

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

DOKTORA TEZİ

**VERİ VE METİN MADENCİLİĞİ İLE HAVA YOLU
İŞLETMELERİNİN SOSYAL MEDYA YORUM VE SKORLARININ
DEĞERLENDİRİLMESİ**

**Hazırlayan
İbrahim BUDAK**

**Danışman
Prof. Dr. Arzu ORGAN**

**Mayıs 2021
DENİZLİ**

**VERİ VE METİN MADENCİLİĞİ İLE HAVA YOLU
İŞLETMELERİNİN SOSYAL MEDYA YORUM VE SKORLARININ
DEĞERLENDİRİLMESİ**

**Pamukkale Üniversitesi
Sosyal Bilimler Enstitüsü
Doktora Tezi
İşletme Ana Bilim Dalı
Genel İşletme Programı**

İbrahim BUDAK

Danışman: Prof. Dr. Arzu ORGAN

**Mayıs 2021
DENİZLİ**

Bu tezin tasarımı, hazırlanması, yürütülmesi, arařtırmalarının yapılması ve bulgularının analizlerinde bilimsel etięe ve akademik kurallara özenle riayet edildiđini; bu arařtırmanın doğrudan birincil ürünü olmayan bulguların, verilerin ve materyallerin bilimsel etięe uygun olarak kaynak gösterildiđini ve alıntı yapılan alıřmalara atıfta bulunulduđunu beyan ederim.

İbrahim BUDAK

ÖNSÖZ

“Veri ve Metin Madenciliği ile Hava Yolu İşletmelerinin Sosyal Medya Yorum ve Skorlarının Değerlendirilmesi” başlıklı tezimin hazırlanmasında emeği geçen danışman hocam Prof. Dr. Arzu ORGAN’a, teşekkür ederim.

Doktora hayatındaki ilk günde ve Tez İzleme Komitesi’nde görev alan desteğini hiçbir zaman esirgemeyen, gerek akademik bakış açısı ve çalışmalarıyla gerekse kişiliğiyle kendime örnek aldığım, verdiği eğitim doğrultusunda çalışmalarımda büyük katkısı olan Prof. Dr. Selçuk Burak HAŞILOĞLU’na sonsuz teşekkür ederim.

Tez İzleme Komitesi’nde verimli ve etkin bir şekilde ilerlememi sağlayan Dr. Öğr. Üyesi Elif HAYTAOĞLU hocama, Tez Savunma Sınavı’nda desteklerini esirgemeyen Doç. Dr. Hüseyin ŞENKAYAS ve Doç. Dr. Kemal VATANSEVER hocalarıma değerli zamanlarını aldığım bilgi, birikim ve tecrübelerini benimle paylaştıkları için teşekkür ederim. Ayrıca akademik hayatım boyunca desteklerini esirgemeyen Mehmet Ali GÖKÇE, Günay KILIÇ, Erkan TURHAN, Zafer DURAN, Suna KÖSE, Nurdanur TAVLAN, Ümmü AHAT ve Berivan İNAN’a ve diğer tüm arkadaşlarıma da teşekkür ederim.

Tezimin yazım aşamasında en fazla havasını soluduğum, yeni dostluklar kazandıran, çalışma ortamı sağlayan ve bilgi birikimlerini aktaran Akdeniz Üniversitesi AKVAM müdürü Prof. Dr. Hale BALSEVEN ve müdür yardımcısı Dr. Öğr. Üyesi Mustafa ÇOBAN’a teşekkür ederim.

Son olarak hayatımda beni yalnız bırakmayan maddi ve manevi her türlü desteği esirgemeyen annem Halime BUDAK, babam Hasan BUDAK, en değerlilerim Dilan BUDAK, Burhan BUDAK ve Mustafa BUDAK’a minnettarlığımı sunarım.

ÖZET

VERİ VE METİN MADENCİLİĞİ İLE HAVA YOLU İŞLETMELERİNİN SOSYAL MEDYA YORUM VE SKORLARININ DEĞERLENDİRİLMESİ

BUDAK, İbrahim

Doktora Tezi

İşletme ABD

Genel İşletme Programı

Tez Yöneticisi: Prof. Dr. Arzu ORGAN

Mayıs 2021, viii+113 sayfa

Veri ve metin madenciliği, anlamlı ilişkileri ve eğilimleri ayırt etmek için kullanıcıların taleplerine göre yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış büyük bir veri miktarını analiz etme sürecidir. İşletmeler, veri ve metin madenciliği teknikleri kullanarak hem kendi işletmeleri içerisinde hem de rakipleri ile rekabette yaşadıkları sorunlarına etkili çözümler üretebilmektedirler. Böylece elde ettikleri bilgiyi rekabet avantajına çevirebilmektedirler. Bu araştırmada, veri ve metin madenciliği algoritmaları kullanılarak rekabete dayalı pazarda müşterilerin istek ve ihtiyaçlarına göre hava yolu firmalarının üstün ve zayıf yönlerinin değerlendirilmesi amaçlanmıştır. Bu araştırmada, sosyal medya sitelerinden olan TripAdvisor'daki çevrimiçi seyahat incelemeleri araştırma kapsamına alınmıştır. Star Alliance küresel hava yolu birliğine üye 26 hava yolu firması değerlendirilmiştir. Araştırmada kullanılan kriterler; her bir kullanıcının yorum ve skorları temel alınarak belirlenmiştir. Duygu Analizi ile müşteri yorumlarından polarite ve polarite değeri belirlendikten sonra Destek Vektör Makineleri, Naive Bayes, Derin Öğrenme Algoritmaları ile sınıflandırma ve tahminleme yapılarak elde edilen sonuçlar karşılaştırılmıştır. Yapılan karşılaştırmada Derin Öğrenme'nin daha iyi sonuç verdiği saptanmıştır. Ayrıca, Karar Ağaçları ile güzergâh, uçuş yapılan yer, kabin sınıfı ve polarite etiketlerine göre kriterlerin önem dereceleri hesaplanarak yorumlanmıştır. Yöntemler karşılaştırılırken aynı zamanda StarAlliance hava yolu işletmeleri de karşılaştırılmıştır. Sonuçta, müşteri yorum ve skorları açısından en iyi hava yolu EVA olarak belirlenmiştir.

Anahtar Kelimeler: Veri Madenciliği, Metin Madenciliği, Hava Yolu, Destek Vektör Makineleri, Naive Bayes, Derin Öğrenme, Karar Ağaçları, Kelime Bulutu

ABSTRACT

EVALUATION ON SOCIAL MEDIA COMMENTS AND SCORES OF AIRLINE COMPANIES BY USING DATA AND TEXT MINING

BUDAK, İbrahim

PhD. Thesis

Business Administration Department

PhD. in Business Administration

Advisor of Thesis: Prof. Dr. Arzu ORGAN

May 2021, viii+113 pages

Data and text mining is the process of analyzing a large amount of semi-structured, unstructured and structured data based on users' demands to distinguish meaningful relationships and trends. By using the data and text mining techniques, companies can produce effective solutions to the problems within their business and with their competitors. Thus, they can turn this information into competitive advantage. In this study, it is aimed to evaluate the strengths and weaknesses of airline companies in competitive market according to the demands and needs of customers through using the data and text mining algorithms. In this research, online user reviews on TripAdvisor which is one of the social media site were included in this research. 26 airlines which are the members of Star Alliance Global Airline Association were evaluated. The criteria used in this research was determined based on the comments and scores of each user. After determination of polarity and polarity value from customer comments with Sentiment Analysis, the results were compared through classification and estimation by using Support Vector Machines, Naive Bayes, and Deep Learning Algorithms. In the comparison made, it was determined that Deep learning gives better results. In addition, with the Decision Trees, the importance levels of the criteria were computed and interpreted according to the route, flight location and cabin class labels. While comparing the methods, Star Alliance airlines were also compared. The results show that the best airline in terms of customer comments and scores was EVA.

Keywords: Data Mining, Text Mining, Airline, Support Vector Machines, Naive Bayes, Deep Learning, Decision Trees, Word Cloud

İÇİNDEKİLER

ÖNSÖZ	ii
ÖZET.....	iii
ABSTRACT.....	iv
İÇİNDEKİLER	v
ŞEKİLLER DİZİNİ.....	vi
TABLolar DİZİNİ	vii
SİMGE VE KISALTMALAR DİZİNİ	viii
GİRİŞ	1

BİRİNCİ BÖLÜM

VERİ VE METİN MADENCİLİĞİ

1.1. Veri Madenciliği.....	5
1.1.1. Veri Madenciliği Teknikleri	7
1.1.1.1. Karar Ağaçları	7
1.1.1.2. Yapay Sinir Ağları	10
1.1.1.3. Regresyon.....	12
1.1.1.4. Market Sepet Analizi.....	14
1.1.1.5. Kümeleme	15
1.2. Metin Madenciliği	16
1.2.1. Metin Madenciliğine Yönelik Alanlar	18
1.2.1.1. Metin Kümeleme ve Sınıflandırma	18
1.2.1.2. Bilgi Çıkarımı	19
1.2.1.3. Bilgi Alımı	19
1.2.1.4. Adlandırılmış Varlık Tanımlama	20
1.2.1.5. Doğal Dil İşleme	20
1.2.1.6. Soru Cevaplama	21
1.2.1.7. Görselleştirme	21
1.3. Veri ve Metin Madenciliği İle Yapılan Bazı Çalışmalar	22

İKİNCİ BÖLÜM

SINIFLANDIRMA VE TAHMİNLEME

2.1. Veri/Metin Madenciliği ile Sınıflandırma ve Tahminleme	28
2.1.1. Duygu Analizi.....	29
2.1.2. Destek Vektör Makineleri.....	35
2.1.3. Naive Bayes Algoritması	39
2.1.4. Derin Öğrenme	43
2.2. Veri ve Metin Madenciliği Uygulama Alanları	48
2.2.1. Müşteri İlişkileri	48
2.2.2. Hizmet Kalitesi	49
2.2.3. Bilgi Yönetimi	51
2.2.4. Finans, Pazarlama, Sağlık ve Diğer Uygulamalar	52

ÜÇÜNCÜ BÖLÜM
VERİ VE METİN MADENCİLİĞİ İLE HAVA YOLU İŞLETMELERİNİN
SOSYAL MEDYA YORUM VE SKORLARINA GÖRE
DEĞERLENDİRİLMESİ VE KARŞILAŞTIRILMASI

3.1. Araştırmanın Amacı	54
3.2. Araştırmanın Kapsamı.....	55
3.3. Araştırmada Kullanılan Araçlar	56
3.4. Veri Seti.....	57
3.5. Veri Analizinin Çerçevesi	58
3.6. Veri Analizi	60
3.6.1. Hava Yollarının Duygu Analizi	61
3.6.2. Hava Yollarının Korelasyon Analizi	63
3.6.3. Hava Yollarının Regresyon Analizi.....	64
3.6.4. Destek Vektör Makineleri ile Hava Yolu Polarite Analizi	64
3.6.5. Naive Bayes ile Hava Yolu Polarite Analizi	66
3.6.6. Derin Öğrenme ile Hava Yolu Polarite Analizi.....	67
3.6.7. Tüm Hava Yollarının Sınıflandırma Sonuçlarının Karşılaştırılması	69
3.6.8. Tüm Hava Yollarının Polarite Değerlerinin RMSE Sonuçlarının Karşılaştırılması	70
3.6.9. Karar Ağaçları ile Kriterlerin Önem Derecelerinin Belirlenmesi.....	72
3.6.10. Hava Yollarının Kelime Bulutu Analizi	83
SONUÇ	98
KAYNAKLAR	102
ÖZGEÇMİŞ	Hata! Yer işareti tanımlanmamış.

ŞEKİLLER DİZİNİ

Şekil 1. Veri Tabanlarında Bilgi Keşfi Adımları.....	6
Şekil 2. Bir Karar Ağacı Örneği.....	8
Şekil 3. Bir Yapay Sinir Ağ Yapısı Örneği.....	10
Şekil 4. Veri ve Metin Madenciliği Bileşenleri.....	18
Şekil 5. Duygu Analizinin Başlıca Görevleri.....	30
Şekil 6. a) Basit Doğrusal DVM; b) Kesikli ve Kesiksiz DVM.....	36
Şekil 7. Rectified Doğrusal Aktivasyon Fonksiyonu.....	45
Şekil 8. Rectified Yayılım Ağı.....	46
Şekil 9. Bir Müşteriye Ait Hava Yolu Değerlendirmesi.....	56
Şekil 10. Web Kazıma Çerçevesi.....	57
Şekil 11. Veri Seti İşlem Çerçevesi.....	58
Şekil 12. Ön İşleme Sürecinin Adımları.....	59
Şekil 13. Metin Sınıflandırma Süreci.....	60
Şekil 14. Algoritmaların Performans Sonuçlarının Karşılaştırılması.....	70
Şekil 15. Algoritmaların RMSE Değerlerinin Karşılaştırılması.....	72
Şekil 16. THY İçin Kelime Bulutu.....	83
Şekil 17. ANA Hava Yolu İçin Kelime Bulutu.....	84
Şekil 18. Aegean Hava Yolu İçin Kelime Bulutu.....	84
Şekil 19. Asiana Hava Yolu İçin Kelime Bulutu.....	85
Şekil 20. Austrian Hava Yolu İçin Kelime Bulutu.....	86
Şekil 21. Avianca Hava Yolu İçin Kelime Bulutu.....	86
Şekil 22. Brussels Hava Yolu İçin Kelime Bulutu.....	87
Şekil 23. Canada Hava Yolu İçin Kelime Bulutu.....	87
Şekil 24. China Hava Yolu İçin Kelime Bulutu.....	88
Şekil 25. Copa Hava Yolu İçin Kelime Bulutu.....	88
Şekil 26. Croatia Hava Yolu İçin Kelime Bulutu.....	89
Şekil 27. Egypt Hava Yolu İçin Kelime Bulutu.....	89
Şekil 28. Ethiopian Hava Yolu İçin Kelime Bulutu.....	90
Şekil 29. EVA Hava Yolu İçin Kelime Bulutu.....	90
Şekil 30. India Hava Yolu İçin Kelime Bulutu.....	91
Şekil 31. LOT Hava Yolu İçin Kelime Bulutu.....	91
Şekil 32. Lufthansa Hava Yolu İçin Kelime Bulutu.....	92
Şekil 33. New Zealand Hava Yolu İçin Kelime Bulutu.....	93
Şekil 34. SAS Hava Yolu İçin Kelime Bulutu.....	93
Şekil 35. Shenzhen Hava Yolu İçin Kelime Bulutu.....	94
Şekil 36. Singapore Hava Yolu İçin Kelime Bulutu.....	94
Şekil 37. South African Hava Yolu İçin Kelime Bulutu.....	95
Şekil 38. Swiss Hava Yolu İçin Kelime Bulutu.....	95
Şekil 39. TAP Hava Yolu İçin Kelime Bulutu.....	96
Şekil 40. Thai Hava Yolu İçin Kelime Bulutu.....	96
Şekil 41. United Hava Yolu İçin Kelime Bulutu.....	97

TABLOLAR DİZİNİ

Tablo 1. Veri Madenciliği ve Metin Madenciliği Karşılaştırılması.....	18
Tablo 2. Veri ve Metin madenciliği İle ilgili Genel Literatür Taraması.....	22
Tablo 3. Duygu Analizi Avantajları ve Dezavantajları.....	33
Tablo 4. DVM Algoritmasının Avantajları ve Dezavantajları	38
Tablo 5. Naive Bayes Algoritmasının Avantaj ve Dezavantajları	41
Tablo 6. Derin Öğrenme Algoritmasının Avantajları ve Dezavantajları	46
Tablo 7. Star Alliance Hava Yolu Üyeleri	55
Tablo 8. Kurallı İfadeler.....	56
Tablo 9. Hava Yolları Yorumlarının Polarite Sayıları	61
Tablo 10. Hava Yollarının Müşteri Yorumlarına Ait Polarite ve Polarite Değeri.....	62
Tablo 11. Tüm Hava Yollarına Ait Değerlendirme Kriterlerinin Korelasyon Sonuçları	63
Tablo 12. Tüm Hava Yollarına Ait Değerlendirme Kriterlerinin Regresyon Sonuçları	64
Tablo 13. THY'nin DVM ile Performans Sınıflandırma Sonucu.....	65
Tablo 14. DVM ile Tüm Hava Yollarının Performans Sınıflandırma Sonucu	65
Tablo 15. THY'nin Naive Bayes ile Performans Sınıflandırma Sonucu	66
Tablo 16. Naive Bayes ile Tüm Hava Yollarının Performans Sınıflandırma Sonucu ...	67
Tablo 17. THY'nin Derin Öğrenme ile Performans Sınıflandırma Sonucu	68
Tablo 18. Derin Öğrenme ile Tüm Hava Yollarının Performans Sınıflandırma Sonucu	68
Tablo 19. Tüm Hava Yollarının Performans Sonuçlarının Karşılaştırılması.....	69
Tablo 20. Tüm Hava Yollarının Polarite Değeri RMSE Sonuçları	71
Tablo 21. Uçuş Güzergâhına Göre Önem Dereceleri	72
Tablo 22. Uçuş Yapılan Yere Göre Önem Dereceleri	75
Tablo 23. Kabin Sınıfına Göre Önem Dereceleri	77
Tablo 24. Polariteye Göre Önem Dereceleri.....	80

SİMGE VE KISALTMALAR DİZİNİ

ANA	Ana Nippon Airways
CRM	Müşteri İlişkileri Yönetimi
DSF	Domain Sensitive Features
DVM	Destek Vektör Makineleri
IBM	Uluslararası İş Makineleri
KNN	K-En Yakın Komşu
LDA	Latent Dirichlet Allocation/ Latent Dirichlet Ayırması
MLP	Çok Katmanlı Algılayıcı
NER	Adlandırılmış Varlık Tanımlama
NLP	Doğal Dil İşleme
PENETS	Pathfinder Network Scaling
RMSE	Kök Ortalama Karesel Hata
SAS	Scandinavian Airlines
SLF	Sentence Level Features
SMO	Sıralı Minimal Optimizasyon
TDVM	Transdüktif Destek Vektör Makineleri
THY	Türk Hava Yolları
TOPSIS	Technique for Order Preference by Similarity to An Ideal Solution
UNICEF	Birleşmiş Milletler Çocuklara Yardım Fonu
VTBK/KDD	Veri Tabanlarında Bilgi Keşfi
WWW	World Wide Web
YSA	Yapay Sinir Ağları

GİRİŞ

Günümüz dünyasında küreselleşme, nüfus artışı ve hızlı kentleşme ile birlikte, ulaşım faaliyetleri, kısa vadede bir yerden bir yere gitmek, bir istek olmaktan çok ihtiyaç olarak görülmektedir. Söz konusu ulaşım faaliyetleri demiryolu, karayolu, hava yolu, denizyolu gibi ulaşım sistemleri ile gerçekleştirilmektedir. Günümüzde hava yolu ile ulaşım sistemi daha fazla tercih edilmektedir ve bu durumda hava yolunun kullanım oranı her geçen gün artmaktadır. İnsanların 'hız' odaklı yeni bir yaşam biçimi benimsemesi ile hava yolu ulaşım sistemlerinin önemi giderek artmaktadır. Hava yolu araçlarının kullanılması kısa vadede hızlı ulaşımın sağlanmasının yanı sıra ülke içinde, bölgesel ve uluslararası ekonomilerin gelişmesinde önemli katkılar sağlamaktadır. Hava yollarının önemi her geçen gün değer kazanmaktadır ve daha fazla kişiye hizmet vermektedir (Turan ve Turan, 2008: 161).

Dünya ekonomisinde ulaşım sektörü önemli bir konuma sahiptir. Katma değeri yüksek ve ulaşım sektöründe stratejik bir öneme, ülkelerin refah düzeyine ulaşmasında ve rekabet gücünde etken bir güce sahip olan hava yolu şirketleri bütün dünyadaki ekonomik büyümeyi, küreselleşmeyi ve liberalleşmeyi geliştirmektedir. Dünya hava yolu trafiği, 2002 yılından beri ücretli yolcu/km açısından % 49 oranında artış göstermiştir. Ülkemizde ise, bu oran daha fazla artış göstermektedir. Son yıllarda Türkiye'de hava yolu taşımacılığı daha hızlı gelişmekle birlikte %20'lere varan artışlar ile büyümeye devam etmektedir. Ülkemizdeki bakanlıkların siyasi çabaları sonucunda son on yedi yılda hava yolu şirketlerinin uçak sayısı %237, koltuk kapasitesi % 276, kargo kapasitesi % 659 artmış, yurt içi ve yurt dışı uçulan nokta sayısı 384'e ulaşmıştır (Sivil Havacılık Faaliyet Raporu, 2019).

Hava yolu taşımacılığında pazar payının artması ile birlikte hava yolu taşımacılığında yoğun bir rekabet söz konusudur. Bu bağlamda hava yolu şirketleri müşteri için en iyi hizmeti sunmak istemektedirler. Müşteriler, hava yolu şirketlerinin sunduğu en iyi hizmetlerden ve fırsatlardan yararlanmak için çevrim içi müşteri değerlendirmelerine önem vermektedirler. Yapılan bu değerlendirmeler doğrultusunda

müşteriler kendilerine en uygun olan hava yolu şirketini seçebilmektedirler. Müşteri değerlendirmelerinin tamamına ulaşmak ve bir arada ele alıp karar vermek için modern teknolojik yöntemlere ihtiyaç duyulmaktadır.

Teknolojik ilerleme ile birlikte büyük miktarda verinin verimli bir şekilde depolanmasını ve alınmasını sağladığını göz önünde bulundurarak, artık mevcut odak noktası artık bilgi edinilmesi ve ham kaynaklardan bilgi yaratma yöntemlerine odaklanması olduğu bilinmektedir. Çevrim içi siteler büyük veri bağlamında gelişmekte olan zorlu bir sektörü temsil etmektedir. İnsanların doğal dil ifadeleri kısa metin mesajları ile kolayca raporlanabilir, karar vermek için eyleme geçirilebilir bilgi oluşturmak için etkili bir şekilde analiz edilmesi gereken büyük boyutlarda benzersiz içerikler hızla oluşturulabilir.

Bilgi ve iletişim teknolojilerindeki gelişmelere paralel olarak, insanların çoğu tüm işlemlerini internet tabanlı bir sistemde gerçekleştirmeye başlamıştır. Bu durumda çoğu insan için, çevrim içi olarak daha fazla iş yapılmasıyla birlikte yeni yaşam tarzları oluşmaktadır. Örneğin, bir müşteri çevrim içi bir ürün satın alması durumunda ürün hakkında duygu ve düşüncelerini belirten bir inceleme notu bırakabilir. Bu söz konusu inceleme notları, ürünler, filmler ve tatiller gibi birçok farklı alanlar için kullanılabilir. Günümüzde bu müşteri tarafından yapılan değerlendirmeler ürün veya hizmetlerle ilgili görüşleri ifade etmenin popüler bir yoludur ve ayrıca kişi ürün ve hizmetler hakkında diğer görüşleri inceleyerek bilgi edinmesi söz konusudur. Bu, yalnızca genel bir derecelendirme ile etiketlenmiş olan inceleme miktarında bir artışa neden olmaktadır. İncelemeleri bütünüyle elde etmek ve değerlendirmek için veri ve metin madenciliği tekniklerinden yararlanılmaktadır. Bu teknikler her müşterinin yaptığı incelemeye ulaşmada kullanılarak analiz edilmesinde fayda sağlamaktadır. Aynı zamanda incelemeleri analiz etmekte kullanılan teknikler araştırmacıya bir öngörü sunmaktadır.

İnceleme sitelerinde yer alan müşteri değerlendirmeleri, ürün veya hizmetin farklı yönlerine yönelik görüşler de dâhil olmak üzere, genel bir derecelendirmeden çok daha faydalı bilgiler içermektedir. Bir film incelemesi, genel olarak 5 yıldıza sahip olabilir, ancak eleştirmen yine de özelliklerinden birine karşı olumsuz bir duygu ifade edebilir. Mevcut tüm değerlendirmeleri okumak çok zaman alabilmektedir ve karşılaştırma yapmak zor olabilir. Bu nedenle, müşteri değerlendirmelerinin otomatik olarak

sınıflandırılması, yönlerinin ve duyarlılığının algılanması, karşılaştırma yapılmasına ihtiyaç duyulmaktadır. Dolayısıyla, veri/metin madenciliğindeki algoritmaların kullanımı bu çerçevede yardımcı olmaktadır.

Veri/metin madenciliği teknikleri ile rekabete dayalı müşterilerin istek ve ihtiyaçlarına göre hava yollarını değerlendirmeye yönelik bu araştırmada, sosyal medya seyahat inceleme web sitesi olan TripAdvisor'daki çevrimiçi müşteri incelemeleri araştırma kapsamına alınmıştır. TripAdvisor, küresel internet trafiğinde ve son 90 gündeki etkileşimde 396. sıradadır (Alexa, 2020). Bu değerlendirmeler doğrultusunda veri/metin madenciliğinde kullanılan algoritmalar ile Star Alliance küresel havacılık sektörünün 26 üye hava yollarının karşılaştırılması yapılmıştır. Değerlendirmede veri/metin madenciliği tekniklerinden Duygu Analizi, Destek Vektör Makineleri, Naive Bayes, Derin Öğrenme ve Karar Ağacı Algoritmaları kullanılmıştır. Duygu Analizi ile müşteri değerlendirmeleri analiz edilerek polarite (pozitif, negatif, nötr) ve polarite değeri (0-1) çıkarılmıştır. Algoritmalar yardımıyla da sınıflandırma, hata tahmini yapılmış, son olarak değerlendirme için kullanılan kriterlerin önem dereceleri belirlenmiştir. Buna ek olarak, kullanılan kriterlerin ilişki ve etkisini göstermek için korelasyon ve regresyon analizi yapılmıştır. Son olarak, müşteri değerlendirmelerinden çıkarılan kelimelerden en çok kullanılan 100 kelime ile her bir hava yolu için kelime bulutları oluşturulmuştur.

Araştırma üç bölümden oluşmaktadır. İlk bölümde kavramsal çerçeve içerisinde veri ve metin madenciliği tanımına, kullanılan tekniklere, kullanım alanlarına ve alan yazındaki ilgili araştırmalara yer verilmiştir. Genel kapsamda Karar Ağaçları, Yapay Sinir Ağları, Regresyon, Market Sepet Analizi ve Kümeleme Algoritmalarına değinilmiştir. Sınıflandırma, bilgi çıkarımı, bilgi alımı, Adlandırılmış Varlık Tanımlama, Doğal Dil İşleme, soru cevaplama ve görselleştirme alanlarına yönelik bilgilere yer verilerek veri ve metin madenciliğinin karşılaştırılması yapılmıştır.

İkinci bölümde, veri/metin madenciliğinde yer alan sınıflandırma ve tahminleme algoritmaları ile uygulama alanlarına yer verilmiştir. Bu bağlamda, Duygu Analizi, Destek Vektör Makineleri, Naive Bayes ve Derin Öğrenme Algoritmalarından bahsedilmiştir. Veri/metin madenciliği uygulama alanlarından olan müşteri hizmetleri, hizmet kalitesi, bilgi yönetimi, finans, pazarlama, sağlık gibi diğer alanlarla olan ilişkisine değinilmiştir.

Son bölümde ise, araştırmanın amacı, kapsamı, kullanılan teknikler doğrultusunda veri ve metin madenciliği teknikleri ile sosyal medya yorum ve skorları kullanılarak her bir hava yoluna ait analizler yapılmıştır. Elde edilen analiz sonuçları ile hava yolları karşılaştırılarak müşterilerin yorumlarından elde edilen kelime bulutlarına yer verilmiştir.

Günümüzde doğru bilgiye ulaşmak ve analiz etmek için veri/metin madenciliği teknikleri kullanılmaktadır. Bu araştırma, seyahat inceleme sitesi temel alınarak hava yolu taşımacılık sektörüne yönelik yapılmıştır. Ek olarak araştırma, sitede yer alan tüm verileri okumadan hangi hava yolunun seçileceği ve müşteri değerlendirmelerinin ne yönde ilerleyeceği konusunda fikir vermektedir. Bu nedenle yapılan araştırma, hava yollarının performansının artırılması konusunda yararlı bilgiler sağlayacaktır.

BİRİNCİ BÖLÜM

VERİ VE METİN MADENCİLİĞİ

Veri ve metin madenciliği, yapılandırılmış, yarı yapılandırılmış, yapılandırılmamış verilerden anlamlı ilişkilerin, kuralların keşfedilmesini ve analiz edilmesi sağlamaktadır. Bunu yaparken veri madenciliğinde, Karar Ağaçları, Yapay Sinir Ağları, Regresyon, Market Sepet Analizi, Regresyon, Market Sepet Analizi ve Kümeleme konuları öne çıkan tekniklerden yararlanmaktadır. Metin madenciliğinin de ise, metin kümeleme ve sınıflandırma, bilgi çıkarımı, bilgi alımı, Adlandırılmış Varlık Tanımlama, Doğal Dil İşleme, soru cevaplama ve görselleştirme alanlarına yer verilecektir. Veri ve metin madenciliği arasındaki temel farklara yer verilerek literatürde yer alan çalışmalara değinilecektir.

1.1. Veri Madenciliği

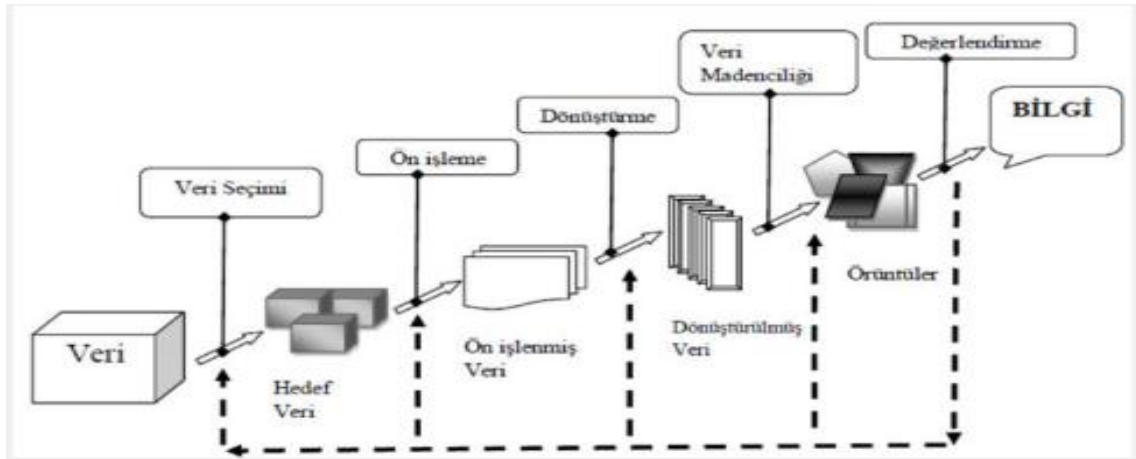
Veri kavramı, uydu görüntüleri, belirli stokların fiyat serileri veya ankete katılanlar tarafından doldurulan işaretlemeler gibi kaydedilmiş ve depolanmış bilgileri ifade etmektedir. İki bilgi türü, verilerle ilişkilendirilmektedir. Birincisi, bireyler tarafından yapılan veri girişleri (örneğin; kaydedilmiş fiyatlar veya anketteki soruların cevapları): İkincisi ise, meta verilerdir. Meta veriler, veriye anlam kazandırmaktadır. Bunun yanı sıra, meta veriler, verinin yer aldığı satır ve sütunların açıklamalarını açık bir şekilde göstermektedir. Meta veride depolanan ilişkileri göz önünde bulundurarak, kayıtlı veritabanlarının geliştirilmesi ve sürdürülmesi veri madenciliği için çok önemlidir (Mirkin, 2005: 48).

Veri madenciliği, anlamlı kalıpları ve kuralları keşfetmek için büyük miktardaki verilerin araştırılmasını ve analizini yapmaktadır. Veri madenciliğinin amacı, bir müşterinin daha iyi anlaşılması yoluyla bir şirketin pazarlama, satış ve müşteri destek işlemlerini iyileştirmesine yardımcı olunması anlamına da gelmektedir. Veri madenciliği tekniklerinin ve araçlarının, astronomi, tıp ve endüstriyel işlem kontrolüne kadar çeşitli alanlarda geçerliliğini sürdürmektedir. Veri madencisi, istatistiklerden, bilgisayar bilimlerinden ve makine öğrenimi araştırmalarından elde edilen bir keşif tekniği kullanmaktadır. Belirli bir durumda uygulanacak belirli bir teknik kombinasyonunun seçimi, veri madenciliği görevinin niteliğine, mevcut verilerin niteliğine, veri madencisinin beceri ve tercihlerine bağlıdır. Veri madenciliğinde kullanılan modeller,

analiz sonuçlarının değerlendirmede skor üretmek için kullanılmaktadır. Skor, bir modelin bulgularını tek bir sayıyla ifade etmenin bir yoludur. Pazarlama alanında skorlar, müşterilerin bir listesini en çok-en sadık veya en çok-en düşük olasılıkla yanıt verme olasılığı veya en çok-en düşük olasılıkla kredi verme şeklinde sıralamak için kullanılabilir (Berry ve Linoff, 1997: 8).

Veri madenciliği, büyük veri tabanlarından daha önce tahmin edilemeyen bilginin çıkarılması ve sonuçların karar vermede uygulanması gibi çok yönlü bir süreçtir. Veri madenciliği teknikleri, verileri kalıptan algılar, modeller ve kurallar çıkarmaktadır. Çıkarılan bilgi daha sonra veri setlerindeki veya veri tabanları arasındaki ilişkileri tanımlayarak sınıflandırma veya tahmin modellerine uygulanabilmektedir. Bu modeller ve kurallar karar vermeyi yönlendirebilir ve bu kararların etkilerini tahmin edebilir (Benoit, 2002: 265).

Veri tabanlarında, depolanan bilgi bakımından zengin verilerden ziyade, karar vericilerin sezgilerine dayanarak önemli kararlar alınmaktadır. Karar vericinin, çok miktarda verinin içine yerleştirilmiş değerli bilgiyi elde etmek için araçları bulunmamaktadır. Ayrıca, bilgiyi manuel olarak bilgi tabanlarına girmek için kullanıcılara veya etki alanı uzmanlarına dayanan mevcut uzman sistem teknolojilerini göz önünde bulundurmamak gerekmektedir. Bu durum, hatalara daha fazla eğilimli olmakla birlikte son derece zaman alıcı ve masraflıdır. Veri analizi yapan, veri madenciliği araçları önemli veri kalıplarını açığa çıkarabilir. Analistler, iş stratejilerine, bilgi tabanlarına, bilimsel ve tıbbi araştırmalara büyük katkı sağlayabilir. Bu yönüyle, bilginin 'altın külçelerine' dönüştürecek veri madenciliği tekniklerinin sistematik bir şekilde geliştirilmesine katkısı olacaktır (Han vd., 2011: 5-6).



Şekil 1. Veri Tabanlarında Bilgi Keşfi (KDD/VTBK) Adımları (Fayyad, 1996: 1590).

Şekil 1’de Veri madenciliği süreci, kolayca anlaşılabilir bir biçimde bilgiyi keşfetmek ve sunmak için Makine Öğrenmesi, istatistik ve görselleştirme tekniklerini kullanmaktadır. VTBK'deki “Bilgi” kelimesi, işlenmiş verilerden çıkarılan kalıpların keşfedilmesi anlamına gelmektedir. Bir model, verilerin bir alt kümesindeki gerçekleri tanımlayan bir ifadedir. Bu nedenle, VTBK ve veri madenciliği arasındaki fark, VTBK, veriden elde edilen bilgileri keşfetme sürecine atıfta bulunurken, veri madenciliği, VTBK işleminin ek adımları olmadan veriden desen çıkarmak için algoritmaların uygulanmasını ifade etmektedir (Dave ve Dave, 2012: 15).

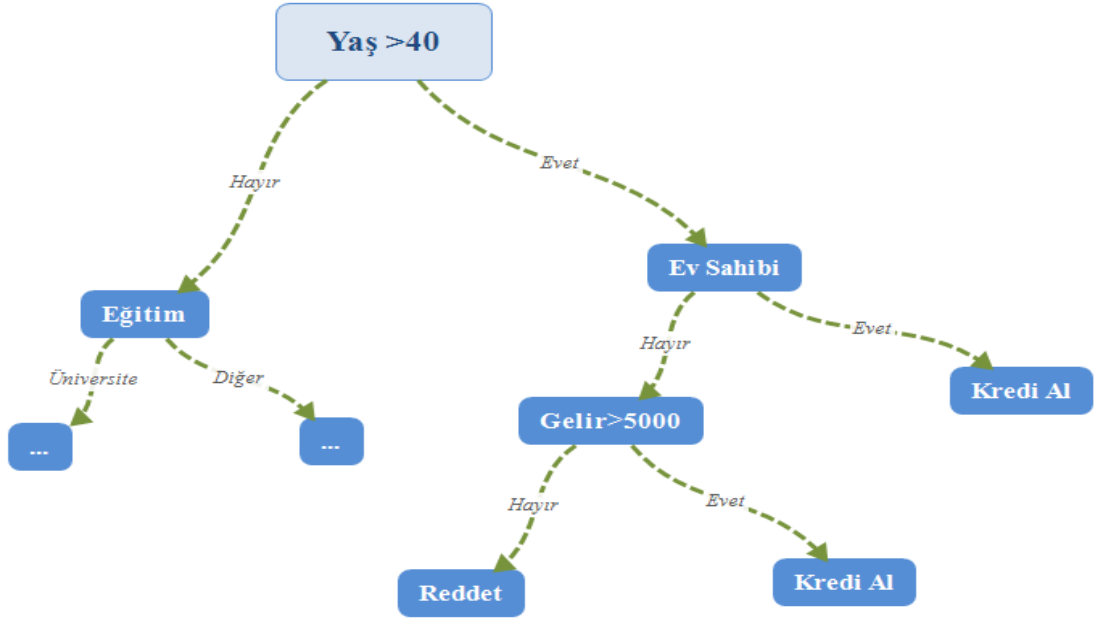
1.1.1. Veri Madenciliği Teknikleri

Her veri madenciliği tekniği, modelleme hedefine bağlı olarak farklı bir amaca hizmet etmektedir. En yaygın iki modelleme hedefi sınıflandırma ve tahmindir. Sınıflandırma modelleri, kategorik etiketleri (ayrık, sırasız) değerlendirirken, öngörü (tahmin) modelleri sürekli değişken değerleriyle ilgilenmektedir. Veri madenciliğinde, Karar Ağaçları, Yapay Sinir Ağları, Regresyon, Market Sepet analizi, Kümeleme algoritmaları gibi teknikler kullanılmaktadır (Palaniappan ve Awang, 2008: 109).

1.1.1.1. Karar Ağaçları

Karar Ağaçları, en yaygın kullanılan yapay zekâ modellerinden biridir. Diğer algoritmalar daha doğru modeller sunmasına rağmen (hedefe daha iyi yaklaşırken), Karar Ağaçları, genellikle çok dikkat çekici kabul edilmektedir. Dikkat çekiciliğinin en önemli sebeplerinden biri anlaşılabilirliktir. Karar Ağaçları, karar fonksiyonlarını tanımlayan bir dizi mantıksal kural şeklinde kolayca ifade edilebilmektedir (Grabczewski, 2014: 1).

Karar Ağaçları, mevcut tüm özellikleri aynı anda kullanan ve her piksel için tek bir üyelik kararı veren geleneksel istatistiksel ve sinirsel / bağlantısal sınıflandırıcıların aksine, Karar Ağacı etiket ataması problemine kademeli veya sıralı bir yaklaşım ile çözüm aramaktadır. Etiketleme işleminin, tek ve karmaşık bir karardan ziyade sıralı testlerin sonuçlarına dayanan basit kararlar zinciri olduğu düşünülmektedir. Karar dizileri ise, düğümlerde uygulanan testlerle Karar Ağacının dallarını oluşturmaktadır (Pal ve Mather, 2003: 556).



Şekil 2. Bir Karar Ağacı Örneği (Pal ve Mather, 2003: 556).

Şekil 2’de yer alan Karar Ağacı örneği, kuralları her nesnenin karar vereceği bir düğümün olduğu hiyerarşik ve sıralı bir yapıda sunmanın bir yolunu göstermektedir. Ağaç, test koşullarını içeren iç karar düğümlerinden ve karar veren yapraklardan oluşmaktadır.

Karar Ağacı Algoritmasındaki temel adım, her düğümdeki özneliğin seçilmesidir. Quinlan, her özellik için bir bilgi kazancının hesaplanmasına dayalı bir seçim ölçüsü sunmuş ve bu bilgi kazancını optimal düzeye çıkaran özellik seçilmiştir. Quinlan Gain Ratio ile öznelikler arasında ilişki kurularak en çok bilgiye sahip kriterlerin önem dereceleri hesaplanmaktadır. X , sonlu bir örnek kümesi ve $\{A_1, \dots, A_p\}$ bir öznelik kümesi olsun. Her A_k özneliği için Karar Ağacı, aşağıdaki Eşitlik (1) ve (2)’deki kullanarak A_k özneliğinin değerlerinde dallanma yoluyla elde edilen bilgileri ölçmektedir.

$$Gain(A_k, X) = I(X) - E(A_k, X) \quad (1)$$

$$I(X) = -\sum_{j=1}^m P_j \log_2 P_j, \quad P_j = \frac{|X \cap F_j|}{|X|} \quad (2)$$

X ‘deki örneklerin m olası sınıfa göre dağılımının rastgeleliğini ölçer. P_j , X ‘deki F_j sınıfına ait örneklerin oranı olarak tanımlanmaktadır. Örneklerin X kümesindeki her bir F_j sınıfının oluşma olasılığı, Eşitlik 3’ deki gibi gösterilmektedir.

$$E(A_K, X) = \sum_{i=1}^n \frac{|X_i|}{|X|} I(X_i) \quad (3)$$

Burada;

n , A_K kriterinin olası değerinin sayısıdır.

