810	beklenti	590	560
şikayet	2.580	2.128	sorun	1.753	1.091	ay	552	373
gün	2.475	1.539	ay	1.662	1.134	şikayet	549	458
kayıt	2.410	2.241	gün	1.654	1.055	sorun	526	325
değişim	2.361	1.677	şikayet	1.611	1.331	gün	515	337
fatura	2.334	2.208	değişim	1.560	1.108	kayıt	489	441
mağaza	2.244	1.273	kayıt	1.520	1.410	bir	470	335
koltuk	2.243	1.196	fatura	1.459	1.390	mağaza	468	266
yatak	2.157	995	mağaza	1.424	826	ulaştı	457	290
soyadı	2.134	2.134	yatak	1.412	651	fatura	450	429
bahsetmek	2.123	2.072	koltuk	1.371	759	koltuk	450	251
ulaştı	2.084	1.353	soyadı	1.340	1.336	yatak	445	229
istemek	2.048	1.452	ulaştı	1.340	873	istemek	442	330
adres	2.046	1.989	bahsetmek	1.331	1.299	değişim	435	313
bir	2.040	1.431	bir	1.305	933	zaman	418	297
bile	2.015	1.512	istemek	1.286	910	soyadı	414	414

Tablo 7’de yer alan marka bazlı kelime listelerini incelediğimizde, kelimelerin toplam frekans ve doküman sayıları açısından birbirlerinden farklılıklar gösterdiği gözlemlenmektedir. A markası için en sık kullanılan kelimeler: “ürün” (6.064), "satın aldı" (4.889), "hizmet" (4.472), "isim" (4.320), "marka" (3.857) olarak şekillenmiştir. Marka A'da en yüksek frekansa sahip kelime "ürün" olarak belirlenmiştir. Bu, markanın ürünleriyle ilgili şikâyet ve geri bildirimlerin çok sayıda olduğunu göstermektedir. "hizmet" ve "satın aldı" gibi kelimeler, ürünlerin satın alma ve satış sonrası hizmet süreçlerinde yaşanan sorunlara işaret etmektedir. "isim" ve "marka" gibi kelimelerin yüksek frekansları, müşterilerin şikayetlerinde marka adını veya ürün isimlerini sıkça belirtme eğiliminde olduklarını gösterir. Listede yer alan "beklemek" (2.945), "şikayet" (2.580), "mağaza" (2.244), “yatak” (2.157) kelimeleri müşterilerin beklentilerini, şikayet konularını ve özellikle yatak ürünleriyle ilgili sıkça dile getirdikleri sorunları yansıtır.

B markası için en sık kullanılan kelimeler: "ürün" (3.961), "satın almak" (3.139), "hizmet" (3.011), "isim" (2.702), "marka" (2.547) olarak şekillenmiştir. Marka B'nin kelime dağılımı, Marka A'ya benzer şekilde "ürün" kelimesinin öne çıktığını gösteriyor.

Ancak, frekanslar A'ya göre daha düşük. "satın aldı" ve "hizmet" kelimeleri, bu markanın ürün satın alma ve satış sonrası hizmetlerde karşılaşılan sorunlarla ilgili sıkça geri bildirim aldığını göstermektedir. Ayrıca, "şikayet" (1.611) kelimesinin sıklığı, bu markanın şikayetlerle ilgili önemli miktarda müşteri geri bildirimini aldığını ortaya koyar. "mağaza" (1.424) ve "yatak" (1.412) gibi kelimeler, mağazalarla veya belirli ürünlerle ilgili yaşanan sorunlara işaret eder. Listede yer alan "ulaşmak" (1.340), "değişim" (1.560) bu kelimeler, teslimat süreçleri ve ürün değişimi ile ilgili geri bildirimlerin önemli olduğunu gösterir.

C markası için en sık kullanılan kelimeler: "ürün" (1.356), "satın almak" (1.035), "hizmet" (876), "isim" (863), "marka" (793) olarak şekillenmiştir. Marka C'de de "ürün" kelimesi en yüksek frekansa sahip olmakla birlikte, diğer markalara kıyasla frekanslar oldukça düşük. Bu, C'nin diğer markalara göre daha az geri bildirim aldığını gösterebilir. "hizmet" ve "satın almak" kelimeleri yine satın alma ve satış sonrası hizmetler alanındaki sorunları işaret etmektedir. "şikayet" (549) kelimesinin sıklığı, C'nin şikayetlerle ilgili geri bildirimlerde belirli bir yoğunluğa sahip olduğunu gösterir. Listede yer alan "problem" (526), "beklemek" (590) bu kelimeler, bu markanın müşterilerinin ürün veya hizmetle ilgili beklentileri karşılayamaması sonucu yaşanan sorunları yansıtmaktadır.

A markası için gelen yorumlar genel anlamda ürünle ilgili sorunlar, ürün kalitesi, işlevsellik veya çeşitlilik gibi daha geniş kapsamlı konular üzerinde daha fazla geri bildirim vermekteyken, B ve C markasına gelen yorumlar daha çok belirli problemlere odaklanıyor. Örneğin, değişim, teslimat, koltuk ve yatak gibi belirli ürün veya süreçlerle ilgili konular daha belirgin. Bu, bu markalarda özellikle belirli ürün kategorilerinde veya teslimat ve iade süreçlerinde sıkça dile getirilen sorunların bulunduğunu gösteriyor. Bu nedenle, B ve C markasındaki geri bildirimlerin belirli ürünler veya süreçlerle daha doğrudan ilişkilendirildiği söylenebilir.

### 3.4.3. Konu Modelleme

Bu bölümde, çağrı merkezine gelen müşteri şikayetlerinin doğru çağrı tipine yönlendirilmesi amacıyla yapılan konu modelleme detaylı bir şekilde ele alınmaktadır. TF-IDF yöntemi, her bir müşteri geri bildiriminde geçen kelimelerin önemini ölçerek çağrı tiplerine özgü anahtar kelimeleri belirlemiştir. Bu süreçte her geri bildirim, kelime frekansları ve önem sıklıklarına göre vektörize edilmiştir. Müşteri şikayetlerinin yalnızca bir çağrı tipine indirgenemeyecek kadar çok yönlü olduğu gözlemlenmiştir. Şikayetlerin birden fazla çağrı tipiyle ilişkilendirilmesini sağlamak amacıyla LDA algoritması kullanılarak her bir şikâyet için çağrı tiplerine göre olasılık dağılımı hesaplanmıştır. Şikayetlerin bu şekilde modellenmesindeki temel amaç; bir müşteri şikayetinin içerdiği farklı temaların birden fazla çağrı tipiyle belirli oranlarda ilişkilendirilebilmesini sağlamaktır. Böylelikle, çoklu yönlendirme sayesinde müşteri şikayetlerinin daha doğru bir şekilde sınıflandırılması mümkün hale gelmiştir. Bu modelin avantajları şunlardır:

- **Karmaşık Şikayetlerin Ele Alınması:** Geri bildirimlerde, belirli bir çağrı tipi ile sınırlı kalmayan, daha geniş kapsamlı şikayetlerin var olduğu anlaşılmıştır. Bu tür şikayetlerde, tek bir çağrı tipine atama yapılması, müşteri sorunlarını tam anlamıyla karşılamayabilir.
- **Departmanlar Arası İş Birliğinin Artırılması:** Bir şikâyetin birden fazla kategoriyle ilişkilendirilmesi, farklı birimlerin (örneğin, teslimat ekibi ve satış sonrası destek birimi) koordinasyon içinde çalışmasını gerektirir. Bu da kurum içi iş birliğini artırır.
- **İyileştirme Alanlarının Belirlenmesi:** Şikayetlerin yalnızca bir çağrı tipiyle değil, birden fazla tipteki sorunlarla ilişkilendirilmesi, şirketin iyileştirilmesi gereken süreçlerini daha iyi belirlemesini sağlar.
- **Müşteri Memnuniyetinin Artırılması:** Çoklu konu içeriğinin dikkate alınması, müşterilerin dile getirdiği tüm sorunların çözüme kavuşmasını sağlar. Bu, müşterilerin kendilerini daha iyi ifade ettiklerini hissetmelerine ve memnuniyetin artmasına katkıda bulunur.

Her bir çağrı tipi için TF-IDF yöntemiyle belirlenen anahtar kelimeler, müşteri geri bildirimlerinin hangi çağrı tipine daha yakın olduğunu saptamada kritik bir rol oynamıştır. Bu anahtar kelimeler, şikayetlerin içeriklerini analiz ederek, çağrı tiplerinin hangi sorun başlıklarında yoğunlaştığını açıkça göstermektedir. Örneğin, "teslimat", "gecikme" gibi kelimeler "Teslimat ve Bayi Sorunları" çağrı tipiyle yüksek TF-IDF skorlarına sahipken; "tamir", "garanti" gibi kelimeler "SSH Sorunları" (Satış Sonrası Hizmetler) çağrı tipinde öne çıkmıştır. Her çağrı tipi için belirlenen anahtar kelimeler Tablo 8’de sunulmuştur:

Tablo 8. Anahtar Kelimeler

Çağrı Tipi	Anahtar Kelimeler	Sorun Temsili
Teslimat ve Bayi Sorunları	Teslimat, gecikme, şube, kargo, gönderi.	Ürün teslimatı sırasında yaşanan gecikmeler, eksik ürünler ve bayiyle ilgili problemler.
SSH Sorunları	Şikâyet, garanti, iade, tamir, destek.	Garanti süreçleri, onarım talepleri ve müşteri memnuniyetsizliğine dair geri bildirimler.
Servis Talebi	Teknik destek, servis, onarım, koltuk, yatak.	Teknik servis ihtiyaçları ve ürün bakımı.
Diğer Hizmetler	Hizmet, bilgilendirme, müşteri hizmetleri, iletişim.	Genel müşteri hizmetleri talepleri ve iletişimle ilgili problemler.

Müşteri geri bildirimleri, LDA algoritması ile çağrı tipleriyle belirli olasılık oranlarında ilişkilendirilmiştir. Bu oranlar, bir şikâyetin içerdiği temaların farklı çağrı tipleriyle ilişkisini göstermektedir. Tablo 9’da bu analizin örnek sonuçları yer almaktadır.

Tablo 9. LDA ile Olasılıklı Atama Örneği

Yorum Numarası	Mevcut Çağrı Tipi	Servis Talebi	Teslimat ve Bayi Sorunları	SSH Sorunları	Diğer Hizmetler	Yeni Çağrı Tipi
2	Teslimat ve Bayi Sorunları	0,17	0,24	0,21	0,38	Diğer Hizmetler
3	Teslimat ve Bayi Sorunları	0,08	0,26	0,36	0,30	SSH Hizmetleri
4	SSH Sorunları	0,03	0,55	0,38	0,04	Teslimat ve Bayi Sorunları
5	Servis Talebi	0,20	0,18	0,12	0,50	Diğer Hizmetler
6	SSH Sorunları	0,07	0,17	0,36	0,40	Diğer Hizmetler

Tablo 9, müşteri şikâyetlerinin çok yönlü analizine dair örnek bir olasılıklı atama tablosu sunmaktadır. Bu tabloda her bir müşteri geri bildirimini, dört ana çağrı tipi (Servis Talebi, Teslimat ve Bayi Sorunları, SSH Sorunları ve Diğer Hizmetler) ile ilişkili olasılık dağılımlarına göre analiz edilmiştir. Her yorum için belirlenen olasılık değerleri, şikâyetin hangi çağrı tipleri ile daha fazla ilişkili olduğunu gösterir. Model, yorumların anahtar kelimelerine ve TF-IDF skorlarına göre her çağrı tipine belirli bir olasılık atayarak, şikâyetlerin kapsamını daha geniş bir perspektifte incelemeyi amaçlamıştır.

Tablo 9'da 3 numaralı yorum, LDA modeline göre büyük ölçüde SSH Sorunları (%35) kategorisiyle ilişkilendirilmiştir. Yüksek olasılık, yorumun içeriğinin satış sonrası hizmetlerde yaşanan sorunlara odaklandığını göstermektedir. Teslimat ve Bayi Sorunları kategorisiyle (%26), diğer hizmetler kategorisi ile (%30) ilişki kurulmuştur. Örnek olarak Tablo 9'da yer alan yorum numarası 3 olan geri bildirim incelenmiştir.

Yorum-3: “*Van İ. T. A. mağazasından ürünleri alalı yaklaşık 2-3 yıl oldu. Koltukların yaylarında ses geliyor. Dış görünümlerinde herhangi bir olumsuzluk olmamasına rağmen rahatsız edici bir şekilde ses var. Bunun giderilmesini talep etmekteyim.*”

Bu geri bildirimde, müşteri, 2-3 yıl önce Van İ. T. A. mağazasından alınan ürünlerin yaylarında ses sorunu olduğunu belirtmiştir. Ürünlerin dış görünümünde sorun olmadığını ifade eden müşteri, rahatsız edici ses probleminin çözülmesini talep etmektedir. Beklentisi onarım/tamir ile ilişkilidir. Bu duruma göre yorumun büyük çoğunluğu, ürünlerin kullanım sonrası ortaya çıkan sorunlarına odaklanmaktadır. Bu, açıkça satış sonrası hizmetler (SSH) kategorisinde değerlendirilmelidir. Özellikle ürünlerin yaylarından gelen rahatsız edici ses ve müşterinin çözüm talebi, SSH ile ilişkilidir.

Tablo 9’da 5 numaralı yorum, LDA modeline göre büyük ölçüde Diğer Hizmetler (%50) kategorisiyle ilişkilendirilmiştir. Bu örnekte, şikâyet Teslimat ve Bayi Sorunları kategorisiyle (%18) oranında ilişkilendirilmiş olsa da LDA modeline göre Diğer Hizmetler (%50) kategorisiyle daha güçlü bir bağ kurmaktadır. Bu durum, şikâyetin yalnızca teslimat süreçlerini değil, aynı zamanda bayi hizmet kalitesi ve genel müşteri memnuniyeti sorunlarını da içeren daha geniş bir çerçeveye sahip olduğunu göstermektedir. Servis Talebi (%19) kategorisi mevcut çağrı tipi olarak belirlenmiş olsa da bu durum yorumun yalnızca bir kısmına hitap etmektedir ve şikâyetin ana temasını yansıtmamaktadır. Şikâyet, aslında daha çok teslimat ve bayilik süreçlerinde yaşanan problemlerle ilgilidir.

