
GAZİ ÜNİVERSİTESİ
Gazi
1926

YENİ TREND BUNGALOV İŞLETMELERİNE İLİŞKİN ÇEVRİMİÇİ YORUMLARIN DUYGU ANALİZİ İLE İNCELENMESİ

Hazal ÇELİK

YÜKSEK LİSANS TEZİ
ENDÜSTRİ MÜHENDİSLİĞİ ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

HAZİRAN 2023

ETİK BEYAN

Gazi Üniversitesi Fen Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

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

Hazal ÇELİK

21/06/2023

YENİ TREND BUNGALOV İŞLETMELERİNE İLİŞKİN ÇEVRİMİÇİ YORUMLARIN DUYGU ANALİZİ İLE İNCELENMESİ

(Yüksek Lisans Tezi)

Hazal ÇELİK

GAZİ ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

Haziran 2023

ÖZET

2019'un son aylarından itibaren tüm dünyayı etkisi altına alan COVID-19 pandemisiyle birlikte sosyal mesafe, bireysellik ve teması azaltma gibi ön plana çıkan hususlar, birçok sektörde olduğu gibi konaklama ve otel tercihlerinin değişmesinde de etkili olmuştur. İnternetin yaygın kullanımı ile birlikte insanlar konaklamayı planladıkları yerlere ilişkin internet sitelerinde veya sosyal medya platformlarında yer alan müşteri yorumlarını ve puanlamalarını dikkate almaktadırlar. İlgili yorum ve puanlamalar hem konaklayacaklar için hem de işletmeler için önemli bir rol oynamaktadır. İşletmelerin müşteri memnuniyetini anlamaları, hizmetlerini iyileştirmeleri ve rekabet avantajı elde etmeleri açısından müşterilerin deneyimlerini paylaştıkları önemli bir geri bildirim kaynağı olan çevrimiçi yorumlar büyük bir değer taşımaktadır. Bu geri bildirimlerde gizli olan duygularla konaklayanların otelden aldıkları hizmete dair yaptıkları puanlamanın örtüşme düzeyinin uygun yöntemlerle analiz edilmesi gerekmektedir. Duygu analizi, internet siteleri ve sosyal medya platformlarındaki yorumların gerçek anlamını ve duygusunu bulmak için yaygın bir yöntem olarak kullanılmaktadır. Bu çalışmada; Sapanca bölgesinde yer alan bungalov işletmelerine yönelik konaklayanların yaptığı Google Places yorumlarının duygularını ve en çok olumlu veya olumsuz yorum yapılan konuları belirlemek amacıyla makine öğrenmesi ve derin öğrenme yöntemleri kullanılmıştır. Yorumların duygu yoğunluğunu bulmak için makine öğrenmesi yöntemlerinden Naive Bayes, Destek Vektör Makineleri ve Rastgele Orman algoritmaları ile derin öğrenme yöntemlerinden BERT (Bidirectional Encoder Representations from Transformers) algoritması kullanılarak başarı oranları hesaplanmış ve sonuçlar karşılaştırılmıştır. Ayrıca, COVID-19 salgınının turizm tercihlerine etkisini değerlendirmek için yorumlar 2020 öncesi ve sonrası olarak ayrılarak, tüketici tercihlerindeki değişimler yorumlanmıştır.

Bilim Kodu	:	90619
Anahtar Kelimeler	:	Yapay zeka, duygu analizi, makine öğrenmesi, derin öğrenme, glamping turizm, bungalov
Sayfa Adedi	:	143
Danışman	:	Doç. Dr. Hatice Ediz ATMACA

EXAMINATION OF ONLINE COMMENTS ON NEW TREND BUNGALOV BUSINESSES BY SENTIMENT ANALYSIS

(M. Sc. Thesis)