$|X_i|$, A_K kriteri için V_i değerine sahip X 'teki örneklerin sayısıdır.

$|X|$, düğümdeki örneklerin sayısıdır.

X_1, \dots, X_n kümelerinin X üzerindeki A_K 'nın n değerleri tarafından oluşturulan bir bölümdür.

$I(X_i)$, olası sınıflar üzerindeki X_i kümesindeki örneklerin dağılımının rastgeleliğini ölçmektedir. Eşitlik (4)'deki gibi gösterilmektedir.

$$I(X_i) = \sum_{j=1}^n \frac{|X_i \cap F_j|}{|X_i|} \log_2 \frac{|X_i \cap F_j|}{|X_i|} \quad (4)$$

Bu nedenle, $E(A_K, X)$ kök olarak A_K olan ağaç için beklenen bilgidir. Bu beklenen bilgi, $I(X_i)$ ölçütü A_K kriterinin n değerleri ağırlıklı ortalamasıdır.

Seçilen özellik bilgi değerini en üst düzeye çıkaran özelliktir. Quinlan, Bias değerinde değişiklik yaparak Gain Ratio için Eşitlik (5) ve (6)'yı kullanmaktadır.

$$IV(A_k) = - \sum_{i=1}^n \frac{|X_i|}{|X|} \log_2 \frac{|X_i|}{|X|} \quad (5)$$

$$G_R(A_K, X) = \frac{I(X) - E(A_K, X)}{IV(A_k)} \quad (6)$$

$IV(A_k)$, özniteliğin (kriterin) bilgi içeriğini ölçer ve Quinlan'a göre "bunun arkasındaki mantık, bir özniteliğin değerini belirleyerek sağlanan bilgilerin olabildiğince çok sınıflandırma amacı için yararlı olması gerektiğidir." Bununla birlikte, değiştirilmiş Bilgi Kazancı Oranı (Gain Ratio), her zaman tanımlanmayabilir (payda sıfır olabilir) ve yüksek kazançlı olanlar yerine düşük kazançlı $IV(A_k)$ olan kriterleri seçebilmektedir (De Mántaras, 1991: 82-83).

Sonuç olarak, Bilgi Kazancı (Information Gain)'nın kullanımından kaynaklanan bias etkisini azaltmak için, Bilgi Kazancı Oranı olarak bilinen bir varyant tanıtılmıştır. Bilgi Kazancı ölçüsü, birçok sonucu olan testlere yöneliktir. Yani, çok sayıda değere sahip kriterleri seçmeyi tercih etmektedir. Bilgi Kazancı Oranı, kriter değerlerinin

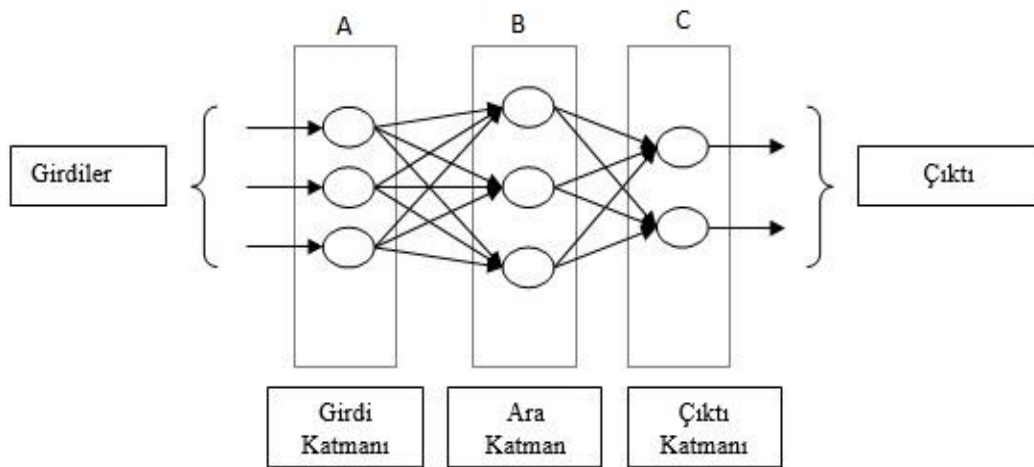
genişliğine ve tekdüzeliğine izin vermek için her kriter için ‘Bilgi Kazancı’ ayarlamaktadır (Shouman vd., 2011: 25).

Karar Ağaçları şeklinde otomatik olarak kurallar oluşturmak, veri arama tekniklerinin geliştirildiği araştırma alanlarının çoğunun incelenmesinin amacı olmuştur. Mühendislik, istatistik, karar teorisi ve daha yakın zamanda yapay zekâ gibi disiplinler, Karar Ağaçlarının üretilmesi ve uygulanmasına adanmış çok sayıda çalışmaya sahiptir (Barros vd., 2015: 7).

1.1.1.2. Yapay Sinir Ağları

Bir Yapay Sinir Ağları (YSA), biyolojik nöronlara benzer bir dizi işleme elementinden (veya nörondan) oluşan bir bilgi işleme sistemidir ve bu elementler arasındaki biyolojik sinir sisteminde sinaptik gücü taklit eden ara bağlantılardan (veya ağırlıklardan) oluşmaktadır. Bir YSA mimarisinde, nöronlar katman adı verilen gruplar halinde düzenlenmektedir. Bir katmandaki her nöron mantıksal paralellikte çalışmaktadır. Veri işlemede, bilgi bir katmandan diğerine iletilmektedir. Bir ağın, bir veya birkaç katmanı olabilmektedir. Bir ağın temel yapısı Şekil 3’deki gibi genellikle üç katmandan oluşmaktadır:

- i. Verinin ağa tanıtıldığı giriş katmanı;
- ii. Verinin işlendiği gizli katman (lar);
- iii. Verilen girdiler için sonuçların üretildiği çıktı katmanı. (Tokar ve Markus, 2000: 157).



Şekil 3. Bir Yapay Sinir Ağ Yapısı Örneği (Bre vd., 2018: 1432).

Şekil 3 bir yapay sinir ağ yapısını göstermektedir. Bu ağ yapısında, girdi katmanı: Dışarıdan gelen verileri ara katmana iletmekle görevli hücrelerden oluşturmaktadır. Ara katman: girdi katmanından aldığı verileri işleyerek çıktı katmanına iletmektedir. Bu katmanda yer alan veriler doğrusal olmayan özellikler taşımaktadır. Veriler, ara katmandaki hücrelerde işlenmektedir. Bir yapay sinir ağında birden fazla gizli veya ara katman bulunabilir. Çıktı katmanı ise, ara katmandan gelen verileri işleyerek girdi setine karşılık gelen çıktı setini meydana getirmektedir (Aydoğan ve Zırhıoğlu, 2018: 585). Şekil 3'teki ağ çalışmasında, tasvir edilen oldukça basitleştirilmiş yapay sinir ağına atıfta bulunularak anlaşılabilir. Bu örnekteki girdi katmanı, üç değişkenden, A, B ve C'den oluşmaktadır. Bu değişkenlerin her biri, ağın ilgili giriş birimine girilmektedir. Bu değerler, ağırlık olarak adlandırılan başlangıçta rastgele bir sayı ile çarpılır. Bu çarpımların ürünleri toplanır ve ara katman biriminin net girdisi haline gelir. Bu değer daha sonra ara katman birimin aktivasyonunu hesaplayan bir lojistik fonksiyona girilir. Ara katman birimin aktivasyonu üçüncü bir ağırlıkla çarpılır. Bu ürün, çıktı biriminin net girdisi olur. Bu toplam daha sonra çıktı biriminin veya "ağ çıktısının" aktivasyonunu hesaplayan aynı lojistik işleve girilir. Ağ, ağırlıkların değeri değiştirilerek eğitilir. Bu değerler, ağırlık değişikliğini belirlemek için ağ çıktısındaki hatayı kullanan bir süreç olan "geri yayılma" yoluyla değiştirilir. Hata, belirli bir eğitim modeli için ağ çıktısı ile bu eğitim modelinin beklenen çıktısı arasındaki fark olarak tanımlanır. Eğitim süreci sırasında, bu ağırlık değerleri belirlenir, böylece ağ tarafından gerçekleştirilen genel hesaplama minimum hata üretecek şekilde optimize edilir. Uygun şekilde eğitilmiş bir ağ, eğitim verilerinde bulunan girdi ve çıktı arasındaki herhangi bir işlevsel veya istatistiksel ilişkiyi çıkarabilir (Baxt, 1991: 844).

Bir YSA, nöronlar olarak bilinen büyük ölçüde birbirine bağlı işlem düğümlerinden oluşmaktadır. Her nöron, ağırlıklı bir girdi kümesini kabul eder ve bir çıktıyla yanıt vermektedir. Böyle bir nöron ilk olarak verilen ağırlıklı girdilerin toplamını oluşturmaktadır. Eşitlik (7)'de gösterildiği gibidir.

$$n = (\sum_{i=1}^P w_i x_i) + b \quad (7)$$

Burada, P ve w_i sırasıyla elaman sayısı ve x_i giriş vektörünün ara bağlantı ağırlığıdır. b nöron için bias değeridir. Bilginin bir dizi bağlantı ağırlığı ve bias olarak saklanmaktadır. Bir önyargıya sahip ağırlıklı girdilerin toplamı f ile temsil edilen bir

etkinleştirme işlevi aracılığıyla işlenmektedir. Hesaplanan çıktı değeri Eşitlik (8)'de gösterilmiştir.

$$f(n) = f[(\sum_{i=1}^P w_i x_i) + b] \quad (8)$$

Temel olarak, nöron modeli, girdileri önemli ölçüde uyarıldığında, yani n yeterince büyük olduğunda biyolojik nöronu temsil etmektedir. Eşik işlevi, sigmoid ve hiperbolik tanjant gibi aktivasyon işlevini tanımlamanın birçok yolu bulunmaktadır (Ertunc ve Hosoz, 2006: 629).

Sonuç olarak, YSA verilerdeki karmaşık modellerin sayısallaştırılmasına ve çoğaltılmasına dayanan güçlü ve parametrik olmayan bir tahmin aracıdır. Çoklu regresyon, diskriminant analizi, lojistik regresyon ve kümeleme analizinden daha iyi performans göstermektedir (Daniel vd., 2003: 48). Bağımlı ve bağımsız değişkenler arasında önceden tanımlanmış herhangi bir temel ilişki gerektirmeyen YSA, finansal analiz, karar problemleri ve örüntü tanımada yaygın olarak kullanılmaktadır. YSA'nın özellikle tahmin ve sınıflandırma problemleriyle başa çıkmada güçlü bir araç olduğu gösterilmektedir. YSA ile 1990'lardan bu yana ulaşım alanında sürücü davranış analizi, kaldırım bakımı, araç hata algılamaları vb. gibi çalışmalar bulunmaktadır (Chang, 2005: 542).

1.1.1.3. Regresyon

Regresyon analizi, iki veya daha fazla kantitatif değişken arasındaki ilişkiyi kullanan istatistiksel bir araçtır. Böylece bir değişken veya diğer değişkenlere bağlı olarak tahmin edilebilir. Örneğin, reklam harcamaları ile satış arasındaki ilişkiyi bilen bir kişi varsa, reklam harcamalarının seviyesi belirlendiğinde satışlar regresyon analizi ile tahmin edilebilir (Neter vd., 1981: 23). Regresyon analizi birçok araştırma projesinin merkezi bir parçasıdır. Regresyon, bir bağımlılık çalışmasıdır (Weisberg, 2005: 1).

Regresyon analizi, sosyal araştırmalarda en sık kullanılan istatistiksel yöntemdir. Bunun nedeni, çoğu sosyal araştırmacının, deneysel olmayan verilerden “nedensel” etkilerin belirlenmesiyle ilgilenmesidir. Regresyon etkilerin ortaya çıkarılmasını sağlayan en belirgin yöntemdir (Brüderl, 2003: 1).

İstatistikte, regresyon analizi, bir bağımlı değişken (aynı zamanda yanıt değişkeni olarak da adlandırılır) ile bir veya daha fazla bağımsız değişken (açıklayıcı değişkenler veya öngörücüler olarak da bilinir) arasındaki ilişkiyi modellemeye yönelik tekniklerden

oluşmaktadır. Eğer bir regresyon modeli, bağımlı değişkeni ile bağımsız değişkenler arasındaki gerçek ilişkiyi yeterince yansıtıyorsa, bu model bağımlı değişkeni tahmin etmek, önemli bağımsız değişkenleri belirlemek ve tepki değişkeni ile bağımsız değişkenler arasında istenen nedensel ilişkiyi kurmak için kullanılabilir. Regresyon analizi yapmak için, araştırmacı ilgili değişkenlere yönelik verileri toplar ve bağımsız değişkenlerin, bağımlı değişkenine olan nicel etkisini tahmin etmek için regresyon modelini kullanır (Yan ve Su, 2009: v).

Çeşitli kriterlere bağlı olarak, regresyon modelleri aşağıdaki şekilde sınıflandırılabilir:

1. Denklemlerin sayısına göre: -Tek bir bağımlı değişkenin tek bir denklemdeki bir dizi açıklayıcı değişkenle açıklandığı tek denklemliler. -İki veya daha fazla bağımlı değişkenin iki veya daha fazla denklemde açıklandığı çoklu denklem modelleri.
2. Açıklayıcı değişkenlerin sayısına göre: - Bağımlı değişkenin bir davranışının yalnızca bir açıklayıcı değişkenle açıklandığı tek değişkenli modeller. - İki veya daha fazla açıklayıcı değişkenin meydana geldiği çok değişkenli (veya çoklu) modeller.
3. Modelin işlevsel (analitik) formuna göre: -Bağımlı değişken ile açıklayıcı değişkenler arasındaki doğrusal ilişkinin uygulandığı doğrusal modeller. - Doğrusal dışındaki matematiksel işlevlerin uygulandığı doğrusal olmayan modeller (örneğin, üstel fonksiyonlar).
4. Bağımlı değişken ile açıklayıcı değişkenler arasındaki ilişkinin niteliğine göre: - Yalnızca bağımlı değişkenle iyi temellendirilmiş nedensel ilişkileri olan açıklayıcı değişkenlerin kullanıldığı nedensel modeller. - Regresyon setinin, bağımlı değişkenle istatistiksel olarak anlamlı korelasyonlar gösteren, ancak nedensel ilişkiler olması gerekmeyen değişkenleri içerdiği semptomatik modeller.
5. Bağımlı değişkenin niteliğine göre: - Bağımlı değişkenin sürekli bir ölçekte ölçüldüğü sürekli değişkenler için modeller (örneğin, yüzde olarak ifade edilen enflasyon). - Bağımlı değişkenin genellikle nitel bir doğaya sahip olduğu ve ikili bir şekilde ifade edildiği ayrı değişkenler için modeller.
6. Modele dâhil edilen değişkenlerin ölçüm esasına göre: - Değişkenlerin bireysel gözlemlerinin zaman içinde değiştiği zaman serisi modelleri (örneğin, bireysel

gözlemlerin çeyrektekilerden çeyreğe farklılık gösterdiği bir şirketin üç aylık bir kar modeli). - Değişkenlerin bireysel gözlemlerinin aynı anda, ancak farklı nesnelere için ölçüldüğü kesitsel modeller (örneğin, 30 Avrupa ülkesindeki suç oranları ile bu ülkelerin kişi başına GSYİH'si arasındaki istatistiksel ilişkiyi yakalayan bir model. Değişkenlerin gözlemleri aynı zamanda ölçülür, ancak ülkeler arasında farklılık gösterir).

Zaman serileri ile kesitsel veriler arasındaki ayrım, regresyon analizi için çok önemlidir, çünkü hem model tahmininde uygulanan metodolojiyi hem de tahmin edilen modellerin pratik uygulamasını etkilemektedir (Welc ve Esquerdo, 2018: 2-3).

1.1.1.4. Market Sepet Analizi

Birliktelik Kuralı (Market Sepet Analizi), 1993'teki başlangıcından bu yana, büyük ilgi görmüştür. Bugün bu tür kuralların madenciliği, VTBK'daki en popüler kalıp bulma yöntemlerinden biridir. Kısaca, bir birliktelik kuralı X ve Y 'nin öge kümeleri olduğu bir $X \rightarrow Y$ ifadesidir. Bu tür kuralların anlamı oldukça sezgiseldir: "Her $T \in D$ işleminin bir dizi öge olduğu" işlemlerin bir veri tabanı verildiğinde, $X \rightarrow Y$, bir T işleminin T yerine X içerdiğinde muhtemelen Y 'yi de içerdiğini ifade etmektedir. Olasılık veya kural güveni, X içeren işlemlerin toplam sayısına göre X 'e ek olarak Y içeren işlemlerin yüzdesi olarak tanımlanmaktadır. Olasılık veya birliktelik kuralı güveni, X içeren işlemlerin toplam sayısına göre X 'e ek olarak Y içeren işlemlerin yüzdesi olarak bilinmektedir. Yani, birliktelik kural güveni koşullu olasılık $p(Y \subseteq T | X \subseteq T)$ olarak anlaşılabilir. Birliği kuralları fikri, " x_1 ve x_2 ürünlerini satın alan bir müşteri de y ürününü % c olasılıkla alacak" gibi kuralların bulunduğu pazar sepeti verilerinin analizinden kaynaklanmaktadır. Birliktelik kuralları araştırılırken, başa çıkılması gereken başlıca iki sorun vardır: Birincisi, algoritmik karmaşıklığıdır. Kuralların sayısı, öğelerin sayısı ile birlikte katlanarak artar. Neyse ki bugünün algoritmaları, bu uçsuz bucaksız arama alanını, kurallar üzerindeki kalite ölçümleri için minimum eşiklere dayalı olarak verimli bir şekilde budayabilir. İkinci olarak, farklı kural yapıları genelleştirilmiş kural setlerinden seçilmelidir. Oluşturulan kural kümeleri normalde oldukça büyük olduğundan bu oldukça maliyetli olabilir. Veri madenciliği uzmanlarının iş sorunlarına doğrudan uygulanabilirliği ile birliktelik kurallarını popüler bir madencilik yöntemi haline getirmiştir (Hipp vd. , 2000: 58).

Birliktelik Kuralı, işletmelerin işlem veri tabanlarından ilişkilendirmeler veya ortak oluşumlar çıkararak müşteri satın alma modellerini keşfetmede faydalı bir yöntemdir. Analizden elde edilen bilgiler pazarlama, satış, servis ve işletme stratejilerinin oluşturulmasında kullanılabildiğinden son derece önemlidir. Bununla birlikte, mevcut yöntemler, söz konusu ürünlerin tüm mağazalarda her zaman rafta olduğu varsayımı nedeniyle, çok mağazalı bir ortamda önemli satın alma modellerini bulamayabilir. Bu yönüyle de, market sepet analizi mevcut yöntemlerden oldukça farklıdır (Chen vd., 2005: 339).

Market Sepet Analizi, büyük perakende ürün çeşitlerinde, alışveriş sepetlerinin bileşiminin anlamlı bir kategori seçimi ile nasıl belirleneceği konusunda fikir vermektedir (Boztuğ ve Reutterer, 2008: 294). Bunun yanı sıra, market sepet analizi, araştırmacıların ürünler veya kategoriler arasındaki açık olmayan genellikle gizli ve sezgisel ilişkileri keşfetmelerini sağlamaktadır. Bu metodolojik yaklaşım, araştırmacıların birlikte ortaya çıkan öğeleri (yani birlikte göründüklerini) sık sık tanımlamalarını ve birlikte ortaya çıkma derecelerini değerlendirmelerini sağlamaktadır. Market sepet analizi, birlikte satın alınan kitap türleri (Amazon.com'da satın alındığı gibi) ile aynı davranışın (VirginWines.com sitesinde satın alındığı gibi) satın alınması muhtemel olan farklı şarap türleri ile ilgili tüketici davranışlarını anlamak için kullanılmıştır. Market sepet analizi, pazarlama alanında ortaya çıktığı ve başlangıçta hangi süpermarket ürünlerinin birlikte satın alındığını anlamak için kullanıldığından, teknik, market sepet analizini benimsemiştir (Aguinis vd., 2013: 1800-1801).

1.1.1.5. Kümeleme

İnsanlar, verilerin gitgide arttığı günümüz dünyasında her gün büyük miktarda bilgi ile karşılaşmaktadırlar. Araştırmacılar, daha fazla analiz ve yönetim için bunları veri olarak depolamaktadırlar. Bu verilerle başa çıkmanın hayati yollarından biri, onları bir dizi kategori veya küme halinde sınıflandırmak veya gruplandırmaktır (Xu ve Wunsch, 2005: 645). Verileri temsil etmenin genel bir yolu, herhangi bir nesne çifti arasındaki benzerliği belirlemektir. İki nesne çok fazla yapı paylaşıyorsa, verileri aynı prototipten çoğaltmak mümkün olmaktadır. Bu fikir, denetlenmeyen algoritmaların zengin bir alt sınıfını oluşturan kümeleme yöntemlerinin temelini oluşturur (Camastra ve Vinciarelli, 2015: 118).

Kümeleme, verileri gruplara, kümelere veya hiyerarşilere ayırarak bir verinin yapısal özelliklerini belirlemeye çalışan bir dizi tekniğe verilen addır. Aynı grup içindeki örnekler, farklı gruplardaki örneklerden birbirine benzemektedir. Kümeleme, keşif amaçlı veri analizi prosedürüdür (Lavine ve Mirjankar, 2006: 2).

Küme analizinin amacı, deney verilerini, her bir kümenin öğelerinin mümkün olduğu kadar benzer ve diğer kümelerinkilerden farklı olması gereken belirli sayıda kümede sınıflandırmaktır. Bu, sınıflandırılacak elemanlar arasında bir mesafe ölçüsü veya benzerlik ölçütünün varlığını ifade etmektedir. Bu gibi sınıfların sayısı önceden belirlenebilir (Ruspini, 1969: 23).

Kümeleme, veri madenciliği, belge alma, görüntü bölütleme ve model sınıflandırması dahil olmak üzere çeşitli keşifsel model analizi, gruplama, karar verme ve makine öğrenimi durumlarında faydalıdır. Bununla birlikte, bu tür birçok problemde, veriler hakkında çok az ön bilgi (örneğin, istatistiksel modeller) vardır ve karar vericinin veriler hakkında olabildiğince az varsayımda bulunması gerekir. Bu kısıtlamalar altında, kümeleme metodolojisi, yapılarının bir değerlendirmesini yapmak için veri noktaları arasındaki karşılıklı ilişkilerin araştırılması için özellikle uygundur (Jain vd., 1999: 265).

1.2. Metin Madenciliği

Metin madenciliği, yapılandırılmamış metin verilerinden anlamlı bilgiler çıkarmaya çalışan yeni bir teknolojidir. Metin madenciliği, veri madenciliğinin metinsel veriye bir uzantısıdır. Çok sayıda metinsel belgeden yararlı bilgileri hızlıca toplamak için, otomatik bilgisayar tekniklerini kullanmak zorunlu hale gelmiştir. Metin madenciliği, metin dosyaları, HTML dosyaları, sohbet mesajları ve e-postalar gibi yapılandırılmamış metin verilerinden faydalı modeller, trendler, desenler veya kurallar bulmaya odaklanmaktadır. Otomatik bir teknik olarak, metin madenciliği “metinlerden bilgiyi verimli ve sistematik olarak belirlemek, çıkarmak, yönetmek, entegre etmek ve kullanmak” için kullanılabilir. Birçok araştırmacı, işletme, sağlık bilimleri ve eğitim alanlarındaki büyük miktarda metin verilerini analiz etmek için metin madenciliği tekniklerini başarıyla kullanmıştır. Witten, Don, Dewsnip ve Tablan, dijital kütüphanedeki belgelerden meta verileri çıkarmak ve metindeki uygun öğeleri işaretleyerek belgeleri zenginleştirmek için metin madenciliği tekniklerini kullanmışlardır (He vd., 2013: 465).

Metin madenciliği, büyük veri tabanlarından ilginç örüntüler bulmaya çalışan veri madenciliği adı verilen bir alandaki bir varyasyondur. Veri madenciliğindeki tipik bir örnek, hangi ürünlerin raflara yakın yerleştirileceğini tahmin etmek veya benzerleri için kuponlar sunarak tüketici satın alma modellerini kullanmaktır. Örneğin, bir el feneri alırsanız, onunla birlikte pil satın almanız da olasıdır. İlgili bir uygulama, kredi kartı kullanımında olduğu gibi sahtekârlığın otomatik tespitidir. Analistler, normal harcama modellerinden sapmaları bulmak için çok sayıda kredi kartı kaydına bakarlar. Klasik bir örnek, denizaşırı bir uçağın uçuşundan sonra az miktarda benzin almak için bir kredi kartının kullanılmasıdır. Talep, ilk satın alma işleminin kartı aktif olduğundan emin olmak için test etmesidir (Hearst, 2003).

Bilgi depolamanın en doğal şekli metin olduğundan, metin madenciliğinin, veri madenciliğinden daha yüksek bir ticari potansiyele sahip olduğuna inanılmaktadır. Aslında, son zamanlarda yapılan bir çalışma, bir şirketin bilgisinin %80'inin metin belgelerinde bulunduğunu göstermiştir. Bununla birlikte, metin madenciliği, doğal olarak yapılandırılmamış ve bulanık olan metin verilerini ele almayı içerdiğinden (veri madenciliğinden) çok daha karmaşık bir iştir. Metin madenciliği, bilgi alma, metin analizi, bilgi çıkarma, kümeleme, kategorizasyon, görselleştirme, veri tabanı teknolojisi, Makine Öğrenmesi ve veri madenciliğini içeren çok disiplinli bir alandır (Tan, 1999: 65).

Veri ve metin madenciliği, makine tarafından okunan materyallerden bilgi türetme işlemidir. Büyük miktardaki materyaller kopyalanır ve veriler çıkarılır. Çıkarılan örnekleri tanımlamak için yeniden birleştirmeler yapılmaktadır. Veri ve metin madenciliği dört aşamadan oluşmaktadır.



Şekil 4. Veri ve Metin Madenciliği Bileşenleri

Şekil 4'de genel olarak bir veri ve metin işlendiğinde potansiyel belgeler tanımlanmaktadır. Bu belgeler daha sonra makinede okunabilir bir formata dönüştürülmektedir. Böylece yapılandırılmış veriler elde edilmektedir. Yararlı bilgiler çıkarılarak yeni bilgiler keşfedilir. Hipotezler test edilir ve yeni ilişkiler tanımlanır. Sonuç olarak, veri ve metin madenciliğinde, araştırma amacıyla çok miktarda materyalden anlamlı sonuçların çıkarıp yorumlanması anlamına gelmektedir

(Libereurope, 2019). Ayrıca veri madenciliği ve metin madenciliğinin karşılaştırılması Tablo 1’de verilmiştir (Slideshare, 2019).

Tablo 1. Veri Madenciliği ve Metin Madenciliği Karşılaştırılması

Veri Madenciliği	Metin Madenciliği
Doğrudan işlem	Dil işleme veya Doğal Dil İşleme
Nedensel ilişkiyi tanımla	Önceden bilinmeyen bilgileri keşfetme
Yapılandırılmış veriler	Yarı yapılandırılmış ve yapılandırılmamış veriler
Veri tabanında bulunan, formatlı veri.	Birden fazla tipte bulunan belli bir alanda bulunmayan veri.

Tablo 1’de incelendiğinde verinin türüne göre veri madenciliği ve metin madenciliği kullanılabilir. Genel olarak, veri madenciliğinde verinin hazır bir veritabanından alındığı ve formatının belli olduğu söylenebilir. Metin madenciliğinde ise birçok veri türü aynı anda daha dağınık veri kaynaklarında bulunabilir.

1.2.1. Metin Madenciliğine Yönelik Alanlar

Akıllı metin analizi, veri/metin madenciliği veya metin içindeki bilgi keşfi olarak da bilinen metin madenciliği, genellikle yapılandırılmamış bir metnin kaynak olduğu ilgi çekici, önemsiz bilgi ve bilginin çıkarılması sürecini ifade etmektedir. Metin madenciliği, metin kümeleme ve sınıflandırma, bilgi çıkarma, bilgi alımı, Adlandırılmış Varlık Tanımlama (NER), Doğal Dil İşleme (NLP), soru cevaplama ve görselleştirme alanlarını kapsamaktadır (De Bruijn ve Martin, 2002: 7).

1.2.1.1. Metin Kümeleme ve Sınıflandırma

Metin kümeleme ve sınıflandırma, kullanıcıların metinsel bilgileri düzenlemelerine ve bağlamsallaştırmasına yardımcı olacak iki yaklaşımdır. Mevcut metin madenciliği sistemleri, tipik olarak, tek terimlerin veya terimlerin kaynaklarının belgeleri temsil etmek için özellikler olarak kullanıldığı, bilgi alımında bilinen bir “the bag of words” modelini kullanmaktadır. Metin madenciliğinde literatürde yer alan son çalışmalar ile gelişmeler meydana gelmektedir. Mühendisler ve alan uzmanları tarafından çalışmalar desteklenmektedir (Bloehdorn vd., 2006: 334).

Metin sınıflandırma, mevcut veri madenciliğinde yaygın olarak kullanılan veri sınıflandırma algoritmalarını kullanmaktadır. Veri sınıflandırma, sınıfa özgü öğrenme yoluyla veri öğrenmek için önceden belirlenmiş sınıflandırma kriterlerini kullanmaktadır. Daha sonra, girdi verilerini önceden belirlenmiş sınıflara konumlandırarak öğrenme sonucu gerçekleştirilir. Geleneksel metin sınıflandırmalarının aksine, metin kümelemesi önceden belirlenmiş kriterleri kullanmamaktadır. Ancak benzer verileri, benzer veri

özelliklerine göre kümelenecek verilen örnekler analiz edilmektedir. Yani, verilen örneklerin bir tahminini yapmak mümkündür (Hong vd., 2015: 24).

1.2.1.2. Bilgi Çıkarımı

Bilgi çıkarma, önceden tanımlanmış bilgi türlerinin metinden çıkarılmasını içermektedir. Buna karşılık, bilgi çıkarma, dokümanları bulmaya odaklanmaktadır. Google veya PubMed arama motorları gibi çok popüler siteler buna örnektir. Bilgi çıkarma, Doğal Dil İşleme konusunda uzmanlaşmış bir alt etki alanı olarak da bilinmektedir. Metin madenciliğinde bilgi çıkarma, yapılandırılmamış verilerden bilgi bulma ve çıkarma süreci olarak tanımlanmaktadır (Meystre vd., 2008: 129).

Bilgi çıkarma, büyük miktarlardaki metinlerden bilgi elde etmenin bir yoludur. Araştırmacılar, literatürde varolan metinsel kalıpları tanımlayabilir, metinleri karşılaştırabilir ve metinleri birbirleri ile ilişkilendirerek kolay bir şekilde analiz edebilirler (Richards vd., 2015: 240).

Bilgi çıkarma, çeşitli uygulamalarda yararlı olabilmektedir. Örneğin; seminer duyuruları, kurs girişleri, iş ilanları ve kiralık daire ilanları gibi. Bilgilerin kolayca aranabileceği bir veri tabanı oluşturmak için metin belgelerinden bilgi çıkarmak önemlidir. Böylece, kullanılan bir metin madenciliği tekniği ile çevrimiçi metne daha kolay ulaşılabilir bir hale gelecektir (Nahm ve Mooney, 2002: 61).

1.2.1.3. Bilgi Alımı

Metin verilerini kullanırken ilk adımlardan biri, ilgili belgelerin geniş bir koleksiyondan çıkarılmasıdır. Bu işlem genellikle bilgi alımı olarak bilinmektedir. Google gibi güçlü arama motorlarına sahip web sayfaları kullanıcıların aradığı bilgiye ilişkin verileri bir belge listesine dönüştürmektedir. Sorgu tabanlı ve belge tabanlı aramalar olarak iki temel arama stratejisi vardır. Sorgu tabanlı aramalarda, kullanıcı tarafından belirlenen anahtar kelime kombinasyonlarını içeren belgeler incelenir. Anahtar kelimeler, AND, OR ve NOT gibi Boolean operatörleri tarafından birleştirilebilir. Belge tabanlı aramalarda ise, birkaç anahtar kelimenin birleşiminden ziyade, verilen bir sorgu belgesine bir bütün olarak benzeyen sıralanmış bir belge listesini incelemeyi amaçlamaktadır. Örneğin; Google'ın popüler arama motoru son zamanlarda, bilimsel makaleler, raporlar ve kitapların alınması için Google Akademik diye özel bir arama aracını bünyesine dahil etmiştir (Krallinger ve Valencia, 2005: 224).

Bilgi alımı, kullanıcıların geniş bir metin verisi koleksiyonundan, belirli bir uygulama problemini çözmek için gerekli olan metin verilerini bulmalarına yardımcı olmaktadır. Böylece büyük miktardaki metinsel veriyi daha kolay işleyebilen küçük metin verisine dönüştürmektedir. Metin madenciliği uygulama sistemlerinde bilgi alımı, kullanıcılara karar verme konusunda yarar sağlamaktadır. Bilgi alımı, metin verilerindeki kalıplardan bilgiyi çıkarmaya ve keşfetmeye destek sağlamaktadır (Zhai ve Massung, 2016: 6).

1.2.1.4. Adlandırılmış Varlık Tanımlama

Adlandırılmış Varlık Tanımlama (NER), bir metnin işlenmesini ve belirli Adlandırma Varlıkları kategorilerine ait olan sözcük veya ifadeleri tanımlamayı içermektedir. NER yazılımı, bilgi çıkarma, bilgi alma ve diğer metin işleme uygulamaları gibi görevler için önemli bir ön işleme aracı olarak işlev görmektedir. NER ile ham metinden daha doğru bir şekilde bilgi aranabilmektedir. Örneğin; Araştırmacı, "Gates" adı verilen bir kişi hakkında tüm belgelerde arama yaparak ilgilenilen bilgiye ulaşılabilir (Mikheev vd., 1999: 1).

Varlık adları bir belgenin ana içeriğini oluşturduğundan, NER daha akıllı bilgi alma ve yönetme yolunda çok önemli bir adımdır. Bir cümlede "kim", "nerede" ve "ne kadar" gibi sorular bilgi çıkarma için önemlidir. NER, bu önemli sorulara cevap veren kalıpları (belirteçleri) bulmaya yardımcı olmaktadır. Buna ek olarak, veriyi belirli dizilerle sınırlandırıp geniş bir yüzey ayrıştırması yaparak bilinmesi gereken ifadeyi ortaya koymaktadır. Metin madenciliğinde NER nispeten basittir ve bir sistemi kolay bir şekilde inşa edilebilir. Ancak, çok sayıda belirsiz durum bir araya geldiğinde istenilen performansa ulaşmak zor olabilmektedir (Zhou ve Su, 2002: 473).

1.2.1.5. Doğal Dil İşleme

Doğal Dil İşleme (NLP), bilgisayarların nasıl doğal dil metnini ve konuşmasını anlayıp kullanılabileceğini araştıran bir araştırma alanıdır. NLP araştırmacıları, insanların dili nasıl anladıkları ve kullandıkları hakkında bilgi edinmeyi amaçlamaktadır. Bu bağlamda, bilgisayar sistemlerinin belirtilen görevleri yerine getirmek için doğal dilleri anlamalarını ve manipüle etmelerini sağlamak için uygun araçlar ve teknikler geliştirilebilir. NLP'nin temelleri, bilgisayar ve bilişim bilimleri, dil bilimi, psikoloji, matematik, elektrik ve elektronik mühendisliği, yapay zekâ ve robotik gibi birçok disiplinin içinde yer almaktadır. NLP'nin uygulamaları arasında makine çevirisi, doğal dil

metin işleme ve özetleme, bilgi alımı, konuşma tanıma, yapay zekâ ve uzman sistemler benzeri birçok çalışma alanı bulunmaktadır (Chowdhury, 2003: 52).

NLP hem bir dizi teori hem de bir dizi teknolojiye dayanan metni analiz etmek için bilgisayarlı bir yaklaşımdır. Aktif bir araştırma ve geliştirme alanı olduğunda birçok tanımı bulunmaktadır. Bir başka tanıma göre NLP, bir dizi görev veya uygulama için insan benzeri dil işlemeyi gerçekleştirmek amacıyla doğal olarak ortaya çıkan metinleri bir veya daha fazla dil bilimsel analiz düzeyinde toplayarak analiz etmektedir. Analiz yaparken kendine özgü hesaplama teknikleri kullanmaktadır (Liddy, 2001: 2).

1.2.1.6. Soru Cevaplama

Soru cevaplama sistemi, bilgi alımına benzemektedir. Amacı, deneyimsiz kullanıcılara bilgiye esnek bir şekilde erişmelerini sağlamak, doğal dilde bir soru yazmak ve cevap içeren bir doküman seti elde etmek değil, verilen sorulara istenilen cevabın bulunmasına yönelik bir uygulamadır. Soru cevaplama, örneğin; George Bush'un kim olduğunu ve UNICEF (Birleşmiş Milletler Çocuklara Yardım Fonu)'in ne olduğunu gibi sorulara cevaplar aramaktadır (Denicia-Carral vd., 2006: 76).

Soru cevaplama teknolojisine sahip birçok web sitesi, son kullanıcıların bilgisayara bir soru sormasına ve bir cevap almasına izin vermektedir. Sorular ve cevaplar birden fazla metin madenciliği tekniğini kullanmaktadır. Soru cevaplama, örneğin; insanlar, yerler, olaylar hakkında cevaplar aramak veya kim, nerede, ne zaman, nasıl vb. soru kategorizasyonları için bilgi çıkarımında kullanılabilir (Gupta ve Lehal, 2009: 66).

Genel olarak, soru cevaplama, doğal olarak formüle edilmiş sorulara kesin cevaplar vermeye çalışan sistemleri açıklamaktadır. Metin madenciliği, yalnızca metinlerde belirtilen gerçeklere doğrudan erişim sağlamakla kalmaz, aynı zamanda literatüre dayalı keşif yaparak metindeki veriler arasındaki dolaylı ilişkileri ortaya çıkarılmasına yardımcı olmaktadır. Böylece metin madenciliği doğrudan aşırı bilgi yükü sorununu ele almaktadır (Zweigenbaum vd., 2007: 359).

1.2.1.7. Görselleştirme

Görselleştirme, veriye görsellik kazandırarak daha anlaşılır olmasını sağlayan bir tekniktir. Metin madenciliği dâhil olmak üzere çeşitli alanlarda güçlü bir araç olduğu kanıtlanmıştır. Metin madenciliği muazzam miktarda veriyi daha küçük bir alt kümeye indirgeyebilse de, bu alt küme genellikle bir analistin makul bir şekilde işlemesi,

kavraması, verinin eğilimlerini saptaması ve uygun bir sonuç çıkarması için çok büyüktür. Metin görselleştirme ve görsel metin madenciliği son işleme araçları, bilginin keşfedilmesini kolaylaştırmanın yanı sıra, çok büyük miktarda verinin büyük bir resme yansıtılması açısından çok önemli olabilmektedir (Puretskiy vd., 2010: 107).

Yıllardır bilim adamları ve mühendisler, birliktelik kurallarının analizini desteklemek için birçok görselleştirme tekniği geliştirdiler. SGI'nin MineSet ve IBM'in QUEST gibi ticari veri madenciliği sistemleri, iş veritabanlarının ilişkilerini görselleştirmek için araçlar sağlamıştır. Bununla birlikte, görselleştirmelerin çoğu, çok sayıda veriyi içeren büyük miktarlarda kurallar ile başa çıkmakta sınırlı kalmaktadır. Bu sınırlama, büyük veritabanlarının bilgilerini anlama ihtiyacı duyan analistler için ciddi zorluklar meydana getirmektedir (Wong vd., 1999: 120).

1.3. Veri ve Metin Madenciliği İle Yapılan Bazı Çalışmalar

Veri ve metin madenciliğini destekleyen algoritmaları içeren literatürdeki çalışmaların bir kısmı aşağıdaki Tablo 2'de verilmiştir. Genel olarak Tablo 2' incelendiğinde veri ve metin madenciliği yapılırken sosyal medya platformları (TripAdvisor, Twitter, Hepsiburada, Foursquare, Facebook, Amazon vb.) kullanılmıştır. Bunun yanı sıra, Duygu (Sentiment) analizi, Destek Vektör Makineleri (DVM), Kümeleme araçları, Naive Bayes gibi algoritmalarla çalışmalar desteklenmiştir. Veri ve metin madenciliğine yönelik öne çıkan bazı çalışmalar Tablo 2'de gösterilmiştir.