Yorum-5: “*C bayisinden köşe koltuk almaya gittim. Öncelikle leke tutmayan ve dokununca iz bırakmayan kumaştan olmasını istedim. Görevli Ürünlerin nano teknoloji olduğunu ve kesinlikle iz olmayacağını söyledi. Sonrasında teşhirde bulunan New Minas koltuk takımını beğendik. Renk seçimi için kartela getirildi ve Antrasit-Krem renklerini seçtik. Ürünün 30-60 gün arsında geleceği söylendi ama 23 Aralık tarihinde ürünü teslim edeceğiz dediler. Ürünü getirdiklerinde bizim istediğimiz teşhir ürünündeki kumaştan farklı bir kumaş olduğunu gördük. Ürünü açmamalarını söyledik ama sözleşmede yazan kumaşın bu olduğu söylenerek açmak zorundayız, siz itirazınızı bayiiye yapın dediler. Bayiiide bulunan görevli bizim teşhirde gördüğümüz ürünün dışında onayımız olmadan ürünün ellerinde olan farklı bir kumaş kaplı olanını bize*

*göndermiş. Sipariş formunda yazan renk bizim istediğimiz ama kumaş bizim istediğimiz değil. Ayrıca ürün kumaşı üzeri tüm çizik doluydu. Servis gönderdiler çizikler gitti. Ürünün istediğimiz kumaşla değişmesini istiyoruz dediğimizde kabul etmiyorlar. Zorla istediğimiz koltuğa değil ellerindeki koltuğa oturtulmak zorunda bırakılmaya çalışılıyor. Çözüm odaklı bir firma değil.”*

- Ürün sergilenenle uyumsuz geldi: Satın alınan ürün, teşhirde görülen özelliklere uymamış, özellikle kumaş farklılıkları müşteri şikâyetinin ana temasını oluşturmuştur.
- Müşteri onayı olmadan farklı ürün teslim edilmesi: Bayi, müşterinin onayı olmadan ellerindeki farklı bir kumaştan ürünü teslim etmiş ve bu durum ciddi bir müşteri memnuniyetsizliği yaratmıştır.
- Ürün üzerindeki çizikler ve servis müdahalesi: Ürün üzerindeki çiziklerin servis tarafından giderilmiş olması, şikâyetin bir boyutunun SSH Sorunları kategorisiyle ilişkilendirilmesine neden olmaktadır.

Bu örnek, bir geri bildirim çoklu çağrı tipleriyle ilişkilendirilmesi gerektiğini ve bu ilişkilendirmenin sorun çözme süreçlerini nasıl daha etkili hale getirebileceğini göstermektedir. Bu örnek, çağrı tipi modelinin şikâyeti yalnızca bir başlık altında sınıflandırması durumunda memnuniyetsizlik yaratan diğer konuların göz ardı edilebileceğini göstermektedir. Çoklu atama yaklaşımı, müşterinin ihtiyaç duyduğu çözümün kapsamını genişletecek ve ilgili departmanlar arasında iş birliğini artırarak şikâyetin daha etkili bir şekilde ele alınmasını sağlayacaktır.

Tablo 10. Çağrı Dağılım Olasılıkları

Çağrı Tipi / Olasılıklı Atama	Servis Talebi	Teslimat ve Bayi Sorunları	SSH Sorunları	Diğer Hizmetler
<b>Servis Talebi</b>	0,043	0,367	0,164	0,426
<b>Teslimat ve Bayi Sorunları</b>	0,038	0,349	0,158	0,454
<b>SSH Sorunları</b>	0,032	0,361	0,156	0,450
<b>Diğer Hizmetler</b>	0,041	0,365	0,157	0,438

Tablo 10, müşteri şikayetlerinin farklı çağrı tipleriyle olan ilişkilendirilme olasılıklarını göstermektedir. Bu analiz, şikayetlerin yalnızca bir çağrı tipiyle sınırlandırılmayacak kadar çok yönlü olduğunu ortaya koymaktadır. Her bir çağrı tipi, diğer çağrı tipleriyle belirli oranlarda ilişkilidir ve bu, müşteri geri bildirimlerinin çok boyutlu yapısını yansıtmaktadır.

**SSH Sorunları:** SSH Sorunları çağrı tipi, kendisi ile %15.6 oranında ilişkilendirilmiştir. Bu durum, satış sonrası hizmetlerin müşteri şikayetlerinde rol oynadığını göstermektedir. Bu kategorinin Teslimat ve Bayi Sorunları %36.1 ve Diğer Hizmetler (%45) ile güçlü bir ilişkisi vardır. SSH sorunlarının teslimat ve bayi sorunlarıyla birlikte açık ve net bir şekilde belirlenmemiş olan diğer kategorisinde yüksek oranla çağrı merkezi yönetiminin bu alanda daha kolay çözümler geliştirmesi ihtiyacını göstermektedir.

**Teslimat ve Bayi Sorunları:** Teslimat ve Bayi Sorunları %34.9 oranıyla yüksek ilişkilendirilen kategoridir. Özellikle Diğer Hizmetler %45.4 ile güçlü ilişkisi olması bu alana yoğunlaşılması gerektiğini göstermektedir. Teslimat gecikmeleri, eksik ürünler veya bayi ile ilgili iletişim problemleri genellikle müşteri memnuniyetsizliğine yol açan ana faktörlerdir. Bu da teslimat sürecindeki iyileştirmelerin farklı şikâyet alanları ile uyumlu şekilde geliştirilmesi gerektiğini işaret eder.

**Servis Talebi:** Teslimat gecikmeleri, eksik ürünler veya bayi ile ilgili iletişim problemleri genellikle müşteri memnuniyetsizliğine yol açan ana faktörlerdir. Bu da satış sonrası hizmetlerin, teslimat süreçleriyle uyumlu şekilde geliştirilmesi gerektiğini işaret eder. Teknik servis taleplerinin %36.7 oranında Teslimat ve Bayi Sorunlarıyla bağlantılı olması, bu iki alanın ayrılmaz bir şekilde birbirine bağlı olduğunu göstermektedir.

**Diğer Hizmetler:** Diğer Hizmetler kategorisindeki yorumlar genellikle müşteri hizmetleri ve iletişim sorunlarını içerir. Bu şikayetlerin %15.7'si SSH Sorunlarıyla ilişkilendirilmiştir. Ayrıca, %36.5'lik bir oranla Teslimat ve Bayi Sorunlarıyla bağlantılıdır.

Tablo 11'de yer alan veriler, müşteri geri bildirimlerinin farklı çağrı tipleriyle ilişkilendirilme olasılıklarını göstermektedir. Bu analiz, modelin her müşteri geri bildirimini yalnızca tek bir çağrı tipiyle sınıflandırılmasının ötesine geçerek, şikayetlerin diğer çağrı tipleri ile de ne ölçüde bağlantılı olduğunu ortaya koyduğunu göstermektedir. Mevcut çağrı tipi ile model tarafından önerilen yeni çağrı tipi arasında yapılan karşılaştırma, şikayetlerin sınıflandırılmasında dikkat edilmesi gereken noktaları vurgulamaktadır.