Hazal ÇELİK

GAZİ UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

June 2023

ABSTRACT

Since the end of 2019, the COVID-19 pandemic has had a global impact, leading to changes in accommodation and hotel preferences, similar to many other sectors. As the use of the internet became widespread, people started considering customer reviews and ratings on internet websites or social media platforms before making their accommodation choices. These reviews and ratings play a significant role for both prospective guests and businesses. Businesses should take these feedback into account to identify complaints and satisfaction levels, improve service quality, and ensure customer satisfaction. It is essential to analyze the alignment between the hidden emotions in customer feedback and the ratings they provide for the services received. Sentiment analysis is a common method used to determine the true meaning and emotions behind reviews on internet platforms and social media. In this study, machine learning and deep learning methods were employed to identify the emotions expressed in Google Places reviews of bungalow accommodations in the Sapanca region. The most frequently mentioned positive and negative aspects were also determined. Machine learning algorithms such as Naive Bayes, Support Vector Machines, and Random Forest, as well as the deep learning algorithm BERT (Bidirectional Encoder Representations from Transformers), were used to measure the intensity of sentiments expressed in the reviews, and their success rates were compared. Additionally, to evaluate the impact of the COVID-19 pandemic on tourism preferences, the reviews were divided into pre- and post-2020 periods, and changes in consumer preferences were analyzed.

Science Code	:	90619
Key Words	:	Artificial intelligence, sentiment analysis, machine learning, deep learning, glamping tourism, bungalow
Page Number	:	143
Supervisor	:	Assoc. Dr. Hatice Ediz ATMACA

TEŞEKKÜR

Tez çalışmam süresince ilgi ve desteğini esirgemeyen, bilgi ve deneyimleri ile bana ışık tutan saygıdeğer hocam Sayın Doç. Dr. Hatice Ediz ATMACA'ya, hayatımın her anında maddi manevi destekleriyle yanımda olan aileme, özverisi, desteği ve anlayışıyla hep yanımda olan eşime ve bu süreçte aramıza yeni katılan canım oğluma en içten teşekkürlerimi sunarım.

İÇİNDEKİLER

Sayfa

ÖZET iv

ABSTRACT v

TEŞEKKÜR vi

İÇİNDEKİLER vii

ÇİZELGELERİN LİSTESİ ix

ŞEKİLLERİN LİSTESİ x

SİMGELER VE KISALTMALAR xi

1. GİRİŞ 1

2. TURİZM VE KONAKLAMA İŞLETMELERİ 5

2.1. Konaklama Türleri 5

2.1.1. Asli (Geleneksel) konaklama işletmeleri 6

2.1.2. Tamamlayıcı konaklama işletmeleri 7

2.2. Konaklamada Hizmetin Önemi 8

2.3. Konaklama İşletmelerinde Müşteri Yorumlarının Önemi 10

3. TÜRKİYE'DE DEĞİŞEN TURİZM VE PANDEMİNİN TURİZME ETKİSİ 13

3.1. Türkiye'de Turizm 13

3.2. Pandemi Dönemi Turizm 14

3.3. Glamping Tanımı ve Özellikleri 16

3.3.1. Glamping'in tarihsel seyri 19

3.3.2. Glamping konaklama türleri 21

3.4. Türkiye'de Glamping Hizmeti Sunan İşletme Örnekleri 23

4. METİN MADENCİLİĞİ, DUYGU ANALİZİ VE MAKİNE ÖĞRENMESİ 25

4.1. Metin Madenciliği Kavramının Tanımı ve Diğer Alanlarla İlişkisi 25

Sayfa

4.1.1. Metin madenciliği uygulama alanları 26
4.1.2. Metin madenciliği aşamaları 26
4.1.3. Metin madenciliği metotları 30
4.2. Duygu analizi 33
4.2.1. Duygu analizi amacı ve önemi 35
4.2.2. Duygu analizi kelime sınıflandırma aşamaları 38
4.2.3. Duygu analizinde kullanılan yöntemler 39
5. LİTERATÜR ARAŞTIRMASI 59
5.1. Konaklama Sektörüne Yönelik Çalışmalar 59
5.2. Glamping Turizm Sektörüne Yönelik Çalışmalar 64
6. UYGULAMA: COVİD-19 PANDEMİSİ SONRASI YENİ TREND BUNGALOV İŞLETMELERİNE İLİŞKİN ÇEVRİMİÇİ YORUMLARIN MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE DUYGU ANALİZİ 67
6.1. Çalışmanın Amacı 67
6.2. Çalışmanın Kapsamı 67
6.3. Veri Setinin Oluşturulması 68
6.3.1. Çalışma verilerinin web ortamından çekilmesi 69
6.3.2. Çalışma verilerinin ön hazırlık süreci 71
6.4. Verilerin Analizi ve Çalışma Modelinin Tanımlanması 75
6.4.1. Yorumların duygu analizi ile incelenmesi 76
6.5. Bulgular 80
7. SONUÇ VE ÖNERİLER 89
KAYNAKLAR 91
EKLER 109
ÖZGEÇMİŞ 143