Tablo 2. Veri ve Metin Madenciliği İle ilgili Genel Literatür Taraması

Yazar	Konu	İçerik
Afzaal vd., (2019: 1-24).	Yorumların değerlendirilmesinde çok yönlü bir sınıflandırma modeli	Facebook, TripAdvisor, Zomato ve Expedia sosyal sitelerinden elde edilen yorumlar değerlendirilmiştir. Değerlendirmede turistler tarafından sıkça ziyaret edilen otel, park, restoran, müze gibi yerlerin veri seti kullanılmıştır. Makine Öğrenme Algoritmalarından yararlanarak yapılan analizde %87 oranında bir başarı elde edilmiştir.
Atan (2016: 1-248).	Metin madenciliği ile Duygu Analizi ve Borsa İstanbul uygulaması	Borsa İstanbul'da işlem gören firmalarla ilgili yayınlanan firmaların finansal durumları ile bu firmalar hakkında yayınlanmış tüm haberler arasındaki ilişkileri ortaya koymaktadır. Bu amaçla, 313 farklı haber kaynağında yayınlanmış 14108 haber Duygu Analizi ile incelenmiştir. Uygulama kısmında ise, metin madenciliği kullanılarak, 2014 yılı içerisinde yayınlanan haberler ile Duygu Analizi gerçekleştirilmiştir. Bu analiz sonucunda elde edilen bulgu, finansal piyasalarla ilgili çıkan haberler arasında önemli ilişkilerin var olduğu görülmüştür.

Balbi vd., (2018: 674-685).	Kullanıcı memnuniyetini ölçmek için farklı değerlendirme sistemlerini birleştirme	Sosyal medyadaki verileri değerlendirme aşamasında hem kullanıcıların sayısal derecelendirmelerini hem de metinsel yorumlar ele alınarak incelenmiştir. Yorumlardan elde edilen polarite değerleri ile derecelendirmeler bir çatı altında birleştirilmiştir. Bu birleştirmeye yönelik bir algoritma önerilmiştir. Araştırmanın son kısmında TripAdvisor'da yayınlanan Floransa'daki Uffizi Galerisi hakkında bir dizi eleştiriyi analiz ederek önerinin etkinliğini gösterilmiştir.
Barushka ve Hajek (2019: 340-350).	Word Embedding ve Deep Neural Network kullanılarak spam tespiti	Çalışma, hem Bag of Words hem de n-grams kullanılarak spam (sahte yorum) tespitine yönelik bir yaklaşım önermektedir. Yaklaşımı doğrulamak için iki otelin olumlu ve olumsuz veri seti kullanılmıştır. Yaklaşımın analizinde ileri beslemeli sinir bir sinir ağı kullanılmıştır. Performans değerlendirmesinde yorumların polarite değerleri incelenmiştir.
Burns vd., (2019: 1-20).	Keşfetme ve sınıflandırma için Sentimet analizi ve Twofold-LDA modeli	TripAdvisor, Mp3 ve Tv veri setleri kullanılarak yapılan değerlendirmede Sentimet analizi ve Twofold-LDA modeli kullanılmıştır. Veri setlerinde yer alan yorumların polarite değerlerine göre sınıflandırması yapılarak karşılaştırmalı bir inceleme yapılmıştır.
Chen vd. (2017: 4629-4647).	Çok boyutlu derecelendirme sistemlerinin değeri ile bir analiz	Çevrim içi ürün derecelendirmeleri, ürün kalitesi hakkında bilgi sunmaktadır. Web sitesinden alınan verilere göre birden fazla boyutta bir analiz yapılmıştır. Çok boyutlu derecelendirme sisteminin tüketicilerin tercihlerine daha uygun ürünleri bulmalarına ve seçimlerinin güvenini artırmalarına yardımcı olduğu gösterilmiştir.
Colladon vd., (2019: 1-11).	Çevrimiçi seyahat forumlarını değerlendirme ve turizm talebini tahmin etmek için sosyal ağ ve semantik ağ analizi	TripAdvisor seyahat forumundan 147 bin kullanıcı tarafından yazılan 2,5 milyondan fazla yorum derlenmiştir. 7 Avrupa başkentindeki turistlerin gelişi, tahmin edilmeye çalışılmıştır. Geleneksel tahmin modellerine entegre edilebilecek yeni önlemler sunulmuştur.
Çam (2019: 1-108).	Metin madenciliğinde kategorik değişkenler için benzetim katsayılarının kullanılması üzerine bir çalışma	Çalışmada, 4 farklı yazarın 6 farklı eseri üzerinde metin madenciliği teknikleri ve benzerlik-uzaklık ölçüleri kullanılarak R programı yardımıyla yazar tanıma ve metin sınıflandırma başarısı elde edilmeye çalışılmıştır. Yazarların eserleri ile ilgili sayısal veriler çıkartılarak bu sayısal veriler ile arasında Öklid uzaklığı hesaplanmış olup KNN algoritması yardımıyla metin sınıflandırması yapılmıştır. Yazarı belli olmayan bir eserin elde edilen verilerden hareket ile hangi metin kümesine ait olduğu tespit edilmeye çalışılmıştır.
Çeliksü (2017: 1-58).	Yabancı dizilerin alt yazı ve Twitter yorumlarının metin madenciliği ile incelenmesi	Belirlenen yabancı dizilerin Türkçe alt yazı ve Twitter yorumlarını açık kaynak R programı ile metin madenciliği açısından incelenmektedir. Bu çalışmada, metin madenciliğinin 'bag of words' yaklaşımı ele alınmıştır. Müşteri şikâyeti, sosyal medya, görüş, spam olarak değerlendirilebilecek e-postalar veya her türlü dijital ya da dijitalleştirilmiş doküman analiz edilmiştir.

- Dhingra ve Yadav (2017: 1-20). Bulanık Değerlendirme Algoritması ve Hadoop kullanarak spam analizi
- Ergün (2012: 1-100). Metin madenciliği yöntemleri ile ürün yorumlarının otomatik değerlendirilmesi
- Gal-Tzur vd., (2018: 1-9). Soru ve cevap forumlarından ulaşılabilir kararlar için gereken bilgileri çıkarmak için geliştirilmiş bir metodoloji
- Han ve Anderson (2018: 231-238). Çevrimiçi yorumlar kullanılarak sosyal ağın etkisini tahmini etme
- Işık (2019: 1-91). E-ticaret markalarına yönelik sosyal medya yorumlarının metin madenciliği yöntemleri ile analizi
- Karamanlı (2019: 1-65). Metin madenciliği ve Duygu Analizi ile müşteri deneyiminin Makine Öğrenme Algoritmaları kullanılarak geliştirilmesi
- Küçük (2019: 1-87). Metin madenciliği tabanlı bildirim takip sistemi
- Amazon ürün yorumları ve örnek bir yorum seti kullanılarak spam tespitine yönelik çalışma yapılmıştır. Bulanık mantık ile 81 kural oluşturulmuştur. Hadoop ile analiz yapılarak %80,77 doğruluk değerine ulaşılmıştır.
- Seçilen bir ürüne ait tüketici yorumları www.hepsiburada.com adı web sitesinden alınarak morfolojik analiz yapılmıştır. Ön işlemeye hazır sözcük türleri ve aldıkları ekler belirlenmiştir. Negatif veya pozitif anlamlar taşıyan bilgilerin çıkarılması için ürün özelliğini belirten kelimeler ve bunları niteleyen sıfatlar önceden tespit edilmiştir. Bir ağaç yapısı oluşturulmuştur. Oluşturulan bu ağaç yapısı üzerinde Derinlik Öncelikli Arama algoritmasıyla arama yaparak sonuca ulaşan bir yazılım geliştirilmiştir.
- TripAdvisor'daki soru ve cevap formlarından veriler kullanılarak sınıflandırmaya yönelik bir metodoloji önerilmiştir. Seyahat talimatları içeren bütün sorular tanımlanmıştır. Sorularda ulaşılmak istenen noktalar belirlenmiştir. Metin madenciliği teknikleri kullanılarak TripAdvisor'da yayınlanan sorularda belirtilen başlangıç noktasını ve varış yerini otomatik olarak ayıklamanın uygulanabilirliğini de gösterilmiştir.
- Sosyal ağın etkisini ölçmek için TripAdvisor'daki yorumlar kullanılmıştır. Model oluşturulduktan sonra değerlendirme dereceleri ile yorumlarda çıkarılan sayısal değerler arasında pozitif bir ilişki olduğu gözlenmiştir. Mevcut incelemelerin, ilk sayfada olduklarında yeni yorumları güçlü bir şekilde etkilediğini gösterilmiştir.
- Makine Öğrenmesi yaklaşımlarından denetimli öğrenme yaklaşımı kullanılarak sosyal medya yorumlarının Duygu Analizi yapılmıştır. Bazı e-ticaret işletmelerine, ürünlerine/hizmetlerine yönelik yapılan yorumlardan oluşturulan veri kümesi Twitter platformu kullanılarak elde edilmiştir. Algoritmalarından Naive Bayes, Sıralı Minimal Optimizasyon (SMO), KNN algoritmaları kullanılmıştır. En iyi performansı gösteren sınıflandırma algoritması ise, %93,52 sınıflandırma doğruluğu ile KNN algoritması olmuştur.
- E-ticaret sitelerinin birinden en çok satılan üç marka akıllı cep telefonu ile ilgili değerlendirmeler alınıp metin madenciliği ile fikir veya duygu barındırıp barındırmadığı ve olumlu-olumsuz yorumlar olup olmadığı saptanmıştır. Veri seti eğitim ve test veri seti olarak ayrılıp, Makine Öğrenmesi sınıflandırma algoritmalarıyla yapılacak yeni yorumun Duygu Analiz sonucu tahmin edilmiştir. DVM (%74,19), Naive Bayes (%67,52) ve KNN Algoritmaları (%81,45) kullanılmıştır. Pozitif ve negatif yorum tahminlemesi yapılmıştır.
- Pamukkale Üniversitesi Pusula Sistemi F1 Bildirim Takip Sistemine geliştiriciler ve

Murakami (2018: 107-114).	Gezi yerlerinin hedef görüntülerinin üç farklı açıdan karşılaştırılması	kullanıcılar tarafından girilen taleplerin projeleri metin madenciliği tekniklerinden Birliktelik Analizi ile otomatik olarak atanması amaçlanmıştır. Bu bağlamda, Pamukkale Üniversitesi Pusula Sistemi F1 Bildirim Takip Sisteminde kullanılan manuel atama işlemi kaldırılarak, zaman ve iş yükü azaltılmıştır.
Nguyen vd., (2017: 267-279).	Akıllı kültürel turizm hizmetleri için coğrafi etiketli sosyal medyadaki kültürel miras kaynaklarını tanımlama ve sıralama	Japonya'daki varış yerinin görüntüsünü turistler açısından incelemek için bir çalışma yapılmıştır. TripAdvisor'da yayınlanan incelemelerden alınan metin verilerini analiz etmek için veri madenciliği yaklaşımı kullanılmıştır. Japonca ve İngilizce metin verileri dikkate alınarak çalışma hazırlanmıştır. Turistlerin hedef görüntülerinde olmayan, fakat yerel tanıtımcılar tarafından bir öneri olarak sunulan bazı gezi etkinlikleri saptanmıştır.
O'Mahony vd., (2018: 480-509).	Yorumlardan bilgi çıkarma	Sosyal medyadan elde edilen çok miktarda veri ile akıllı turizm uygulamaları ve hizmetleri, kullanıcıların eğilimlerini anlamak için incelenmiştir. Çalışma kültürel turizme odaklanmıştır. Bu çalışmanın temel amacı, coğrafi etiketli sosyal medyadan faydalı kültürel miras kaynaklarını belirlemek ve bunları kullanıcı eğilimine göre sıralamaktır.
Özkul (2019: 1-89).	Akıllı ulaşım sistemleri algısının metin madenciliği yöntemlerinden Duygu Analizi ile ölçülmesi	Twitter, Facebook, Amazon ve TripAdvisor gibi hizmetler, çok sayıda ürün ve hizmet hakkında olumlu ve olumsuz, zengin bir kullanıcı görüşleri kaynağı sağlamaktadır. Ürün görüşlerine yönelik fikirlerin nasıl çıkarılabileceğini ve bir öneri sunmak için verilerin nasıl kullanılabileceğini gösterilmiştir. Bu kapsamda, birtakım vaka çalışması sunulmuştur.
Park (2017: 147-163).	Çevrimiçi tüketicinin etkisini tahmini	Günlük hayatımızda her an karşılaştığımız akıllı ulaşım sistemleri hakkında bireylerin duygu ve düşüncelerinin hangi yönde olduğunu belirlemeye çalışılmıştır. Rapidminer programı kullanılarak tweetlerin Duygu Analizi yapılmıştır. Sonuç olarak bireylerin akıllı ulaşım sistemi uygulamaları üzerinde % 62 oranında olumlu bir görüşe sahip olduğu çıkarılmıştır.
Ren vd., (2019: 333-344).	Çevrimiçi müşterilerin yorumlarını ölçmek için Pathfinder Network Scaling (PENETS) kullanımı	Londra ve New York'ta bulunan yaklaşık 45 restoranın 5090 çevrim içi verisi incelenmiştir. Sonuçlar, kullanıcı derecelendirmeleri ve eğlence arasındaki eğrisel ilişkileri ortaya koymaktadır. Ek olarak, bu araştırmanın bulguları, araştırmacıların, sayım verilerinin özelliklerini yönetmelerine izin veren doğrusal regresyon ve Poisson regresyona göre modellenmiştir.
Rintyarna vd., (2019: 84).	Ürün için yapılan yorumların Duygu Analizi ile değerlendirilmesi	TripAdvisor'dan Kaliforniya'daki dünyanın ilk Disneyland bölgesi hakkında 14.142 etkili inceleme toplamıştır. PENETS ile benzerlik karşılaştırılması yapılarak kullanıcıların farklı segmentlere yönelik ilişkileri ortaya çıkarılmaya çalışılmıştır.
		Cümlelerin özelliklerinin ortaya konması açısından bag of words kullanılmıştır. Duygu Analizi ile cümlelerin polaritesi ortaya koyulmuştur. Duygu Analiz sonucuna ek olarak

- Sharma vd., (2019: 1-11). Duygu Analizi ve TOPSIS yöntemi ile otellerin sıralanması
- Song vd., (2018: 93-105). Farklı ülkelerden turistlerin ihtiyaçlarını keşfetmek için seyahat yorumlarının içerik analizi
- Song vd., (2019: 1-31). Seyahat değerlendirmeleriyle turist memnuniyetini belirleme
- Sönmez (2017: 1-135). İstanbul'daki alışveriş merkezleri için yapılan çevrimiçi yorumların metin madenciliği ile analizi
- Tan (2018: 1-68). Sosyal ağlarda metin madenciliği teknikleri ile bilgi keşfi
- Tran vd., (2019: 393:405). Setiment analizine dayalı olarak otel yorumlarının incelenmesi
- SLF (Sentence Level Features) ve DSF (Domain Sensitive Features) yöntemleri ile cümlelerin anlamlarına yönelik bilgi çıkarılmıştır.
- TripAdvisor web sitesi derecelendirmelerine dayanarak bir otel sıralama modeli önerilmiştir. Hizmet, temizlik, değer, uyku kalitesi, oda ve konumdan oluşan kriterler dikkate alınmıştır. Duygu Analizi ile yorumlar pozitif, negatif ve nötr olarak sınıflandırılmıştır. Ardından sınıflandırmadaki değerler kullanılarak 5 otel TOPSIS yöntemi ile sıralanmıştır.
- Farklı ülkelerden insanların ihtiyacını anlamak için etkili ve hedefli bir tanıtım ve pazarlama stratejisi geliştirmek önemlidir. TripAdvisor'da yer alan İngilizce, Japonca ve Çince yorumlar değerlendirilerek metin madenciliği teknikleri ile analiz yapılmıştır. Anlamsal bir sonuç çıkarılarak veriler yorumlanmıştır.
- Seyahat eden turistlerin yorumları ile bir turizm araştırmasının sonuçlarını değerlendirilmiştir. TripAdvisor'da yayımlanan ve yedi farklı bölgeden gelen turistler tarafından yapılan 1058 yorum incelenmiştir. Çalışmada metin madenciliği teknikleri, Pearson analizi ve Fisher testi kullanılmıştır. Turist memnuniyetine yorumların karşılaştırmalı analizi için Duygu Analizinden faydalanılmıştır.
- En etkili sosyal medya platformlarından biri olan Foursquare üzerinde İstanbul'daki alışveriş merkezlerine bırakılan yorumlar ile veriler analiz edilerek mekân popülerlikleri analizi ve alışveriş merkezleri için metin madenciliği yöntemlerinden biri olan konu modelleme yapılmıştır. Yorumlar Foursquare API kullanılarak bir program aracılığıyla elde edilmiştir. Yorumların yanında toplam check-in sayısı, tekil kullanıcı sayısı, puan bilgileri gibi veriler de alınmıştır. MALLEET yazılımı kullanılarak Konu Modelleme yöntemlerinden biri olan LDA kullanılmıştır. Çalışma sonucunda, 6 konu tespit edilmiştir.
- Karar destek sistemi ile geliştirilen olumlu ve olumsuz değerlendirmeler gerçek zamanlı olarak saptanmakta ve Duygu Analizleri otomatik bir şekilde gerçekleştirilmektedir. Bu bağlamda sosyal medya kullanıcılarının büyük miktardaki yorumu okumadan gitmeyi planladıkları yerler hakkında karar vermelerine olanak sağlanmıştır. Elde edilen olumlu ve olumsuz değerlendirmelere göre işletme sahiplerinin eksikliklerini gidermeleri ve hizmet kalitelerini arttırmaları yönünde adım atmaları amaçlanmıştır. Çalışmada Duygu Analizi ve Doğal Dil İşleme kullanılmıştır. Uygulanan yöntemlerin başarılı sonuçlar verdiği gözlenmiştir.
- Otellerin hizmetleri, tesis imkânları ve pazarlama kampanyalarını değerlendirmek için TripAdvisor yorumları incelenmiştir. Hem sayısal veri hem de müşterilerin yaptıkları yorumlar alınarak bir çerçeve oluşturulmuştur. 11 farklı alanda 405 otel

Tsao vd., (2019: 283-300).	TripAdvisor kullanıcılarının Duygu Analizi ile incelenmesi	Duygu Analizine göre değerlendirilmiştir. Ardından değerlendirilen oteller gruplandırılmıştır. Gözden Geçirme değerleri ile derecelendirme arasındaki ilişki incelenmiştir. Asimetrik bir ilişki var mı? Çalışma, marka gücünü, önemli bir moderatör rolü olarak görmektedir. Çünkü belirli bir ürün için mevcut incelemelerin ortalama derecesi karar vericiler için sezgisel bir ipucudur. Dolayısıyla, bu çalışmanın amacı, yorumun içeriği ile derecelendirme arasındaki asimetrik bir ilişkinin marka gücüne bağlı olacağını iddia etmektir. Çalışmanın sonucunda; ‘Güçlü bir marka için negatif içeriğin sayısal değeri, pozitif içeriğin sayısal değerinden daha büyüktür. Zayıf bir marka için ise, pozitif içeriğin sayısal değeri, negatif içeriğin sayısal değerinden daha büyüktür. ‘ bilgiler elde edilmiştir.
Yousif vd., (2019: 1805-1838).	Duygu Analizi ile bilimsel alıntılar üzerine bir araştırma	Bilimsel alıntılar içindeki duyguları analiz etmeyi amaçlamışlardır. Çalışma akademik kütüphanelerde yayınlanan farklı alanlardaki bilimsel alıntılar üzerine kapsamlı bir araştırma sunmaktadır. Alıntılar üzerine Duygu Analizi ile sınıflandırma yapılarak Makine Öğrenme tekniklerinden DVM ve Bayesian ağları ile yapılan değerlendirmenin performansı ölçülmüştür.

İKİNCİ BÖLÜM

SINIFLANDIRMA VE TAHMİNLEME

Duygu Analizi, bireylerin günlük hayatta yaptıkları yorumların okunmadan belli bir sözlüğe entegre olarak polarite ve polarite değerlerinin çıkarıldığı bir analizdir. Çıkan analiz doğrultusunda ileriye yönelik fikir vermeye yardımcı olması amacıyla sınıflandırma ve tahminleme için kullanılması tasarlanan Destek Vektör Makineleri, Naive Bayes ve Derin Öğrenme Algoritmalarına yer verilecektir. Algoritmaların müşteri hizmetleri, hizmet kalitesi, bilgi kalitesi, bilgi yönetimi, finans, pazarlama, sağlık ve diğer birçok alanlardaki kullanım alanlarına değinilerek literatürde yapılan araştırmalardan bahsedilecektir.

2.1. Veri/Metin Madenciliği ile Sınıflandırma ve Tahminleme

Veri madenciliği, mevcut veritabanlarından anlamlı öngörülerin deşifre edilmesi ve işletme kullanıcıları tarafından tüketilmek üzere sonuçların analiz edilmesi süreci olarak tanımlanabilir. Çeşitli kaynaklardan gelen verileri analiz etmek ve anlamlı bilgiler halinde özetlemek, istatistiksel bilgi keşfinin bir parçasıdır. Bu durum sadece işletme kullanıcılarına değil, aynı zamanda istatistiksel analistler, danışmanlar ve veri bilimcileri gibi birden fazla topluluğa da yardımcı olmaktadır. Çoğu zaman, veritabanlarından bilgi keşif süreci beklenmedik sonuçlar çıkarabilir ve bu sonuçlar birçok farklı şekilde yorumlanabilir. Artan sayıda cihaz, tablet, akıllı telefon, bilgisayar, sensör ve diğer çeşitli dijital cihazlar, veri üretmek ve her zamankinden daha hızlı bir şekilde veri toplamaya yardımcı olmaktadır. Günümüz bilgisayarlarının kabiliyetiyle, artan veriler, herhangi bir işin karar verme süreciyle ilgili çeşitli soruları cevaplamak için önceden işlenebilir ve modellenabilir. Veri madenciliği, istatistiksel metodolojiler, makine öğrenme teknikleri, görselleştirme ve örüntü tanıma teknolojileri kullanılarak ayrık veri tabanları ve bilgi havuzları arasında bilgi yoğun bir arama olarak ortaya çıkmaktadır (Makhabel vd., 2017: 8).

Bilgi yoğun bir arama olan veri madenciliği teknikleri problemler için karar desteği sağlayabileceği gibi metin madenciliği teknikleri de karar desteği sağlamak adına iyi ve öngörülebilir interaktif yazılımlar kullanmaktadır. Bunun yanı sıra, metin madenciliği gelecekteki eğilimlerin başarılı bir tahmin yapılması konusunda önemli

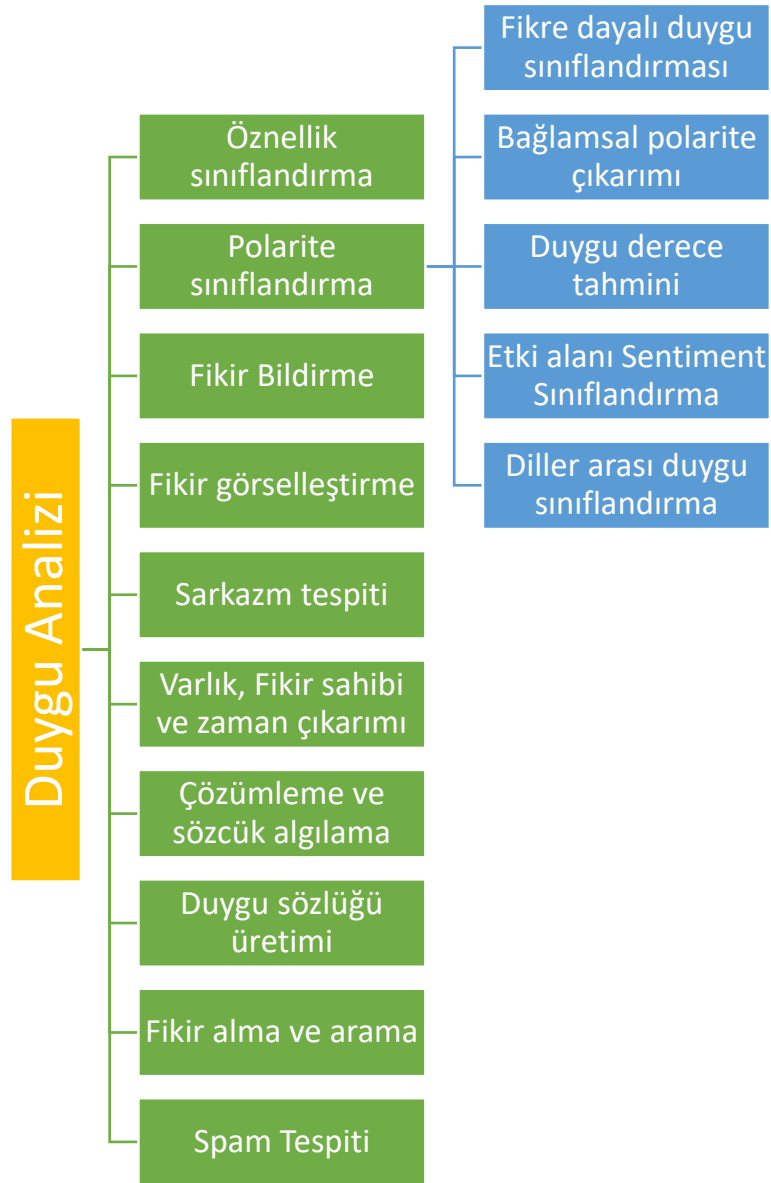
adımlar atmaktadır. Şöyle ki, metin madenciliği teknikleri bilgiyi daha verimli kullanarak, pazarda meydana gelen değişiklikleri, pazardaki ürünün düşük veya yüksek performansı ile ilgili önleyici kararlar alarak iş zekâsı ile birlikte koordine çalışmaya olanak tanımaktadır. Bu bağlamda, metin madenciliği, işletme yönetiminin optimum kararlar verebilmesinde satış, istihdam memnuniyeti, pazarlama stratejisi, rekabet yarışında üstünlük sağlayabilmesi gibi birçok açıdan avantajlar sağlayabilmesi için tasarlanmıştır (Melek, 2012: 33). Veri/metin madenciliğinde en yaygın kullanılan karar verme tekniklerinin bazıları şunlardır: Duygu Analizi, DVM, Naive Bayes Algoritmaları ve Derin Öğrenme.

2.1.1. Duygu Analizi

Fikir madenciliği olarak da adlandırılan Duygu Analizi, kişilerin yazılı metinlerde ifade edilen varlıklara yönelik fikirleri, duyguları, değerlendirmeleri, tutumları ve duyguları analiz eden çalışma alanıdır. Varlıklar; ürünler, hizmetler, kuruluşlar, bireyler, etkinlikler, sorunlar veya konular olabilir. Bu alan, geniş bir problem alanını temsil etmektedir. Fikir madenciliği, fikir analizi, fikir çıkarma, öznellik analizi, etki analizi ve inceleme madenciliği artık Duygu Analizi çatısı altındadır. Duygu Analizi terimi ilk olarak Nasukawa ve Yi'de (2003), fikir madenciliği terimi önce Dave ve ark. (2003) tarafından ortaya atıldığı bilinmektedir (Liu, 2015: 1-3). Nasukawa ve Yi çalışmasında, Duygu Analizini duygu ifadeleri, duygu ifadelerinin kutupluğu ve gücü, duygu ifadelerinin konu ile ilişkisinin olduğunu ortaya koymuşlardır. Örneğin, "XXX YYY'yi yener" cümlesinde, "yener" ifadesi XXX'e karşı olumlu bir duygu ve YYY'ye karşı olumsuz bir duygu anlamına gelir. Bu örnek çerçevesinde Duygu Analizinin temeli atılmıştır (Nasukawa ve Yi, 2003: 71). Dave ve ark. Çalışmasında ise, yakınlık ölçüleri ve kelime listelerini kullanarak verileri şablonlara sığdırmak ve kullanıcı görüşlerini modellemek için ortaya koymuşlardır (Dave vd., 2003: 520). Ancak, duygu ve fikir madenciliğine yönelik araştırmalar daha önce başlamıştır. Yıllar geçtikçe, web'deki sosyal medya sistemleri yeni katılımcılarla birlikte birçok platform oluşturmakta ve gelişmektedir. YouTube, Facebook ve Twitter başta olmak üzere birçok farklı blog ve forumlara kullanıcılar tarafından görüşler yazılmaktadır. Duygu Analizi bu görüşlerin nasıl işlenip yarar sağlayacağı konusunda fikir vermektedir (Liu, 2015: 1-3).

Duygu Analizi, 2000'in başından beri NLP'de en aktif araştırma alanlarından biri olmuştur. Duygu Analizinin amacı, bir karar destek sistemi veya bir karar verici

tarafından kullanılacak yapısal ve eyleme geçirilebilir bilgi oluşturmaktır. Analiz, doğal dilde oluşturulan metinlerden öznel bilgi çıkarabilen otomatik araçlar tanımlamaktadır. Bununla birlikte, Duygu Analizi, metinlerden pozitif, negatif veya nötr duyguları (polarite olarak da adlandırılır) ortaya çıkarmaktadır. En popüler Duygu Analizi görevlerinin sınıflandırılması Şekil 5’de verilmektedir (Pozzi vd., 2016: 1-4):



Şekil 5. Duygu Analizinin Başlıca Görevleri

Şekil 5’deki Duygu Analizinde öznellik sınıflandırması ilk adımlardan birisi olarak bilinmektedir. Öznellik sınıflandırması; tam metinleri, cümleleri ve hatta daha kısa metinsel belgeleri öznel ve nesnel olarak sınıflandırmaktadır (Belisário vd., 2020: 261). Fikir bildirme, Duygu Analizde bir metnin özetinin oluşturulmasını anlamına

gelmektedir. Yani, belirlenen görüşlerin kısa bir şekilde kullanıcıya sunulmasını ifade etmektedir (Hu ve Liu, 2004: 168). Fikir görselleştirme, dilbilim ve NLP'deki teorik araştırmalardan sosyal medyaya, haber izlemeye kadar değişen Duygu Analizi araştırmalarında ve uygulamalarında var olan çeşitli veri alanlarını ve kullanıcı görevlerini görsel olarak temsil etmektedir (Kucher vd., 2018: 73). Sarkazm, küçümsemeyi ifade etmek için ironiyi kullanan dilbilimsel bir araçtır. Bir tartışma forumunda sarkazmı tespit etmek söz konusu olduğunda, yalnızca önceki yorumların bağlamını anlamayı değil, aynı zamanda tartışma konusu hakkında gerekli arka plan bilgisine de ihtiyaç duymaktadır. Argo sözcüklerin ve gayri resmi dilin kullanılması, sözcüksel ipuçlarına olan güveni de azaltmaktadır. Bu özel sarkazm türünü tespit etmek güçtür (Hazarika vd., 2018: 1837). Varlık, fikir sahibi ve zaman çıkarımı NER'in en klasik sorunlarından biridir. NER'in temel amacı, kelimeleri yer adı, kişi adı, tarih, saat gibi belirli kategorilerde sınıflandırmaktır. Bu sınıflandırmanın ana yaklaşımı istatistiksel ve matematiksel kurallara dayanmaktadır (Zaggia, 2016: 38). Çözümleme ve sözcük algılama, Doğal Dil İşleme'nin önemli görevlerinden biridir. Metnin anlamını doğru bir şekilde anlama amacına ulaşmak için metindeki kelimelerin belirsizliğini giderme ile ilgilenirler. Metnin anlamındaki belirsizliği gidermeden farklı olarak, uygun kelime listesi ekler ve hedef kelimelerin anlamlarını dikkate almaktadır (Wu vd., 2018: 6223). Duygu sözlüğü üretimi; Sosyal medyadaki Duygu Analizi uygulamalarının çoğu, makine öğrenimi yaklaşımı yerine sözlüğe dayalı yaklaşıma dayanmaktadır. Sözlük üretimi için insan tarafından etiketlenmiş büyük miktarda eğitim verisine ihtiyaç duyulmaktadır. Sözcük üretiminde birçok alan, belirli terimler ve jargonlar gibi kendi alt dilleriyle karakterize edilerek oluşturulmaktadır (Tai ve Kao, 2013: 53). Fikir alma ve arama, iki tipik sorgu türü üzerine kurulmuştur. Birincisi, belirli bir kuruluş veya kuruluşun bir yönü hakkında kamuoyu görüşlerini bulmak. Örneğin; bir dijital kamera veya kameranın resim kalitesi hakkında müşteri görüşlerini bulmak. İkincisi ise, bir kişi veya fikir sahibi bir kuruluşun bir yönü hakkında görüşlerini bulmak. Örneğin; Barack Obama'nın kürtaj hakkındaki görüşünü bulmak. Bu iki örnek çerçevesinde fikir alma ve arama analizi özetlenebilir (Liu, 2012: 118). Spam tespiti, daha çok istenmeyen e-posta algılama problemleri için kullanılmaktadır. Spam tespitinde kullanıcı ürün inceleme aşında bir ağ sınıflandırma görevi formüle edilir. Bu görevde, kullanıcı ürün ağına düşen veriler sahte veya gerçek olarak sınıflandırılır (Rayana ve Akoglu, 2015: 986). Son olarak polarite sınıflandırması, daha gelişmiş analizlerin bir alt görevi olarak görülebilir.

Örneğin, bir ürünün değerlendirmelerini etkileyen ve bu tür kararları daha güvenilir kılan artıları ve eksileri değerlendirmek için bireysel incelemelerde kullanılacak ‘pro’ (lehte) ve ‘con’ (alehte) ifadelerinin tanımlanmasında uygulanabilir (Cambria vd., 2017: 4).

Bir ürünü satın alma veya bir filmi izleme kararı almak için, bir kullanıcının çevrim içi olarak başkalarının bıraktığı binlerce fikri manuel olarak okuması neredeyse imkânsızdır. Ya yapay zekâ, arzu ettiğimiz ürün hakkında genel düşünceyi bize vermemize yardımcı olursa ne olur? Duygu Analizi, fikirleri analiz eden ve kullanıcıların duyarlılığını çıkaran yapay zekâ yöntemidir. Duygu Analizi, polarite sınıflandırması yardımıyla bir ürünün iyi mi-kötü mü veya pozitif mi-negatif mi olduğunu, farklı özelliklerinin neler olduğunu, ürünü nereden satın almanız gerektiğini söyleyebilir (Poria vd., 2018: 15).

Duygu Analizi skoru belirlenmeden önce cümle içerisinde geçen duygu ifadeleri, duygu taşıyan sözcükler, duygu değişimleri bir dizi ön işlemden geçirilmektedir. Tespit edilebilecek aksan ve özel karakterler kaldırılmaktadır. Ardından duygu değişimine neden olabilecek kelime türleri belirlenmektedir. Duygu taşıyan kelimelerin ortalama duygu skorunu hesaplamak için sözcük tabanlı bir duygu skoru ile derecelendirme yapılmaktadır. Bu doğrultuda, bir ‘d’ belgesindeki ‘ s_i ’ bölümündeki ‘ i ’. duygu skoru, ‘ s_i ’ bölümündeki her bir ‘ e_{ij} ’ duygu ifadesinin duygu skor fonksiyonu ile hesaplanabilir. Ya da ‘ s_i ’ bölümündeki her bir duygu taşıyan ‘ w_{ij} ’ ve o duyguyu değiştiren ‘ m_{ij} ’ (eğer bu değişken metinde bulunduğu da varsayılan olarak ‘1’ alınır) ile duygu skoru hesaplanabilir. Eşitlik 1’de bir metin bölümündeki duygu skoru formüle edilmiştir.

$$sent(s_i) = \begin{cases} \frac{\sum_{j=1}^{v_i} sent(e_{ij})}{v_i} & \text{if } v_i > 0, \\ \frac{\sum_{j=1}^{t_i} (sent(w_{ij}).sent(m_{ij}))}{t_i} & \text{else (varsa)}, \end{cases} \quad (1)$$

Burada; ‘ v_i ’: metin bölümündeki duyguyu ifade eden görsel (emoji) ipucuların sayısını; ‘ t_i ’: metin bölümündeki duygu taşıyan metinsel ipucu sayısını göstermektedir.

Her bir metin bölümü tarafından iletilen duygu skoru belirlendikten sonra, tüm metin bölümleri tek bir belgede yeniden birleştirilir. Dokümanın duygu skoru daha sonra tüm bölüm seviyesinde duygu skorlarının ağırlıklı ortalaması alınarak hesaplanmaktadır. Burada ağırlıklar ilgili bölümdeki duygu taşıyan kelime veya ifadelerin sayısının nispi

oranlarına karşılık gelmektedir. Böylelikle bir ‘d’ dökümanı için duygu skoru Eşitlik (2)’deki gibi hesaplanmaktadır.

$$sent(d) = \frac{\sum_{i=1}^p (sent(s_i) \cdot (v_i + (a_i \cdot t_i)))}{\sum_{i=1}^p (v_i + (a_i \cdot t_i))} \quad (2)$$

Burada; ‘p’: d belgesindeki bölüm sayısını; ‘a_i’: s_i metin bölümünün ipucuları üzerinde tam bir Duygu Analizi yapılmasının gerekli olup olmadığını belirten bir Boolean değişkenini (0 veya 1) ifade etmektedir. Eşitlik (3), Boolean değişkenini göstermektedir.

$$a_i = \begin{cases} 0 & \text{if } v_i > 0, \\ 1 & \text{else.} \end{cases} \quad (3)$$

Böylece belgenin duygu skoru elde edilmektedir. Negatif bir puan tipik olarak negatif bir belgeyi ‘-1’ gösterirken, diğer puanlar ‘1’ pozitif bir sınıflandırma sağlamaktadır. Belgenin pozitif veya negatif sınıflandırma formülasyonu Eşitlik (4)’deki gibidir (Hogenboom vd., 2013: 706-707).

$$class(d) = \begin{cases} 1 & \text{if } sent(d) \geq 0, \\ -1 & \text{else.} \end{cases} \quad (4)$$

Özetle Duygu Analizi, metin madenciliğinde güncel bir araştırma alanını oluşturmaktadır. Doğal Dil İşleme veya makine öğrenme yöntemlerinin köküdür. Karar vermenin önemli kaynakları, çevrim içi yorumlardan çıkarılabilir, tanımlanabilir ve değerlendirilebilir (Vadivukarassi vd., 2017: 54). Tablo 3’de Duygu Analizinin avantajları ve dezavantajları verilmiştir (Dashtipour vd., 2016: 769).

Tablo 3. Duygu Analizi Avantajları ve Dezavantajları

Avantajlar	Dezavantajlar
Uygulaması çok basit	Özellik seçimi etkisiz
Diğer dillere kolayca evrilebilir	Hesaplama yapmak zor olabilir
Hem küçük hem de büyük veri kümeleri için kullanışlıdır	Uyumluluk dereceleri hesaplamak yoğun işlem gerektirebilir
Borsa değerlerinin Duygu Analizi için otomatik olarak bir sözlük oluşturur	Borsa değerleri dışında diğer alanlar için eksik uygulanması güç olabilir
Birden fazla dil için kullanılabilir	Çok dilli Duygu Analizi için kaynak bulunmayabilir

Literatürde Duygu Analizi ile ilgili önde gelen çalışmaların bazıları aşağıdaki gibidir:

Kharde ve Sonawane (2016) tarafından yapılan çalışmada, Twitter verilerinin Duygu Analizi için bir teknik araştırma ele alınmıştır. Bu çalışma esas olarak, görüşlerin oldukça yapılandırılmamış, heterojen olduğu ve bazı durumlarda pozitif, negatif veya nötr olduğu tweetlerdeki bilgileri analiz etmeye yardımcı olan Twitter verilerinin Duygu

Analizine odaklanmaktadır. Çalışma sonucunda, Naive Bayes, Maksimum Entropi ve DVM gibi çeşitli makine öğrenme algoritmaları kullanılarak, Twitter veri akışları hakkında performans karşılaştırması yapılmıştır (Kharde ve Sonawane, 2016: 5-15).

Medhat vd. (2014) tarafından yapılan çalışmada, Duygu Analizi algoritmaları ve uygulamaları ele alınmıştır. Bu araştırma makalesi, bu alandaki son güncellemeyle ilgili kapsamlı bir genel bakışa odaklanmıştır. Yakın zamanda önerilen algoritmaların geliştirmeleri ve çeşitli Duygu Analizi uygulamaları bu çalışmada kısaca araştırılmış ve sunulmuştur. Çalışmada, son zamanlarda yayınlanmış ve alıntı yapılan makalelerin 53'ü kategorize edilmiş ve özetlenmiştir. Ayrıca, Naive Bayes DVM'nin, duygu sınıflandırma problemini çözmek için en sık kullanılan makine öğrenme algoritmaları olduğu belirtilmiştir (Medhat vd., 2014: 1093-1113).

Ortigosa vd. (2014) tarafından yapılan çalışmada, Facebook'da Duygu Analizi ve e-öğrenme uygulaması incelenmiştir. Kullanıcıların yazdıkları mesajlarda iletiği şekliyle duygu polaritesi (pozitif, nötr veya negatif) hakkında bilgi elde etmek ve kullanıcıların olağan duygu durumlarını modelleyerek önemli duygusal değişiklikleri tespit etmek çalışmanın hedefini oluşturmaktadır. Çalışmada, bir Facebook uygulaması olan SentBuk kullanılmıştır. SentBuk, Facebook'ta kullanıcılar tarafından yazılan mesajları alır ve bunları etkileşimli bir arayüz aracılığıyla kullanıcılara göstererek kutuplarına göre sınıflandırmaktadır. Yöntem, duygusal değişiklik tespiti, arkadaşınızın duygu bulmasını, mesajlara göre kullanıcı sınıflandırmasını ve istatistiklerini de desteklemektedir. SentBuk'ta uygulanan sınıflandırma yöntemi hibrit bir yaklaşımı izlemektedir. Çalışmanın sonucunda kullanılan yöntemin doğruluğunun tespiti için sözcük tabanlı yaklaşım, Naive Bayes, DVM ve Karar Ağacı gibi Makine Öğrenme Algoritmaları kullanılmıştır. Elde edilen doğruluk oranlarının tatmin edici sonuçlar verdiği görülmüştür (Ortigosa vd., 2014: 527-541).