Tablo 11. Çağrı Tipi Eşleşmeleri

Kategori	Yorum Sayısı	Oran (%)
Yüksek Olasılıkla Doğru Eşleşme	2.203	22,18%
Yüksek Olasılıkla Yanlış Eşleşme	7.727	77.81%

**Yüksek Olasılıkla Doğru Eşleşme:** Müşteri geri bildirimlerinin %22.18'i, mevcut çağrı tipiyle modelin önerdiği çağrı tipi arasında yüksek düzeyde uyum göstermektedir. Bu durum, modelin birincil olarak belirlenen çağrı tipini anlamlı bir şekilde desteklediğini gösterir. Ancak, doğru eşleşme sağlanan bu geri bildirimlerde dahi, diğer çağrı tipleriyle düşük seviyede de olsa bağlantılar bulunabilir.

**Yüksek Olasılıkla Yanlış Eşleşme:** Şikayetlerin %77.81’inde, modelin önerdiği çağrı tipi, mevcut çağrı tipinden farklıdır. Bu durum çağrı tipi uyumunun düşük olduğunu ve müşteri geri bildirimlerinin birden fazla çağrı tipiyle ilişkilendirilmesi gerektiğini ve şikayetlerin çok boyutlu analiz edilmesinin önemini vurgular. Yanlış eşleşme olarak nitelendirilen bu durumlar, aslında geri bildirim çok yönlü doğasını yansıtarak daha geniş kapsamlı bir çözüm ihtiyacını ortaya koymaktadır.

### 3.5. Sınıflandırma

Konu modelleme çalışması, gelen yorumların mevcut çağrı tipleri ile uyumunun düşük olduğunu göstermektedir. Elde ettiğimiz bu sonuçları doğrulamak amacıyla hem mevcut veride yer alan çağrı tipinin hedef değişken olması durumunda hem de LDA ile belirlenen olasılıkların en büyük olduğu değeri yeni çağrı tipi belirlenerek sınıflandırılması durumunda model performansları karşılaştırılmıştır. Sınıflandırma işlemlerinin uygulanabilmesi için veri setinde bulunan öznitelikler TF-IDF sonucunda oluşan belge-kelime tablosu girdi özellikleri olarak kullanılmıştır.

Çalışmada kullanılan algoritmaların amacı, sınıflandırma problemlerini çözmektir. Bu algoritmalar ile veri setindeki anlamlı desenleri ve ilişkileri öğrenerek, yeni ve görülmemiş verilere ilişkin tahminlerde bulunmaktır. Şekil 10’da ve Şekil 11’de yer alan veri setleriyle sürekli TF-IDF öznitelik değerlerinden hareketle gelen yorumun hangi çağrı tipi ile ilişkilendirilebileceğini tahmin edilmesi amaçlanmıştır.

TF-IDF Öznitelik Matris Değişkenleri						Hedef Değişken
Öznitelik 1	Öznitelik 2	Öznitelik 3	*****	*****	Öznitelik n	Eski Çağrı Tipi

Şekil 10. Mevcut Sınıflandırma Veri Seti

TF-IDF Öznitelik Matris Değişkenleri						Hedef Değişken
Öznitelik 1	Öznitelik 2	Öznitelik 3	*****	*****	Öznitelik n	Yeni Çağrı Tipi

Şekil 11. Yeni Sınıflandırma Veri Seti

Sınıflandırma için aşağıdaki algoritmalar, çapraz doğrulama yöntemiyle veriler eğitim ve test verisi olarak ayrılarak değerlendirilmiştir. Elde edilen sonuçlara göre, en yüksek

doğruluk değerine sahip olan algoritma çağrı atama modelinde kullanılmak üzere seçilmiştir. Sınıflandırma için kullanılan modeller aşağıdaki gibidir;

- Naive Bayes
- Yapay Sinir Ağları
- Destek Vektör Makinesi
- Karar Ağaçları
- LightGBM Algoritması
- AdaBoost
- Rastgele Orman

### 3.5.1. Naive Bayes

Python'da Naive Bayes sınıflandırıcısının sonuçları aşağıdaki gibidir (Tablo 12):

Tablo 12. Naive Bayes Sonuçları

Hedef Değişken	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Kappa
Yeni Çağrı Tipi	0.802	0.611	0.578	0.878	0.666
Eski Çağrı Tipi	0.421	0.215	0.244	0.230	-0.016

Doğruluk: Yukarıda tanımlanan Denklem (3) kullanılarak, yeni hedef değişkenle %80.27 doğruluk elde edilirken, eski hedef değişkenle bu oran %42.12'ye düşmüştür. Eski hedef değişkenle doğruluk, rastgele tahmine çok yakındır, bu da modelin sınıflar arasında net bir ayrım yapamadığını gösterir.

F1 Skoru: Yeni hedef değişkenle %61.21 F1 skoru, modelin iyi bir genel performansa sahip olduğunu ve doğru tahminlerin hem kesinlik hem de hatırlama açısından dengeli olduğunu gösteriyor. Eski hedef değişken ile F1 skoru ise %21.59 bu modelin sınıflandırmasında hem kesinlik hem de hatırlama açısından zayıf performans sergilediğini ve sınıflandırmanın pek güvenilir olmadığını gösterir.

Kappa değeri yeni hedef değişkende 0.666 ile modelin sınıflar arasında net bir ayrım yapabildiğini gösterir. Eski değişken ile yapılan sınıflandırmada Kappa değeri (-0.016),

şansa dayalı tahminin bile altına düştüğünü ve modelin sınıflar arasında anlamlı bir ayrım yapamadığını gösterir.

### 3.5.2. Yapay Sinir Ağları

Python'da Yapay Sinir Ağları sınıflandırıcısının sonuçları aşağıdaki gibidir (Tablo 13):

Tablo 13. Yapay Sinir Ağları Sonuçları

Hedef Değişken	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Kappa
Yeni Çağrı Tipi	0.885	0.850	0.837	0.867	0.820
Eski Çağrı Tipi	0.382	0.248	0.249	0.251	-0.007

**Doğruluk:** Yukarıda tanımlanan Denklem (3) kullanılarak, yeni hedef değişkenle %88.5 doğruluk elde edilmiştir, bu da modelin sınıflar arasında net bir ayrım yapabildiğini ve başarılı bir performans sergilediğini göstermektedir. Eski hedef değişkenle doğruluk oranı %38.2'ye düşmüş ve rastgele tahmine çok yakın bir seviyede kalmıştır; bu durum, modelin sınıflar arasında anlamlı bir ayrım yapamadığını ve düşük performans gösterdiğini ifade etmektedir.

**F1 Skoru:** Yeni hedef değişkenle %85.3 F1 skoru elde edilmiştir; bu, modelin genel performansının yüksek olduğunu ve doğru tahminlerin hem kesinlik hem de hatırlama açısından dengeli olduğunu göstermektedir. Eski hedef değişkenle F1 skoru ise %24.8 seviyesinde kalmış, bu da modelin sınıflandırma konusunda hem kesinlik hem de hatırlama açısından zayıf performans sergilediğini ve sonuçların güvenilir olmadığını ifade etmektedir.