ÇİZELGELERİN LİSTESİ

Çizelge **Sayfa**

Çizelge 4.1. Karmaşıklık matrisi 51

Çizelge 6.1. Her algoritma için elde edilen karmaşıklık matrisleri 81

Çizelge 6.2. Duygu analizi sonucunda elde edilen performans değerleri..... 81

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil 3.1. Kavramsal olarak glamping turizmi 18

Şekil 4.1. Duygu analizi süreç akışı..... 34

Şekil 4.2. Duygu analizi kelime sınıflandırma şeması..... 38

Şekil 4.3. Duygu analizi yöntemleri 40

Şekil 4.4. Destek vektör makineleri algoritması kesikli çizgiler ve düz çizgi 45

Şekil 4.5. Rastgele orman algoritması çalışma prensibi 46

Şekil 4.6. BERT modeli genel yapısı..... 49

Şekil 6.1. Rating değerlerine göre yorumların dağılımı 69

Şekil 6.2. Bungalov işletmeleri linklerinin tutulduğu liste görüntüsü 70

Şekil 6.3. Bir bungalov işletmesine ait yorumların CSV dosyasındaki örnek görüntüsü..... 71

Şekil 6.4. Temizleme ve köklerine ayırma işlemleri sonucu elde edilen CSV dosyası örnek görüntüsü 73

Şekil 6.5. Vektör uzay modeli için kullanılan kod..... 74

Şekil 6.6. Çalışmanın modeli 75

Şekil 6.7. Belirteç dizisi uzunluğu histogram grafiği 79

Şekil 6.8. Naive Bayes algoritmasına ait ROC eğrisi 82

Şekil 6.9. Destek vektör Makineleri algoritmasına ait ROC eğrisi..... 83

Şekil 6.10. Rastgele orman algoritmasına ait ROC eğrisi 84

Şekil 6.11. BERT algoritmasına ait ROC eğrisi 84

Şekil 6.12. Olumlu yorumlara ait kelime bulutu..... 85

Şekil 6.13. Olumsuz yorumlara ait kelime bulutu 86

Şekil 6.14. Pandemi öncesi yorum sayıları..... 87

Şekil 6.15. Pandemi sonrası yorum sayıları..... 87

SİMGELER VE KISALTMALAR

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

Kısaltmalar	Açıklamalar
BERT	Bidirectional Encoder Representations from Transformers
FN	False Negative
FP	False Positive
TF-IDF	Term Frequency — Inverse Document Frequency
TN	True Negative
TP	True Positive
UNWTO	Birleşmiş Milletler Dünya Turizm Örgütü

1. GİRİŞ

2019 yılının son çeyreğinde Çin'in Wuhan kentinde ortaya çıkan ve kısa sürede tüm dünyayı etkisi altına alan Covid-19 pandemisi, insanların yaşamlarında ve sektörlerin işleyişinde derin değişimlere neden olmuştur (Kaygısız, 2021). Özellikle seyahat ve turizm sektörleri, pandeminin etkilerini derinden hisseden sektörler arasında yer almıştır. İnsanlar, konaklamalarını güvende hissedebilecekleri, sosyal mesafe kurallarına uygun ve daha az kalabalık ortamları tercih etmeye başlamışlardır (Kıvılcım, 2020). Bu süreçte ön plana çıkan turizm tercihlerinden biri de konaklama deneyimini lüks otellerdeki konfor ve hizmet kalitesiyle birleştirirken aynı zamanda doğayla iç içe olma fırsatı sunan glamping turizmdir (Düzgün, 2021).

Glamping turizm, konaklayan kişilere kamp deneyimi yerine lüks ve konforlu bir tatil seçeneği sunan bir turizm türüdür. Bu yeni tatil trendinde doğayla bütünleşmek isteyenler, kalabalık otellerin yerine daha özel ve izole bir konaklama deneyimi arayanlar için bungalov evleri, glamping turizm için ideal bir seçenek olarak öne çıkmaktadır. Bungalov evler, doğanın içinde konumlanmış. Yüksek hizmet kalitesi sunan, konforlu olanaklara sahip olan ve aynı zamanda fiziksel mesafe kurallarına uygun bir tatil imkânı oluşturan konaklama birimleridir (Ergüven Yılmaz ve Kutlu, 2015: 256-257).