Boiy ve Moens (2009) tarafından yapılan çalışmada, çok dilli web metinlerinde Duygu Analizi için makine öğrenimi yaklaşımı ele alınmıştır. Web'de bulunan İngilizce, Felemenkçe ve Fransızca yazılan blog, inceleme ve forum metinlerinde Duygu Analizi ile ilgili Makine Öğrenme araştırması sunulmuştur. İnsanların belirli tüketim ürünleri ile ilgili ifade ettikleri duygularla ilgilenilmiştir. Ürünlere yönelik metinler pozitif, negatif ve nötr olarak sınıflandırılmasının yanı sıra her bil dil için doğrulukları eğitilip test edilmiştir. DVM, Naive Bayes ve Maksimum Entropi algoritmaları ile İngilizce (%83)

metinlerdeki doğruluk oranının Felemenkçe (%70) ve Fransızca (%68) metinlerde ulaşılan doğruluk oranından yüksek olduğu görülmüştür. Yeterli eğitim verisi ilave edilmesi durumunda ve nötr duyguların filtrelenmesiyle doğruluk oranının her bir dil için artacağı iddia edilmiştir (Boiy ve Moens, 2009: 526-558).

Kuhamanee vd. (2017) tarafından yapılan çalışmada, çevrim içi sosyal ağlar üzerinden veri madenciliği yaklaşımını kullanarak Tayland, Bangkok'a yabancı turistlerin duyarlılığını analiz etmektedir. Amaç, Bangkok turizm endüstrisini geliştirmek ve teşvik etmek için yabancı turistlerin duyarlılığı hakkında bilgi edinmektir. Bu çalışma için 2017 yılında Twitter'dan 10.000 veri seti almıştır. Bu veri setinden, İngilizce dili kullanılarak yazılmış belgeler alınmış ve ön işlemden geçirilmiştir. Daha sonra, Bangkok'u ziyaret etmek için gelen kişiler seyahat, iş, aile ziyareti, eğitim ve sağlık, tedavi olarak beş kategoride sınıflandırılmıştır. Ardından metindeki duygulara yönelik etiketler çıkartıldıktan sonra Karar Ağacı, Naive Bayes, DVM ve YSA ile test edilip analiz edilmiştir. RapidMiner uygulaması ile yapılan analiz sonucunda en yüksek doğruluğa sahip algoritmaların YSA (%80,33) ve DVM (%80, 11) olduğu görülmüştür (Kuhamanee vd., 2017: 1068-1073).

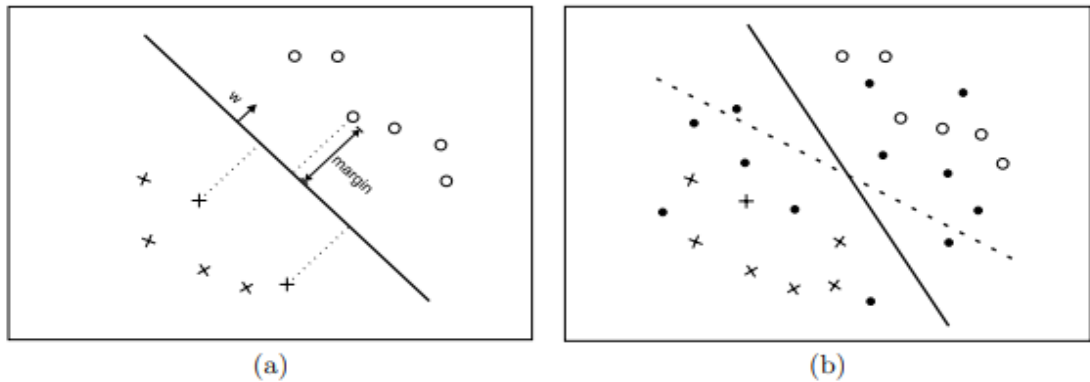
2.1.2. Destek Vektör Makineleri

DVM, 1992'de Vapnik ve arkadaşları tarafından hem regresyon hem de örüntü tanıma için uygulanan yeni bir evrensel öğrenme makinesi önerilmiştir. DVM, girdi alanındaki verileri, sorunun doğrusal olarak ayrılabilir olduğu yüksek boyutlu bir özellik alanına eşlemek için çekirdek (kernel) eşleme adı verilen bir aygıt kullanmaktadır. Bir DVM'nin karar fonksiyonu sadece destek vektörlerin sayısı ve ağırlıkları ile değil, destek vektörü kernel olarak adlandırılan "apriori" ile de ilgilidir (Zhang vd., 2004: 34). DVM temel bir Makine Öğrenme teknolojisidir. Güçlü teorik temelleri ve mükemmel ampirik başarılarla sahiptir. El yazısı karakterler, basamak tanımlama, nesne tanıma ve metin sınıflandırma gibi görevlere uygulanmıştır (Tong ve Chang, 2001: 108).

DVM, istatistiksel öğrenme teorisine dayanır ve örüntüleri tanıma yeteneğine sahiptir. DVM, yapısal risk minimizasyonu çerçevesinde ve Vapnik-Chervonenkis (VC) sınırları teorisinde yeni öğrenme tekniklerini temsil etmektedir. DVM, sınıflandırma, regresyon, zaman serisi tahmini ve fonksiyon tahmin problemlerini çözmek için çeşitli uygulamalarda başarıyla kullanılmaktadır. DVM ile yapılan sınıflandırma ve regresyon performansı diğer sınıflandırma ve regresyon tekniklerinden daha iyi olduğu

görülmektedir. DVM'nin uygulanması, popüler bir DVM sınıflandırma ve regresyon yazılımı olan LIBSVM yazılımı kullanılarak gerçekleştirilmektedir (Terzic, 2013: 4-5).

DVM ile yapılan bir uygulamadaki veriler daha yüksek boyutları ile eşleştirilerek optimum bir ayırma köprüsü oluşturulur. Bu temelde sinir ağı mimarileri için birçok yerel minimum varlığı tanımada muzdarip iken DVM, kuadratik programlama problemlerinin çözülmesini mümkün kılmaktadır. İşlevler ve parametreler bir boyuttaki bağlantı sayısını minimize edecek şekilde seçilmektedir. Ardından yöntem, fonksiyonun tahmin problemini çözmek için genişletilmektedir. Bu amaçla, Vapnik'in epsilon duyarsız kayıp fonksiyonu ve Huber'ın kayıp fonksiyonu kullanılmaktadır. Doğrusal durumun yanı sıra, polinomlara, spline'lara, radyal tabanlı fonksiyon ağlarına ve çok katmanlı algılayıcılara dayanan DVM'ler başarıyla uygulanmaktadır (Suykens ve Vandewalle, 1999: 293). DVM, iki boyutlu bir uzayda a ve b olarak Şekil 6'de görüldüğü gibidir. Koyu olan daireler etiketlenmemiş örnekleri temsil etmektedir (Tong ve Koller, 2001: 47).



Şekil 6. a) Basit doğrusal DVM; b) Kesikli ve Kesiksiz DVM

Şekil 6'da ilk olarak hem etiketli hem de etiketlenmemiş bir veri kümesi verilmektedir. Öğrenme görevi, etiketlenmemiş verilere mümkün olduğunca doğru bir şekilde etiket atamaktır. DVM'ler, hem etiketli hem de etiketlenmemiş verilere göre marjı maksimize eden hiper düzlemi bularak transdüksiyon gerçekleştirebilir. Şekil 6b'de görüldüğünde son zamanlarda, Transdüktif Destek Vektör Makineleri (TDVM) metin sınıflandırması için kullanılmış ve hassasiyette bazı iyileştirmeler elde edilmiştir (Joachims, 1998: 138).

DVM, verileri yüksek boyutlu uzaya taşıyarak sınıflandırma yapmak için tasarlanmıştır. Daha sonra tahmin yapabilmek için yöntem genişletilmiştir. DVM, birçok Makine Öğrenmenin tersine yapısal ve deneysel hataları azaltmak için test

örneklerindeki gösterdiği başarı oranı, eğitim örneklerinde gösterdiği başarı oranına yakındır (Erişti vd., 2007: 3; Uçar, 2006: E2-1).

$$d(x) = w^T \varphi(x) + b \quad (5)$$

Burada, $w \in R^n$ ve $b \in R$ parametreleri çok boyutlu düzlemi temsil etmektedir. $\varphi(x)$ ise verileri yüksek boyutlu uzaya taşımak için kullanılan bir dönüşümü ifade etmektedir. Eşitlik (6), verilen birincil optimum problemin eğitim hatasını ve Eşitlik (7) ise, problemin genelleme hatasını minimize etmeye yardımcı olmaktadır.

$$L(w, \xi_i) = \frac{C}{L} \sum_{i=1}^L \xi_i + \frac{1}{2} \|w\|^2 \quad (6)$$

$$y_i [w^T \varphi(x_i) + b] \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (7)$$

Burada, L eğitim örneklerinin sayısını, C eğitim ve genelleme hatası arasındaki ödünleşimi ifade etmektedir. ξ_i ise $w^T \varphi(x) + b$ ve y_i arasındaki mutlak hatayı nitelemektedir. Eşitlik (6)'da yer alan ikinci parametrenin minimize edilmesi, öz nitelik uzayında karşıt sınıfın iki en yakın verisi arasındaki mesafe olarak bilinen kısmın maksimize edilmesine karşılık gelmektedir. Verileri genellikle yüksek boyutlu uzaya dönüştüren $\varphi(x)$ bilinmediği durumlarda, işlemlerde kolaylık sağlaması amacıyla bir çekirdek (kernel) yardımıyla örtük olarak yüksek boyutlu uzaya taşınmaktadır. Bu doğrultuda, birincil optimum problemin, Lagrange çarpanları yardımı ile ikincil forma dönüştürülmektedir. Eşitlik (8) ve (9)'da gösterildiği gibidir.

$$L_{ikincil}(\lambda) = -\frac{1}{2} \sum_{i,j=1}^L \lambda_i \lambda_j y_i y_j K(x_i, x_j) + \sum_{i=1}^L \lambda_i \quad (8)$$

Kısıtlar:

$$\sum_{i=1}^L y_i \lambda_i = 0, 0 \leq \lambda_i \leq \frac{C}{L} \quad i = 1, 2, \dots, L. \quad (9)$$

Burada, λ_i Lagrange çarpanlarını ve $K(x_i, x_j)$ çekirdeği $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ ile hesaplanan iç çarpımı ifade etmektedir. Sonuç olarak, karar verilmesi aşamasında çekirdek ve Lagrange çarpanları yardımıyla hesaplama yapılmaktadır. İlgili hesaplama Eşitlik (10) gösterilmektedir.

$$d(x) = \text{sing}(\sum_{\text{destek vektörler}} y_i \lambda_i K(x_i, x_j) + b) \quad (10)$$

Burada, destek vektörler $\lambda_i \geq 0$ Lagrange çarpanlarına karşılık gelen x_i değerlerini ifade etmektedir (Erişti vd., 2007: 3).

Sonuç olarak, bu yöntem yardımıyla binlerce örnekten oluşan veri setlerinden oluşan DVM optimizasyon problemini çözmek oldukça kolaydır. Ampirik olarak, en son teknoloji DVM öğrenme algoritmalarının çalışma süreleri, yaklaşık olarak kuadratik olarak ölçeklenmektedir. Yani DVM'ye iki kat daha fazla veri verdiğinizde, çalışması dört kat daha uzun sürecektir. DVM'ler yaklaşık bir milyon örnek içeren veri kümeleri üzerinde başarılı bir şekilde eğitilebilir. (Noble, 2006: 1567). Tablo 4'de DVM algoritmasının avantajları ve dezavantajları verilmiştir (Singh ve Husain, 2014: 16).

Tablo 4. DVM Algoritmasının Avantajları ve Dezavantajları

Avantajlar	Dezavantajlar
Deneysel sonuçlarda çok iyi performans	DVM'nin bir dezavantajı, kategorik veya eksik değer olması durumunda, önceden işlenmesi gerekesidir
Veri seti boyutluluğuna düşük bağımlılık	Ortaya çıkan modelin zor yorumlanması

Literatürde DVM ile ilgili önde gelen çalışmaların bazıları aşağıdaki gibidir:

Zainuddin ve Selamat (2014) tarafından yapılan çalışmada, DVM ve Duygu Analizi ele alınmıştır. Duygu Analizi, bir metnin yönünü pozitif veya negatif olarak sınıflandırdığı için bir sınıflandırma görevi olarak ele alınmaktadır. Bu çalışma, bir duygu sınıflandırıcısını eğitmek için DVM uygulayan deneysel sonuçları açıklamaktadır. Deneysel analiz, Chi-Square özellik seçimini kullanarak sınıflandırma doğruluğunda önemli bir iyileşme sağlayabildiğini ortaya koymaktadır. Çalışma, Pang ve Taboada Corpus (film, kitap, araba, telefon gibi araçlarla ilgili yorumlar) veri kümesini kullanmıştır. DVM algoritması ile eğitilip test edilen her bir veri kümesine ilişkin başarı %80'in üstünde olduğu gözlemlenmiştir (Zainuddin ve Selamat, 2014: 333-337).

Zheng ve Ye (2009) tarafından yapılan çalışmada, DVM kullanılarak Çin gezginlerinin yorumlarına yönelik sınıflandırma yapılmıştır. İnternette, kullanıcı tarafından oluşturulan yorumların nasıl analiz edileceği ve farklı duygu sınıflarında nasıl sınıflandırılacağı bir sorun haline gelmektedir. Gezgin yorumları bu sorunun büyük bir parçasını oluşturmaktadır. Çalışmada, Çin'de bir seyahat acentesi olan 'ctrip' internet adresinden veriler derlenmiştir. Duygu Analizi ile pozitif ve negatif olarak etiketlenen veriler DVM ile eğitilip test edilmiştir. Test sonucunda çince (% 91,15) ve ingilizce yorumlar (%85,14) için başarı oranları verilmiştir (Zheng ve Ye, 2009: 335-338).

Povoda vd. (2016) tarafından yapılan çalışmada, DVM ve büyük verilere dayalı Duygu Analizi yapılmıştır. Metin belgelerine dayanan bu çalışmada birden fazla web sayfasında yer alan farklı ürün segmentinde geri bildirimler otomatik olarak elde

edilmiştir. İngilizce, Almanca, Çekçe ve İspanyolca olmak üzere farklı dillerdeki metin verileri derlenmiştir. İlgili metinler Duygu Analizi ile pozitif ve negatif olarak etiketlenmiştir. Ardından Makine Öğrenme Algoritmalarından DVM ile 7000 örnek eğitilip test edilmiştir. Her bir dil için test sonucunda elde edilen başarı %86'nın üstünde olduğu görülmüştür (Povoda vd., 2016: 543-545).

BholaneSavita ve Gore (2016) tarafından yapılan çalışmada, DVM kullanarak Twitter verilerinde Duygu Analizi incelenmiştir. Çalışmada, E-ticaretin en önemli kısımlarından biri Duygu Analizi olarak gösterilmektedir. Milyonlarca kullanıcı tarafından en çok ziyaret edilen sosyal ağ sitelerinden biri Twitter'dır. Burada siyaset, markalar, ürünler ve ünlüler ile ilgili çeşitli konular dikkat çekmektedir. Bu çalışmada, tweetler kullanılarak duygu, LDA (Latent Dirichlet Allocation) ve DVM analizleri yapılmıştır. Duygu Analizinden çıkarılan pozitif ve negatif etiketlerin yanı sıra, LDA ile Duygu Analizdeki değişimlerin arkasındaki nedenler araştırılmıştır. Son olarak, DVM analiz ile Duygu Analizinden çıkan etiketler eğitilip test edilmiştir. Test sonucunda başarı % 97,54 olarak bulunmuştur (BholaneSavita ve Gore, 2016: 365-370).

Mullen ve Collier (2004) tarafından yapılan çalışmada, farklı bilgi kaynaklarına sahip DVM kullanarak Duygu Analizi yapılmıştır. Çalışmada, Epinions.com'un film inceleme verileri derlenmiştir. İncelemede, zarflar ve sıfatlar için çeşitli uygunluk ifadeleri ve metne konu olan farklı bilgiler dâhil edilmiştir. Analizde DVM ile geçmişte etkili olduğu gösterilen unigram modeller ve unigram modellerin lemmatize versiyonları birleştirilmiştir. Yapılan araştırmalar, unigram özellik tabanlı DVM'lerin, gerçek değerlere sahip olumlu ölçümlere dayanan hibrit DVM'lerden üstün performans gösterdiği gözlemlenmiştir. Çalışmanın sonucunda elde edilen başarı %80'in üstünde olduğu vurgulanmıştır. Konu için kararlaştırılmış daha küçük bir müzik incelemeleri veri kümesi de analiz edilerek çalışmaya eklenmiştir (Mullen ve Collier, 2004: 412-418).

2.1.3. Naive Bayes Algoritması

Naive Bayes Algoritması, veri boyutunun yüksek olduğu, kriterlerin bağımsız olduğu veya hesaplamaları basitleştirmek adına kriterlerin bağımsız varsayıldığı, diğer yöntemlere kıyasla daha verimli sonuçlar beklenildiği, büyük veritabanlarına uygulandığında yüksek doğruluk ve hız sağlanması istenildiği durumlarda tercih edilmektedir (Cherian ve Bindu, 2017: 70). Bir sınıflandırma yöntemi olarak Naive Bayes, düşük sapma ile çalışabilen, yeni eğitim verileri elde edildikçe hızlı

güncellenebilen, veri kümesindeki gürültüyü işleme yeteneği yüksek, veri kümesindeki eksik değerleri işleme yeteneği olan diğer sınıflandırma yöntemleri ile kıyaslandığında yüksek hesaplama verimliliği gibi özelliklerle atfedilmektedir. Ayrıca, Naive Bayes uygulamasında gerekli ayarlama parametreleri veya etki alanı bilgisi bulunmamaktadır. Naive Bayes'in en büyük dezavantajı sadece özelliklerin bağımsızlığı varsayımında yatmaktadır. Buna rağmen, yöntem genellikle rekabetçi sınıflandırma doğruluğu sağlar ve pratikte özellikle kıyaslama sonuçları olarak yaygın olarak uygulanmaktadır (Taha, 2013: 3).

Bayes Teoremi, belirli koşullu olasılıkların hesaplanmasına izin veren olasılık tabanlı bir sınıflandırma yöntemidir. Koşullu olasılıklar, bir olayın başka bir olayın olasılığı üzerindeki etkisini yansıtan olasılıklardır. Bayes teoreminde genellikle kullanılan terim, gerçek olasılık ve koşullu olasılıktır. Bir hipotez veya olayın önceki olasılığı, herhangi bir ek bilgiye ihtiyaç duyulmadığında elde ediliyorsa gerçek olasılıktır. Koşullu olasılık ise, elde edilen bazı ek bilgiler veya kanıtlar kullanılarak hipotezin gözden geçirilmiş olasılığıdır. Bayes teoremi Eşitlik (11)'deki gibi yazılabilir:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (11)$$

Burada: $P(A)$, A olayının gerçekleşme olasılığı; $P(B)$, B olayının gerçekleşme olasılığı; $P(A|B)$, B olayı gerçekleştiğinde A olayının gerçekleşme olasılığı; $P(B|A)$, A olayı gerçekleştiğinde B olayının gerçekleşme olasılığını göstermektedir. A hipotezi doğru veya yanlış olabileceğinden, Bayes teoremi Eşitlik (12)'deki gibi yazılabilir (Sharma vd., 2015: 705-706):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)} * P(A) + P(B - A) * P(-A) \quad (12)$$

Naive Bayes sınıflandırıcıları, ister sürekli ister kategorik olsun, rastgele sayıda bağımsız değişkeni işleyebilir. $X = (x_1, x_2, x_3, \dots, x_d)$ veri seti verildiğinde, bir dizi olası sonuç olan $C = (c_1, c_2, c_3, \dots, c_d)$ veri seti arasından C_j koşullu olasılığını oluşturmak istenebilir. Daha açık bir ifade ile X , belirleyicidir ve C , bağımlı değişkende bulunan kategorik düzeyler kümesidir. Bu doğrultuda Bayes kuralı Eşitlik (13)'deki gibi oluşturulabilir.

$$p(C_j|x_1, x_2, \dots, x_d) \propto p(x_1, x_2, \dots, x_d|C_j)p(C_j) \quad (13)$$

Burada; $p(C_j|x_1, x_2, \dots, x_d)$ sınıf üyeliğinin koşullu olasılığıdır. Bir başka deyişle, X 'in C_j 'ye ait olma olasılığıdır. Naive Bayes, bağımsız değişkenlerin koşullu olasılıklarının istatistiksel olarak bağımsız olduğunu varsaydığından, bir terimde yer alan ürününün olasılığını Eşitlik (14) ve koşullu olasılığını Eşitlik (15)'deki gibi ayrıştırabiliriz:

$$p(X|C_j) \propto \prod_{k=1}^d p(x_k|C_j) \quad (14)$$

$$p(C_j|X) \propto p(C_j) \prod_{k=1}^d p(x_k|C_j) \quad (15)$$

Yukarıdaki Bayes kuralını kullanarak, en yüksek koşullu olasılığa ulaşan C_j sınıfında yeni bir X durumu etiketlenebilir (Tripathi vd., 2015: 669).

Buna ek olarak Naive Bayes, metin sınıflandırmalarında iyi çalışma eğiliminde olan ve genellikle DVM gibi modellere kıyasla eğitilmesi için daha az zaman alan çok basit bir olasılıksal modeldir. Naive Bayes sınıflandırma, güçlü bağımsızlık varsayımı ile birlikte Bayes kuralına dayanmaktadır. Naive Bayes modeli, koşullu bir bağımsızlık varsayımı içermektedir. Bu varsayım, metin sınıflandırmasındaki doğruluğu çok fazla etkilememektedir. Ancak problem için hızlı sınıflandırma algoritmalarını uygulanabilir kılmaktadır. Naive Bayes'in metin sınıflandırma üzerindeki performansı Eşitlik (16)'de gösterilmektedir (Narayanan vd. 2013: 195).

$$P(x_i|c) = \frac{\text{Count of } x_i \text{ in documents of class } c}{\text{Total no of words in documents of class } c} \quad (16)$$

Burada: *Count of x_i in documents of class c* : c sınıfı belgelerde x_i sayısı; *Total no of words in documents of class c* : c sınıfı belgelerdeki toplam kelime sayısını göstermektedir. Tablo 5'de Naive Bayes Algoritmasının avantaj ve dezavantajları verilmiştir (Ang vd., 2016: 2010).

Tablo 5. Naive Bayes Algoritmasının Avantaj ve Dezavantajları

Avantajlar	Dezavantajlar
Farklı boyutlardaki eğitim veri setlerini sınıflandırmada ve iyi sınıflandırma sonuçları üretmede son derecede kararlıdır	Kriterlerin bağımsızlığı varsayımı altında eğitim veri setlerinin nedensel ilişkilerini temsil etmek pratik değildir.
Yapısı basittir	
Sınıflandırma sonuçlarını üretmede çok hızlıdır	

Literatürde Naive Bayes ile ilgili önde gelen çalışmaların bazıları aşağıdaki gibidir:

Liu vd. (2013) tarafından yapılan çalışmada, Naive Bayes sınıflandırıcısı kullanılarak büyük veri analizi için ölçeklenebilir duygu sınıflandırması ele alınmıştır. Duygu Analizinde, bir mesajdan duygu veya düşüncüyü çıkarılmaktadır. Makine öğrenim teknolojileri, karar vermeyi göreceli olarak yüksek doğrulukla tahmin etmek veya desteklemek için eğitim veri kümesinden “öğrenme” yetenekleri nedeniyle duygu sınıflandırmasında yaygın olarak kullanılmaktadır. Ancak, veri kümesi büyük olduğunda, bazı algoritmalar iyi ölçeklenemeyebilir. Bu çalışmada Amazon film veri setindeki yorumlar Naive Bayes Algoritması ile değerlendirilmiştir. Yorumlar Duygu Analizi yardımıyla pozitif ve negatif olarak etiketlenmiştir. Çalışmanın sonucunda, etiketlenen veri kümesi boyutu arttığında Naive Bayes Algoritmasının % 82 doğruluk ile sınıflandırma yapabildiği gösterilmiştir (Liu vd., 2013: 99-104).

Go vd. (2009) tarafından çalışmada, Twitter mesajlarının duyarlılığını otomatik olarak sınıflandırmak için yeni bir yaklaşım sunulmuştur. Ürünlerin satın almadan önce duyarlılığını araştırmak isteyen tüketiciler veya markalarının kamuoyunu izlemek isteyen şirketlerin faydalanması için tweetler pozitif ve negatif olarak sınıflandırılmıştır. Sınıflandırmanın doğruluğunu ölçmek için Makine Öğrenme Algoritmalarından Naive Bayes, DVM ve Maximum Entropy kullanılmıştır. Eğitim verilerinde, gürültülü etiket olarak kullanılan Twitter ifadeleri içeren mesajlar olmasına rağmen algoritmaların %80'nin üzerinde başarı gösterdiği tespit edilmiştir. Buna ek olarak çalışma, yüksek başarı elde etmek için gereken bir Makine Öğrenme Algoritmasının yapması gereken ön işleme adımlarını açıklamıştır (Go, 2009).

Ye vd. (2009) tarafından çalışmada, denetimli makine öğrenimi yaklaşımları ile seyahat yerleri için yapılan çevrim içi yorumlar incelenmiştir. Bu yorumlar, bloglar, Wiki veya web siteleri gibi farklı biçimlerde görünebilir. Bu yorumlardaki bilgiler, hem gezginler hem de yöneticiler için çeşitli anlayış ve planlama süreçleri için değerlidir. Bununla birlikte sitelerde, kullanıcılar mevcut tüm bilgileri okuyamadıklarından aşırı bilgi yüklemesi söz konusudur. Yahoo ve Google gibi arama motorlarındaki sorgu işlevleri kullanıcılara ihtiyaç duydukları bilgilerin fazlasını sunmaktadır. Bu bağlamda çalışma, ABD ve Avrupa'daki yedi popüler seyahat destinasyonu için seyahat bloglarındaki yorumların duygu sınıflandırması yapılmıştır. Naive Bayes, DVM ve karakter tabanlı N-gram modelinin üç denetimli Makine Öğrenme Algoritması ile duygu sınıflandırması test edilip algoritmalar karşılaştırılmıştır. Çalışmanın sonucunda, üç

yaklaşımın sınıflandırma başarıları % 80'nin üstünde olduğu görülmüştür (Ye vd. 2009: 6527-6535).

Harisinghaney vd. (2014) tarafından çalışmada, KNN, Naive Bayes ve Reverse DBSCAN algoritmaları kullanılarak metin ve görüntü tabanlı spam e-posta sınıflandırması yapılmıştır. E-postalar, kısa mesajlar ve çevrim içi mesajlaşma, hayatımızın bir parçasını oluşturmaktadır. Tüm bu iletişimlerden, e-postalar daha fazla dikkat çekmektedir. Bu nedenle, çeşitli e-posta sağlayıcıları e-postaları, spam ve jambonlara göre filtrelemek için algoritmalar kullanmaktadır. Bu çalışmada, Enron veri kümesindeki e-postalar ve spamlar ele alınmıştır. KNN (%83), Naive Bayes (%87) ve Reverse DBSCAN (% 74) algoritmalarının doğruluk oranlarının tatmin edici olduğu görülmüştür (Harisinghaney vd., 2014: 153-155).

An vd. (2017) tarafından yapılan çalışmada, şarkı sözlerine dayalı duygu sınıflandırması için Naive Bayes Algoritması tercih edilmiştir. Müzik kaynaklarının etkin yönetimi, insanların algılarını yansıtan duyguların değerlendirilmesi açısından önemlidir. Çalışmada, Baidu müzik adlı popüler bir web sitesi üzerinden dört farklı veri seti şeklinde toplam 3316 şarkı sözü derlenerek dört farklı sınıflandırma etiketi kullanılmıştır. Metin sınıflandırma yöntemlerinden en iyi algoritmalarından birisi olan Naive Bayes Algoritması ile veriler eğitilip test edilmiştir. Çalışmanın sonucunda her bir sınıflandırma ile elde edilen başarının %60'nın üstünde olduğu görülmüştür (An vd., 2017: 635-638).

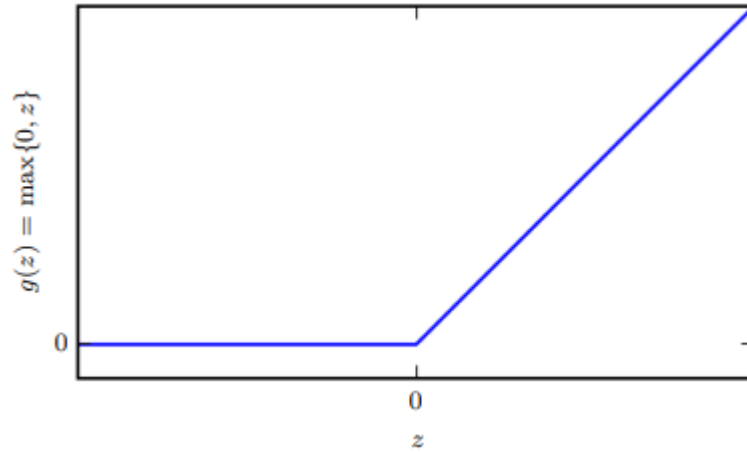
2.1.4. Derin Öğrenme

Derin Öğrenme, bilgisayarların deneyimden öğrenmesini ve dünyayı kavramlar hiyerarşisi açısından anlamasını sağlayan bir makine öğrenim şeklidir. Bilgisayar deneyimlerden bilgi topladığından, bir bilgisayar kullanıcısı operatörünün bilgisayar için gerekli olan tüm bilgileri resmi olarak belirtmesine gerek yoktur. Kavramlar hiyerarşisi, bilgisayarın karmaşık kavramları daha basit olanlardan yapılandırarak öğrenmesini sağlamaktadır. Bu hiyerarşilerin bir grafiği birçok katman derinliğindedir. Genellikle sinir ağları veya çok katmanlı algılayıcılar (MLP'ler) olarak da isimlendirilen derin ileri beslemeli ağlar, tipik derin öğrenme modelleridir. İleri beslemeli ağlar makine öğrenimi uygulayıcıları için son derece önemlidir. Birçok önemli ticari uygulamanın temelini oluşturmaktadır. Endüstride yer alana uzmanlar başta derin ileri beslemeli ağlar olmak üzere optimizasyon algoritmaları, evrimsel (convolutional) ağlar, modelleme, Doğal Dil İşleme, konuşma tanıma, çevrim içi öneri sistemleri, biyoinformatik, robotik, arama

motorları, çevrim içi reklamcılık ve video oyunları gibi uygulamaları inceleyip ilgili problemlere uygulamaktadırlar (Kim, 2016: 351-352).

Derin Öğrenme, çoklu işlem katmanlarından oluşan hesaplama modellerinin, çoklu soyutlama seviyesine sahip verilerin temsillerini öğrenmesine olanak tanımaktadır. Bu modeller görsel nesne tanıma, konuşma tanıma, nesne algılama ve ilaç keşfi gibi diğer birçok alandaki son teknolojiye uyarlanıp geliştirilmektedir. Derin Öğrenme, bir makinenin her katmandaki temsili yapıyı hesaplamak için kullanılan iç parametrelerin önceki katmandaki temsilindeki yapının nasıl değiştirmesi gerektiğini belirtmek için geri yayılım algoritmasını kullanarak büyük veri kümelerindeki karmaşık yapıyı keşfetmektedir. Derin evrimsel ağlar video, görüntü, konuşma ve sesin işlenmesinde önemli roller oynarken, tekrarlanan (recurrent) ağlar, metin ve konuşma gibi sıralı verilere ışık tutmaktadır (LeCun vd., 2015: 436).

İleri beslemeli ağlar, iletişim ağı veya ağ olarak adlandırılmaktadır. Çünkü tipik olarak birçok farklı fonksiyonun bir araya getirilmesiyle oluşmaktadır. Model, fonksiyonların nasıl bir araya getirildiğini açıklayan yönlendirilmiş bir döngüsüz (çevrimsiz) grafikte ilişkilidir. Örneğin, $f(x) = f^{(3)}\left(f^{(2)}\left(f^{(1)}(x)\right)\right)$ oluşturmak için bir zincire bağlı üç $f^{(1)}$, $f^{(2)}$ ve $f^{(3)}$ fonksiyonuna sahip olduğunu düşünelim. Bu zincir yapıları sinir ağlarının en sık kullanılan yapılarıdır. Bu durumda, $f^{(1)}$ ağın ilk katmanı olarak adlandırılır, $f^{(2)}$ ikinci katman olarak adlandırılır ve bu böyle devam etmektedir. Zincirin toplam uzunluğu modelin derinliğini vermektedir. Bu terminolojiden, “Derin Öğrenme” adı doğmaktadır. İleri beslemeli bir ağın son katmanına çıktı katmanı denmektedir. İleri beslemeli sinir ağlarının çoğunda kullanılması tavsiye edilen etkinleştirme işlevi ‘Rectified doğrusal aktivasyon fonksiyonu’dur. Tavsiye edilen fonksiyon $g(z) = \max\{0, z\}$ olarak tanımlanır ve Şekil 7’deki gibi gösterilmektedir (Goodfellow vd., 2016:168-175):



Şekil 7. Rectified Doğrusal Aktivasyon Fonksiyonu (Goodfellow vd., 2016:175).

Rectified Derin Öğrenme Ağları görüntü tanıma ve metin sınıflandırma görevlerinde çok daha iyi performans göstermektedir. Rectified fonksiyonunun matematiksel modeli Eşitlik (17)'de verilmektedir.

$$h^{(i)} = \max(w^{(i)T}x, 0) = \begin{cases} w^{(i)T}x & w^{(i)T}x > 0 \\ 0 & \text{else} \end{cases} \quad (17)$$

Rectified fonksiyonu 0'ın üzerinde etkinleştirildiğinde, kısmi türevi 1'dir. Bu doğrultuda, Rectified fonksiyonu tam olarak 0 ile doyurulmaktadır. Böylelikle fonksiyon, sınıflandırma yapılırken gizli katmana bağlanacak her giriş katmandaki kriterlere yardımcı olmaktadır (Maas vd., 2013: 3-4). Son katmanda ise, Derin Öğrenmenin sınıflandırma problemlerine yönelik çözüm için 'softmax' fonksiyonu kullanılır. Softmax fonksiyonu $\sum_{k=1}^K p_k$ olarak adlandırılan K sınıfları için ayrı bir olasılık dağılımı belirtir. Bir sinir ağının sondan bir önceki katmanında 'x' gibi bir aktivasyon ve softmax katmanındaki ağırlık parametrelerini ' θ ' alındığında, softmax katmanının girişi Eşitlik 18'deki gibi gösterilebilir.

$$c = \sum_i^{n-1} \theta_i x_i \quad (18)$$

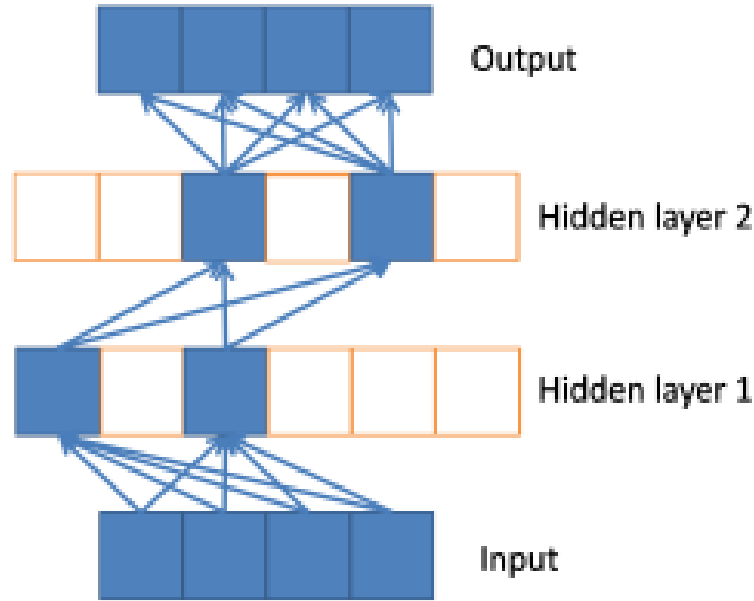
Tahmin edilecek sınıf için son olarak Eşitlik (19) ve Eşitlik (20) kullanılarak aşağıdaki gibi formüle edilebilir (Agarap, 2018):

$$p_k = \frac{e^{c_k}}{\sum_{k=0}^{n-1} e^{c_k}} \quad (19)$$

$$\hat{y} = \arg \max p_i \quad i \in 1, \dots, N \quad (20)$$

Rectified etkinleştirme fonksiyonu, bir ağın kolayca aralıklı (sparse) gösterimler elde etmesini sağlamaktadır. Örneğin, giriş katmanındaki ağırlıkların homojen olarak

verilmesinden sonra, gizli katmandaki birimlerin yaklaşık % 50'si sürekli çıktı katmandaki birimlere gönderilerek sıfırlamaktadır. Ardından bu katman (fraksiyon), aralıklı olarak (sparse) sonuç çıkartan düzenleme ile kolayca artırılabilir. Şekil 8'de Rectified ağındaki aktivasyonların ve düğümlerin aralıklı olarak yayılışı gösterilmektedir.



Şekil 8. Rectified Yayılm Ağı

Şekil 8'de gösterildiği gibi, ağıdaki tek doğrusallık, aktif olan veya olmayan tek tek nöronlarla ilişkili yol seçimi verilmektedir. Belirli bir girdi için sadece bir nöron alt kümesi aktiftir. Girdi katmanı aktif nöronların bir alt kümesini seçer ve bu alt kümede hesaplama doğrusaldır. Çıktı katmanı, girdi katmanının doğrusal bir işlevidir. Bu doğrusallık nedeniyle, gradyanlar nöronların aktif yollarında hızlıca ilerler (sigmoid veya tanh birimlerinin aktivasyon doğrusallıklarından dolayı gradyan yok etme etkisi yoktur) ve matematiksel hesaplama daha kolaydır (Glorot vd., 2011: 318). Tablo 6'da Derin Öğrenme Algoritmasının avantajları ve dezavantajları verilmiştir (Li, 2017: 25).

Tablo 6. Derin Öğrenme Algoritmasının Avantajları ve Dezavantajları

Avantajlar	Dezavantajlar
Örüntü tanıma problemlerinde iyi	Çıkarım yapmak ve karar vermek zor olabilir.
Veriye dayalı ve birçok problemde performans yüksektir	Semboller doğrudan işleyemez
Sistem yapımında çok az alan bilgisi gereklidir veya hiç bilgiye ihtiyaç yoktur	Büyük veriye ihtiyaç duymaktadır ve bu nedenle veri boyutu küçüldüğünde performansı düşebilir
Veriyi öğrenme ve işlemler arası geçiş mükemmeldir	Model genellikle bir kara kutudur ve anlaşılması zordur
Öğrenme algoritması basittir	Hesaplamalı öğrenmenin maliyeti yüksektir

Esas olarak denetimli öğrenme yöntemi desteklemektedir	Denetimsiz öğrenme yöntemlerinin geliştirilmesi gerekiyor
--	---

Literatürde Derin Öğrenme ile ilgili önde gelen çalışmaların bazıları aşağıdaki gibidir:

Zhang vd. (2018) tarafından yapılan çalışmada, Derin Öğrenmeye genel bir bakış açısı sunarak Duygu Analizindeki mevcut uygulamalar için bir araştırma yapılmıştır. Yapılan araştırmanın sonucunda, Derin Öğrenme Algoritmalarının birçoğu, Duygu Analizi için en gelişmiş sonuçları verdiği gözlemlenmiştir. Yakın gelecekte Derin Öğrenme konusundaki ilerlemelerle Duygu Analizi için heyecan verici çalışmalar verilebileceği öngörülmüştür (Zhang vd., 2018: e1253).

Rojas-Barahona (2016) tarafından yapılan çalışmada, Duygu Analizinden elde edilen etiketlerin Derin Öğrenme Algoritmaları ile test edilmesi işlenmiştir. Sosyal ağların ortaya çıkması, bloglara, önerilere ve incelemelere erişme olasılığını açmıştır. Buradaki zorluk, fikir madenciliği veya Duygu Analizi yardımıyla verilerden polariteyi çıkarmaktır. Bu analizin doğasında spesifik zorluklar, özel yorumlama ve kelimelerin kutupsallığından kaynaklı dilsel fenomenlerle ilgili konular bulunmaktadır. Son zamanlarda, Derin Öğrenme Algoritmalarıyla Duygu Analizi sonuçlarına yönelik etiketler test edilip sınıflandırılmaktadır. Bu çalışmada da, 10662 film yorumunun pozitif ve negatif etiketleri çıkarılarak Derin Öğrenme algoritmaları ile %80 başarı elde edilmiştir (Rojas-Barahona, 2016: 701-719).