**Kappa Değeri:** Yeni hedef değişkenle elde edilen 0.820 Kappa değeri, modelin sınıflar arasında net bir ayrım yapabildiğini ve oldukça başarılı bir performans sergilediğini göstermektedir. Eski hedef değişkenle yapılan sınıflandırmada ise Kappa değeri (-0.007) olarak bulunmuş, bu da modelin şansa dayalı tahminin bile altına düştüğünü ve sınıflar arasında anlamlı bir ayrım yapamadığını ifade etmektedir.

### 3.5.3. Destek Vektör Makinesi

Python’da Destek vektör Makinesi sınıflandırıcısının sonuçları aşağıdaki gibidir (Tablo 14):

Tablo 14. Destek Vektör Makinesi Sonuçları

Hedef Değişken	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Kappa
Yeni Çağrı Tipi	0.891	0.852	0.836	0.870	0.828
Eski Çağrı Tipi	0.424	0.227	0.249	0.210	-0.00251

**Doğruluk:** Yukarıda tanımlanan Denklem (3) kullanılarak, yeni hedef değişkenle %89.1 doğruluk elde edilmiştir, bu da modelin sınıflar arasında net bir ayrım yapabildiğini ve başarılı bir performans sergilediğini göstermektedir. Eski hedef değişkenle doğruluk oranı %42.4’e düşmüş ve rastgele tahmine çok yakın bir seviyede kalmıştır; bu durum, modelin sınıflar arasında anlamlı bir ayrım yapamadığını ve düşük performans gösterdiğini ifade etmektedir.

**F1 Skoru:** Yeni hedef değişkenle %85.2 F1 skoru elde edilmiştir; bu, modelin genel performansının yüksek olduğunu ve doğru tahminlerin hem kesinlik hem de hatırlama açısından dengeli olduğunu göstermektedir. Eski hedef değişkenle F1 skoru ise %22.7 seviyesinde kalmış, bu da modelin sınıflandırma konusunda hem kesinlik hem de hatırlama açısından zayıf performans sergilediğini ve sonuçların güvenilir olmadığını ifade etmektedir.

**Kappa Değeri:** Yeni hedef değişkenle elde edilen 0.828 Kappa değeri, modelin sınıflar arasında net bir ayrım yapabildiğini ve oldukça başarılı bir performans sergilediğini göstermektedir. Eski hedef değişkenle yapılan sınıflandırmada ise Kappa değeri (-0.00251) olarak bulunmuş, bu da modelin şansa dayalı tahminin bile altına düştüğünü ve sınıflar arasında anlamlı bir ayrım yapamadığını ifade etmektedir.

### 3.5.4. Karar Ağaçları

Python’da Karar Ağaçları sınıflandırıcısının sonuçları aşağıdaki gibidir (Tablo 15):

Tablo 15. Karar Ağaçları Sonuçları

Hedef Değişken	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Kappa
Yeni Çağrı Tipi	0.676	0.620	0.621	0.621	0.495
Eski Çağrı Tipi	0.380	0.251	0.251	0.252	-0.000436

**Doğruluk:** Yukarıda tanımlanan Denklem (3) kullanılarak, yeni hedef değişkenle %67.6 doğruluk elde edilmiştir, bu da modelin sınıflar arasında ortalama performansta belirgin bir ayırım yapabildiğini ve performansının iyi olduğunu göstermektedir. Eski hedef değişkenle doğruluk oranı %38’e düşmüş ve rastgele tahmine çok yakın bir seviyede kalmıştır; bu durum, modelin sınıflar arasında anlamlı bir ayırım yapamadığını ve düşük performans gösterdiğini ifade etmektedir.

**F1 Skoru:** Yeni hedef değişkenle %62 F1 skoru elde edilmiştir; bu, modelin genel performansının ortalama olduğunu ve doğru tahminlerin hem kesinlik hem de hatırlama açısından ortalama olduğunu göstermektedir. Eski hedef değişkenle F1 skoru ise %25.1 seviyesinde kalmış, bu da modelin sınıflandırma konusunda kesinlik ve hatırlama açısından sınırlı bir performans sergilediğini ve sonuçların güvenilir olmadığını ifade etmektedir.

**Kappa Değeri:** Yeni hedef değişkenle elde edilen 0,495 Kappa değeri, modelin sınıflar arasında ayırım yapma konusunda ortalama performansta olduğunu ve performansının şansa dayalı tahmine göre iyi olduğunu göstermektedir. Eski hedef değişkenle yapılan sınıflandırmada ise Kappa değeri (-0.000436) olarak bulunmuş, bu da modelin sınıflar arasında herhangi bir anlamlı ayırım yapamadığını ve tamamen rastgele tahmine yakın bir performans sergilediğini ifade etmektedir.

### 3.5.5. LightGBM Algoritması

Python’da LightGBM sınıflandırıcısının sonuçları aşağıdaki gibidir (Tablo 16):

Tablo 16. LightGBM Sonuçları

Hedef Değişken	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Kappa
Yeni Çağrı Tipi	0.863	0.819	0.798	0.847	0.783
Eski Çağrı Tipi	0.421	0.231	0.249	0.271	-0.0016

**Doğruluk:** Yukarıda tanımlanan Denklem (3) kullanılarak, yeni hedef değişkenle %86.3 doğruluk elde edilmiştir, bu da modelin sınıflar arasında ortalama performansta belirgin bir ayırım yapabildiğini ve performansının iyi olduğunu göstermektedir. Eski hedef değişkenle doğruluk oranı %42.1’e düşmüş ve rastgele tahmine çok yakın bir seviyede kalmıştır; bu durum, modelin sınıflar arasında anlamlı bir ayırım yapamadığını ve düşük performans gösterdiğini ifade etmektedir.

**F1 Skoru:** Yeni hedef değişkenle %81.9 F1 skoru elde edilmiştir; bu, modelin genel performansının ortalama olduğunu ve doğru tahminlerin hem kesinlik hem de hatırlama açısından ortalama olduğunu göstermektedir. Eski hedef değişkenle F1 skoru ise %23.1 seviyesinde kalmış, bu da modelin sınıflandırma konusunda kesinlik ve hatırlama açısından sınırlı bir performans sergilediğini ve sonuçların güvenilir olmadığını ifade etmektedir.

**Kappa Değeri:** Yeni hedef değişkenle elde edilen 0,783 Kappa değeri, modelin sınıflar arasında ayırım yapma konusunda ortalama performansta olduğunu ve performansının şansa dayalı tahmine göre iyi olduğunu göstermektedir. Eski hedef değişkenle yapılan sınıflandırmada ise Kappa değeri (-0.0016) olarak bulunmuş, bu da modelin sınıflar arasında herhangi bir anlamlı ayırım yapamadığını ve tamamen rastgele tahmine yakın bir performans sergilediğini ifade etmektedir.