Pandemiyle birlikte tercih edilirliği artan bungalov işletmelerine yönelik bir bilgi alma arayışında olan insanların, bilgilere ulaşmak için internet siteleri ve sosyal medya platformlarında yer alan yorumlardan faydalandıkları görülmektedir (Düzgün, 2021). Bu yorumlar, tüketiciler kadar işletmeler için de son derece önemlidir**.** İşletmeler, şikâyetleri ve memnuniyetleri belirlemek, hizmet kalitesini artırmak ve müşteri memnuniyetini sağlamak için bu yorumlardaki geri bildirimleri dikkate almalıdır.

Bu geri bildirimleri analiz etmek ve değerli bilgiler elde etmek amacıyla yaygın olarak duygu analizi yöntemi kullanılmaktadır. Duygu analizi metinlerdeki duygusal ifadeleri tespit etmek ve sınıflandırmak için kullanılan bir yöntemdir (Eröz, 2011). Bu yöntem, metinlerdeki kelimelerin, cümle yapılarının ve tonlamaların analiz edilerek pozitif, negatif veya nötr duyguların belirlenmesini sağlar (Bird, Wiles, Okalik, Kilabuk ve Egeland, 2009).

Bu çalışmanın amacı Sapanca'da bulunan bungalov işletmelerine yönelik Google Places üzerinden yapılan yorumları duygu analizi yöntemleriyle incelemek ve işletme sahiplerine yorumlarda en sık geçen olumlu ve olumsuz noktaları göstererek öneriler sunmaktır.

Toplamda yedi bölümden oluşan çalışmanın, ilk bölümü olan giriş bölümünde Covid-19 pandemisinin ciddi düzeyde etkilediği sektörlerden birisinin seyahat ve turizm sektörü olduğu, salgın nedeniyle insanların daha az kalabalık ortamları tercih etmeye başladığı ve bu tür bir turizm imkânı veren glamping turizminin ve buna bağlı bungalov evlerinin bu dönemde rağbet görmeye başladığı belirtilmiş ve bu tür işletmelere yönelik çevrimiçi yorumların analiz edilmesinin önemi üzerinde durulmuştur.

İkinci bölümde turizm ve konaklama işletmeleri ele alınmış, bu kapsamda konaklama türlerinden asli (geleneksel) konaklama ve tamamlayıcı konaklama işletmeleri, konaklamada hizmetinin ve konaklama işletmelerinde müşteri yorumlarının öneminden bahsedilmiştir.

Üçüncü bölümde Türkiye'de değişen turizm ve pandeminin turizme etkisi üzerinde durulmuş, bu kapsamda Türkiye'de turizm sektörü, pandemi döneminde turizmin durumu, turizm çeşitlerinden glamping turizminin tanımı, özellikleri, tarihsel seyri ve konaklama türleri üzerinde durulmuştur. Ayrıca bu bölümde Türkiye'de glamping hizmeti sunan işletmelere de yer verilmiştir.

Dördüncü bölüm metin madenciliği, duygu analizi ve makine öğrenmesi başlığını taşımaktadır. Bu bölümde metin madenciliği kavramının tanımı ve diğer alanlarla ilişkisi, metin madenciliği uygulama alanları, metin madenciliği aşamaları ve metotları, duygu analizi, duygu analizinin amacı ve önemi, duygu analizi kelime sınıflandırma aşamaları ve duygu analizinde kullanılan yöntemler üzerinde durulmuştur.

Beşinci bölümde ise konaklama işletmeleriyle ilgili çevrimiçi platformlarda yer alan müşteri değerlendirmelerine yönelik yapılan çalışmalar ele alınmıştır. Bu alanda konuyla ilgili çalışmalar 2013-2023 tarihleri arasında iki başlık olarak incelenmiştir.

Altıncı bölüm ise çalışma ile ilgili elde edilen verilerin analizinin yapıldığı uygulama bölümüdür. Bu bölümde veri setinin oluşturulması sürecinden, elde edilen verilerin

analizinden bahsedilmiş, araştırma modelinin tanımlaması yapılmış ve veri analizinden elde edilen bulgular verilmiştir.