Li ve Chen (2014) tarafından yapılan çalışmada, Duygu Analizi tabanlı Derin Öğrenme kullanarak kayıt dışı ekonomide en çok satan satıcıları belirlemek amaçlanmıştır. Kayıt dışı ekonomi, siber suç ekosisteminde önemli bir bileşendir. Çünkü siber suçunun tüm aşamalarını kolaylaştıran kötü amaçlı araçlar ve hizmetlerin alışverişinde bulunmaları için kullanıcılara bir pazar sunmaktadır. Bu nedenle, karaborsa satıcıları siber güvenlik araştırmacıları ve uygulayıcıları için özel bir öneme sahiptir. Bu çalışmada, rus bir kayıt dışı forumu ve kartopu örnekleme kullanılarak analiz edilmiştir. Derin Öğrenme Algoritmasının yanı sıra DVM, Naive Bayes, Maksimum Entropy ve KNN de kullanılmıştır. Tüm algoritmaların başarısı %89 üstünde olduğu görülmüştür (Li ve Chen, 2014: 64-67).

Shirani-Mehr (2014) tarafından yapılan çalışmada, film incelemeleri kullanılarak Duygu Analizine Derin Öğrenme Algoritmaları ve Naive Bayes Algoritması

uygulanmıştır. Duygu Analizi, NLP'deki temel zorluklardan biridir. Son zamanlarda, Derin Öğrenme uygulamaları farklı NLP görevlerinde etkileyici sonuçlar göstermiştir. Bu çalışmada, Stanford Duygu Treebank'ı ana veri seti kullanılmıştır. Veri seti 11855 yorumu barındırmaktadır. Kullanılan algoritmalar ile veri seti eğitilip test edilmiştir. Elde edilen sonuçların başarı ve hata oranları karşılaştırılmıştır (Shirani-Mehr, 2014: 1-8).

Day ve Lin (2017) tarafından yapılan çalışmada, Google Play tüketici incelemesinde Duygu Analizi için Derin Öğrenme ele alınmıştır. Sosyal medyadaki görüş kutupluluğunu anlamak için tüketici incelemelerinde Duygu Analizine olan ilgi artmıştır. Çince Google Play tüketici incelemeleri ele alınmıştır. Google Play'de 196.651 yorumu toplamak için bir web madenciliği tekniği uygulanmıştır. Çalışmanın sonucunda Duygu Analizi için DVM, Naive Bayes ve bir Derin Öğrenme Algoritması kullanılarak başarı oranları karşılaştırılmıştır. En iyi performansı Derin Öğrenmenin Algoritmasının verdiği görülmüştür (Day ve Lin, 2017: 382-388).

2.2. Veri ve Metin Madenciliği Uygulama Alanları

Bu bölümde veri ve metin madenciliğinin genel olarak hangi alanlarla ilişkili olduğu ortaya koyulmuştur. Bu alanlar ilgili ilk olarak müşteri ilişkileri, hizmet kalitesi, bilgi yönetimine değinilmiş ve son olarak veri ve metin madenciliğinin diğer uygulamalara yönelik çalışma alanlarından bahsedilmiştir.

2.2.1. Müşteri İlişkileri

Müşteri ilişkileri yönetimi (CRM), bir işletmenin müşterilerini anlamaya çalışan insan, süreç ve teknolojinin birleşimidir. Müşteriyi elde tutma ve ilişki geliştirmeye odaklanarak ilişkileri yönetmeye yönelik entegre bir yaklaşımdır. CRM, bilgi teknolojisindeki ilerlemelerden ve müşteri odaklı süreçlerdeki örgütsel değişikliklerden evrimleşmiştir. CRM'yi başarıyla entegre eden işletmeler, müşteri sadakati ve uzun vadede kârlılık ödülleri alacaktır. Başarılı bir CRM uygulamasını yönetmek, teknoloji, süreç, entegre ve dengeli bir yaklaşım gerektirir. Bu sebeple CRM, tüm müşteri temas noktalarından alınan müşteri verilerini edinen, dikkate değer müşterilerin profillerini açığa çıkarırken ve satın alma modellerini tahmin ederken bir müşterinin tek ve kapsamlı bir görünümünü oluşturan daha karmaşık bir uygulamadır. Müşteri davranışını gözlemleyen ve analiz eden teknoloji, işletmelerin en iyi müşterileri daha kolay belirlemelerini, pazarlama çabalarına odaklanmalarını ve sık sık satın alma olasılığı yüksek olanları ödüllendirmelerini sağlamaktadır (Chen ve Popovich, 2003: 676).

Müşteri ilişkileri yönetiminde veri/metin madenciliğinin en büyük amacı, müşterilere etkili servisi sağlamak için gerekli bilgiyi edinmektir. Bu amaca ulaşmak için çeşitli teknikler (birliktelik kuralları, kümeleme algoritmaları ve sınıflandırma yöntemleri) kullanılmaktadır. Tekniklerin başarılı bir şekilde uygulanması büyük avantajlar sunmaktadır. Özellikle web’de nümerik bilginin yanı sıra daha fazla metin verisi bulunmaktadır. Metin verileri, yapılandırılmamış veriler olarak nitelendirilmektedir. Son zamanlarda, metin madenciliği teknikleri ile yapılandırılmamış veriler CRM’de daha başarılı sonuçlar elde etmek için kullanılmaktadır. CRM, veri analiz sonuçları, kampanya yönetimi, gerçek zamanlı karar destek sistemi ve bütünleşik yönetimin etkili bir şekilde yürütülmesi ile dizayn edilmelidir (Ahn vd., 2003: 329-330).

Metodoloji olarak değerlendirilen metin madenciliği, metin içerdiği sürece her türlü dijital belge için geçerli evrensel bir kavramdır. Metin madenciliğinin en popüler uygulamaları, World Wide Web (WWW) yoluyla erişilebilen veya web’in kendisinde bulunan büyük bir metin belgesi grubu olarak kabul edilen internet belge havuzlarıyla ilgilidir. Çoğunlukla müşteriler (veya iş ortakları) ve çalışanlar arasındaki iletişimin e-posta mesajları, telefon görüşmesi transkriptlerini içeren büyük miktarda dokümantasyon CRM çözümlerine destek veren metin veritabanlarını ifade etmektedir. Tipik bir CRM sisteminde depolanan belge sayısı orta ölçekli bir şirket için bile manuel analiz için genellikle çok fazladır. Müşterinin tutumu hakkındaki önemli bilgilerin bu depodan çıkarılabilmesi ve metin madenciliği teknikleri ile işlenmesi bu sistem için önemli bir katkı sağlayacaktır (Gawrysiak vd., 2004: 869-870).

2.2.2. Hizmet Kalitesi

Kalite, "müşterinin gereksinimlerini karşılamak" veya "amaca uygunluk" olarak tanımlanmaktadır. Deming, Feigenbaum, Ishikawa, Juran ve arkadaşları tarafından ortaya konan kalite tanımları bu kategoriye girmektedir. Bu yaklaşım, müşterilerin gereksinimlerini belirleme ve ardından bu gereksinimleri karşılama yeteneğine dayanmaktadır. Bunun nedeni, müşterilerin gereksinimlerinin tasarım aşamasında hizmete dâhil edilmesidir, ancak dönüşüm aşamasında uyumluluk derecesi belirlenmektedir. Bu çerçevede, hizmet kalitesi rekabetçiliğin kritik bir belirleyicisi olarak kabul edilmektedir. “Hizmet kalitesine” dikkat edilmesi, bir kuruluşun kendisini diğer kuruluşlardan ayırt etmesine yardımcı olabilir ve bu sayede kalıcı bir rekabet avantajı elde edilebilir. Hizmet Kalitesi; eğitim, sağlık, sosyal güvenlik, polis, ordu,

ulaştırma, hayır kurumları, kiliseler, araştırma vakıfları, kamu hizmetleri, oteller, hava yolları, mimarlar, avukatlar, restoranlar, perakendeciler, eğlence kurumları, bankalar, sigorta şirketleri, reklam ajansları, danışmanlık firmaları, pazar araştırma şirketleri, iletişim şirketleri gibi birçok alanı kapsamaktadır (Ghobadian vd., 1994: 48).

1984'ten bugüne kadar hizmet kalitesi modelindeki gelişmelerin, hizmetlerin geleneksel yapıdan bilişim tabanlı hizmetlere doğru yol aldığı gözlenmektedir. Bilişim teknolojilerinin müşteriye anlama ve tanımaya yönelik güvenilirlik, erişim, iletişim, yeterlilik, yanıt verme gibi belirli hizmet kalitesi boyutlarını iyileştirmek için kullanılmaktadır. Bazı vaka çalışmaları yoluyla bilişim teknolojileri müşterilere yönelik bilgileri toplama, işlemleri izleme ve servisi kolaylaştırma gibi faaliyetleri de yürütmektedir (Seth vd., 2005: 945).

Bilgi teknolojilerin gelişimi ile birlikte, internetin benzeri görülmemiş ve artan yaygınlığı, hem tüketiciler hem de işletmeler için birincil bilgi kaynağı haline gelen çeşitli sosyal medya platformlarında (Facebook, Twitter, YouTube, TripAdvisor, Yelp, Amazon, Travelocity vb.) kullanıcı tarafından oluşturulan içeriğin hızla büyümesini sağlamıştır. Özellikle, internette yayınlanan ürün ve hizmetler için kullanıcı tarafından oluşturulan incelemeler, dijital müşteri-müşteri etkileşimleri için en güçlü bilgi paylaşım aracı haline gelmektedir. İşletmeler, bilgileri kendilerinin ve rakiplerinin ürün ve performanslarının geri bildirimlerini ve değerlendirmelerini toplamak, tüketici davranışlarını ve beklentilerini daha iyi anlamak ve pazar fırsatlarını tanımlamak için kullanmanın önemini giderek daha fazla kabul etmektedir. Kullanıcılar tarafından oluşturulan veriler ile hizmet kalitesi anlamlandırılabilir. Mevcut literatürde hizmet kalitesi ve performansı kullanıcı incelemelerinin metinsel içeriğini incelemek ve bunları farklı boyutlara sınıflandırmak için Duygu Analizi tekniğini kullanmaktadır. Mevcut metin madenciliği çalışmalarında Duygu Analizi sonuçları, hizmet kalitesi boyutlarının yakalanmasında ve ölçülmesinde yüksek düzeyde başarı getirmektedir (Duan vd., 2013: 3119-3120).

2.2.3. Bilgi Yönetimi

Bilgi, klasik Yunan döneminden bu yana batı felsefesinde epistemolojik tartışmayı tanımlayan geniş ve soyut bir kavramdır. Bununla birlikte, son birkaç yılda, bilgiyi önemli bir organizasyonel kaynak olarak ele almak konusunda artan bir ilgi bulunmaktadır. Bilgi yönetim sistemlerinin amacı, kuruluşlarda bilginin oluşturulmasını, aktarılmasını ve uygulanmasını desteklemektir. Bilgi ve bilgi yönetimi karmaşık ve çok yönlü kavramlardır. Bu nedenle, bilgi yönetim sistemlerinin etkili bir şekilde geliştirilmesi ve uygulanması, temel bir gerekliliktir. Bilgi, metin, grafik, kelimeler veya diğer sembolik formlar şeklinde sunulmaktadır. Bireylerin aynı veri veya bilgi anlayışına ulaşması için belirli bir bilgi tabanını paylaşmaları gerekmektedir. Bilgi yönetim sistemleri, kurumsal bilginin yönetimine uygulanan bir bilgi sistemleri sınıfını ifade etmektedir. Yani, bilgi oluşturma, depolama / alma, aktarma ve uygulama gibi organizasyonel süreçleri desteklemektedir. Bilişim teknolojileri, bilgi yönetimini desteklemektedir. Örneğin; çevrim içi dizinleri kullanarak veya bir veritabanında kayıtlı bir bilgi kaynağı bulmak, sanal ekiplerde birlikte çalışma, geçmiş projeler hakkındaki bilgilere erişim ve işlem verilerini analiz ederek müşteri ihtiyaçları ve davranışları hakkında bilgi edinme gibi birçok örnek verilebilir (Alavi ve Leidner, 2001: 114).

İnternet ve WWW gelişimi ile birlikte bilgi yönetim sistemleri sürekli faaliyetlerini kısa mesafede birçok kişiye ulaştırabilmektedir. Bu doğrultuda bilgi, teknolojinin birkaç kombinasyonu ve permütasyonu yapay zekâ araçları, akıllı veri ambarları, genetik algoritma araçları, sinir ağları, uzman sistemler, vaka tabanlı akıl yürütme uygulamaları, kural tabanlı ve akıllı ajanlar üzerine kuruludur. Bu teknolojilerden hangilerinin bilgi yönetim sistemine uyduğunu ve nasıl entegre edilebileceğini anlamak için, bilgi yönetimi bağlamındaki rolleri anlamak önemlidir (Tiwana, 2000: 182).

Bilgi yönetimi sistemleri, veri tabanındaki bilgileri yapılandırılmış veya yapılandırılmamış kayıtlar biçiminde depolar ve içeriği kullanıcının sorgusunu tatmin eden tüm kayıtları elde etmek için sorgulama yöntemleri sağlamaktadır. Veri madenciliği olarak da bilinen veri tabanlarındaki bilgi keşfi, büyük miktarda verinin bilgisayar yardımıyla araştırılmasına ve bunların içindeki ilginç modellerin keşfedilmesine odaklanmaktadır. Veri tabanlarındaki bilgi keşfi üzerindeki çoğu çalışma yapılandırılmış veri tabanları ile ilgili olsa da, yalnızca yapılandırılmamış metin biçiminde mevcut olan

büyük miktarda bilgi bulunmaktadır. Metin madenciliği de, yapılandırılmamış veriler ile ilgilenmektedir. Metin madenciliğine yönelik yaklaşımların çoğu, bir metin belgesi koleksiyonu verildiğinde, her bir belgeyle ilişkili etiketler üzerinde bilgi bulma işlemleri gerçekleştirmektedir. Bu alandaki çalışmalar, veritabanlarındaki kalıpların otomatik olarak keşfedilmesine yönelik Makine Öğrenme ve istatistiksel analiz tekniklerinin uygulanmasını ve verilerin araştırılması için kullanıcı rehberli ortamların sağlanmasına yardımcı olmaktadır (Feldman vd., 1998: 1-2).

2.2.4. Finans, Pazarlama, Sağlık ve Diğer Uygulamalar

Metin madenciliği finans, pazarlama, sağlık, biyoloji, güvenlik, internet, akademik uygulamalar gibi birçok alana uygulanmaktadır. Bunların bir kısmı, aşağıda görüldüğü üzere örneklendirilmiştir.

Metin madenciliği ile bir yazarın el yazısını tanımlamak, yapay zekâ ve örüntü tanıma alanlarında oldukça zorlu problemlerden biridir. Geleneksel olarak, el yazısı tanımlama sistemleri bir veri toplama, veri ön işleme, özellik çıkarma ve sınıflandırma dizisini takip etmektedir. El yazısı tanımlaması ile ilgili araştırma iki kategoriye odaklanmıştır: çevrim dışı ve çevrim içi. Çevrim dışı el yazısı olan materyaller, yalnızca taranmış görüntü bilgileri içerdikleri için daha genel ancak tanımlanması daha zor olarak kabul edilmektedir. Buna karşılık, çevrim içi el yazısı ile ilgilenen sistemler popülerdir ve el yazısının bilgi açısından zengin olduğu kabul edilerek daha iyi performans elde etmesi beklenmektedir. El yazısı tanımlama sistemlerini kategorilere ayırmanın bir başka yolu, metne bağımlı veya metne bağımsız olmalarıdır. Metne bağlı yöntemler yüksek doğruluk sağlar, ancak metin içeriğinin olmadığı durumlarda uygulanamazlar. Metinden bağımsız yöntemler içeriğe karşı dayanıklıdır, ancak genel uygulanabilirliklerini sağlamak için büyük miktarda veri gerektirir (Yang vd., 2016: 46).

Veri/Metin madenciliğinde borsa analizi, ilginç araştırma alanlarından biridir. Birçok yatırımcı borsada yer almaktadır ve daha başarılı yatırımlar yapabilmek için piyasanın geleceği hakkında daha fazla bilgi edinmek istemektedirler. Etkili piyasa tahmini, ticaret önerileri olan yatırımcılara yardımcı olabilir veya otomatik tüccar araçları içinde bir bileşen olarak kullanılabilir. Bazen tahmin sistemleri, gelecekteki piyasa yönü gibi destekleyici bilgiler sunarak yatırımcılara dolaylı olarak yardımcı olmaktadır. Örneğin, seçilen bir hisse senedinin 24 saat boyunca yönünün "yukarı" olması öngörülüyorsa hisse senedini satın almak kârlı bir işlem olarak görülmektedir. Son

zamanlarda bazı arařtırmacılar, haberlerin borsayı etkileyen en etkili kaynaklardan biri olduđunu ve daha dođru tahminlere ulařmak için gerekli olduđunu bulmuřlardır (Nikfarjam vd., 2010: 256).

Veri/metin madenciliđinin tıbbi metne uygulanması, hem modern tıbbi metinler hem de diđer konu alanlarına ait tarihsel belgeler tıbbi arařtırma yapmak için kaynak konumundadır. Bu dođrultuda tıp alanında birkaç açıklamaalı ‘corpora’ yaratılmıřtır. Corpora, modern klinik kayıtlardan, yani doktorlar tarafından normalde sadece diđer doktorlar tarafından okunması amaçlanan bireysel hastalar hakkında yazılan raporlardan oluřmaktadır. Klinik kayıtlar genellikle resmi olmayan bir tarzda yazılmaktadır. Metin madenciliđi yöntemleri, kayıtlarda yer alan ilaç isimleri, tıbbi kořulları tanımlayarak ve kavramlar arasında iliřki kurarak kullanıcıların hangi ilaçları hangi tıbbi durumlarda kullanması gerektiđi ile ilgi bilgi sađlamaktadır (Thompson vd., 2016: 1).

ÜÇÜNCÜ BÖLÜM

VERİ VE METİN MADENCİLİĞİ İLE HAVA YOLU İŞLETMELERİNİN SOSYAL MEDYA YORUM VE SKORLARINA GÖRE DEĞERLENDİRİLMESİ VE KARŞILAŞTIRILMASI

Veri ve metin madenciliği ile hava yolu işletmelerinin sosyal medya yorum ve skorlarının değerlendirilmesi ve karşılaştırılması yapılacaktır. Bu doğrultuda araştırmanın amacı, kapsamı, kullanılan araçlar, veri seti, veri setinin çerçevesi ve veri analizinden bahsedilecektir. Sosyal medya yorum ve skorlarının sınıflandırılması ve tahminlemesinin yanı sıra kullanılan değerlendirme skorlarının korelasyon ve regresyon analizine yer verilmiştir. Hava yollarının uçuş güzergâhları, uçuş yapılan yer, kabin sınıfı ve polarite sınıfları esas alınarak değerlendirme skorlarının önem dereceleri Karar Ağaçları Algoritması ile hesaplanacaktır.

3.1. Araştırmanın Amacı

Araştırmanın amacı, veri ve metin madenciliği ile hava yolu işletmelerinin sosyal medya yorum ve skorlarına değerlendirilmesi yapılmıştır. Değerlendirmede, Star Alliance hava yolları için yapılan çevrimiçi müşteri yorumlarından polarite ve polarite değerlerinin bulunarak veri/metin madenciliği algoritmaları ile sınıflandırılması ve tahmin edilmesi amaçlanmaktadır. Bu doğrultuda, dünyanın en büyük hava yolu topluluğu olan Star Alliance küresel hava yolu birliğinin 26 üye hava yolundan alınmış ve seyahat platformu olan TripAdvisor'da müşterilerin yaptığı yorumlar değerlendirilmiştir (TripAdvisor, 2019). Değerlendirme kapsamında hava yolunu kullanan müşterilerin yorumları baz alınmıştır. Alınan yorumların Duygu Analizi yapılarak polaritesi (pozitif, negatif veya nötr) ve polarite değeri (0-1) belirlenmiştir. Veri ve metin madenciliği algoritmalarından DVM, Naive Bayes, Derin Öğrenme ile polarite (sınıflandırma) performans değerleri belirlenmiştir. DVM ve Derin Öğrenme Algoritması ile polarite değerlerine yönelik tahmin yapılarak RMSE (Kök Ortalama Karesel Hata) değeri hesaplanmış ve algoritmadan çıkan sonuçları karşılaştırılmıştır. Araştırma değerlendirme skorlarına yönelik korelasyon ve regresyon analizi yapılarak bir hava yolunu değerlendiren kullanıcının müşteri hizmetleri, fiyat/fayda oranı, check-in ve binış olduğu ortaya koyulmuştur. Ardından araştırmada kullanılan kriterlerin Karar Ağaçları

ile her etiket bazında önem dereceleri belirlenerek genel olarak en önemli kriterin müşteri yorumlarından çıkarılan polarite değerinin olduğu belirtilmiştir. Araştırmanın sonunda ise, ileriki çalışmalara temel olması için yorumlardan çıkarılan en sık kullanılan kelimelerin görselleştirilmesi yapılarak kelime bulutu analizi yapılmıştır. Elde edilen tüm analiz sonuçları ülkemizde yer alan Türk Hava Yolları (THY) ile karşılaştırılmıştır.

3.2. Araştırmanın Kapsamı

Sanal ortamda kullanıcıların hizmetine sunulan sosyal medya sitesi olan 2000 yılında kurulan dünyanın en büyük seyahat topluluğu olan TripAdvisor sitesi kullanılmıştır (Ayдын, 2016: 17). Star Alliance küresel hava yolu birliğinin 26 üye hava yolu şirketi ele alınmıştır. TripAdvisor’da incelenen hava yolları Tablo 7’de verilmiştir.

Tablo 7. Star Alliance Hava yolu Üyeleri

No	Hava Yolları	No	Hava Yolları
1	Aegean Airlines,	14	Eva Air
2	Air Canada	15	Air India
3	Air China	16	LOT Polish Airlines
4	Air New Zealand	17	Lufthansa
5	Ana Nippon Airways (ANA)	18	Scandinavian Airlines (SAS)
6	Asiana Airlines	19	Shenzhen Airlines
7	Austrian Airlines	20	Singapore Airlines
8	Avianca	21	South African Airways
9	Brussels Airlines	22	Swiss International Air Lines
10	Copa Airlines	23	Tap Portugal
11	Croatia Airlines	24	Thai
12	Egyptair	25	Türk Hava Yolları (THY)
13	Ethiopian	26	United Airlines

Tablo 7’deki tüm hava yollarına ilişkin TripAdvisor sitesinde 2020 yılına kadar yapılan toplam 197690 müşteri yorumu incelenmiştir. Bu yorumların yanı sıra müşteriler tarafından yapılan farklı değerlendirme skorları araştırmaya dâhil edilmiştir. Bu değerlendirme kriterleri şunlardır: genel skor, ayak koyma alanı, koltuk konforu, uçak içi eğlence (wifi, tv, filmler vb.), müşteri hizmetleri, fiyat/fayda oranı (paranın karşılığı/değeri), temizlik, check-in ve biniş, yiyecek ve içecek’tir. Kriterlere verilen değerlendirme puanları 10-50 arasında değişmektedir. Yorumlardan elde edilen polarite (pozitif, negatif, nötr) ve polarite değeri (0-1), uçuşlar, uçuş güzergahları (Avrupa, Uluslararası, Meksika, Amerika, Afrika vb.) ve kabin sınıfı (ekonomi, avantajlı ekonomi, birinci sınıf, öncelikli sınıf) etiketleri ile kriterlerin önem dereceleri belirlenmiştir. Şekil 9’de TripAdvisor’da bir müşteriye ait hava yolu için kriter değerlendirilmesi verilmiştir.

Şekil 9. Bir Müşteriye Ait Hava yolu Değerlendirmesi

Şekil 9’da TripAdvisor seyahat inceleme sitesinde hava yolu için değerlendirme yapan bir müşteriye ait değerlendirmede tez kapsamı doğrultusunda ele alınan kriterler vurgulanmıştır. Müşterinin yaptığı yorumdan polarite ve polarite değeri olarak iki kriter üretilerek DVM, Naive Bayes, Derin Öğrenme ve Karar Ağaçları Algoritmaları ile analizler yapılmıştır.

3.3. Araştırmada Kullanılan Araçlar

Müşteri yorumlarının derlenerek veritabanına aktarılmasında WebHarvy 6.0.1.173 programı kullanılmıştır. WebHarvy, görsel bir web kazıma aracıdır. Verileri kazımak için kesinlikle herhangi bir komut dosyası veya kod yazmaya gerek yoktur. Web sayfalarında gezinmek için WebHarvy'nin yerleşik tarayıcısı kullanılabilir. Kazınacak veriler, fare tıklamaları ile seçilebilir (Haddaway, 2015: 188). Kazınacak verileri kolay bir şekilde derlemek için ‘kurallı ifadeler (regular expression)’ kullanılmıştır. Bu kurallı ifadeler Tablo 8’de gösterilmiştir.

Tablo 8. Kurallı İfadeler

Kriterler	Kurallı İfadeler
Genel skor	bubble_(\d)+
Müşteri yorumu	wrote a review ([^<]*)
Ayak koyma alanı	bubble_(\d)+"[a-zA-Z <>/;=]"*Legroom
Uçak içi eğlence	bubble_(\d)+"[a-zA-Z <>/;=]"*In-flight Entertainment
Paranın değeri	bubble_(\d)+"[a-zA-Z <>/;=]"*Value for money
Check-in ve binış	bubble_(\d)+"[a-zA-Z <>/;=]"*Check-in and boarding

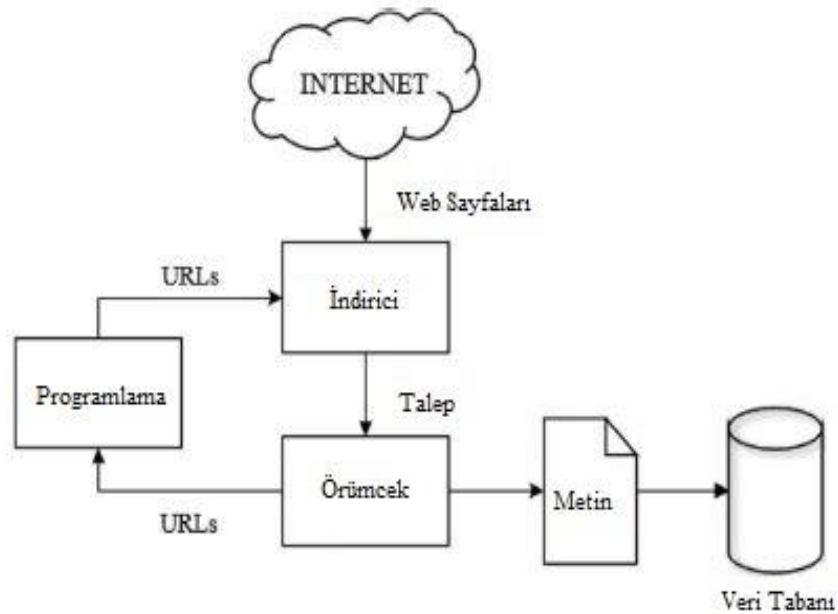
Koltuk konforu	bubble_(\d+)"[a-zA-Z </>/:;=]"*Seat comfort
Müşteri hizmetleri	bubble_(\d+)"[a-zA-Z </>/:;=]"*Customer service
Temizlik	bubble_(\d+)"[a-zA-Z </>/:;=]"*Cleanliness
Yiyecek ve içecek	bubble_(\d+)"[a-zA-Z </>/:;=]"*Food and Beverage

Yorum ve kriterlerin değerlendirilip analiz edilmesinde ise, açık kaynak kodlu RapidMiner 9.6.0 uygulaması kullanılmıştır. RapidMiner, kullanıcının veri analizi işlemlerini operatörleri tarafından ‘tak-çalıştır’ tarzında tasarlamasına olanak tanıyan kod kullanılmasına gerek olmadan çalışan bir veri analizi platformudur. Ayrıca, RapidMiner Marketplace’te sunulan uzantılar geliştirilerek RapidMiner’e işlevsellik eklenebilir (Ristoski vd., 2015: 142).

Tezdeki verilerin toplanıp bir araya getirilmesi için de Microsoft Company tarafından geliştirilen Excel 2016 kullanılmıştır.

3.4. Veri Seti

Teze ilişkin veri seti oluşturulurken Web Kazıma (Scraping) kullanılmıştır. Web Kazıma, genellikle verileri belirli ihtiyaçlar ve amaçlar için analiz eden HTML veya XHTML gibi biçimlendirme dilleri kullanan web sitesi sayfalarında, internetten otomatik veri ve bilgi toplama işlemidir (Slamet vd., 2018: 1). Elde edilen veriler TripAdvisor sitesinden 26 Star Alliance hava yolunun müşteri yorumları derlenmiştir. Şekil 10’de yorumların Web kazıma ile gösterimi verilmiştir.

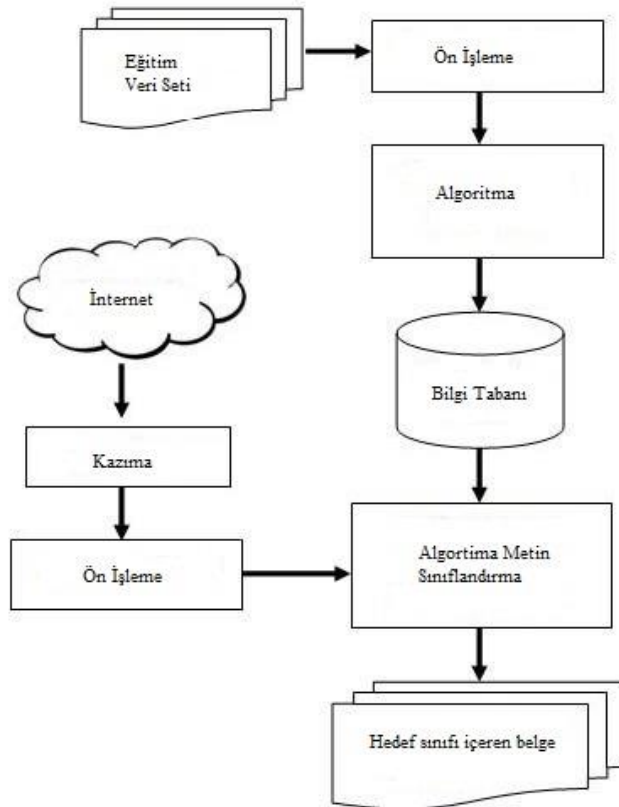


Şekil 10. Web Kazıma Çerçevesi

Şekil 10'da Web kazıma, bir URL listesi biçiminde tohumlarla başlamaktadır. Planlayıcı, önceliğe karar vererek ve yinelenen parçaları ortadan kaldırarak URL'lerin sırasını yönetmektedir. Daha sonra, indirici, web sayfalarını internetten almaktan ve bunları, sayfaları ayrıştırmak ve hedeflenen içerikleri çıkarmak için kullanılan örümceğe sağlamaktan sorumludur. Sonuç olarak elde edilecek veriler iki bölümden oluşmaktadır. Birincisi web sitelerinden alınan tarihi içeren metinsel yorumlar ve ardından alınan veriler veri tabanında saklanmaktadır; diğeri sayfalarda bulunan URL'lerdir ve ardından URL'ler programlayıcıya taşınmaktadır. Prosedürler, hedeflenen tüm metinsel belgeleri ele geçirene kadar tekrar edilmektedir. Belgelerin her biri veritabanında zaman, başlık ve içerik olarak görüntülenmektedir. Web kazıma ile ilgili tüm işlem adımları WebHarvy programı kullanılarak yapılmıştır.

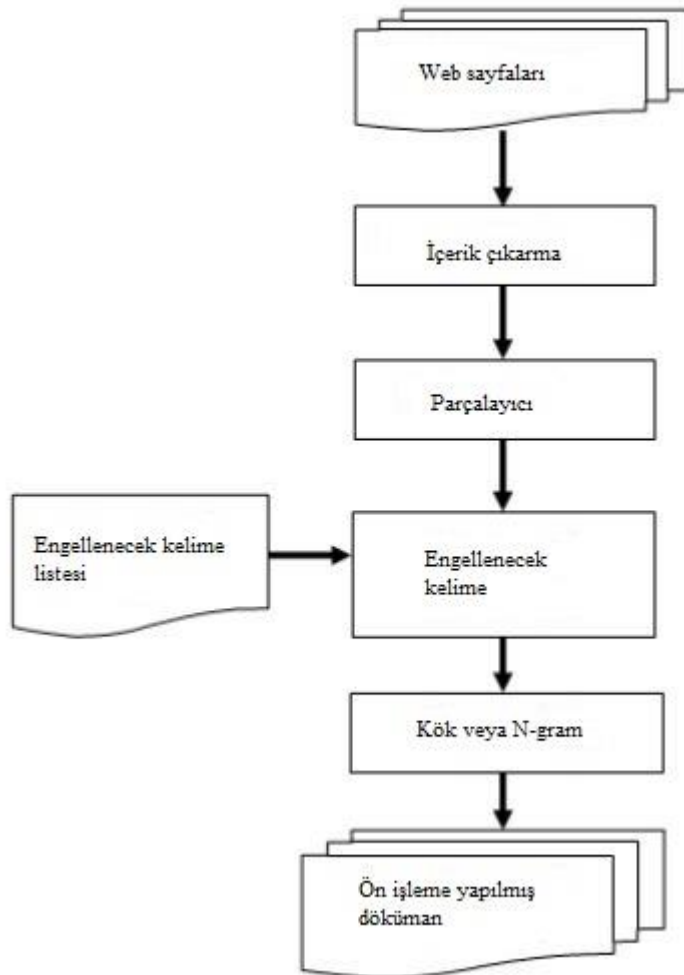
3.5. Veri Analizinin Çerçevesi

Veri setinin analizi için Derin Öğrenme, DVM, Naive Bayes ve Karar Ağaçları Algoritmaları kullanılmıştır. Bu doğrultuda, veri setini eğitim ve test kümesine ayrılarak yapılan sınıflandırma ve tahmin için doğruluk değerleri ve hata değerleri ortaya koyulmuştur. Şekil 11 veri setinin işlem çerçevesini göstermektedir.



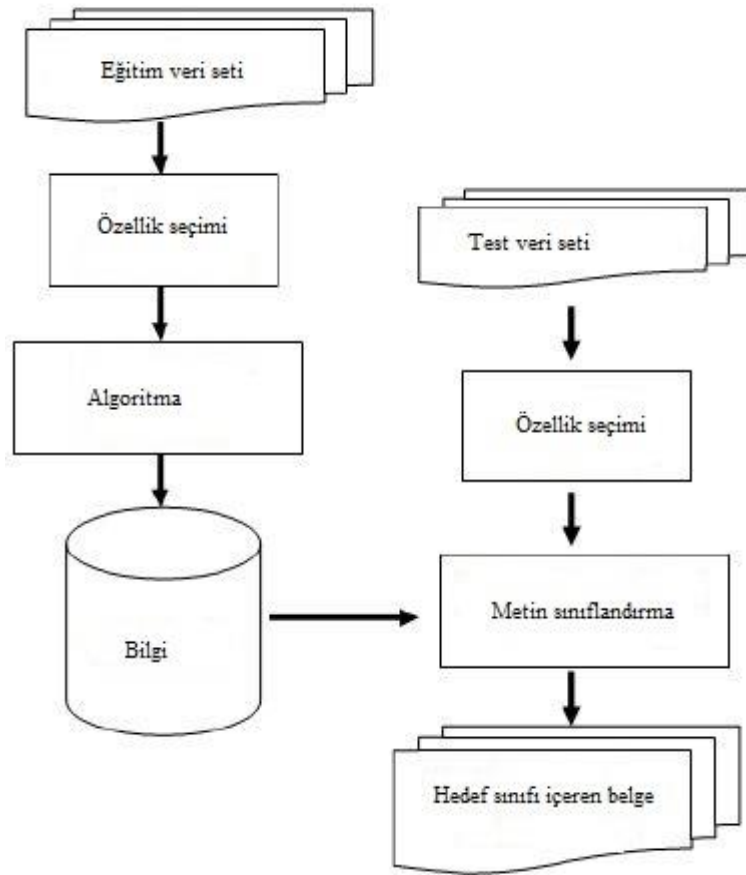
Şekil 11. Veri Seti İşlem Çerçevesi

Şekil 11’de Sınıflandırma sürecimiz iki aşamaya ayrılmıştır. İlk aşama, sınıflandırıcı bilgisi oluşturmak için belgeyi tanımlayarak öğrenmektir. İkinci aşama, yeni verilerin belirli bir sınıfa sınıflandırılmasıdır. Ön işleme, sınıflandırma veya öğrenme aşamalarının her birisi için kullanılmaktadır. Şekil 12’de ise Ön işleme sürecinin gösterimi verilmiştir.



Şekil 12. Ön İşleme Sürecinin Adımları

Şekil 12’de ön işleme 4 adımdan oluşmaktadır. İlk adım, içerik çıkarma ile ilgilidir. Bu adım, web sayfasındaki önemsiz kısımları kaldırarak web sitesi içeriğini çıkarmak için kullanılır. İkinci adım, parçalayıcı (Tokenization)’dır. Bu adım, belgeyi kelime, sayı veya noktalama işaretleri gibi belirteçlere ayırmak için kullanılmaktadır. Üçüncü adım, engellenecek kelimelerin (stopwords) kaldırılmasıdır. Genellikle belgede görünen bazı yaygın sözcükler, içindeki bazı sözlük listeleri kullanılarak kaldırılmaktadır. Ön işleme aşamamızdaki dördüncü adım, köklenme (stem) veya n-gram işlemidir. Bu işlem, kelimenin farklı morfolojik değişimi temel kelimeye eşlemek için kullanılmaktadır. Şekil 13’de metin sınıflandırma süreci verilmiştir.



Şekil 13. Metin Sınıflandırma Süreci

Şekil 13’de metin sınıflandırma aşaması iki adıma ayrılmıştır. İlk adımda, bazı bilgileri oluşturmak için öğrenme adımıdır ve ikinci adım, belge sınıflandırmasıdır. Sınıflandırıcının belgeyi öğrenmesine veya sınıflandırmasına yardımcı olmak için özellik çıkarımının yapılması gerekmektedir. Sınıflandırma algoritmamızda özellik olarak standart kelime torbası temsili olarak kullanılmaktadır.

3.6. Veri Analizi

Araştırmada, TripAdvisor’dan edinilen Star Alliance hava yolu firmaları için yapılan müşterilerin yorumları WebHarvy programı ile derlenmiştir. Yorumlar derlenirken ortak dünya dili olan İngilizce yorumlar baz alınmıştır. Edinilen veriler doğrultusunda 197690 müşteri yorumunun polaritesi ve polarite değeri RapidMiner programında yer alan ‘Aylien text analysis’ aracı ile çıkarılmıştır. Çıkarılan polarite ve polarite değerine göre algoritmalarından çıkan sonuçlar analiz edilmiştir. Analiz yapılırken verilerin %70’i eğitim, kalan %30’u test için seçilmiştir. Veriler karıştırılmış örnekleme ile seçilmiştir (Liu ve Cocea, 2017: 358). Bunun yanı sıra, hava yolları kriterlerinin önem

dereceleri belirlenmiştir. Araştırmanın sonunda ise, hava yolları için müşterilerin yaptıkları yorumlarının kelime bulutları oluşturulmuştur.

3.6.1. Hava Yollarının Duygu Analizi

TripAdvisor'dan edilen yorumlar baz alınarak Duygu Analizi ile her bir hava yolunun polaritesi (pozitif, negatif, nötr), polarite sayısı ve polarite değeri çıkarılmıştır. Tablo 9'da her bir hava yoluna ait polarite (pozitif, negatif ve nötr) yorum sayıları verilmiştir.

Tablo 9. Hava Yolları Yorumlarının Polarite Sayıları

Hava yolu	Pozitif	Negatif	Nötr	Toplam
ANA	2500	324	124	2948
Aegean	3661	731	152	4544
Asiana	1198	240	92	1530
Austrian	2328	910	177	3415
Avianca	2018	1242	156	3416
Brussels	1332	961	139	2432
Canada	11163	8922	1051	21136
China	1487	1271	306	3064
Copa	1996	790	147	2933
Croatia	425	207	35	667
Egypt	1422	661	170	2253
Ethiopian	2022	1080	206	3308
EVA	3396	449	180	4025
India	3511	2285	382	6178
LOT	1257	1081	139	2477
Lufthansa	11046	4265	874	16185
New Zealand	9699	1645	475	11819
SAS	3199	1205	305	4709
Shenzhen	100	54	17	171
Singapore	15607	2431	846	18884
South African	2339	1064	251	3654
Swiss	4179	1484	288	5951
TAP	3240	2582	308	6130
Thai	6535	1531	593	8659
THY	10086	3630	643	14359
United	24127	16585	2131	42843

Tablo 9'da görüldüğü üzere, hava yollarından en fazla yorum alan hava yolu United (42843), en az yorum alan Shenzhen (171)'dir. THY ise toplam 14359 yorum almıştır. Tüm hava yollarında pozitif yorum sayısı, negatif veya nötr yorum sayısından fazla olduğu görülmektedir. Tablo 10'da hava yollarına ait bir kısım müşteri yorumlarının polarite ve polarite değeri verilmiştir.