### 3.5.6. AdaBoost

Python'da AdaBoost sınıflandırıcısının sonuçları aşağıdaki gibidir (Tablo 17):

Tablo 17. AdaBoost Model Sonuçları

Hedef Değişken	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Kappa
Yeni Çağrı Tipi	0.709	0.669	0.660	0.682	0.542
Eski Çağrı Tipi	0.435	0.192	0.250	0.225	0.0026

**Doğruluk:** Yukarıda tanımlanan Denklem (3) kullanılarak, yeni hedef değişkenle %70.92 doğruluk elde edilirken, eski hedef değişkenle bu oran %43.55'e düşmüştür. Eski hedef değişkenle doğruluk, rastgele tahmine çok yakındır, bu da modelin sınıflar arasında net bir ayırım yapamadığını gösterir.

**F1 Skoru:** Yeni hedef değişkenle elde edilen %66.9 F1 skoru, modelin genel performansının sınırda olduğunu göstermektedir. Bu değer, doğru tahminlerin hem kesinlik hem de hatırlama açısından sınırlı bir seviyede kaldığını işaret etmektedir. Eski çağrı tipi hedef değişkeniyle yapılan sınıflandırmada ise F1 skoru %19.2 seviyesinde kalmıştır. Bu oldukça düşük bir performans olup, modelin kesinlik ve hatırlama açısından ciddi eksiklikler sergilediğini ve güvenilir bir sınıflandırma yapamadığını ortaya koymaktadır.

**Kappa Değeri:** Yeni hedef değişkenle elde edilen 0.542 Kappa değeri, modelin sınıflar arasında anlamlı bir ayırım yapabildiğini ancak performansın ideal seviyenin altında olduğunu göstermektedir. Eski çağrı tipi hedef değişkeninde ise 0.0026 gibi çok düşük bir Kappa değeri elde edilmiştir. Bu, modelin sınıflar arasında neredeyse rastgele tahmin yaptığını ve anlamlı bir ayırım gücüne sahip olmadığını ifade etmektedir.

### 3.5.7. Rastgele Orman

Python’da Rastgele Orman sınıflandırıcısının sonuçları aşağıdaki gibidir (Tablo 18):

Tablo 18. Rastgele Orman Model Sonuçları

Hedef Değişken	Doğruluk	F1 Skoru	Kesinlik	Hatırlama	Kappa
Yeni Çağrı Tipi	0.802	0.726	0.683	0.813	0.678
Eski Çağrı Tipi	0.415	0.225	0.244	0.280	-0.0171

**Doğruluk:** Yukarıda tanımlanan Denklem (3) kullanılarak, yeni hedef değişkenle %80.27 doğruluk elde edilirken, eski hedef değişkenle bu oran %41.5'e düşmüştür. Eski hedef değişkenle doğruluk, rastgele tahmine çok yakındır, bu da modelin sınıflar arasında net bir ayrım yapamadığını gösterir.

**F1 Skoru:** Yeni hedef değişkenle elde edilen %72.6 F1 skoru, modelin genel performansının sınırda olduğunu göstermektedir. Bu değer, doğru tahminlerin hem kesinlik hem de hatırlama açısından sınırlı bir seviyede kaldığını işaret etmektedir. Eski çağrı tipi hedef değişkeniyle yapılan sınıflandırmada ise F1 skoru %22.5 seviyesinde kalmıştır. Bu oldukça düşük bir performans olup, modelin kesinlik ve hatırlama açısından ciddi eksiklikler sergilediğini ve güvenilir bir sınıflandırma yapamadığını ortaya koymaktadır.

**Kappa Değeri:** Yeni hedef değişkenle elde edilen 0.678 Kappa değeri, modelin sınıflar arasında anlamlı bir ayrım yapabildiğini ancak performansın ideal seviyenin altında olduğunu göstermektedir. Eski çağrı tipi hedef değişkeninde ise (-0.0171) gibi çok düşük bir Kappa değeri elde edilmiştir. Bu, modelin sınıflar arasında neredeyse rastgele tahmin yaptığını ve anlamlı bir ayrım gücüne sahip olmadığını ifade etmektedir.

Tablo 19. Karşılaştırma Tablosu

	Naive Bayes	Yapay Sinir Ağları	Destek Vektör Makinesi	Karar Ağaçları	LightGBM	AdaBoost	Rastgele Orman
<b>Doğruluk Oranı</b>	0.802	0.885	<b>0.891</b>	0.676	0.863	0.709	0.802
<b>Kappa İstatistiği</b>	0.666	0.820	<b>0.828</b>	0.495	0.783	0.542	0.678
<b>Etkinlik Analizi</b>	<b>F1 Skoru</b>	<b>F1 Skoru</b>	<b>F1 Skoru</b>	<b>F1 Skoru</b>	<b>F1 Skoru</b>	<b>F1 Skoru</b>	<b>F1 Skoru</b>
	0.611	0.850	<b>0.852</b>	0.620	0.819	0.669	0.726
	<b>Kesinlik</b>	<b>Kesinlik</b>	<b>Kesinlik</b>	<b>Kesinlik</b>	<b>Kesinlik</b>	<b>Kesinlik</b>	<b>Kesinlik</b>
	0.578	<b>0.837</b>	0.836	0.621	0.798	0.660	0.683
	<b>Hatırlama</b>	<b>Hatırlama</b>	<b>Hatırlama</b>	<b>Hatırlama</b>	<b>Hatırlama</b>	<b>Hatırlama</b>	<b>Hatırlama</b>
<b>0.878</b>	0.867	0.870	0.621	0.847	0.682	0.813	

Tablo 19’da yer alan sınıflandırma sonuçları incelendiğinde SVM algoritmasının model testine ait %89.1 doğruluk oranı ile en iyi sonucu ürettiği söylenebilir. Doğruluk ölçütü oldukça önemli bir kriterdir. Bu ölçüte göre SVM algoritmasını sırasıyla Yapay Sinir Ağları, LightGBM, Naive Bayes, Rastgele Orman, AdaBoost ve Karar Ağaçları algoritmaları izlemektedir. Bu ölçütü F1 skoru ile ele alırsak sıralama; SVM, Yapay Sinir Ağları, LightGBM, Rastgele Orman, AdaBoost, Karar Ağaçları ve Naive Bayes olarak benzer dağılım göstermektedir.

Sınıf değişkenlerinin kesinlik değerlerinin ortalamasına göre 1’e en yakın değeri (0,837) olarak Yapay Sinir Ağları en iyi sonucu oluşturmuş olup, diğer algoritmalar bu ölçüte göre SVM, LightGBM, Rastgele Orman, AdaBoost, Karar Ağaçları ve Naive Bayes şeklinde sıralanabilir. Bu ölçütü sınıf değişkenlerinin hatırlama (duyarlılık) değerlerinin ortalamalarıyla beraber ele alırsak tablodan görüleceği üzere algoritmalar, duyarlılık ölçütüne göre Naive Bayes, SVM, Yapay Sinir Ağları, LightGBM, Rastgele Orman, AdaBoost ve Karar Ağaçları olarak sıralanabilir. Görüleceği üzere, kesinlik ölçütü ve duyarlılık ölçütü birbiriyle yakın olan bir sıralama ortaya koymuştur.