Tablo 10. Hava Yollarının Müşteri Yorumlarına Ait Polarite ve Polarite Değeri

Hava Yolu	Yorum	Polarite	Polarite Değeri
Aegean	Perfect in all occasion! The best after Turkish airlines. Helpful crew, the doos is excellent and you don't have to pay it.	Pozitif	0,991
ANA	Old and dated plane, air hostess, forgot to complete meal service and laughed it off. Did not feel welcome at all.	Negatif	0,780
Asiana	Traveling with Asiana is a breeze. The FAS are very warm and helpful and the amenities are ok. In terms of food, it's average.	Nötr	0,640
Austrian	1 hour late out of Tirana. Missed my connection and no other options available so had to stay overnight in Vienna. This airport is one of their hubs. It is much better staffed that at the origin of my flight.	Negatif	0,723
Avianca	Nicer people working for Avianca	Pozitif	0,465
Brussels	Flight left on time and arrived slightly early in Manchester. Very efficient staff. Leg room ok for economy. The service needs to be improved. To much money for extra leg space. Limited quantity of soft drinks.	Nötr	0,409
Canada	Routine service.	Pozitif	0,796
China	Poor quality of comfort, service, cleanliness with average meals. Got impression staff wanted to be invisible	Negatif	0,649
Copa	Flight was crowded, I supposed due to the good deals they offered flying from Brazil. Nothing special about it, regular service, food was not so good.	Nötr	0,524
Croatia	Direct flight and frendly crew. Snack was innovative and nice. A little bit more space for legs will be appretiated.	Pozitif	0,905
Egypt	Avoid. Sloppy. Passengers are treated as if they are an inconvenience, The flight was however cheap.	Negatif	0,951
Ethiopian	Manutention of the plane was poor. Service average. Entertainment average. Staff good. Seats needs refurbishing. The best value in my opinion. I always used to fly Thai but their prices are tooo high for me now. Good service, food and drink. Normally ontime.	Nötr	0,395
EVA	Great staff, pleasant aircraft. What more could you want?	Pozitif	0,953
India	Flight got delayed by 3 hours or even more. Way too annoyed with the service. This sucks, you can't waste our time like this.	Negatif	0,999
LOT	Good legroom comfortable for long journey. Entertainment system is average food is above average but not the best.	Nötr	0,663
Lufthansa	Really good flight with a well known airline. Boarding was prompt, didn't go by groups, waited a bit in the bus since the aircraft wasn't ready yet.	Pozitif	0,838
New Zealand	found staff to be rude, food very ordinary. plane was old and not in good repair. interior very worn.	Negatif	0,806
SAS	Late to take off but made up time en route. Jet was fine, crew were very friendly Small plane but okay.	Nötr	0,875
Shenzhen	Great experience	Pozitif	0,977
Singapore	once did they ask us did we want water, English was very poor and hard to understand food not to bad, very old plane	Negatif	0,938
South African	Very pleased with flight. Food was okay service was satisfactory.	Nötr	0,416
Swiss	Entertainment was relatively good. Left and came back on schedule In combination with their Edelweiss service this was a very good and cheap option to visit Vietnam. Good service, decent aircraft.	Pozitif	0,532
TAP	It's gets you there but they are bad about luggage. Carry on your bag. If they lose it, they don't care.	Negatif	0,734
Thai	Nothing out of place. Boarded on time. In-flight service average: average food, average seats, average cleanliness, outdated films on the server, no internet. All-in-all very average.	Nötr	0,609
THY	Turkish Airlines... Always good specially because of its connection flights to almost every country.	Pozitif	0,958

United	Pretty ordinary service in a very cramped regional aircraft. Useless in flight wifi with no internet connection available.	Negatif	0,851
--------	--	---------	-------

Tablo 10’da Star Alliance hava yollarına ait 26 hava yolunun pozitif, negatif ve nötr yorumların polarite ve polarite değerleri örnek olarak verilmiştir. Polarite (pozitif, negatif ve nötr) üç sınıftan oluşmaktadır. Polarite değeri ise, 0-1 arasında sayısallaştırılmıştır. Tezin kapsamı doğrultusunda kullanılan kriterlerin korelasyon ve regresyon analizleri yapılmıştır. Bunun yanı sıra Derin Öğrenme, DVM, Naive Bayes, Karar Ağaçları Algoritmaları kullanılarak hava yollarına yönelik polariteye göre sınıflandırma, polarite değerine göre tahmin ve kullanılan kriterlerin ağırlıkları analiz edilerek hava yollarına yönelik sonuçlar karşılaştırılmıştır.

3.6.2. Hava Yollarının Korelasyon Analizi

Star Alliance hava yolu birliğinin 26 hava yolu üyesinin TripAdvisor sitesinde yer alan genel skor, ayak koyma alanı, koltuk konforu, uçak içi eğlence (Wifi, tv, filmler vb.), müşteri hizmetleri, fiyat/fayda oranı, temizlik, check-in ve biniş, yiyecek ve içecek kriterlerine yönelik verileri alınarak korelasyon sonuçları Tablo 11’de gösterilmiştir. Korelasyon sonuçları Excel 2016 kullanarak elde edilmiştir.

Tablo 11. Tüm Hava Yollarına Ait Değerlendirme Kriterlerinin Korelasyon Sonuçları

Kriterler	Genel skor	Ayak koyma alanı	Uçak içi eğlence	Fiyat/fayda oranı	Check-in ve biniş	Koltuk konforu	Müşteri hizmetleri	Temizlik	Yiyecek ve içecek
Genel skor	1	0,634*	0,610*	0,809*	0,731*	0,701*	0,837*	0,719*	0,709*
Ayak koyma Alanı	0,634*	1	0,504*	0,610*	0,514*	0,826*	0,567*	0,588*	0,574*
Uçak içi eğlence	0,610*	0,504*	1	0,594*	0,499*	0,586*	0,573*	0,569*	0,610*
Fiyat/fayda oranı	0,809*	0,610*	0,594*	1	0,658*	0,675*	0,765*	0,703*	0,686*
Check-in ve biniş	0,731*	0,514*	0,499*	0,658*	1	0,559*	0,714*	0,661*	0,584*
Koltuk konforu	0,701*	0,826*	0,586*	0,675*	0,559*	1	0,628*	0,655*	0,632*
Müşteri hizmetleri	0,837*	0,567*	0,573*	0,765*	0,714*	0,628*	1	0,703*	0,681*
Temizlik	0,719*	0,588*	0,569*	0,703*	0,661*	0,655*	0,703*	1	0,660*
Yiyecek ve içecek	0,709*	0,574*	0,610*	0,686*	0,584*	0,632*	0,681*	0,660*	1

*Korelasyon 0,01 düzeyinde anlamlıdır.

Tablo 11’de 9 kriterin birbiri ile olan korelasyon sonuçları verilmiştir. Genel skor ile en yüksek korelasyon ilişkisi olan kriter müşteri hizmetleridir. Ardından koltuk konforu ile ayak koyma alanı, genel skor ile fiyat/fayda oranı, check-in ve biniş ile müşteri hizmetleri gibi kriterlerin birbiri ile ilişkisi 0,7 üzerindedir. Sonuç olarak, genel skor ile diğer kriterlerin birbiri ile ilişkisinin daha yüksek olduğu görülmektedir.

3.6.3. Hava Yollarının Regresyon Analizi

Hava Yollarının korelasyon analizi sonuçları doğrultusunda genel skorun diğer kriterler ile ilişkisinin yüksek olduğu görülmüştür. Bu bağlamda, genel skorun bağımlı değişken; ayak koyma alanı, koltuk konforu, uçak içi eğlence (wifi, tv, filmler vb.), müşteri hizmetleri, fiyat/fayda oranı, temizlik, check-in ve biniş, yiyecek ve içecek kriterlerinin de bağımsız değişken olarak seçilerek regresyon analizi yapılmıştır. Regresyon sonuçları Excel 2016 kullanarak hesaplanmış, tüm kriterlerin genel skoru açıklamadaki çoklu R değeri 0,888, R kare değeri 0,788, ayarlı R kare değeri ise, 0,788 olarak elde edilmiştir. Genel skoru açıklayan kriterlerin değerleri Tablo 12’de gösterilmiştir.

Tablo 12. Tüm Hava Yollarına Ait Değerlendirme Kriterlerinin Regresyon Sonuçları

	Katsayılar	Standart Hata	t Stat	P-değeri	Düşük %95	Yüksek %95
Kesişim	-1,329	0,070	-18,866	0,000	-1,467	-1,191
Ayak Koyma Alanı	0,029	0,003	10,758	0,000	0,023	0,034
Uçak İçi Eğlence	0,038	0,002	21,689	0,000	0,034	0,041
Fiyat/fayda oranı	0,236	0,002	97,786	0,000	0,232	0,241
Check-in ve biniş	0,155	0,002	74,954	0,000	0,151	0,160
Koltuk konforu	0,124	0,003	41,880	0,000	0,119	0,130
Müşteri Hizmetleri	0,317	0,002	140,856	0,000	0,312	0,321
Temizlik	0,054	0,003	20,187	0,000	0,049	0,060
Yiyecek ve içecek	0,089	0,002	41,205	0,000	0,084	0,093

Tablo 12’de genel skoru açıklayan en önemli kriterlerin müşteri hizmetleri, check-in ve biniş, paranızın karşılığıdır. Herhangi bir hizmet alınmadığında genel skorun keşisim değeri sıfırın altın 1,329 olduğu görülmektedir. Tablo 12’de katsayıların dışında standart hata değeri, t değeri, p değeri, düşük %95 ve yüksek %95 değerleri de verilmiştir.

3.6.4. Destek Vektör Makineleri ile Hava Yolu Polarite Analizi

DVM ile her bir hava yolunun polarite sonuçlarının sınıflandırılmasından önce özellik seçimi yapılmıştır. DVM algoritmasının performans sonuçlarını görmek açısından verilerin %70’i eğitim, kalan %30’u test için seçilmiştir. Veriler karıştırılmış örnekleme ile seçilmiştir. Kullanılan algoritma, eksik veri ile işlem yapamadığından dolayı, eksik veriler için bulunduğu sütundaki ortalama değer hesaplanarak analiz edilmiştir. Genel skor, ayak koyma alanı, koltuk konforu, uçak içi eğlence (wifi, tv, filmler vb.), müşteri hizmetleri, fiyat/fayda oranı, temizlik, check-in ve biniş, yiyecek ve içecek, polarite (pozitif, negatif, nötr) kriterleri performans ölçümü için kullanılmıştır. Buradaki kriterlerden polarite değeri etiket olarak seçilmiştir. Araştırmada, THY verilerinin

analizine odaklanarak geniş perspektifte değerlendirilmesi yapılmıştır. THY'den ortaya çıkan performans sonuçları ile diğer Star Alliance kuruluşu hava yollarının sonuçları karşılaştırılmıştır. Tablo 13'de THY'nin DVM algoritması ile sınıflandırma performans sonucu verilmiştir.

Tablo 13. THY'nin DVM ile Performans Sınıflandırma Sonucu

Doğruluk Performans Sonucu: %86,51	Gerçek Pozitif (GP)	Gerçek Negatif (GN)	Gerçek Nötr (GNÖ)	Sınıf Doğruluk
Tahmin Pozitif (TP)	2907	245	152	87,98%
Tahmin Negatif(TN)	144	820	40	81,67%
Tahmin Nötr (TNÖ)	0	0	0	00,00%
Sınıf Geribildirim	95,28%	77,00%	00,00%	

Tablo 13'de sınıflandırmanın başarı değeri %86,51 ve sınıflandırmanın hata değeri %13,49 olarak bulunmuştur. Tablo 13'de görüldüğü üzere bir sonraki müşterinin pozitif bir yorum ihtimali %87,98 negatif bir ihtimali %81,67 ve nötr bir yorum yazma ihtimali ise, % 00,00 olarak analiz edilmiştir.

Tablo 13'deki performans sonucundaki doğruluk, Gerçek Pozitif (GP), Gerçek Negatif (GN) ve Gerçek Nötr (GNÖ) toplam sınıflandırma sayısı üzerindeki orandır. Köşegen sayıları ile hesaplanan karşılık matrisi üzerindeki toplam elemanlar doğruluk oranını göstermektedir. Doğruluk oranı, Eşitlik 21'de gösterilmiştir (Hussein vd., 2012: 258).

$$\text{Doğruluk} = (GP + GN + GNÖ / (GP + GN + GNÖ + TP + TN + TNÖ)) \quad (21)$$

DVM algoritmasının tüm hava yollarına ait polarite sınıflandırma doğruluk sonuçları ve sınıflandırma hata değeri Tablo 14'de verilmiştir.

Tablo 14. DVM ile Tüm Hava yollarının Performans Sınıflandırma Sonucu

Hava Yolları	Sınıflandırma Performans Değeri %	Sınıflandırma Hata Değeri %	Sıralama
EVA	89,81	10,19	1
Aegean	89,07	10,93	2
ANA	88,46	11,54	3
New Zealand	87,51	12,49	4
Singapore	86,95	13,05	5
THY	86,51	13,49	6
Autrian	85,94	14,06	7
Asiana	85,84	14,16	8
Avianca	84,88	15,12	9
Swiss	83,87	16,13	10
Copa	83,52	16,48	11
Thai	83,03	16,97	12
Lufthansa	82,97	17,03	13
TAP Portugal	82,38	17,62	14
Ethiopian	81,25	18,75	15
Brussels	80,14	19,86	16

Canada	79,62	20,38	17
SAS	79,19	20,81	18
Egypt	79,14	20,86	19
India	78,95	21,05	20
South African	78,28	21,72	21
LOT	77,93	22,07	22
United	77,72	22,28	23
Croatia	76,00	24,00	24
China	74,65	25,35	25
Shenzhen	58,82	41,18	26

Tablo 14’de DVM algoritması ile sınıflandırma performans değerinin en yüksek (%89,81) ve en düşük hata değeri (%10,19) olan hava yolu EVA Air olarak bulunmuştur. Performans değerinin en düşük (%58,82) ve sınıflandırma hata değerinin en yüksek (%41,18) olan hava yolu Shenzhen olarak gösterilmiştir. THY’nin performans değeri ise % 86,51 iken, sınıflandırma hata değeri % 13,49’dur.

3.6.5. Naive Bayes ile Hava Yolu Polarite Analizi

Olasılık hesaplamasına dayalı olarak performans değerlendirmesi yapan Naive Bayes Algoritması ile hava yollarının başarı değeri hesaplanmıştır. Hava yolu ile derlenen verilerin %70’i eğitim, kalan %30’u test için seçilmiştir. Veriler karıştırılmış örnekleme ile seçilmiştir. Kullanılan algoritma, olasılık tabanlı olduğu için eksik veri ile işlem yapabilmektedir. Genel skor, ayak koyma alanı, koltuk konforu, uçak içi eğlence (wifi, tv, filmler vb.), müşteri hizmetleri, fiyat/fayda oranı, temizlik, check-in ve biniş, yiyecek ve içecek, polarite (pozitif, negatif, nötr) kriterleri performans ölçümü için kullanılmıştır. Buradaki kriterlerden polarite etiket olarak seçilmiştir. Araştırmada, THY verilerinin analizine odaklanarak geniş perspektifte değerlendirmesi yapılmıştır. THY’den ortaya çıkan performans sonuçları ile diğer Star Alliance kuruluşu hava yollarının sonuçları karşılaştırılmıştır. Tablo 15’de THY’nin DVM algoritması ile sınıflandırma performans sonucu verilmiştir.

Tablo 15. THY’nin Naive Bayes ile Performans Sınıflandırma Sonucu

Doğruluk Performans Sonucu: % 83,52	Gerçek Pozitif (GP)	Gerçek Negatif (GN)	Gerçek Nötr (GNÖ)	Sınıf Doğruluk
Tahmin Pozitif (TP)	2724	173	102	90,83%
Tahmin Negatif(TN)	212	841	57	75,77%
Tahmin Nötr (TNÖ)	115	51	33	16,58%
Sınıf Geribildirim	89,28%	78,97%	17,19%	

Tablo 15’de sınıflandırmanın başarı değeri %83,52 ve sınıflandırmanın hata değeri %16,48 olarak bulunmuştur. Tablo 15’de görüldüğü üzere bir sonraki müşterinin pozitif bir yorum ihtimali %90,83, negatif bir ihtimali %75,77 ve nötr bir yorum yazma

ihtimali ise, % 16,58 olarak analiz edilmiştir. Naive Bayes Algoritmasının tüm hava yollarına ait polarite sınıflandırma doğruluk sonuçları ve sınıflandırma hata değeri Tablo 16’da verilmiştir.

Tablo 16. Naive Bayes ile Tüm Hava Yollarının Performans Sınıflandırma Sonucu

Hava Yolları	Sınıflandırma Performans Değeri %	Sınıflandırma Hata Değeri %	Sıralama
EVA	87,49	12,51	1
Aegean	86,21	13,79	2
ANA Nippon	85,07	14,93	3
New Zealand	85,03	14,97	4
Singapore	84,06	15,94	5
THY	83,52	16,48	6
Autrian	83,30	16,70	7
Avianca	83,22	16,78	8
Asiana	82,14	17,86	9
Swiss	82,02	17,98	10
Copa	81,48	18,52	11
TAP Portugal	81,19	18,81	12
Lufthansa	80,49	19,51	13
Thai	79,91	20,09	14
Brussels	79,59	20,41	15
Canada	79,06	20,94	16
India	78,58	21,42	17
Croatia	78,50	21,50	18
Ethiopian	78,12	21,88	19
LOT	77,12	22,88	20
United	76,50	23,50	21
South African	76,46	23,54	22
SAS	75,94	24,06	23
Egypt	74,26	25,74	24
China	72,69	27,31	25
Shenzhen	60,78	39,22	26

Tablo 16’da Naive Bayes Algoritması ile sınıflandırma performans değerinin en yüksek (%87,49) ve en düşük hata değeri (%12,51) olan hava yolu EVA Air olarak bulunmuştur. Performans değerinin en düşük (%60,78) ve sınıflandırma hata değerinin en yüksek (%39,22) olan hava yolu Shenzhen olarak gösterilmiştir. THY’nin performans değeri ise % 83,52 iken, sınıflandırma hata değeri % 16,48’dir.

3.6.6. Derin Öğrenme ile Hava Yolu Polarite Analizi

Derin Öğrenme Algoritmasından, Rectified doğrusal aktivasyon fonksiyonu kullanılarak hava yollarının performans değerlendirmesi yapılmıştır. Hava yolu temel alınarak derlenen verilerin %70’i eğitim, kalan %30’u test için seçilmiştir. Veriler karıştırılmış örnekleme ile seçilmiştir. Kullanılan algoritma, aktivasyon fonksiyonu yardımıyla eksik veri ile işlem yapabilmektedir. Genel skor, ayak koyma alanı, koltuk konforu, uçak içi eğlence (wifi, tv, filmler vb.), müşteri hizmetleri, fiyat/fayda oranı, temizlik, check-in ve biniş, yiyecek ve içecek, polarite (pozitif, negatif, nötr) kriterleri

performans ölçümü için kullanılmıştır. Buradaki kriterlerden polarite etiket olarak seçilmiştir. Araştırmada, THY verilerinin analizine odaklanarak geniş perspektifte değerlendirmesi yapılmıştır. THY'den ortaya çıkan performans sonuçları ile diğer Star Alliance kuruluşu hava yollarının sonuçları karşılaştırılmıştır. Tablo 17'de THY'nin Derin Öğrenme Algoritması ile sınıflandırma performans sonucu verilmiştir.

Tablo 17. THY'nin Derin Öğrenme ile Performans Sınıflandırma Sonucu

Doğruluk Performans Sonucu: % 85,96	Gerçek Pozitif (GP)	Gerçek Negatif (GN)	Gerçek Nötr (GNÖ)	Sınıf Doğruluk
Tahmin Pozitif (TP)	2930	292	161	86,61%
Tahmin Negatif(TN)	121	773	31	83,57%
Tahmin Nötr (TNÖ)	0	0	0	0,00%
Sınıf Geribildirim	96,00%	72,58%	0,00%	

Tablo 17'de sınıflandırmanın başarı değeri %85,96 ve sınıflandırmanın hata değeri %16,48 olarak bulunmuştur. Tablo 17'de görüldüğü üzere bir sonraki müşterinin pozitif bir yorum ihtimali %86,61, negatif bir ihtimali %83,57 ve nötr bir yorum yazma ihtimali ise, % 00,00 olarak analiz edilmiştir. Derin Öğrenme Algoritmasının tüm hava yollarına ait polarite sınıflandırma doğruluk sonuçları ve sınıflandırma hata değeri Tablo 18'de verilmiştir.

Tablo 18. Derin Öğrenme ile Tüm Hava Yollarının Performans Sınıflandırma Sonucu

Hava Yolları	Sınıflandırma Performans Değeri %	Sınıflandırma Hata Değeri %	Sıralama
EVA	90,97	09,03	1
Aegean	89,14	10,86	2
ANA Nippon	88,69	11,31	3
New Zealand	87,42	12,58	4
Asiana	87,15	12,85	5
Singapore	86,97	13,03	6
THY	85,96	14,04	7
Avianca	85,17	14,83	8
Autrian	84,38	15,62	9
Swiss	84,20	15,80	10
Copa	83,41	16,59	11
Lufthansa	83,17	16,83	12
TAP Portugal	82,87	17,13	13
Thai	82,83	17,17	14
Croatia	81,00	19,00	15
Brussels	80,96	19,04	16
Ethiopian	80,34	19,66	17
India	79,49	20,51	18
SAS	79,33	20,67	19
Canada	79,10	20,90	20
LOT	78,87	21,13	21
South African	78,74	21,26	22
United	77,76	22,24	23
Egypt	77,22	22,78	24
China	74,43	25,57	25
Shenzhen	62,75	37,25	26

Tablo 18’de Derin Öğrenme Algoritması ile sınıflandırma performans değerinin en yüksek (%90,97) ve en düşük hata değeri (%9,03) olan hava yolu EVA Air olarak bulunmuştur. Performans değerinin en düşük (%62,75) ve sınıflandırma hata değerinin en yüksek (%37,25) olan hava yolu Shenzhen olarak gösterilmiştir. THY’nin performans değeri ise % 85,96 iken, sınıflandırma hata değeri % 14,04’dır.

3.6.7. Tüm Hava Yollarının Sınıflandırma Sonuçlarının Karşılaştırılması

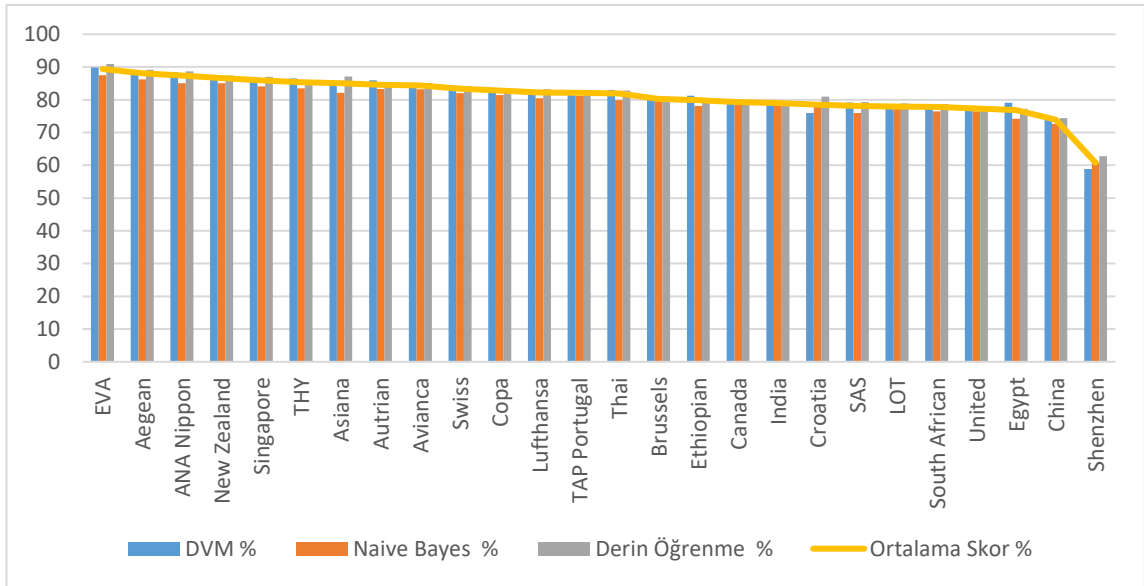
Her bir hava yolunun DVM, Naive Bayes ve Derin Öğrenme Algoritmaları ile yapılan sınıflandırma başarı sonuçları ve ortalamaları Tablo 19’de verilmiştir.

Tablo 19. Tüm Hava Yollarının Performans Sonuçlarının Karşılaştırılması

Hava Yolları	DVM %	Naive Bayes %	Derin Öğrenme %	Ortalama Skor %	Sıralama
EVA	89,81	87,49	90,97	89,42	1
Aegean	89,07	86,21	89,14	88,14	2
ANA	88,46	85,07	88,69	87,41	3
New Zealand	87,51	85,03	87,42	86,65	4
Singapore	86,95	84,06	86,97	85,99	5
THY	86,51	83,52	85,96	85,33	6
Asiana	85,84	82,14	87,15	85,04	7
Autrian	85,94	83,30	84,38	84,54	8
Avianca	84,88	83,22	85,17	84,42	9
Swiss	83,87	82,02	84,20	83,36	10
Copa	83,52	81,48	83,41	82,80	11
Lufthansa	82,97	80,49	83,17	82,21	12
TAP Portugal	82,38	81,19	82,87	82,15	13
Thai	83,03	79,91	82,83	81,92	14
Brussels	80,14	79,59	80,96	80,23	15
Ethiopian	81,25	78,12	80,34	79,90	16
Canada	79,62	79,06	79,10	79,26	17
India	78,95	78,58	79,49	79,01	18
Croatia	76,00	78,50	81,00	78,50	19
SAS	79,19	75,94	79,33	78,15	20
LOT	77,93	77,12	78,87	77,97	21
South African	78,28	76,46	78,74	77,83	22
United	77,72	76,50	77,76	77,33	23
Egypt	79,14	74,26	77,22	76,87	24
China	74,65	72,69	74,43	73,92	25
Shenzhen	58,82	60,78	62,75	60,78	26

Tablo 19’de algoritmalardan çıkan sınıflandırma performans değerleri karşılaştırılmıştır. Algoritmaların başarı değerlerinin ortalaması dikkate alınarak sıralama

yapılmıştır. Performans başarı değeri en yüksek (%89,42) EVA Air iken, en düşük (%60,78) performans başarı değeri Shenzhen hava yoludur. Üç algoritmanın ortalamasına göre THY'nin performans başarı değeri ise % 85,33 olarak hesaplanmıştır. Bunun yanı sıra Shenzhen hava yolu dışında diğer tüm hava yollarının başarı ortalaması birbirine yakın ve %70'in üzerinde olduğu görülmektedir. Shenzhen hava yolunun diğer hava yollarının performans başarı değerinden düşük olmasının sebebi olarak bu hava yoluna yapılan yorum sayısının diğer hava yollarından düşük olduğu düşünülmektedir. Şekil 14'de, Tablo 15'deki verilerin görselleştirilmesi yapılmıştır.



Şekil 14. Algoritmaların Performans Sonuçlarının Karşılaştırılması

Şekil 14 'de DVM, Naive Bayes ve Derin Öğrenme Algoritmalarının ortalamaları çubuk grafiği çizilmiştir. Çizgi grafiğinin üstünde kalan ve Shenzhen hava yolu dışında diğer hava yolları için yapılan polarite tahmin sonucunda en yüksek başarı değeri ile tahmin eden algoritmaların DVM ve Derin Öğrenme olduğu görülmektedir.

3.6.8. Tüm Hava Yollarının Polarite Değerlerinin RMSE Sonuçlarının Karşılaştırılması

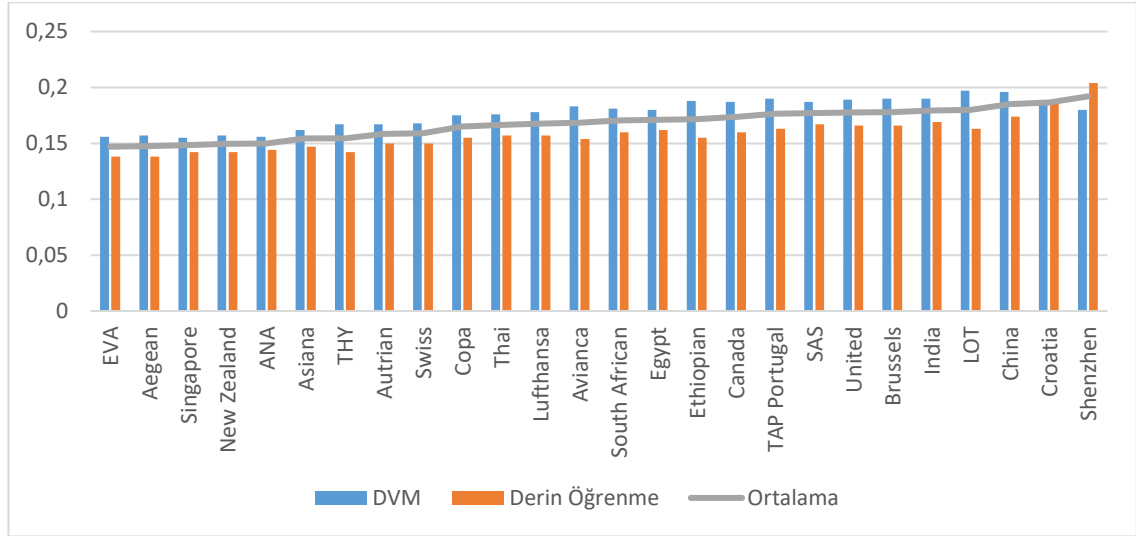
Hava yolları için polarite yani sınıflandırmanın pozitif, negatif veya nötr olduğunu tahmin etmenin yanı sıra aldığı sayısal yani polarite değerini tahmin etmek araştırma açısından önemlidir. Hava yolu temel alınarak derlenen verilerin %70'i eğitim, kalan %30'u test için seçilmiştir. Veriler karıştırılmış örnekleme ile seçilmiştir. Genel skor, ayak koyma alanı, koltuk konforu, uçak içi eğlence (wifi, tv, filmler vb.), müşteri hizmetleri, fiyat/fayda oranı, temizlik, check-in ve biniş, yiyecek ve içecek, polarite

değeri (0-1) kriterleri performans ölçümü için kullanılmıştır. Buradaki kriterlerden polarite değeri etiket olarak seçilmiştir. Tüm hava yollarının DVM ve Derin Öğrenme Algoritmaları ile polarite değerlerinin tahmini yapıldığında çıkan hata değeri yani RMSE sonuçları ve algoritmaların sonuçlarının ortalamaları Tablo 20’de verilmiştir.

Tablo 20. Tüm Hava Yollarının Polarite Değeri RMSE Sonuçları

Hava Yolları	DVM	Derin Öğrenme	Ortalama	Sıralama
EVA	0,156	0,138	0,147	1
Aegean	0,157	0,138	0,148	2
Singapore	0,155	0,142	0,149	3
New Zealand	0,157	0,142	0,150	4
ANA	0,156	0,144	0,150	5
Asiana	0,162	0,147	0,155	6
THY	0,167	0,142	0,155	7
Autrian	0,167	0,150	0,159	8
Swiss	0,168	0,150	0,159	9
Copa	0,175	0,155	0,165	10
Thai	0,176	0,157	0,167	11
Lufthansa	0,178	0,157	0,168	12
Avianca	0,183	0,154	0,169	13
South African	0,181	0,160	0,171	14
Egypt	0,180	0,162	0,171	15
Ethiopian	0,188	0,155	0,172	16
Canada	0,187	0,160	0,174	17
TAP Portugal	0,190	0,163	0,177	18
SAS	0,187	0,167	0,177	19
United	0,189	0,166	0,178	20
Brussels	0,190	0,166	0,178	21
India	0,190	0,169	0,180	22
LOT	0,197	0,163	0,180	23
China	0,196	0,174	0,185	24
Croatia	0,187	0,186	0,187	25
Shenzhen	0,180	0,204	0,192	26

Tablo 20’de hava yollarının polarite değerlerinin tahminine yönelik çıkan RMSE değerlerinin sıfıra yakın olduğu görülmektedir. Değerlerin sıfıra yakın olması başarılı bir tahmin gerçekleştirdiği anlamına gelmektedir. Bu doğrultuda algoritmaların RMSE ortalaması temel alınarak yapılan sıralamada en düşük (0,147) hata değeri olan hava yolu EVA iken, en yüksek (0,192) hata değeri olan hava yolu Shenzhen olduğu görülmektedir. THY’nin ise, yedinci sırada olduğu ve 0,155 hata değerine sahip olduğu dikkat çekmektedir. Şekil 15’de, Tablo 20’deki verilerin görselleştirilmesi yapılmıştır.



Şekil 15. Algoritmaların RMSE Değerlerinin Karşılaştırılması

Şekil 15’de DVM ve Derin Öğrenme Algoritmalarının ortalamaları çubuk grafiği çizilmiştir. Çizgi grafiğinin altında kalan ve Shenzhen hava yolu dışında diğer hava yolları için yapılan tahmin sonucunda en düşük hata değeri ile tahmin eden algoritmanın Derin Öğrenme olduğu görülmektedir.

3.6.9. Karar Ağaçları ile Kriterlerin Önem Derecelerinin Belirlenmesi

Hava yollarının polarite ve polarite değerlerinin tahmin edilmesinin yanı sıra genel skor, ayak koyma alanı, koltuk konforu, uçak içi eğlence (wifi, tv, filmler vb.), müşteri hizmetleri, fiyat/fayda oranı, temizlik, check-in ve biniş, yiyecek ve içecek, polarite değeri kriterlerinin önem dereceleri uçuş güzergâhları, uçuşlar, kabin sınıfı ve polarite etiketlerine göre belirlenmiştir. Kriterlerin önem dereceleri için Karar Ağacı Algoritması kullanılmıştır. Karar Ağacı Algoritmasında yer alan ‘gain ratio’ özelliği kullanılarak hava yolları için bilgi değeri en yüksek olan kriterlerin önem dereceleri hesaplanmıştır. Tüm hava yolları için sonuçlar derlenmiş ve THY ile karşılaştırılmıştır. Tablo 21’de uçuş güzergâhı etiketine göre hava yollarının önem dereceleri verilmiştir.

Tablo 21. Uçuş Güzergâhına Göre Önem Dereceleri

Kriter	Önem Derecesi	Kriter	Önem Derecesi
THY		Brussels	
Polarite Değeri	0,375	Polarite Değeri	0,307
Paranın Değeri	0,222	Uçak İçi Eğlence	0,288
Yiyecek ve İçecek	0,148	Yiyecek ve İçecek	0,121
Ayak Koyma Alanı	0,146	Paranın Değeri	0,113
Uçak İçi Eğlence	0,057	Ayak Koyma Alanı	0,074
Genel Skor	0,036	Müşteri Hizmetleri	0,052
Temizlik	0,011	Check-in ve Biniş	0,017
Check-in ve Biniş	0,002	Koltuk Konforu	0,014
Koltuk Konforu	0,002	Genel Skor	0,013

Aegean		Austrian	
Polarite Değeri	0,311	Genel Skor	0,286
Check-in ve Biniş	0,231	Polarite Değeri	0,238
Koltuk Konforu	0,214	Ayak Koyma Alanı	0,146
Genel Skor	0,103	Yiyecek ve İçecek	0,137
Ayak Koyma Alanı	0,083	Uçak İçi Eğlence	0,087
Paranın Değeri	0,032	Paranın Değeri	0,047
Uçak İçi Eğlence	0,016	Koltuk Konforu	0,037
Yiyecek ve İçecek	0,005	Temizlik	0,013
Temizlik	0,003	Check-in ve Biniş	0,004
Müşteri Hizmetleri	0,002	Müşteri Hizmetleri	0,004
Avianca		Canada	
Check-in ve Biniş	0,367	Genel Skor	0,314
Polarite Değeri	0,226	Temizlik	0,299
Yiyecek ve İçecek	0,093	Check-in ve Biniş	0,146
Koltuk Konforu	0,078	Polarite Değeri	0,096
Temizlik	0,072	Koltuk Konforu	0,054
Müşteri Hizmetleri	0,054	Uçak İçi Eğlence	0,036
Genel Skor	0,037	Ayak Koyma Alanı	0,033
Paranın Değeri	0,029	Paranın Değeri	0,013
Ayak Koyma Alanı	0,026	Müşteri Hizmetleri	0,007
Uçak İçi Eğlence	0,018	Yiyecek ve İçecek	0,003
ANA		Copa	
Polarite Değeri	0,243	Polarite Değeri	0,396
Yiyecek ve İçecek	0,208	Temizlik	0,192
Genel Skor	0,185	Müşteri Hizmetleri	0,173
Check-in ve Biniş	0,148	Ayak Koyma Alanı	0,173
Paranın Değeri	0,084	Koltuk Konforu	0,034
Ayak Koyma Alanı	0,075	Uçak İçi Eğlence	0,024
Temizlik	0,047	Yiyecek ve İçecek	0,005
Uçak İçi Eğlence	0,010	Paranın Değeri	0,002
China		EVA	
Polarite Değeri	0,311	Müşteri Hizmetleri	0,290
Genel Skor	0,236	Polarite Değeri	0,189
Temizlik	0,199	Check-in ve Biniş	0,179
Check-in ve Biniş	0,088	Uçak İçi Eğlence	0,098
Yiyecek ve İçecek	0,080	Ayak Koyma Alanı	0,086
Ayak Koyma Alanı	0,032	Koltuk Konforu	0,075
Uçak İçi Eğlence	0,022	Yiyecek ve İçecek	0,073
Paranın Değeri	0,020	Temizlik	0,006
Koltuk Konforu	0,013	Paranın Değeri	0,004
Ethiopian		Shenzhen	
Polarite Değeri	0,440	Ayak Koyma Alanı	0,290
Genel Skor	0,204	Check-in ve Biniş	0,198
Yiyecek ve İçecek	0,090	Müşteri Hizmetleri	0,154
Ayak Koyma Alanı	0,088	Polarite Değeri	0,144
Uçak İçi Eğlence	0,087	Yiyecek ve İçecek	0,110
Temizlik	0,060	Koltuk Konforu	0,055
Müşteri Hizmetleri	0,026	Temizlik	0,050
Paranın Değeri	0,003		
India		SAS	
Temizlik	0,369	Polarite Değeri	0,375
Müşteri Hizmetleri	0,184	Uçak İçi Eğlence	0,189
Ayak Koyma Alanı	0,124	Müşteri Hizmetleri	0,116
Paranın Değeri	0,124	Yiyecek ve İçecek	0,112
Polarite Değeri	0,109	Genel Skor	0,067
Uçak İçi Eğlence	0,075	Koltuk Konforu	0,063
Genel Skor	0,007	Temizlik	0,049
Yiyecek ve İçecek	0,004	Ayak Koyma Alanı	0,023

Koltuk Konforu	0,004	Paranın Değeri	0,006
LOT		Lufthansa	
Polarite Değeri	0,277	Polarite Değeri	0,412
Koltuk Konforu	0,155	Genel Skor	0,171
Paranın Değeri	0,144	Check-in ve Biniş	0,161
Ayak Koyma Alanı	0,133	Temizlik	0,077
Genel Skor	0,108	Ayak Koyma Alanı	0,064
Uçak İçi Eğlence	0,046	Paranın Değeri	0,056
Check-in ve Biniş	0,042	Uçak İçi Eğlence	0,024
Yiyecek ve İçecek	0,041	Koltuk Konforu	0,019
Müşteri Hizmetleri	0,040	Müşteri Hizmetleri	0,008
Temizlik	0,014	Yiyecek ve İçecek	0,008
New Zealand		South African	
Polarite Değeri	0,308	Polarite Değeri	0,327
Check-in ve Biniş	0,212	Check-in ve Biniş	0,197
Genel Skor	0,156	Temizlik	0,110
Ayak Koyma Alanı	0,083	Genel Skor	0,089
Uçak İçi Eğlence	0,079	Ayak Koyma Alanı	0,069
Müşteri Hizmetleri	0,049	Uçak İçi Eğlence	0,066
Temizlik	0,035	Müşteri Hizmetleri	0,064
Paranın Değeri	0,033	Yiyecek ve İçecek	0,052
Koltuk Konforu	0,031	Paranın Değeri	0,019
Yiyecek ve İçecek	0,013	Koltuk Konforu	0,007
Swiss		Thai	
Polarite Değeri	0,476	Polarite Değeri	0,341
Genel Skor	0,183	Temizlik	0,138
Paranın Değeri	0,174	Ayak Koyma Alanı	0,126
Uçak İçi Eğlence	0,069	Genel Skor	0,119
Koltuk Konforu	0,033	Koltuk Konforu	0,095
Yiyecek ve İçecek	0,029	Paranın Değeri	0,080
Temizlik	0,026	Uçak İçi Eğlence	0,053
Ayak Koyma Alanı	0,006	Yiyecek ve İçecek	0,038
Müşteri Hizmetleri	0,005	Check-in ve Biniş	0,009
TAP		United	
Polarite Değeri	0,248	Ayak Koyma Alanı	0,300
Check-in ve Biniş	0,178	Genel Skor	0,265
Yiyecek ve İçecek	0,154	Polarite Değeri	0,168
Müşteri Hizmetleri	0,147	Koltuk Konforu	0,119
Genel Skor	0,136	Check-in ve Biniş	0,065
Ayak Koyma Alanı	0,047	Paranın Değeri	0,058
Uçak İçi Eğlence	0,032	Müşteri Hizmetleri	0,011
Temizlik	0,032	Uçak İçi Eğlence	0,011
Koltuk Konforu	0,020	Temizlik	0,003
Paranın Değeri	0,004	Yiyecek ve İçecek	0,001
Asiana		Croatia	
Polarite Değeri	0,710	Polarite Değeri	1,000
Koltuk Konforu	0,290		
Egypt		Singapore	
Polarite Değeri	1,000	Polarite Değeri	1,000

Tablo 21’de uçuş güzergahına (Avrupa, Uluslararası, Meksika, Amerika, Afrika vb.) göre kriterlerin önem dereceleri hesaplandığında müşteri yorumlarından elde edilen pozitif, negatif ve nötr sınıfların sayısal değeri olan polarite değerinin her hava yolu için belirleyici olduğu görülmektedir. Ardından genel skor, ayak koyma alanı, check-in ve biniş, temizlik kriterlerinin hava yollarını değerlendirmede yüksek önem derecelerine

sahip olduğu dikkat çekmektedir. THY için belirleyici olan kriterlerin başında polaritenin değeri (0,375), paranın değeri (0,222) ve yiyecek ve içecek (0,148) kriterleri önemli rol oynamaktadır. THY dışında diğer hava yollarında Austrian hava yolu için genel skor (0,286), Avianca hava yolu için check-in biniş (0,367), EVA hava yolu için müşteri hizmetleri (0,290), United hava yolu için ayak koyma alanı (0,290), India hava yolu için temizlik (0,369) kriterleri birinci öneme sahip olduğu görülmektedir. Ek olarak Croatia, Egypt ve Singapore hava yolları için tek önemli kriterin polarite değeri olduğu yer almaktadır. Tablo 22’de uçuş yapılan yer (bir A noktasından bir B noktasına) etiketine göre her bir hava yolu için kriterlerin önem dereceleri hesaplanmıştır.

Tablo 22. Uçuş Yapılan Yere Göre Önem Dereceleri

Kriter	Önem Derecesi	Kriter	Önem Derecesi
THY		Aegean	
Polarite Değeri	0,487	Polarite Değeri	0,260
Genel Skor	0,100	Ayak Koyma Alanı	0,147
Temizlik	0,100	Yiyecek ve İçecek	0,120
Paranın Değeri	0,086	Genel Skor	0,093
Müşteri Hizmetleri	0,051	Uçak İçi Eğlence	0,085
Yiyecek ve İçecek	0,051	Check-in ve Biniş	0,078
Ayak Koyma Alanı	0,050	Paranın Değeri	0,073
Uçak İçi Eğlence	0,038	Koltuk Konforu	0,063
Koltuk Konforu	0,025	Müşteri Hizmetleri	0,063
Check-in ve Biniş	0,013	Temizlik	0,019
ANA		EVA	
Polarite Değeri	0,423	Polarite Değeri	0,388
Genel Skor	0,127	Check-in ve Biniş	0,146
Temizlik	0,124	Genel Skor	0,111
Uçak İçi Eğlence	0,095	Yiyecek ve İçecek	0,088
Yiyecek ve İçecek	0,069	Ayak Koyma Alanı	0,078
Paranın Değeri	0,058	Temizlik	0,072
Ayak Koyma Alanı	0,041	Müşteri Hizmetleri	0,053
Koltuk Konforu	0,040	Uçak İçi Eğlence	0,032
Check-in ve Biniş	0,023	Koltuk Konforu	0,031
Asiana		Avianca	
Polarite Değeri	0,380	Polarite Değeri	0,444
Genel Skor	0,144	Genel Skor	0,122
Check-in ve Biniş	0,127	Paranın Değeri	0,094
Uçak İçi Eğlence	0,108	Ayak Koyma Alanı	0,094
Paranın Değeri	0,090	Yiyecek ve İçecek	0,073
Ayak Koyma Alanı	0,055	Müşteri Hizmetleri	0,063
Müşteri Hizmetleri	0,036	Uçak İçi Eğlence	0,039
Temizlik	0,022	Koltuk Konforu	0,031
Koltuk Konforu	0,019	Temizlik	0,022
Yiyecek ve İçecek	0,019	Check-in ve Biniş	0,020
Austrian		China	
Polarite Değeri	0,735	Polarite Değeri	0,945
Ayak Koyma Alanı	0,133	Genel Skor	0,055
Genel Skor	0,133		
Brussels		Canada	
Polarite Değeri	0,270	Polarite Değeri	0,407
Genel Skor	0,154	Genel Skor	0,145
Temizlik	0,115	Yiyecek ve İçecek	0,101

Check-in ve Biniş	0,111	Paranın Değeri	0,091
Müşteri Hizmetleri	0,086	Temizlik	0,079
Uçak İçi Eğlence	0,065	Ayak Koyma Alanı	0,057
Koltuk Konforu	0,062	Müşteri Hizmetleri	0,042
Paranın Değeri	0,057	Koltuk Konforu	0,039
Ayak Koyma Alanı	0,053	Check-in ve Biniş	0,020
Yiyecek ve İçecek	0,027	Uçak İçi Eğlence	0,020
Copa		Croatia	
Polarite Değeri	0,240	Polarite Değeri	0,377
Uçak İçi Eğlence	0,217	Check-in ve Biniş	0,238
Genel Skor	0,217	Temizlik	0,200
Ayak Koyma Alanı	0,109	Uçak İçi Eğlence	0,130
Temizlik	0,109	Koltuk Konforu	0,055
Paranın Değeri	0,109		
Egypt		Ethiopian	
Polarite Değeri	0,503	Polarite Değeri	0,370
Genel Skor	0,104	Ayak Koyma Alanı	0,119
Ayak Koyma Alanı	0,103	Koltuk Konforu	0,119
Uçak İçi Eğlence	0,061	Uçak İçi Eğlence	0,094
Paranın Değeri	0,059	Yiyecek ve İçecek	0,077
Müşteri Hizmetleri	0,056	Genel Skor	0,069
Check-in ve Biniş	0,043	Temizlik	0,059
Yiyecek ve İçecek	0,036	Müşteri Hizmetleri	0,052
Koltuk Konforu	0,026	Check-in ve Biniş	0,027
Temizlik	0,009	Paranın Değeri	0,015
India		Shenzhen	
Polarite Değeri	0,417	Polarite Değeri	0,438
Ayak Koyma Alanı	0,171	Uçak İçi Eğlence	0,156
Yiyecek ve İçecek	0,085	Genel Skor	0,125
Uçak İçi Eğlence	0,085	Check-in ve Biniş	0,094
Müşteri Hizmetleri	0,085	Temizlik	0,051
Koltuk Konforu	0,085	Yiyecek ve İçecek	0,047
Genel Skor	0,070	Paranın Değeri	0,045
		Ayak Koyma Alanı	0,044
LOT		Singapore	
Ayak Koyma Alanı	0,298	Polarite Değeri	0,575
Check-in ve Biniş	0,298	Genel Skor	0,271
Genel Skor	0,236	Paranın Değeri	0,072
Polarite Değeri	0,168	Check-in ve Biniş	0,052
		Temizlik	0,030
Lufthansa		SAS	
Polarite Değeri	0,531	Polarite Değeri	0,429
Genel Skor	0,171	Genel Skor	0,158
Uçak İçi Eğlence	0,061	Müşteri Hizmetleri	0,085
Yiyecek ve İçecek	0,058	Ayak Koyma Alanı	0,062
Ayak Koyma Alanı	0,041	Paranın Değeri	0,049
Koltuk Konforu	0,039	Check-in ve Biniş	0,048
Paranın Değeri	0,030	Yiyecek ve İçecek	0,048
Müşteri Hizmetleri	0,029	Uçak İçi Eğlence	0,046
Temizlik	0,020	Temizlik	0,041
Check-in ve Biniş	0,019	Koltuk Konforu	0,034
New Zealand		South African	
Polarite Değeri	0,449	Polarite Değeri	0,363
Genel Skor	0,206	Uçak İçi Eğlence	0,315
Uçak İçi Eğlence	0,124	Temizlik	0,074
Ayak Koyma Alanı	0,075	Genel Skor	0,071
Temizlik	0,052	Koltuk Konforu	0,058
Paranın Değeri	0,042	Paranın Değeri	0,050
Yiyecek ve İçecek	0,036	Check-in ve Biniş	0,034

Check-in ve Biniş	0,008	Yiyecek ve İçecek	0,018
Müşteri Hizmetleri	0,008	Ayak Koyma Alanı	0,016
Swiss		TAP	
Polarite Değeri	0,478	Polarite Değeri	0,253
Genel Skor	0,125	Genel Skor	0,212
Ayak Koyma Alanı	0,076	Ayak Koyma Alanı	0,145
Yiyecek ve İçecek	0,064	Paranın Değeri	0,098
Uçak İçi Eğlence	0,061	Koltuk Konforu	0,093
Check-in ve Biniş	0,057	Yiyecek ve İçecek	0,082
Paranın Değeri	0,050	Check-in ve Biniş	0,033
Koltuk Konforu	0,042	Müşteri Hizmetleri	0,030
Temizlik	0,027	Uçak İçi Eğlence	0,027
Müşteri Hizmetleri	0,020	Temizlik	0,025
Thai		United	
Polarite Değeri	0,329	Polarite Değeri	0,591
Uçak İçi Eğlence	0,159	Genel Skor	0,149
Yiyecek ve İçecek	0,117	Ayak Koyma Alanı	0,090
Check-in ve Biniş	0,088	Yiyecek ve İçecek	0,055
Genel Skor	0,086	Check-in ve Biniş	0,038
Koltuk Konforu	0,066	Koltuk Konforu	0,027
Ayak Koyma Alanı	0,048	Müşteri Hizmetleri	0,020
Temizlik	0,044	Paranın Değeri	0,019
Paranın Değeri	0,042	Temizlik	0,010
Müşteri Hizmetleri	0,022		

Tablo 22’de uçuş yapılan yere göre kriterlerin önem dereceleri hesaplandığında müşteri yorumlarından elde edilen pozitif, negatif ve nötr sınıfların sayısal değeri olan polarite değerinin her hava yolu için belirleyici olduğu görülmektedir. Ardından genel skor, ayak koyma alanı, uçak içi eğlence, check-in ve biniş, temizlik kriterlerinin hava yollarını değerlendirmede yüksek önem derecelerine sahip olduğu dikkat çekmektedir. THY için belirleyici olan kriterlerin başında polaritenin değeri (0,487), genel skor (0,100) ve temizlik (0,100) kriterleri önemli rol oynamaktadır. THY hava yolu dışında, diğer hava yollarında LOT için ayak koyma alanı (0,298), China ve Austrian hava yolu için polarite değeri ile birlikte genel skor kriterinin de yüksek önem değerine sahip olduğu görülmektedir. Tablo 23’de kabin sınıfı (ekonomi, avantajlı ekonomi, birinci sınıf, öncelikli sınıf) etiketine göre her bir hava yolu için kriterlerin önem dereceleri hesaplanmıştır.

Tablo 23. Kabin Sınıfına Göre Önem Dereceleri

Kriter	Önem Derecesi	Kriter	Önem Derecesi
THY		Asiana	
Polarite Değeri	0,505	Temizlik	0,190
Paranın Değeri	0,104	Paranın Değeri	0,190
Check-in ve Biniş	0,104	Polarite Değeri	0,169
Temizlik	0,100	Yiyecek ve İçecek	0,149
Yiyecek ve İçecek	0,056	Müşteri Hizmetleri	0,120
Uçak İçi Eğlence	0,052	Genel Skor	0,049
Müşteri Hizmetleri	0,039	Ayak Koyma Alanı	0,039
Ayak Koyma Alanı	0,023	Koltuk Konforu	0,039
Koltuk Konforu	0,011	Uçak İçi Eğlence	0,034

Genel Skor	0,005	Check-in ve Biniş	0,021
ANA		Aegean	
Polarite Değeri	0,476	Polarite Değeri	0,468
Uçak İçi Eğlence	0,242	Yiyecek ve İçecek	0,207
Paranın Değeri	0,186	Uçak İçi Eğlence	0,171
Koltuk Konforu	0,028	Check-in ve Biniş	0,063
Ayak Koyma Alanı	0,025	Genel Skor	0,039
Temizlik	0,021	Koltuk Konforu	0,030
Genel Skor	0,015	Müşteri Hizmetleri	0,014
Yiyecek ve İçecek	0,006	Ayak Koyma Alanı	0,007
Avianca		Brussels	
Polarite Değeri	0,431	Polarite Değeri	0,510
Müşteri Hizmetleri	0,215	Koltuk Konforu	0,105
Genel Skor	0,178	Yiyecek ve İçecek	0,082
Yiyecek ve İçecek	0,072	Müşteri Hizmetleri	0,081
Paranın Değeri	0,041	Genel Skor	0,065
Temizlik	0,023	Check-in ve Biniş	0,062
Uçak İçi Eğlence	0,018	Uçak İçi Eğlence	0,062
Ayak Koyma Alanı	0,015	Paranın Değeri	0,024
Koltuk Konforu	0,007	Ayak Koyma Alanı	0,008
Austrian		China	
Polarite Değeri	0,295	Polarite Değeri	0,720
Uçak İçi Eğlence	0,284	Uçak İçi Eğlence	0,143
Genel Skor	0,196	Yiyecek ve İçecek	0,061
Check-in ve Biniş	0,160	Ayak Koyma Alanı	0,046
Müşteri Hizmetleri	0,038	Temizlik	0,016
Yiyecek ve İçecek	0,018	Koltuk Konforu	0,007
Ayak Koyma Alanı	0,005	Check-in ve Biniş	0,007
Koltuk Konforu	0,004		
Canada		Copa	
Polarite Değeri	0,538	Uçak İçi Eğlence	0,229
Genel Skor	0,153	Ayak Koyma Alanı	0,224
Müşteri Hizmetleri	0,134	Polarite Değeri	0,185
Paranın Değeri	0,051	Genel Skor	0,071
Temizlik	0,050	Koltuk Konforu	0,070
Uçak İçi Eğlence	0,029	Paranın Değeri	0,065
Ayak Koyma Alanı	0,025	Check-in ve Biniş	0,050
Koltuk Konforu	0,011	Temizlik	0,041
Check-in ve Biniş	0,004	Yiyecek ve İçecek	0,038
Yiyecek ve İçecek	0,004	Müşteri Hizmetleri	0,029
China		Croatia	
Polarite Değeri	0,468	Polarite Değeri	0,713
Temizlik	0,285	Check-in ve Biniş	0,160
Check-in ve Biniş	0,077	Paranın Değeri	0,083
Genel Skor	0,075	Temizlik	0,019
Uçak İçi Eğlence	0,036	Koltuk Konforu	0,013
Müşteri Hizmetleri	0,026	Uçak İçi Eğlence	0,012
Ayak Koyma Alanı	0,013		
Koltuk Konforu	0,010		
Paranın Değeri	0,010		
EVA		Egypt	
Polarite Değeri	0,377	Genel Skor	0,270
Paranın Değeri	0,180	Polarite Değeri	0,178
Check-in ve Biniş	0,147	Uçak İçi Eğlence	0,174
Koltuk Konforu	0,087	Temizlik	0,140
Genel Skor	0,073	Paranın Değeri	0,059
Yiyecek ve İçecek	0,054	Yiyecek ve İçecek	0,057
Müşteri Hizmetleri	0,047	Check-in ve Biniş	0,036
Temizlik	0,016	Ayak Koyma Alanı	0,034

Ayak Koyma Alanı	0,015	Müşteri Hizmetleri	0,032
Uçak İçi Eğlence	0,004	Koltuk Konforu	0,019
Ethiopian		Shenzhen	
Polarite Değeri	0,403	Polarite Değeri	0,479
Genel Skor	0,156	Temizlik	0,274
Ayak Koyma Alanı	0,156	Müşteri Hizmetleri	0,142
Müşteri Hizmetleri	0,149	Uçak İçi Eğlence	0,042
Paranın Değeri	0,137	Koltuk Konforu	0,041
		Paranın Değeri	0,022
India		Lufthansa	
Polarite Değeri	0,390	Polarite Değeri	0,404
Uçak İçi Eğlence	0,253	Yiyecek ve İçecek	0,204
Paranın Değeri	0,115	Koltuk Konforu	0,081
Müşteri Hizmetleri	0,096	Uçak İçi Eğlence	0,075
Genel Skor	0,070	Temizlik	0,075
Check-in ve Biniş	0,043	Genel Skor	0,072
Temizlik	0,015	Müşteri Hizmetleri	0,052
Koltuk Konforu	0,011	Paranın Değeri	0,019
Ayak Koyma Alanı	0,007	Ayak Koyma Alanı	0,018
LOT		New Zealand	
Ayak Koyma Alanı	0,318	Polarite Değeri	0,175
Polarite Değeri	0,282	Müşteri Hizmetleri	0,150
Check-in ve Biniş	0,094	Yiyecek ve İçecek	0,149
Koltuk Konforu	0,092	Genel Skor	0,146
Genel Skor	0,054	Uçak İçi Eğlence	0,110
Müşteri Hizmetleri	0,047	Temizlik	0,109
Paranın Değeri	0,042	Paranın Değeri	0,073
Yiyecek ve İçecek	0,031	Koltuk Konforu	0,045
Uçak İçi Eğlence	0,024	Check-in ve Biniş	0,029
Temizlik	0,016	Ayak Koyma Alanı	0,014
SAS		Singapore	
Polarite Değeri	0,411	Polarite Değeri	0,359
Genel Skor	0,124	Genel Skor	0,107
Koltuk Konforu	0,124	Müşteri Hizmetleri	0,106
Temizlik	0,087	Ayak Koyma Alanı	0,104
Uçak İçi Eğlence	0,086	Check-in ve Biniş	0,087
Check-in ve Biniş	0,066	Uçak İçi Eğlence	0,065
Müşteri Hizmetleri	0,053	Yiyecek ve İçecek	0,062
Yiyecek ve İçecek	0,026	Koltuk Konforu	0,060
Ayak Koyma Alanı	0,016	Paranın Değeri	0,032
Paranın Değeri	0,008	Temizlik	0,018
South African		Swiss	
Polarite Değeri	0,346	Polarite Değeri	0,520
Uçak İçi Eğlence	0,184	Temizlik	0,193
Yiyecek ve İçecek	0,097	Paranın Değeri	0,100
Ayak Koyma Alanı	0,089	Genel Skor	0,058
Müşteri Hizmetleri	0,089	Yiyecek ve İçecek	0,056
Paranın Değeri	0,085	Ayak Koyma Alanı	0,031
Genel Skor	0,056	Uçak İçi Eğlence	0,019
Koltuk Konforu	0,028	Müşteri Hizmetleri	0,013
Temizlik	0,018	Koltuk Konforu	0,010
Check-in ve Biniş	0,008		
Thai		United	
Polarite Değeri	0,383	Genel Skor	0,242
Paranın Değeri	0,175	Polarite Değeri	0,234
Temizlik	0,095	Check-in ve Biniş	0,191
Koltuk Konforu	0,077	Müşteri Hizmetleri	0,124
Yiyecek ve İçecek	0,064	Yiyecek ve İçecek	0,096
Genel Skor	0,064	Uçak İçi Eğlence	0,032

Ayak Koyma Alanı	0,047	Temizlik	0,031
Uçak İçi Eğlence	0,047	Paranın Değeri	0,019
Müşteri Hizmetleri	0,045	Ayak Koyma Alanı	0,018
Check-in ve Biniş	0,002	Koltuk Konforu	0,012

Tablo 23’de kabin sınıfına göre kriterlerin önem dereceleri hesaplandığında müşteri yorumlarından elde edilen pozitif, negatif ve nötr sınıfların sayısal değeri olan polarite değerinin her hava yolu için belirleyici olduğu görülmektedir. Ardından uçak içi eğlence, genel skor, müşteri hizmetleri, temizlik kriterlerinin hava yollarını değerlendirmede yüksek önem derecelerine sahip olduğu dikkat çekmektedir. THY için belirleyici olan kriterlerin başında polaritenin değeri (0,505), paranın değeri (0,104) ve check-in biniş (0,104) kriterleri önemli rol oynamaktadır. THY dışında diğer hava yollarında Asiana hava yolu için temizlik ve paranın değeri (0,190), Copa hava yolu için uçak içi eğlence (0,229), United hava yolu için genel skor (0,242), LOT hava yolu için ayak koyma alanı (0,318) kriterleri birinci öneme sahip olduğu görülmektedir. Tablo 24’de polarite (pozitif, negatif, nötr) etiketine göre her bir hava yolu için kriterlerin önem dereceleri hesaplanmıştır.

Tablo 24. Polariteye Göre Önem Dereceleri

Kriter	Önem Derecesi	Kriter	Önem Derecesi
THY		ANA	
Polarite Değeri	0,370	Polarite Değeri	0,367
Koltuk Konforu	0,150	Ayak Koyma Alanı	0,138
Ayak Koyma Alanı	0,138	Check-in ve Biniş	0,136
Paranın Değeri	0,122	Yiyecek ve İçecek	0,093
Genel Skor	0,063	Paranın Değeri	0,090
Yiyecek ve İçecek	0,046	Genel Skor	0,055
Temizlik	0,035	Müşteri Hizmetleri	0,038
Check-in ve Biniş	0,033	Koltuk Konforu	0,034
Müşteri Hizmetleri	0,025	Uçak İçi Eğlence	0,029
Uçak İçi Eğlence	0,019	Temizlik	0,020
Aegean		Austrian	
Polarite Değeri	0,356	Polarite Değeri	0,530
Uçak İçi Eğlence	0,166	Genel Skor	0,127
Müşteri Hizmetleri	0,110	Check-in ve Biniş	0,121
Temizlik	0,105	Paranın Değeri	0,074
Genel Skor	0,065	Koltuk Konforu	0,062
Paranın Değeri	0,063	Temizlik	0,032
Ayak Koyma Alanı	0,044	Müşteri Hizmetleri	0,025
Yiyecek ve İçecek	0,040	Yiyecek ve İçecek	0,014
Check-in ve Biniş	0,026	Ayak Koyma Alanı	0,010
Koltuk Konforu	0,025	Uçak İçi Eğlence	0,005
Asiana		Brussels	
Polarite Değeri	0,406	Polarite Değeri	0,421
Yiyecek ve İçecek	0,173	Müşteri Hizmetleri	0,185
Temizlik	0,144	Ayak Koyma Alanı	0,178
Genel Skor	0,097	Yiyecek ve İçecek	0,055
Koltuk Konforu	0,070	Genel Skor	0,051
Check-in ve Biniş	0,064	Temizlik	0,046
Uçak İçi Eğlence	0,024	Paranın Değeri	0,034

Müşteri Hizmetleri	0,021	Check-in ve Biniş	0,028
		Koltuk Konforu	0,002
Avianca		Canada	
Polarite Değeri	0,427	Polarite Değeri	0,513
Genel Skor	0,141	Koltuk Konforu	0,121
Temizlik	0,102	Genel Skor	0,094
Uçak İçi Eğlence	0,085	Uçak İçi Eğlence	0,078
Ayak Koyma Alanı	0,067	Yiyecek ve İçecek	0,053
Müşteri Hizmetleri	0,053	Temizlik	0,049
Paranın Değeri	0,046	Ayak Koyma Alanı	0,043
Koltuk Konforu	0,036	Paranın Değeri	0,031
Check-in ve Biniş	0,033	Müşteri Hizmetleri	0,011
Yiyecek ve İçecek	0,012	Check-in ve Biniş	0,006
China		EVA	
Polarite Değeri	0,337	Polarite Değeri	0,251
Temizlik	0,172	Ayak Koyma Alanı	0,224
Paranın Değeri	0,141	Müşteri Hizmetleri	0,158
Ayak Koyma Alanı	0,113	Uçak İçi Eğlence	0,109
Koltuk Konforu	0,096	Yiyecek ve İçecek	0,079
Genel Skor	0,046	Genel Skor	0,060
Müşteri Hizmetleri	0,044	Koltuk Konforu	0,053
Uçak İçi Eğlence	0,033	Paranın Değeri	0,046
Check-in ve Biniş	0,012	Check-in ve Biniş	0,016
Yiyecek ve İçecek	0,006	Temizlik	0,005
Copa		Croatia	
Polarite Değeri	0,243	Polarite Değeri	0,407
Ayak Koyma Alanı	0,170	Yiyecek ve İçecek	0,134
Check-in ve Biniş	0,117	Genel Skor	0,098
Yiyecek ve İçecek	0,112	Müşteri Hizmetleri	0,096
Genel Skor	0,099	Check-in ve Biniş	0,084
Paranın Değeri	0,096	Paranın Değeri	0,075
Uçak İçi Eğlence	0,090	Koltuk Konforu	0,054
Müşteri Hizmetleri	0,064	Ayak Koyma Alanı	0,031
Temizlik	0,010	Uçak İçi Eğlence	0,022
Egypt		Ethiopian	
Polarite Değeri	0,358	Polarite Değeri	0,473
Check-in ve Biniş	0,134	Check-in ve Biniş	0,192
Uçak İçi Eğlence	0,095	Ayak Koyma Alanı	0,089
Yiyecek ve İçecek	0,089	Paranın Değeri	0,055
Ayak Koyma Alanı	0,071	Genel Skor	0,053
Genel Skor	0,070	Uçak İçi Eğlence	0,044
Paranın Değeri	0,060	Koltuk Konforu	0,042
Müşteri Hizmetleri	0,055	Müşteri Hizmetleri	0,033
Temizlik	0,053	Yiyecek ve İçecek	0,010
Koltuk Konforu	0,016	Temizlik	0,007
India		LOT	
Polarite Değeri	0,611	Polarite Değeri	0,264
Check-in ve Biniş	0,084	Check-in ve Biniş	0,200
Uçak İçi Eğlence	0,058	Koltuk Konforu	0,109
Ayak Koyma Alanı	0,044	Ayak Koyma Alanı	0,106
Temizlik	0,041	Yiyecek ve İçecek	0,103
Müşteri Hizmetleri	0,038	Müşteri Hizmetleri	0,087
Paranın Değeri	0,037	Genel Skor	0,062
Koltuk Konforu	0,036	Temizlik	0,059
Genel Skor	0,027	Paranın Değeri	0,007
Yiyecek ve İçecek	0,024	Uçak İçi Eğlence	0,003
Lufthansa		New Zealand	
Polarite Değeri	0,484	Polarite Değeri	0,408
Genel Skor	0,105	Ayak Koyma Alanı	0,186

Müşteri Hizmetleri	0,068	Temizlik	0,077
Koltuk Konforu	0,060	Koltuk Konforu	0,075
Uçak İçi Eğlence	0,057	Müşteri Hizmetleri	0,073
Check-in ve Biniş	0,057	Genel Skor	0,049
Ayak Koyma Alanı	0,055	Paranın Değeri	0,042
Paranın Değeri	0,045	Check-in ve Biniş	0,041
Temizlik	0,042	Uçak İçi Eğlence	0,034
Yiyecek ve İçecek	0,028	Yiyecek ve İçecek	0,015
SAS		Singapore	
Polarite Değeri	0,275	Polarite Değeri	0,400
Check-in ve Biniş	0,146	Uçak İçi Eğlence	0,107
Yiyecek ve İçecek	0,137	Ayak Koyma Alanı	0,103
Genel Skor	0,131	Temizlik	0,068
Paranın Değeri	0,123	Check-in ve Biniş	0,067
Koltuk Konforu	0,053	Paranın Değeri	0,055
Ayak Koyma Alanı	0,049	Müşteri Hizmetleri	0,053
Uçak İçi Eğlence	0,046	Koltuk Konforu	0,051
Müşteri Hizmetleri	0,034	Genel Skor	0,050
Temizlik	0,006	Yiyecek ve İçecek	0,045
Shenzhen		South African	
Polarite Değeri	0,433	Polarite Değeri	0,342
Paranın Değeri	0,263	Yiyecek ve İçecek	0,207
Koltuk Konforu	0,142	Ayak Koyma Alanı	0,142
Temizlik	0,052	Temizlik	0,125
Ayak Koyma Alanı	0,049	Genel Skor	0,061
Check-in ve Biniş	0,030	Check-in ve Biniş	0,043
Müşteri Hizmetleri	0,030	Paranın Değeri	0,034
Swiss		TAP	
Polarite Değeri	0,423	Polarite Değeri	0,491
Ayak Koyma Alanı	0,082	Ayak Koyma Alanı	0,127
Genel Skor	0,082	Genel Skor	0,088
Yiyecek ve İçecek	0,077	Check-in ve Biniş	0,072
Koltuk Konforu	0,072	Uçak İçi Eğlence	0,067
Uçak İçi Eğlence	0,068	Koltuk Konforu	0,049
Müşteri Hizmetleri	0,061	Yiyecek ve İçecek	0,038
Temizlik	0,049	Temizlik	0,026
Check-in ve Biniş	0,048	Müşteri Hizmetleri	0,024
Paranın Değeri	0,038	Paranın Değeri	0,017
Thai		United	
Polarite Değeri	0,429	Polarite Değeri	0,524
Uçak İçi Eğlence	0,109	Koltuk Konforu	0,098
Genel Skor	0,106	Temizlik	0,087
Check-in ve Biniş	0,104	Ayak Koyma Alanı	0,084
Paranın Değeri	0,094	Uçak İçi Eğlence	0,048
Yiyecek ve İçecek	0,082	Genel Skor	0,044
Ayak Koyma Alanı	0,041	Müşteri Hizmetleri	0,033
Koltuk Konforu	0,016	Check-in ve Biniş	0,032
Müşteri Hizmetleri	0,009	Paranın Değeri	0,030
Temizlik	0,008	Yiyecek ve İçecek	0,019

Tablo 24’de polariteye göre kriterlerin önem dereceleri hesaplandığında müşteri yorumlarından elde edilen pozitif, negatif ve nötr sınıfların sayısal değeri olan polarite değerinin her hava yolu için belirleyici olduğu görülmektedir. Ardından ayak koyma alanı, genel skor, check-in ve biniş, genel skor, paranın değeri kriterlerinin hava yollarını

Şekil 18’de Aegean hava yolu için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin “uçuş (flying), deneyim (experience), uçuş (flying), içecek (drinks), seyahat (travel), rahat (comfortable), ücretsiz (free), havalimanı (airport), mükemmel (excellent)” olduğu görülmektedir. Şekil 19’da Asiana hava yolu için oluşturulan kelime bulutu verilmiştir.



Şekil 19. Asiana Hava Yolu İçin Kelime Bulutu

Şekil 19’da Asiana hava yolu için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin “kabin (crew), görevli (attendants), birinci sınıf (business class), ekonomi (economy), gezi (trip), havaalanı (airport), kontrol (check), yemek (meal), mükemmel (excellent), salon (lounge)” olduğu görülmektedir. Şekil 20’de Austrian hava yolu için oluşturulan kelime bulutu verilmiştir.

(premium economy), deneyim (experience), mükemmel (excellent), deęişim (change)” olduęu görünmektedir. Şekil 24’de China hava yolu için oluşturulan kelime bulutu verilmiştir.



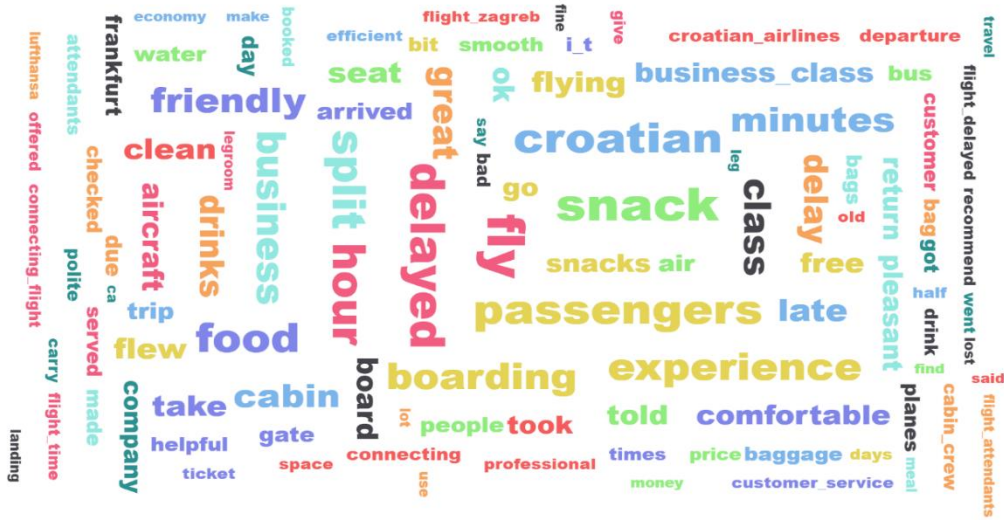
Şekil 24. China Hava Yolu İçin Kelime Bulutu

Şekil 24’de China hava yolu için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin “eğlence (entertainment), birinci sınıf (business), kötü (bad), bagaj (luggage), otel (hotel), yemek (meal), yolcu (passengers), fiyat (price), müşteri hizmetleri (customer service), deneyim (experience), gecikmeli (delayed)” olduęu görünmektedir. Şekil 25’de Copa hava yolu için oluşturulan kelime bulutu verilmiştir.



Şekil 25. Copa Hava Yolu İçin Kelime Bulutu

Şekil 25’de Copa hava yolu için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin “Panama şehri (panama city), birinci sınıf (business class), uçuş (flying), bagaj (luggage), uçuş kapı (gate), bilet (ticket), rahat (comfortable), ücretsiz (free), müşteri hizmetleri (customer service), bağlantı (connection), koltuk (seat)” olduğu görülmektedir. Şekil 26’da Croatia hava yolu için oluşturulan kelime bulutu verilmiştir.



Şekil 26. Croatia Hava Yolu İçin Kelime Bulutu

Şekil 26’da Croatia hava yolu için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin “gecikmeli (delayed), atıştırılacak (snack), uçuş (flying), birinci sınıf (business class), yiyecek (food), yolcu (passengers), rahat (comfortable), gezi (trip), müşteri hizmetleri (customer service), bağlantı (connecting), alkol (smooth)” olduğu görülmektedir. Şekil 27’de Egypt hava yolu için oluşturulan kelime bulutu verilmiştir.

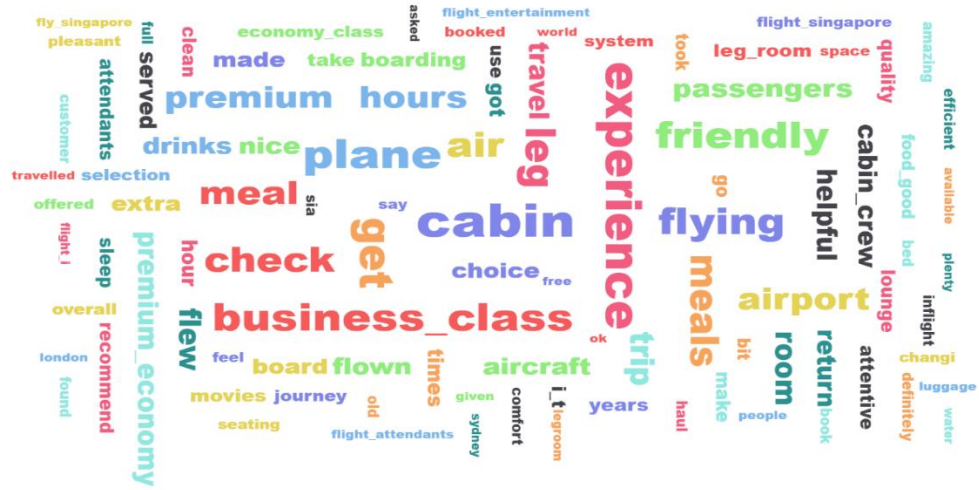


Şekil 27. Egypt Hava Yolu İçin Kelime Bulutu



Şekil 35. Shenzhen Hava Yolu İçin Kelime Bulutu

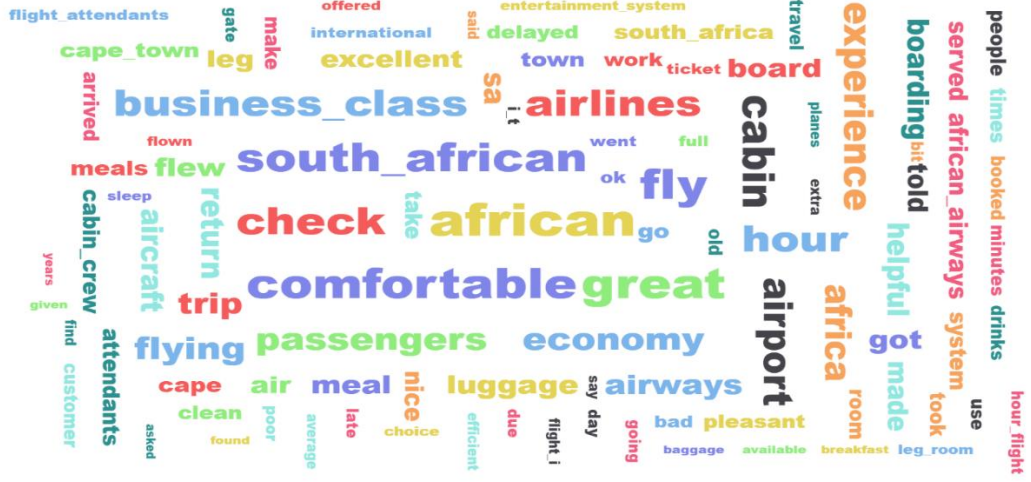
Şekil 35’de Shenzhen hava yolu için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin “deneyim (experience), güzel hava yolu (air nice), uluslararası (international), koltuk (seats), sınıf (class), kabin ekibi (crew), yolcu (passengers), arkadaş canlısı (friendly), gecikmeli (delayed), ekonomi sınıfı (economy), eğlence (entertainment)” olduğu görülmektedir. Şekil 36’da Singapore hava yolu için oluşturulan kelime bulutu verilmiştir.



Şekil 36. Singapore Hava Yolu İçin Kelime Bulutu

Şekil 36’da Singapore hava yolu için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin “eğlence (experience), birinci sınıf (business class), yemek (meals), kabin (cabin), sınıf (class), kontrol (check), avantajlı ekonomi sınıfı (premium economy), arkadaş canlısı (friendly), içecek (drinks), yardımsever (helpful), tavsiye edilen

(recommend)” olduğu görünmektedir. Şekil 37’de South African hava yolu için oluşturulan kelime bulutu verilmiştir.



Şekil 37. South African Hava Yolu İçin Kelime Bulutu

Şekil 37’de South African hava yolu için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin “deneyim (experience), birinci sınıf (business class), kontrol (check), kabin (cabin), harika (great), rahat (comfortable), ekonomi sınıfı (economy), Afrika (African), bagaj (luggage), ortalama (average), biniş (boarding)” olduğu görünmektedir. Şekil 38’de Swiss hava yolu için oluşturulan kelime bulutu verilmiştir.



Şekil 38. Swiss Hava Yolu İçin Kelime Bulutu

Şekil 38’de Swiss hava yolu için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin “deneyim (experience), birinci sınıf (business class), güzel (nice), kabin (cabin), mükemmel (excellent), rahat (comfortable), ekonomi sınıfı (economy), arkadaş

ekonomi sınıfı (economy), mükemmel (excellent), arkadaş canlısı (friendly)” olduğu görülmektedir. Şekil 41’de United hava yolu için oluşturulan kelime bulutu verilmiştir.



Şekil 41. United Hava Yolu İçin Kelime Bulutu

Şekil 41’de Thai hava yolu için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin “deneyim (experience), ekstra (extra), güzel (nice), uçuş görevlisi (flight attendants), birinci sınıf (business), yolcu (passengers), rahat (comfortable), eğlence (entertainment), müşteri hizmetleri (customer service), bagaj (luggage), arkadaş canlısı (friendly)” olduğu görülmektedir.

Kelime bulutu, belirli bir metin veya konuda kullanılan kelimelerden oluşan ve her kelimenin boyutunun sıklığını veya önemini belirttiği bir resim veya görsel olduğu görülmektedir. Kullanıcılar veya işletmeler, bu görsel ile yardımıyla hızlı karar alabilmektedir. Araştırmada hava yolları için yapılan yorumların görselleştirilmesi yapılmıştır. Kelime bulutu ile yapılan bu analiz ilerde yapılacak çalışmalara temel olması niteliğinde düşünülmektedir.

SONUÇ

Veri ve metin madenciliği sayesinde işletmeler, karmaşık ve büyük veri kümelerini basit, hızlı ve etkili bir şekilde analiz etmektedirler. Aynı zamanda şirketler, manuel ve tekrar eden görevlerinin bir kısmını azaltmak için bu güçlü araçtan yararlanmaktadır. Böylelikle ekiplerine değerli zamanlar kazandırmakta ve müşteri destek temsilcilerinin daha iyi yaptıkları işe odaklanmalarına izin vermektedirler. İşletme müşterilerinin elektronik ortamda neyi beğendiğini veya neyi eleştirdiğini anlamak isteyebilmektedir. Bu bağlamda veri/metin madenciliği teknikleri müşteri yorumlarında ortaya çıkan en popüler konuları ve insanların işletmeler hakkında ne düşündüğünü tespit edilmesinde yardımcı olmaktadır. Buradan da işletmeler, eksik ve zayıf yönlerini görerek kendi hizmet standartlarını düzenleyip rekabet durumundaki rakiplerine karşı avantaj sağlayabilirler.

Metin madenciliği tekniklerinden Duygu Analizi ile müşteri yorumlarının ‘pozitif mi, negatif mi yoksa nötr mü?’ sorularına cevap verilebilmektedir. Ayrıca belirli konuyla ilgili olarak müşteriler tarafından bahsedilen ana anahtar kelimeleri de bulmaya yardımcı olmaktadır. Veri madenciliği tekniklerinden özellikle Karar Ağaçları, DVM, Naive Bayes ve Derin Öğrenme Algoritmaları kullanılarak sınıflandırma ve tahmin yapılabilmektedir. Özetle veri/metin madenciliği, işletmelerin verilerinden en iyi şekilde yararlanmasına yardımcı olmaktadır ve bu durum daha iyi veri odaklı iş kararlarına yol açmaktadır.