Doğruluk ölçütü ve etkinlik ölçütleri dikkate alındığında doğru sınıflandırma performansı yüksek olan algoritma SVM algoritmasıdır. Veri seti için SVM algoritması kullanılarak yapılacak olan sınıflandırmalar diğer yöntemlere göre daha yüksek oranda doğruluk içeren sonuçlar verecektir.

Sınıflandırma algoritmalarının sonuçlarına göre en yüksek doğruluk değerine sahip sınıflandırma yöntemi %89.1 doğruluk ile Destek Vektör Makinesi olarak seçilmiştir ve ileriki zamanlarda gelebilecek bir müşteri geri bildirimlerinin çağrı tipi olasılık dağılımları ve hedef değişken sınıfının tahmini için kullanımının en iyi sonucu verebileceği bulgusu elde edilmiştir. Şekil 12’de müşteri şikayetlerinin sınıfının tahmini için bir örneğe yer verilmiştir. Şekil 12’de yer alan çağrı atama modeline göre gelen çağrılar için geçmiş dönem verileri ile eğitilmiş olan model LDA Olasılık değerlerini oluşturarak, çağrıda gizli bilgileri keşfedecek ve sonucunda atama gerçekleştireceği sınıf ile öncelikli olarak çağrıyı ilgili birime yönlendirecektir.

Yorum	TF-IDF Özellikler Matrisi	LDA Olasılıklar	Sınıf Değişkeni
***	***	?	Teslimat ve Bayi Sorunları / SSH Sorunları / Servis Talepleri / Diğer Hizmetler

Şekil 12. Çağrı Atama Modeli Örneği

## 4. BÖLÜM SONUÇ ve ÖNERİLER

Bu çalışmada, müşteri geri bildirimlerinin metin madenciliği ve veri madenciliği yöntemleri ile analiz edilmesi sonucunda müşteri hizmetleri yönetim sürecine rehberlik edecek bir çağrı atama modeli önerilmiştir. Çalışmada, müşteri memnuniyetini artırmak ve müşteri şikayetlerini yönetme süreçlerini optimize etmek için veri madenciliği ve metin madenciliği yöntemleri birlikte uygulanmıştır. Bu bağlamda, mobilya sektöründe faaliyet gösteren üç farklı markaya ait toplam 9.930 müşteri şikâyeti detaylı bir şekilde analiz edilmiştir. Yapılan analizler, müşteri şikayetlerinin yalnızca tek bir kategoride sınıflandırılmayacak ve genellenemeyecek kadar karmaşık olduğunu ortaya koymuştur.

LDA yöntemi ile yapılan konu modelleme çalışmaları, müşteri şikayetlerinin farklı konulara olasılıksal dağılımlarına ulaşılmasına ve dört ana çağrı tipi ile olasılıklı olarak eşleştirileceğini göstermiştir. İncelenen çalışmalarda metin madenciliği alanında uygulanan konu modelleme, duygu analizi ve bilgi çıkarımı amaçlı yaklaşımlar genelde belirli bir amaç için uygulanırken, bu çalışmada konu modellemesi, mevcut sistemin performansını açığa çıkarmak ve herhangi bir manuel müdahale olmadan çağrı atamalarını sınıflandırmak amacıyla belge kelime listesinden de yararlanılmıştır. Ayrıca, LDA ile elde edilen olasılık dağılımları doğrultusunda oluşturulan çağrılar, yeni bir sınıf değişkeni belirlenerek TF-IDF girdi matrisi kullanılarak tekrar sınıflandırılmış ve bu yöntemle LDA'nın yarattığı olasılıkların doğruluğu test edilmiştir.

Müşteri hizmetleri biriminden elde edilen geri bildirimlerin çağrı atama uyumluluğu incelendiğinde, modelin olasılıklı dağılım sonuçlarına göre uyum düzeyinin düşük olduğu (%22,18) ve bu verilerin içerisinde gizlenen bilgilerin keşfedilmesi gerekliliğini ortaya çıkarttığı görülmüştür.

Sınıflandırma analizleri sonucunda, SVM sınıflandırıcısı ile elde edilen %89.1 doğruluk oranı, modelin müşteri şikayetlerini doğru sınıflandırma ve gelen çağrıları çağrı tiplerine atama yeteneğini ortaya koymaktadır. Bu başarı oranı, metin madenciliği ve makine öğrenmesi gibi yöntemlerin müşteri şikâyet yönetim süreçlerinde etkili olduğunu göstermektedir.

Gelecek çalışmalarda, farklı sektörlerden ve veri kaynaklarından (sosyal medya, e-posta, anket) elde edilen geri bildirimlerin kullanımı, modelin geliştirilebilirliğini artırabilir. Transformer tabanlı modeller (BERT, GPT) ve duygu analizi entegre edilerek sınıflandırma performansı geliştirilebilir. Ayrıca konu modelleme ile şikayetlerin alt kategorilere ayrılması sağlanabilir. LDA ile oluşturulan olasılıkların gerçek zamanlı sistemlerde test edilmesi ve TF-IDF tabanlı sınıflandırma sonuçları ile karşılaştırılması, modelin operasyonel uygulanabilirliği ve iş yükü azaltma potansiyelini değerlendirme açısından değerli olacaktır.

## KAYNAKÇA

1. Wibowo, A., & Chen, S. C., 2020. "Customer Behavior as an Outcome of Social Media Marketing". *Sustainability*, 13(1), 189.
2. Yücel, M., 2013. "Toplam Hizmet Kalitesinin SERVQUAL Analizi ile Ölçümü: Bankacılık Sektöründe Bir Araştırma". *Elektronik Sosyal Bilimler Dergisi*, 12(43), 143-158.
3. Chen, Y., & Xie, J., 2008. "Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix". *Management Science*, 54(3), 477–491.
4. Costa, A. C. F., Capelo, N. F., & Espuny, M., 2024. "Digitalization of Customer Service in Small and Medium-Sized Enterprises". *International Journal of Entrepreneurial Behavior & Research*.
5. Chernova, D. V., & Sharafutdinova, N. S., 2020. "The Transformation of Customer Value in Retail Network Services under Digitalization". In *Digital age: Chances and challenges*.
6. Varol, H. C., & İşler, D. B., 2023. "Dijitalleşme Sürecinde Müşteri Deneyimi ve Sadakat İlişkisi". *Uluslararası Global Turizm Araştırmaları Dergisi*, 8(1), 1-20.
7. Kobayashi, Y., Uchida, T., Inoue, T., Iwasaki, Y., & Ito, R., 2024. "Analysis of 50 Years of Health Food Research Trends Using Natural Language Processing and Generative AI". *Procedia Computer Science*, 207, 123-134.
8. Blei, D. M., Ng, A. Y., & Jordan, M. I., 2003. "Latent Dirichlet Allocation". *Journal of Machine Learning Research*, 3, 993–1022.
9. Hong, L., & Davison, B. D., 2010. "Empirical Study of Topic Modeling in Twitter". *Proceedings of the First Workshop on Social Media Analytics (SOMA)*, 80–88.
10. Budak, İ., & Sökmen, A., 2022. "Otel Hizmetlerinin Değerlendirilmesinde LDA ile Analiz". *Journal of Tourism and Gastronomy Studies*, 10(4), 2942-2954.
11. Kaynar, O., & Akdoğan, M. Ş., 2021. "Otellere İlişkin Çevrimiçi Geribildirimlerin Makine Öğrenmesi Yöntemleriyle Duygu Analizi". *Journal of Social Sciences Institute*, 2021(48), 55–68.