İş dünyasında doğru karar vermek ve uygun politikalar geliştirmek için hız faktörünün önemli olduğu bir sektörde hava yolu taşımacılığıdır. Hava yolu taşımacılığı, hız avantajının yanı sıra sürekli artan kapasite, uzun menzil ve emniyet özellikleriyle çağdaş bir ulaşım sistemidir. Hava yolu endüstrisi özel coğrafi yollara ihtiyaç duymadan, yeryüzüyle en az temas yoluyla, yolcuları ve kargoları çıkış noktasından varış noktasına ulaştırmaktadır. Ulaştırma türleri içinde en son keşfedilen ve en hızlı gelişen hava yolu taşımacılığı, uzun mesafede yolcu ve acil değerli kargo taşımacılığında neredeyse tek seçenektir. Coğrafi sahası geniş, yerleşim yerleri dağınık, doğal koşulları yüzey ulaşım ağı için elverişli olmayan yerlerde hava yolu taşımacılığı bütünleştirici bir faktör olarak görülmektedir.

Seyahatlerinde uçak yolculuğunu tercih eden kişilerin ihtiyaçlarına daha kusursuz bir şekilde yanıt verebilmek ve hizmet kalitesini her geçen gün bir adım daha öteye taşıyabilmek amacı ile kurulmuş olan Star Alliance 26 hava yolu üyesi ile bu araştırmada

ele alınmıştır. Hava yollarını değerlendirmek ve karşılaştırmak için seyahat inceleme sitelerinden olan TripAdvisor'daki müşterilerin yapmış olduğu İngilizce yorumlar ve sayısal skorlar kriter olarak alınmıştır. Bunlar: genel skor, ayak koyma alanı, koltuk konforu, uçak içi eğlence (wifi, tv, filmler vb.), müşteri hizmetleri, fiyat/fayda oranı, temizlik, check-in ve biniş, yiyecek ve içecek'tir. Kriterlere verilen değerlendirme puanları 10-50 arasında değişmektedir. Bu kriterlere ek olarak yorumlardan elde edilen polarite (pozitif, negatif, nötr) ve polarite değeri (0-1), uçuşlar, uçuş güzergahları (Avrupa, Uluslararası, Meksika, Amerika, Afrika vb.) ve kabin sınıfı (ekonomi, avantajlı ekonomi, birinci sınıf, öncelikli sınıf) kriterleri uygulamada kullanılmıştır.

Araştırmada, ele alınan kriterler bağlamında veri ve metin madenciliği tekniklerinden Duygu Analizi, DVM, Naive Bayes, Derin Öğrenme ve Karar Ağacı Algoritmaları kullanılmıştır. Buna ek olarak, kriterler arasındaki ilişki ve etkiyi incelemek amacıyla da korelasyon ve regresyon analizi yapılmıştır. İlk olarak Duygu Analizi ile müşteri yorumlarını temel alınarak hava yollarının polarite (pozitif, negatif, nötr) ve polarite değerleri çıkarılmıştır. Korelasyon analizi yapılarak genel skor kriterinin diğer kriterler ile ilişkisinin daha yüksek değerlere sahip olduğu görülmüştür. Bu doğrultuda genel skor ile sırasıyla müşteri ilişkileri (0,837), fiyat/fayda oranı (0,809), check-in ve biniş (0,731), temizlik (0,719), yiyecek ve içecek (0,709), koltuk konforu (0,701), ayak koyma alanı (0,634) ve uçak içi eğlence (0,610); ayak koyma alanı ile koltuk konforu (0,826); müşteri hizmetleri ile fiyat/fayda oranı (0,765); müşteri hizmetleri ile check-in ve biniş (0,714); fiyat/fayda oranı ile temizlik (0,703) pozitif yönlü bir ilişkiye sahiptir. Kriterlerin ilişki değerleri genel skor ile daha çok bağlantılı olduğundan regresyon analizi yapılırken genel skor kriteri bağımlı değişken ve diğer kriterler bağımsız değişken olarak seçilmiştir. Kriterlerin genel skoru açıklamadaki çoklu R değeri 0,888 dir. Genel skor kriterini en iyi açıklayan bağımsız kriterler sırasıyla müşteri hizmetleri (0,317), fiyat/fayda oranı (0,236), check-in ve biniş (0,155), koltuk konforu (0,124), yiyecek ve içecek (0,089), temizlik (0,054), uçak içi eğlence (0,038) ve ayak koyma alanı (0,029)'dir. Analiz edilen tüm değerlendirme skorları arasında en önemli olan iki kriterin sırasıyla müşteri hizmetleri ve fiyat/fayda oranı olduğu görülmüştür.

Duygu Analizinden çıkan polarite etiketlerinin sınıflandırma tahmini için DVM, Naive Bayes ve Derin Öğrenme Algoritmaları kullanılmıştır. DVM algoritmaları ile yapılan performans doğruluk sonucunda EVA (% 89,81) ilk sıradayken, Shenzhen

(%58,82) son sırada yer almaktadır. THY (%86,51) ise, altıncı sırada yer almaktadır. Naive Bayes Algoritmaları ile yapılan performans doğruluk sonucunda EVA (% 87,49) ilk sıradayken, Shenzhen (%60,78) son sırada yer almaktadır. THY (%83,52) ise, altıncı sırada yer almaktadır. Derin Öğrenme Algoritmaları ile yapılan performans doğruluk sonucunda EVA (% 90,97) ilk sıradayken, Shenzhen (%62,75) son sırada yer almaktadır. THY (%85,96) ise, yedinci sırada yer almaktadır. Üç algoritmanın performans değerlendirme sonucunda genel olarak en iyi algoritmanın Derin Öğrenme Algoritması olduğu görülmektedir.

Duygu Analizinden çıkan polarite değerinin hata tahmini için DVM ve Derin Öğrenme Algoritmaları kullanılmıştır. DVM algoritmaları ile yapılan tahminde hata değeri olan en düşük olan hava yolu EVA (0,156) iken, en yüksek hata değeri olan hava yolu LOT (0,197)'dur. THY (0,167) ise, yedinci sırada yer almaktadır. Derin Öğrenme Algoritması ile yapılan tahminde hata değeri olan en düşük olan hava yolları EVA ve Aegean (0,156) iken, en yüksek hata değeri olan hava yolu Shenhzen (0,204)'dir. THY (0,142) ise, yedinci sırada yer almaktadır. İki algoritmanın RMSE değeri sonucunda genel olarak en iyi algoritmanın Derin Öğrenme Algoritması olduğu görülmektedir.

Kriterlerin önem derecelerinin belirlenmesi için Karar Ağaçları Algoritması kullanılmıştır. Kriterlerin önem dereceleri için uçuş güzergahları, uçuşlar, kabin sınıfı ve polarite (pozitif, negatif, nötr) etiketleri baz alınarak analiz edilmiştir. Uçuş güzergahı temel alındığında THY için sırasıyla, polarite değeri (0,375), paranın değeri (0,222) en yüksek önem değerine sahip kriter iken, koltuk konforu (0,002) en düşük önem değerine sahiptir. EVA hava yolu için, müşteri hizmetleri (0,290) en yüksek önem değerine sahip iken, paranın değeri (0,004) en düşük önem değerine sahiptir. Uçuş yapılan yer baz alındığında THY için, polarite değeri (0,487) olarak en yüksek önem değerine sahip iken, check-in ve biniş (0,013) en düşük önem değerine sahiptir. EVA hava yolu için, polarite değeri (0,388) en yüksek önem değerine sahip iken, koltuk konforu (0,031) en düşük önem değerine sahiptir. Kabin sınıfı baz alındığında THY için polarite değeri (0,505) en yüksek önem değerine sahip iken, EVA hava yolu için, polarite değeri (0,377) en yüksek önem değerine sahip iken, uçak içi eğlence (0,004) en düşük önem değerine sahiptir. Polarite baz alındığında THY için, polarite değeri (0,370) en yüksek önem değerine sahip iken, uçak içi eğlence (0,019) en düşük önem değerine sahiptir. EVA hava yolu için,

polarite değeri (0,251) en yüksek önem değerine sahip iken, temizlik (0,005) en düşük önem değerine sahiptir.

Araştırmanın son bölümünde, her bir hava yolu için kelime bulutu analizi yapılmıştır. Müşterilerin yaptığı yorumlardan elde edilen verileri derleyerek çıkarılan kelimelerden en çok kullanılan 100 kelime görselleştirilmiştir. Hava yollarına ilişkin olarak yapılan analizde kelimelerin sıklığına göre şu şekilde sıralandığı görülmektedir; “rahat, deneyim, birinci sınıf, arkadaş canlısı, mükemmel, kabin ekibi, gezi, bagaj, gecikmeli, eğlence, fiyat, müşteri hizmetleri, temizlik, uluslararası, kontrol, bilet, yemek, yardımsever, ücretsiz, koltuk, ekonomi sınıfı, yolcu”. Kelime bulutu ile yapılan bu analiz, ileride yapılacak çalışmalara temel olması niteliği taşıyacağı düşünülmektedir.

İleriki araştırmalarda, veri/metin madenciliği teknikleri kullanarak TripAdvisor seyahat sitesi dışında farklı veriler de kullanılarak karşılaştırma yapılabilir. Veriler sohbetler, e-postalar, anketler, elektronik tablolar, veritabanları, sosyal medya, inceleme siteleri, haber kaynakları ve diğer web sitelerinden bilgiler olabilir. Araştırmada kullanılan kriterlerin yanı sıra farklı kriterler de eklenebilir ve farklı algoritmaların kullanımıyla analiz zenginleştirilebilir. Ayrıca bu araştırmada, global bir bakış açısı sağlamak adına İngilizce dili esas alınmıştır. Müşterilerin işletmelere yönelik yaptığı İngilizce dildeki yorumlar dışında diğer dillerde yapılan yorumlar da analiz edilip araştırmaya dâhil edilebilir. Araştırmada kullanılan algoritmalar irdelendiğinde veri boyutu büyüdüğünde genel olarak derin öğrenmenin daha iyi sonuç verdiğini ve ileriki araştırmalara katkı sunacağı düşünülmektedir.

KAYNAKLAR

- Afzaal, M., Usman, M., Fong, A. C., & Fong, S. (2019). "Multiaspect-based Opinion Classification Model for Tourist Reviews", *Expert Systems*, 36(2), e12371, 1-24.
- Agarap, A. F. (2018). "Deep Learning Using Rectified Linear Units (ReLU)", *arXiv preprint arXiv:1803.08375*.
- Aguinis, H., Forcum, L. E. & Joo, H. (2013). "Using Market Basket Analysis in Management Research", *Journal of Management*, 39(7), 1799-1824.
- Ahn, J. Y., Kim, S. K., & Han, K. S. (2003). "On the Design Concepts for CRM System", *Industrial Management & Data Systems*, 324-331.
- Alavi, M., & Leidner, D. E. (2001). "Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues", *MIS Quarterly*, 107-136.
- Alexa, https://www.alexandata.com/siteinfo/tripadvisor.com#section_traffic (02.05.2020).
- An, Y., Sun, S., & Wang, S. (2017, May). "Naive Bayes Classifiers for Music Emotion Classification Based on lyrics", *In 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, IEEE, 635-638
- Ang, S. L., Ong, H. C., & Low, H. C. (2016). "Classification using the General Bayesian Network", *Pertanika J. Sci. Technol*, 24(1), 205-211.
- Aydın, B. (2016). "Sosyal Medyada Restoran İmajı: Tripadvisor Örneği", *Disiplinlerarası Akademik Turizm Dergisi*, 1(1), 13-30.
- Aydoğan, İ. & Zırhlıoğlu, G. (2018). "Öğrenci Başarılarının Yapay Sinir Ağları ile Kestirilmesi", *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 15(1), 577-610.
- Atan, S. (2016). *Metin Madenciliği ile Sentiment Analizi ve Borsa İstanbul Uygulaması*, (Doktora Tezi), Sosyal Bilimler Enstitüsü, Ankara Üniversitesi, Ankara.
- Balbi, S., Misuraca, M., & Scepi, G. (2018). "Combining Different Evaluation Systems on Social Media for Measuring User Satisfaction", *Information Processing & Management*, 54(4), 674-685.
- Barros, R. C., De Carvalho, A. C. ve Freitas, A. A. (2015). *Automatic Design of Decision-Tree Induction Algorithms*, Springer International Publishing.
- Barushka, A., & Hajek, P. (2019, May). "Review Spam Detection Using Word Embeddings and Deep Neural Networks", *In IFIP International Conference on Artificial Intelligence Applications and Innovations*, Springer, Cham, 340-350
- Baxt, W. G. (1991). "Use of an Artificial Neural Network for the Diagnosis of Myocardial Infarction", *Annals of Internal Medicine*, 115(11), 843-848.
- Belisário, L. B., Ferreira, L. G., & Pardo, T. A. S. (2020, March). "Evaluating Methods of Different Paradigms for Subjectivity Classification in Portuguese", *In International Conference on Computational Processing of the Portuguese Language*, Springer, Cham, 261-269.

- Benoit, G. (2002). "Data Mining", *Annual Review of Information Science and Technology*, 36, 265–310.
- Berry, M. J. & Linoff, G. (1997). *Data Mining Techniques: for Marketing, Sales, and Customer Support*, John Wiley & Sons, Inc.
- BholaneSavita, D., & Gore, D. (2016). "Sentiment Analysis on Twitter Data Using Support Vector Machine", *International Journal of Computer Science Trends and Technology (IJCST)*–4, 365-370.
- Bloehdorn, S., Cimiano, P. & Hotho, A. (2006). *Learning Ontologies to Improve Text Clustering and Classification*, In *From Data and Information Analysis To Knowledge Engineering* (pp. 334-341). Springer, Berlin, Heidelberg.
- Boiy, E., & Moens, M. F. (2009). "A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts", *Information Retrieval*, 12(5), 526-558.
- Boztuğ, Y. ve Reutterer, T. (2008). "A Combined Approach for Segment-Specific Market Basket Analysis", *European Journal of Operational Research*, 187(1), 294-312.
- Bre, F., Gimenez, J. M. & Fachinotti, V. D. (2018). "Prediction of Wind Pressure Coefficients on Building Surfaces Using Artificial Neural Networks", *Energy and Buildings*, 158, 1429-1441.
- Brüderl, J. (2003). *Applied Regression Analysis Using Stata*. Skriptum Zur Vorles. Im Frühjahr.
- Burns, N., Bi, Y., Wang, H., & Anderson, T. (2019). "Enhanced Twofold-LDA Model for Aspect Discovery and Sentiment Classification", *International Journal of Knowledge-Based Organizations (IJKBO)*, 9(4), 1-20.
- Camastra, F. & Vinciarelli, A. (2015). *Machine Learning for Audio, Image and Video Analysis: Theory and Applications*, Springer.
- Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (Eds.). (2017). *A Practical Guide to Sentiment Analysis*, Cham, Switzerland: Springer International Publishing.
- Chang, L. Y. (2005). "Analysis of Freeway Accident Frequencies: Negative Binomial Regression Versus Artificial Neural Network", *Safety Science*, 43(8), 541-557.
- Chen, I. J., & Popovich, K. (2003). "Understanding Customer Relationship Management (CRM)", *Business Process Management Journal*, 672-688.
- Chen, P. Y., Hong, Y., & Liu, Y. (2017). "The Value of Multidimensional Rating Systems: Evidence from a Natural Experiment and Randomized Experiments". *Management Science*, 64(10), 4629-4647.
- Chen, Y. L., Tang, K., Shen, R. J., & Hu, Y. H. (2005). "Market Basket Analysis in a Multiple Store Environment", *Decision Support Systems*, 40(2), 339-354.
- Cherian, V., & Bindu, M. S. (2017). Heart Disease Prediction Using Naïve Bayes Algorithm and Laplace Smoothing Technique. *International Journal of Computer Science Trends and Technology (IJCST)*, 5(2), 68-73.
- Chowdhury, G. G. (2003). "Natural Language Processing", *Annual Review of Information Science and Technology*, 37(1), 51-89.

- Colladon, A. F., Guardabascio, B., & Innarella, R. (2019). "Using Social Network and Semantic Analysis to Analyze Online Travel Forums and Forecast Tourism Demand", *Decision Support Systems*, 113075.
- Daniel, K. W., Tripathi, N. K., & Honda, K. (2003). "Artificial Neural Network Analysis of Laboratory and in Situ Spectra for the Estimation of Macronutrients in Soils of Lop Buri (Thailand)", *Soil Research*, 41(1), 47-59.
- Dashtipour, K., Poria, S., Hussain, A., Cambria, E., Hawalah, A. Y., Gelbukh, A., & Zhou, Q. (2016). "Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques", *Cognitive Computation*, 8(4), 757-771.
- Dave, D. K. & Dave, P. K. (2012). "Knowledge Discovery in Databases (KDD) in Online Educational System Through LON-CAPA System", *System*, 1(4), 15-19.
- Dave, K., Lawrence, S. & Pennock, D. M. (2003, May). "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews", *In Proceedings of the 12th International Conference on World Wide Web* (pp. 519-528).
- Day, M. Y. & Lin, Y. D. (2017, August). "Deep Learning for Sentiment Analysis on Google Play Consumer Review", *In 2017 IEEE International Conference On Information Reuse And Integration (IRI), IEEE*, 382-388.
- De Bruijn, B. & Martin, J. (2002). "Getting to the (C)ore of Knowledge: Mining Biomedical Literature", *International Journal of Medical Informatics*, 67(1-3), 7-18.
- De Mántaras, R. L. (1991). "A Distance-Based Attribute Selection Measure for Decision Tree Induction", *Machine Learning*, 6(1), 81-92.
- Denicia-Carral, C., Montes-y-Gómez, M., Villaseñor-Pineda, L. & Hernández, R. G. (2006, August). *A Text Mining Approach for Definition Question Answering*. *In International Conference on Natural Language Processing (in Finland)* (76-86). Springer, Berlin, Heidelberg.
- Dhingra, K., & Yadav, S. K. (2017). "Spam Analysis of Big Reviews Dataset Using Fuzzy Ranking Evaluation Algorithm and Hadoop", *International Journal of Machine Learning and Cybernetics*, 1-20.
- Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013, January). "Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality", *In 2013 46th Hawaii International Conference on System Sciences, IEEE*, 3119-3128.
- Ergün, K. (2012). *Metin Madenciliği Yöntemleri ile Ürün Yorumlarının Otomatik Değerlendirilmesi*, (Doktora Tezi), Fen Bilimleri Enstitüsü, Sakarya Üniversitesi, Sakarya.
- Erişti, H., Uçar, A., & Demir, Y. (2007). "Destek Vektör Makineler ve Dalgacık Dönüşümü Kullanarak Güç Sistemlerinde Geçici Rejim Bozulmalarının Sınıflandırılması", *TMMOB Elektrik Mühendisleri Odası*, 1- 4.
- Ertunc, H. M., & Hosoz, M. (2006). "Artificial Neural Network Analysis of a Refrigeration System with an Evaporative Condenser", *Applied Thermal Engineering*, 26(5-6), 627-635.

- Fayyad, U. M. (1996). "Data Mining And Knowledge Discovery in Databases: Applications in Astronomy and Planetary Science", *American Association for Artificial Intelligence*, Menlo Park, CA (United States), 1590-1592.
- Feldman, R., Fresko, M., Hirsh, H., Aumann, Y., Liphsta, O., Schler, Y. & Rajman, M. (1998). "Knowledge Management: A Text Mining Approach.", In Proc. *The 2nd Int. Conf, On Practical Aspects of Knowledge Management (PAKM98)*, 1-9.
- Gal-Tzur, A., Rechavi, A., Beimel, D., & Freund, S. (2018). "An Improved Methodology for Extracting Information Required for Transport-Related Decisions from Q&A Forums: A Case Study of TripAdvisor", *Travel Behaviour and Society*, 10, 1-9.
- Gawrysiak, P., Rybinski, H., Gajda, D., & Golebski, M. (2004). "Extending Open Source Software Solutions for CRM Text Mining", *In ICWI*, 869-872.
- Ghobadian, A., Speller, S., & Jones, M. (1994). "Service Quality", *International Journal of Quality & Reliability Management*, 43-66.
- Glorot, X., Bordes, A., & Bengio, Y. (2011, June). "Deep Sparse Rectifier Neural Networks", *In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, 315-323.
- Go, A., Bhayani, R., & Huang, L. (2009). "Twitter Sentiment Classification Using Distant Supervision", *CS224N Project Report, Stanford*, 1(12).
- Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning*, MIT press.
- Grąbczewski, K. (2014). *Meta-Learning In Decision Tree Induction*, Springer International Publishing.
- Gupta, V. & Lehal, G. S. (2009). "A Survey of Text Mining Techniques and Applications", *Journal of Emerging Technologies in Web Intelligence*, 1(1), 60-76.
- Haddaway, N. R. (2015). "The Use of Web-Scraping Software in Searching for Grey Literature", *Grey Journal*, 11(3), 186-190.
- Han, J., Pei, J. & Kamber, M. (2011). *Data Mining: Concepts and Techniques*, Elsevier.
- Han, S., & Anderson, C. K. (2018, November). "Estimating the Effect of Social Influence on Subsequent Reviews", *In INFORMS International Conference on Service Science, Springer, Cham*, 231-238.
- Harisinghaney, A., Dixit, A., Gupta, S., & Arora, A. (2014, February). "Text and Image Based Spam Email Classification Using KNN, Naive Bayes and Reverse DBSCAN Algorithm", *In 2014 International Conference on Reliability Optimization and Information Technology (ICROIT), IEEE*, 153-155
- Hazarika, D., Poria, S., Gorantla, S., Cambria, E., Zimmermann, R., & Mihalcea, R. (2018). "Cascade: Contextual Sarcasm Detection in Online Discussion Forums", *In Proceedings of the 27th International Conference on Computational Linguistics*, 1837-1848.
- He, W., Zha, S. & Li, L. (2013). "Social Media Competitive Analysis and Text Mining: A Case Study in the Pizza Industry", *International Journal of Information Management*, 33(3), 464-472.
- Hearst, M. (2003). *What is Text Mining*, SIMS, UC Berkeley.

- Hipp, J., Güntzer, U., & Nakhaeizadeh, G. (2000). "Algorithms for Association Rule Mining a General Survey and Comparison", *ACM Sigkdd Explorations Newsletter*, 2(1), 58-64.
- Hogenboom, A., Bal, D., Frasinca, F., Bal, M., de Jong, F., & Kaymak, U. (2013, March). "Exploiting Emoticons In Sentiment Analysis", *In Proceedings of the 28th Annual ACM Symposium on Applied Computing*, 703-710.
- Hong, S. S., Lee, W. & Han, M. M. (2015). "The Feature Selection Method Based on Genetic Algorithm for Efficient of Text Clustering and Text Classification", *Int. J. Advance Soft Compu. Appl*, 7(1), 22-40.
- Hu, M., & Liu, B. 2004. "Mining and Summarizing Customer Reviews", *In KDD*, 168-177.
- Hussein, S. M., Ali, F. H. M., & Kasiran, Z. (2012, May). "Evaluation Effectiveness of Hybrid IDs Using Snort with Naive Bayes to Detect Attacks", *In 2012 Second International Conference on Digital Information and Communication Technology and it's Applications (DICTAP)*, IEEE, 256-260.
- Işık, N. (2019). *Metin Madenciliği Yöntemleri ile E-Ticaret Markalarına Yönelik Sosyal Medya Yorumlarının Analizi*, (Yüksek Lisans Tezi), Sosyal Bilimler Enstitüsü, Marmara Üniversitesi, İstanbul.
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). "Data Clustering: a Review", *ACM Computing Surveys (CSUR)*, 31(3), 264-323.
- Joachims, T. (1998, April). *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*, In European Conference on Machine Learning, Springer, Berlin, Heidelberg.
- Karamanlı, E. (2019). *Makine Öğrenmesi Algoritmaları Kullanarak, Metin Madenciliği ve Duygu Analizi ile Müşteri Deneyiminin Geliştirilmesi*, (Yüksek Lisans Tezi), Sosyal Bilimler Enstitüsü, İstanbul Üniversitesi, İstanbul.
- Kharde, V. & Sonawane, P. (2016). "Sentiment Analysis of Twitter Data: a Survey of Techniques", *arXiv preprint arXiv:1601.06971*, 5-15.
- Kim, K. G. (2016). "Book Review: Deep learning", *Healthcare Informatics Research*, 22(4), 351-354.
- Krallinger, M. & Valencia, A. (2005). "Text Mining and Information-Retrieval Services for Molecular Biology", *Genome Biology*, 6(7), 224.
- Kucher, K., Paradis, C., & Kerren, A. (2018, February). "The State of the Art in Sentiment Visualization", *In Computer Graphics Forum*, 37 (1), 71-96.
- Kuhamanee, T., Talmongkol, N., Chaisuriyakul, K., San-Um, W., Pongpisuttinun, N. & Pongyupinpanich, S. (2017, July). "Sentiment Analysis of Foreign Tourists to Bangkok Using Data Mining Through Online Social Network", *In 2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*, IEEE. 1068-1073.
- Küçük, M. A. (2020). *Metin Madenciliği Tabanlı Bildirim Takip Sistemi*, (Yüksek Lisans Tezi), Fen Bilimleri Enstitüsü, Pamukkale Üniversitesi, Denizli.

- Lavine, B. K. & Mirjankar, N. (2006). "Clustering and Classification of Analytical Data". *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, 1-20.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). "Deep Learning", *Nature*, 521(7553), 436-444.
- Li, H. (2017). "Deep Learning for Natural Language Processing: Advantages and Challenges", *National Science Review*, 5(1), 24-26.
- Li, W., & Chen, H. (2014, September). "Identifying Top Sellers in Underground Economy Using Deep Learning-Based Sentiment Analysis", *In 2014 IEEE Joint Intelligence and Security Informatics Conference, IEEE*, 64-67.
- Libereurope, <https://libereurope.eu/wpcontent/uploads/Text%20and%20Data%20Mining%20Factsheet.pdf>, (11.11.2019).
- Liddy, E. (2001). *Natural Language Processing*, 2nd. Encyclopedia of Library and Information Science, Marcel Decker.
- Liu, B. (2012). "Sentiment Analysis and Opinion Mining", *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
- Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments and Emotions*, Cambridge University Press.
- Liu, B., Blasch, E., Chen, Y., Shen, D., & Chen, G. (2013, October). "Scalable Sentiment Classification for Big Data Analysis Using Naive Bayes Classifier", *In 2013 IEEE International Conference on Big Data, IEEE*, 99-104
- Liu, H., & Cocea, M. (2017). "Semi-random Partitioning of Data into Training and Test Sets in Granular Computing Context", *Granular Computing*, 2(4), 357-386.
- Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013, June). "Rectifier Nonlinearities Improve Neural Network Acoustic Models", *In Proc. ICML*, 30 (1), 3-8.
- Makhabel, B., Mishra P., Danneman, N. & Heimann, R. (2017). *R: Mining Spatial, Text, Web and Social Media Data*, Packt Publishing.
- Medhat, W., Hassan, A., & Korashy, H. (2014). "Sentiment Analysis Algorithms and Applications: A Survey", *Ain Shams Engineering Journal*, 5(4), 1093-1113.
- Melek, C. (2012). *Metin Madenciliği Teknikleri ile Şirketlerin Vizyon İfadelerinin Analizi (Doktora Tezi)*, DEÜ Sosyal Bilimleri Enstitüsü, Kütahya.
- Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C. & Hurdle, J. F. (2008). "Extracting Information from Textual Documents in the Electronic Health Record: A Review of Recent Research", *Yearbook of Medical Informatics*, 17(01), 128-144.
- Mikheev, A., Moens, M. ve Grover, C. (1999, June). "Named Entity Recognition without Gazetteers", *In Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics*, 1-8
- Mirkin, B. (2005). *Clustering for Data Mining: A Data Recovery Approach*, Chapman & Hall – CRC.
- Mullen, T., & Collier, N. (2004, July). "Sentiment Analysis Using Support Vector Machines with Diverse Information Sources", *In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 412-418.

- Murakami, K. H. (2018). "A Comparison of Destination Images from Three Different Perspectives", *Journal of Global Tourism Research*, 3(2), 107-114.
- Nahm, U. Y. & Mooney, R. J. (2002, March). "Text Mining with Information Extraction". *In Proceedings of the AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases*, 60-67.
- Narayanan, V., Arora, I., & Bhatia, A. (2013, October). "Fast and Accurate Sentiment Classification Using an Enhanced Naive Bayes Model", *In International Conference on Intelligent Data Engineering and Automated Learning*, 194-201.
- Nasukawa, T., & Yi, J. (2003, October). "Sentiment Analysis: Capturing Favorability Using Natural Language Processing", *In Proceedings of the 2nd International Conference on Knowledge Capture*, 70-77.
- Neter, J., Wasserman, W., & Kutner, M. H. (1981). *Applied Linear Regression Models*, Richard D, Irwin, Inc.
- Nguyen, T. T., Camacho, D., & Jung, J. E. (2017). "Identifying and Ranking Cultural Heritage Resources on Geotagged Social Media for Smart Cultural Tourism Services", *Personal and Ubiquitous Computing*, 21(2), 267-279.
- Nikfarjam, A., Emadzadeh, E., & Muthaiyah, S. (2010, February). "Text Mining Approaches for Stock Market Prediction", *In 2010 the 2nd International Conference on Computer and Automation Engineering (ICCAE), IEEE*, 4, 256-260.
- Noble, W. S. (2006). "What is a Support Vector Machine?", *Nature Biotechnology*, 24(12), 1565-1567.
- O'Mahony, M. P., & Smyth, B. (2018). "From Opinions to Recommendations", *In Social Information Access Springer, Cham*, 480-509.
- Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). "Sentiment Analysis in Facebook and its Application to E-Learning", *Computers in Human Behavior*, 31, 527-541.
- Özkul, M. (2019). *Metin Madenciliği Yöntemi ile Akıllı Ulaşım Sistemleri Algısının Duygu Analizi İle Ölçülmesi*, (Yüksek Lisans Tezi), Sosyal Bilimler Enstitüsü, Bandırma Onyediy Eylül Üniversitesi, Balıkesir.
- Pal, M., ve Mather, P. M. (2003). "An Assessment of the Effectiveness of Decision Tree Methods for Land Cover Classification", *Remote Sensing of Environment*, 86(4), 554-565.
- Palaniappan, S. & Awang, R. (2008, March). "Intelligent Heart Disease Prediction System Using Data Mining Techniques", *In Computer Systems and Applications, AICCSA, IEEE/ACS International Conference on, IEEE*, 108-115
- Park, S. (2017). "Estimating the Effect of Online Consumer Reviews: an Application of Count Data Models", *In Analytics in Smart Tourism Design, Springer, Cham*, 147-163.
- Poria, S., Hussain, A., & Cambria, E. (2018). *Multimodal Sentiment Analysis*, Cham, Switzerland: Springer.
- Povoda, L., Burget, R., & Dutta, M. K. (2016, June). "Sentiment Analysis Based On Support Vector Machine and Big Data", *In 2016 39th International Conference on Telecommunications and Signal Processing (TSP), IEEE*, 543-545.

- Pozzi, F. A., Fersini, E., Messina, E., & Liu, B. (2016). *Sentiment Analysis in Social Networks*, Morgan Kaufmann.
- Puretskiy, A. A., Shutt, G. L. & Berry, M. W. (2010). "Survey of Text Visualization Techniques", *Text Mining: Applications And Theory*, 105-127.
- Rayana, S., & Akoglu, L. (2015, August). "Collective Opinion Spam Detection: Bridging Review Networks and Metadata", *In Proceedings of the 21th Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 985-994.
- Ren, Q., Xu, F., & Ji, X. (2019). "Use of the Pathfinder Network Scaling to Measure Online Customer Reviews: A Theme Park Study", *Strategic Change*, 28(5), 333-344.
- Richards, J. D., Tudhope, D. & Vlachidis, A. (2015). "Text Mining in Archaeology: Extracting Information from Archaeological Reports", *Mathematics and Archaeology*, 240-253.
- Rintyarna, B. S., Sarno, R., & Fatichah, C. (2019). "Evaluating the Performance of Sentence Level Features and Domain Sensitive Features of Product Reviews on Supervised Sentiment Analysis Tasks", *Journal of Big Data*, 6(1), 84.
- Ristoski, P., Bizer, C., & Paulheim, H. (2015). "Mining the Web of Linked Data with RapidMiner", *Journal of Web Semantics*, 35, 142-151.
- Rojas-Barahona, L. M. (2016). "Deep Learning for Sentiment Analysis", *Language and Linguistics Compass*, 10(12), 701-719.
- Ruspini, E. H. (1969). "A New Approach to Clustering", *Information and Control*, 15(1), 22-32.
- Seth, N., Deshmukh, S. G., & Vrat, P. (2005). "Service Quality Models: a Review", *International Journal of Quality & Reliability Management*, 913-949.
- Sharma, H., Tandon, A., Kapur, P. K., & Aggarwal, A. G. (2019). "Ranking Hotels Using Aspect Ratings Based Sentiment Classification and Interval-Valued Neutrosophic TOPSIS", *International Journal of System Assurance Engineering and Management*, 1-11.
- Sharma, P., Singh, D., & Singh, A. (2015, February). "Classification Algorithms on a Large Continuous Random Dataset Using RapidMiner Tool", *In 2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, IEEE, 704-709.
- Shirani-Mehr, H. (2014). "Applications of Deep Learning to Sentiment Analysis of Movie Reviews", *Technical Report*, 1-8.
- Shouman, M., Turner, T., & Stocker, R. (2011, December). "Using Decision Tree for Diagnosing Heart Disease Patients". *In Proceedings of the Ninth Australasian Data Mining Conference-121*, 23-30.
- Singh, P. K., & Husain, M. S. (2014). "Methodological Study of Opinion Mining and Sentiment Analysis Techniques", *International Journal on Soft Computing*, 5(1), 11-21.
- Sivil Havacılık Genel Müdürlüğü, <http://web.shgm.gov.tr/> (17.01.2021).

- Slamet, C., Andrian, R., Maylawati, D. S. A., Darmalaksana, W., & Ramdhani, M. A. (2018). "Web Scraping and Naive Bayes Classification for Job Search Engine". *In IOP Conference Series: Materials Science and Engineering*, IOP Publishing, 288, (1), 012038.
- Song, S., Kawamura, H., Uchida, J., & Saito, H. (2019). "Determining Tourist Satisfaction from Travel Reviews", *Information Technology & Tourism*, 1-31.
- Song, S., Saito, H., & Kawamura, H. (2018). "Content Analysis of Travel Reviews: Exploring the Needs of Tourists from Different Countries", *In Information and Communication Technologies in Tourism 2018*, Springer, Cham. 93-105
- Sönmez, N. (2017). *Çevrimiçi Yorumların Metin Madenciliği ile Analizi: İstanbul'daki Alışveriş Merkezleri Üzerine Bir Çalışma*, (Doktora Tezi), İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.
- Suykens, J. A., & Vandewalle, J. (1999). "Least Squares Support Vector Machine Classifiers", *Neural Processing Letters*, 9(3), 293-300.
- Taha, A. M., Mustapha, A., & Chen, S. D. (2013). Naive Bayes-Guided Bat Algorithm for Feature Selection", *The Scientific World Journal*, 1-10.
- Tai, Y. J., & Kao, H. Y. (2013, December). "Automatic Domain-Specific Sentiment Lexicon Generation with Label Propagation", *In Proceedings of International Conference on Information Integration and Web-based Applications & Services*, 53-62.
- Tan, A. H. (1999, April). "Text Mining: The State of the Art and the Challenges", *In Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases*, 8, 65-70.
- Tan, F. G. (2018). *Metin Madenciliği Teknikleri ile Sosyal Ağlarda Bilgi Keşfi*, (Yüksek Lisans Tezi), Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.
- Terzic, J., Terzic, E., Nagarajah, R. & Alamgir, M. (2013). *Ultrasonic Fluid Quantity Measurement in Dynamic Vehicular Applications*, Springer International Pu.
- Thompson, P., Batista-Navarro, R. T., Kontonatsios, G., Carter, J., Toon, E., McNaught, J., ... & Ananiadou, S. (2016). "Text Mining The History of Medicine", *PloS one*, 11(1), e0144717,1-33.
- Tiwana, A. (2000). *The Knowledge Management Toolkit: Practical Techniques for Building A Knowledge Management System*, Prentice Hall PTR.
- Tokar, A. S. & Markus, M. (2000). "Precipitation-Runoff Modeling Using Artificial Neural Networks and Conceptual Models", *Journal of Hydrologic Engineering*, 5(2), 156-161.
- Tong, S., & Chang, E. (2001, October). "Support Vector Machine Active Learning For Image Retrieval", *In Proceedings of the Ninth ACM International Conference on Multimedia*, 107-118.
- Tong, S., & Koller, D. (2001). "Support Vector Machine Active Learning With Applications to Text Classification", *Journal of Machine Learning Research*, 2, 45-66.
- Tran, T., Ba, H., & Huynh, V. N. (2019, March). "Measuring Hotel Review Sentiment: An Aspect-Based Sentiment Analysis Approach", *In International Symposium on*

Integrated Uncertainty in Knowledge Modelling and Decision Making, Springer, Cham, 393-405.

TripAdvisor, <https://www.tripadvisor.com>, (03.04.2020).

Tripathi, P., Vishwakarma, S. K., & Lala, A. (2015, December). "Sentiment Analysis of English Tweets Using RapidMiner", In *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, IEEE. 668-672.

Tsao, H. Y., Chen, M. Y., Lin, H. C. K., & Ma, Y. C. (2019). "The Asymmetric Effect of Review Valence on Numerical Rating: a Viewpoint From a Sentiment Analysis of Users of TripAdvisor", *Online Information Review*, 43(2), 283-300.

Turan, F. & Turan, S. K., (2008), "Havaalanlarının Sosyal Etkileri: Sabiha Gökçen Havaalanı Örneği", *Ankara Üniversitesi, Türkiye Coğrafyası Araştırma ve Uygulama Merkezi, 5.Ulusal Coğrafya Sempozyumu*, 161-166.

Slideshare, <https://www.slideshare.net/Vikky991991/text-mining-71746150>, (11.11.2019).

Uçar, A. (2006). *Destek Vektör Makine Tabanlı Bulanık Sistemler, Yeni Bir Gürbüz Sınıflayıcı ve Regresör Tasarımı*, (Doktora Tezi), Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Elazığ.

Vadivukarassi, M., Puviarasan, N., & Aruna, P. (2017). "Sentimental Analysis of Tweets Using Naive Bayes Algorithm", *World Applied Sciences Journal*, 35(1), 54-59.

Weisberg, S. (2005). *Applied Linear Regression*, John Wiley & Sons.

Welc, J., & Esquerdo, P. J. R. (2018). *Applied Regression Analysis for Business*. Springer Books.

Wong, P. C., Whitney, P. & Thomas, J. (1999, October). "Visualizing Association Rules for Text Mining", In *Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis' 99)*, IEEE. 120-123.

Wu, G., He, Y., & Hu, X. (2018). "Entity Linking: an Issue to Extract Corresponding Entity with Knowledge Base", *IEEE Access*, 6, 6220-6231.

Xu, R. & Wunsch, D. C. (2005). "Survey of Clustering Algorithms", *IEEE Transactions On Neural Networks*, 3(16), 645-678.

Yan, X. & Su, X. (2009). *Linear Regression Analysis: Theory and Computing*, World Scientific.

Yang, W., Jin, L., & Liu, M. (2016). "Deepwriterid: An End-to-End Online Text-Independent Writer Identification System", *IEEE Intelligent Systems*, 31(2), 45-53.

Ye, Q., Zhang, Z., & Law, R. (2009). "Sentiment Classification of Online Reviews to Travel Destinations by Supervised Machine Learning Approaches", *Expert Systems With Applications*, 36(3), 6527-6535.

Yousif, A., Niu, Z., Tarus, J. K., & Ahmad, A. (2019). "A Survey on Sentiment Analysis of Scientific Citations", *Artificial Intelligence Review*, 52(3), 1805-1838.

Zaggia, C. (2016). *Approaches and Methods of Sentiment Analysis Applied on Dainese Spa*, (Bachelor's thesis), Università Ca'Foscari, Venezia.

- Zainuddin, N., & Selamat, A. (2014, September). "Sentiment Analysis Using Support Vector Machine", *In 2014 International Conference on Computer, Communications, and Control Technology (I4CT), IEEE*, 333-337.
- Zhai, C. & Massung, S. (2016). *Text Data Management and Analysis: A Practical Introduction to Information Retrieval And Text Mining*, Morgan & Claypool.
- Zhang, L., Wang, S., & Liu, B. (2018). "Deep Learning for Sentiment Analysis: A Survey", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253.
- Zhang, L., Zhou, W., & Jiao, L. (2004). "Wavelet Support Vector Machine", *Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), IEEE* 34(1), 34-39.
- Zheng, W., & Ye, Q. (2009, November). Sentiment Classification of Chinese Traveler Reviews by Support Vector Machine Algorithm. *In 2009 Third International Symposium on Intelligent Information Technology Application* (Vol. 3, pp. 335-338). *IEEE*.
- Zhou, G. & Su, J. (2002, July). "Named Entity Recognition Using an HMM-Based Chunk Tagger", *In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 473-480.
- Zweigenbaum, P., Demner-Fushman, D., Yu, H. & Cohen, K. B. (2007). "Frontiers of Biomedical Text Mining: Current Progress", *Briefings in Bioinformatics*, 8(5), 358-375.