12. Sutherland, I., & Kiatkawsin, K., 2020. "Topic Modeling of Online Accommodation Reviews using LDA". *Sustainability*, 12(5), 1821.
13. Göker, H., & Tekedere, H., 2017. "FATİH Projesine Yönelik Görüşlerin Metin Madenciliği Yöntemleri ile Otomatik Değerlendirilmesi". *Bilişim Teknolojileri Dergisi*, 10(4), 1–10.
14. Büyükeke, A., Sökmen, A., & Gencer, C., 2020. "Metin Madenciliği ve Duygu Analizi Yöntemleri ile Sosyal Medya Verilerinden Rekabetçi Avantaj Elde Etme: Turizm Sektöründe Bir Araştırma". *Journal of Tourism & Gastronomy Studies*, 8(2), 123–140.
15. Melek, C., 2012. "Metin Madenciliği Teknikleri ile Şirketlerin Vizyon İfadelerinin Analizi". Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İzmir, 111 s.
16. Sözen, E., & Bardak, T., 2021. "Mobilya Üretiminde Kullanılan Farklı Malzemelerin Web Madenciliği Yöntemleri ile Değerlendirilmesi". *Mobilya ve Ahşap Malzeme Araştırmaları Dergisi*, 1(1), 15–28.
17. Erduran, Y. G., 2017. "Online Müşteri Şikayetlerinin Veri Madenciliği ile İncelenmesi". Trakya Üniversitesi Sosyal Bilimler Enstitüsü, Doktora Tezi, Edirne, 205 s.
18. Ahmetoğlu, M., & Duygun, A., 2021. "Mobilya Sektörüne Yönelik Müşteri Şikayetlerinin İncelenmesi". *Mobilya ve Ahşap Endüstrisi Dergisi*, 5(2), 45–62.
19. Han, J., Kamber, M., Pei, J., 2011. "Data Mining: Concepts and Techniques. (3rd Edition)". Morgan Kaufmann Burlington, MA, 744 pp.
20. Krishnaiah, V., & Narsimha, G., 2014. "Survey of Classification Techniques in Data Mining". *International Journal of Computer Science and Engineering*, 4(9), 42-50.
21. Şahinaslan, Ö., & Dalyan, H., 2022. "Naive Bayes Sınıflandırıcısı Kullanılarak YouTube Verileri Üzerinden Çok Dilli Duygu Analizi". *Bilişim Teknolojileri Dergisi*, 15(2).
22. Hearst, M. A., 1999. "Untangling Text Data Mining". *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 37(1), 3–10.
23. Nahm, U. Y., & Mooney, R. J., 2000. "A Mutually Beneficial Integration of Data Mining and Information Extraction". *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 37, 627–634.

24. Parker, N., & Kızıllırmak, B., 2023. "Çevrim İçi Müşteri Yorumlarını Etkileyen Faktörler Üzerine Keşifsel Bir Çalışma: Trendyol Örneği". Anadolu Üniversitesi Sosyal Bilimler Dergisi, 23(4), 1393-1414.
25. Derdiyok, T., Doğru, Ç., & Ünal, S., 2020. "Dijital Çağda Entelektüel Sermayenin Dönüşümü ve Şirketlerin Rekabet Güçlerine Etkileri". Ufuk Üniversitesi Sosyal Bilimler Dergisi.
26. Taşkın, K., 2022. "Sanal Pazar Uygulamalarının Kullanılabilirliğinin Duygu Analizi Temelli Değerlendirilmesi". Premium e-Journal of Social Sciences, 3(1), 45-60.
27. Atılğan, K. Ö., & Yoğurtcu, H., 2021. "Kargo Firması Müşterilerinin Twitter Gönderilerinin Duygu Analizi". Çağ Üniversitesi Sosyal Bilimler Dergisi, 12(4), 35-48.
28. Yücel, N., & Cömert, Ö., 2023. "Müşteri Duyarlılığını Keşfetmek için Yapay Zekâ Destekli Analiz ile Çevrimiçi Ürün İncelemelerinden Anlamlı Bilgiler Elde Etme". Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 5(2), 88-104.
29. Güneş, Y., & Arıkan, M., 2023. "Twitter Veri Seti İçeriğinin Tanımlayıcı Analiz ile Keşfi: Çevrimiçi Yemek Siparişi Üzerine Bir Uygulama". Bilişim Teknolojileri Dergisi, 7(3), 22-37.
30. Şahin, S., 2023. "Fiziksel Kanıtlar ile Gastronomi Deneyiminin Memnuniyet Üzerindeki Etkisi: Diyarbakır'da Bir Araştırma". JATOS, 4(2), 82.
31. Hoffmann, N., 2022. "Brand Position in the Eyes of Customers: Assessment of Selected Airlines by the Passengers Online Reviews". Catallaxy, 7(1), 7-21.
32. Irmak, H., & Ecevit, S. Z., 2022. "Covid-19 Pandemisi Döneminde Öğrencilerin Uzaktan Eğitimde Yaşadıkları Sorunların Kelime Bulutu ve Gizli Anlamsal Dizinleme Yöntemleriyle Tespit Edilmesi". Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, 6(2), 162-170.
33. Ödev, G., & Peker, S., 2022. "Müşteri Şikâyet Yönetiminde Firmaların Performanslarının Değerlendirilmesi: Kümeleme Analizi İncelemesi". DÜMF MD.
34. Şahin, S., 2016. "Turistlerin Şikâyet Davranışları ve Şikayetlerin Çözümüne Yönelik Beklenti Farklılıkları". Journal of International Social Research, 9(42), 2033-2033.

35. Muslu, D., 2009. "Sigortacılık Sektöründe Risk Analizi: Veri Madenciliği Uygulaması". İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul, 120 s.



## ÖZGEÇMİŞ

### KİŞİSEL BİLGİLER

**Adı Soyadı:** Sevgi PEHLİVAN

**Uyruğu:** T.C.

**Doğum Tarihi ve Yeri:**

**Medeni Durum:**

**e-mail:**

**Yazışma Adresi:**

### EĞİTİM

Derece	Kurum	Mezuniyet Tarihi
Yüksek Lisans	Erciyes Üniversitesi, Endüstri Mühendisliği	2024
Lisans	Erciyes Üniversitesi, Endüstri Mühendisliği	2021
Lise	Seyide Daloğlu Anadolu Lisesi, Kayseri	2017

### İŞ DENEYİMLERİ

Yıl	Kurum	Görev
2021-Halen	Bellona Mobilya A.Ş.	İnsan Kaynakları Uzmanı

### YABANCI DİL

İngilizce

### YAYINLAR

- Pehlivan, S., Torun, H., "Text Mining for Customer Feedback Analysis: Case Study of Furniture Companies", ICDEC 2024, Kayseri, TÜRKİYE.