Çalışmanın yedinci ve son bölümü ise sonuç ve öneriler bölümüdür. Bu bölümde bulguların değerlendirilmesi sonucunda ulaşılan sonuçların değerlendirmesi yapılmış ve bungalov işletmelerine önerilerde bulunulmuştur.

2. TURİZM VE KONAKLAMA İŞLETMELERİ

Turistler, deneyimlerinden uzaklaşarak farklı milletlerin tarihi ve kültürel zenginliklerini ve kendilerine ait değerleri değerlendirmek için isteklidirler. Bu gereksinimleri karşılamak için turizm işletmeleri kurulmuştur. Turizm konaklama, yeme-içme, seyahat ve diğer turizm konaklamaları dahil olmak üzere 4 farklı kategoriye ayrılmış ve sundukları hizmetlere göre gruplandırılmıştır (Yılmaz, 2022: 9).

Konaklama işletmeleri, seyahat halinde olan veya belirli bir süre için evden uzakta bulunan insanların öncelikle konaklama ihtiyaçlarını karşılamakla birlikte bazı sosyal ihtiyaçlarını da yerine getirebilecek kapasitede olan işletmelerdir. Bu nedenle konaklama sektörü, ticari amaçlarla oluşturulmuş ve seyahat eden insanların konaklama, yeme-içme ve rekreasyon giderlerini karşılayacağı çok farklı işletmeleri içeren bir sektördür (Okuyucu ve Akgiş, 2016). Farklı bir tanıma göre konaklama işletmeleri turizm sektöründe faaliyet gösteren tesislerdir ve konaklama, yeme-içme ve diğer sosyal ihtiyaçları karşılamak için tasarlanmışlardır. Bu işletmeler, planlama, üretim ve pazarlama süreçlerinde gereken örgütlenme ve işletme uzmanlığına sahip olmak amacıyla kurulmuşlardır (Baker, Huyton ve Bradley, 2000; Kozak N., Kozak M.A. ve Kozak M., 2014).

2.1. Konaklama Türleri

Konaklama işletmelerinin sınıflandırılması yapılırken çeşitli faktörler göz önünde bulundurulmuştur. Konaklama işletmelerinin sınıflandırılması, seyahat eden kişilerin gelir düzeyleri, ihtiyaçları ve zevkleri ile birlikte bölgenin turizm şekli ve yapısını dikkate alarak belirlenmektedir. Konukların seyahat amacı, profilleri ve ziyaret ettikleri bölgeye göre çeşitli konaklama tesisleri bulunmaktadır. Bu tesisler, konaklama türüne uygunluk, sunduğu olanaklar, mimarisi, işlevleri ve ticari faaliyetleri gibi faktörlere göre kategorize edilmiştir (Kalıpçı, 2021).

"Turizm Teşvik Kanunu ve Turizm Yatırım ve İşletmeleri Nitelikleri Yönetmeliği"ne göre konaklama işletmeleri iki ana kategoride değerlendirilmektedir: asli konaklama işletmeleri ve yardımcı konaklama işletmeleri (Şener, 2009: 8). Çalışmanın bu bölümünde literatürde yaygın olarak kullanılan "asli/geleneksel" ve "yardımcı/tamamlayıcı" sınıflandırması biçimine yer verilecektir.

2.1.1. Asli (Geleneksel) konaklama işletmeleri

"Kültür ve Turizm Bakanlığı Mevzuatı"na göre asli konaklama işletmeleri: oteller, moteller, tatil köyleri ve pansiyonlar'dır (Kültür Bakanlığı, 2023).

Asli konaklama işletmeleri; "aslî fonksiyonları müşterilerin konaklama ihtiyaçlarını sağlamak olan bu hizmetin yanında, yeme-içme, spor ve eğlence ihtiyaçları için yardımcı ve tamamlayıcı birimleri de bünyelerinde bulundurabilen işletmeler" olarak tanımlanmaktadır (Ünal ve Bayar, 2021: 75). Yönetmelikte belirtildiği üzere, asli konaklama işletmeleri arasında oteller, moteller, tatil köyleri ve pansiyonlar gibi çeşitli tesisler yer almaktadır. Bu işletmelerin genel özellikleri, konaklama birimlerinin düzenlenmesi, personel ve mekan düzenlemeleri, bedensel engelli konuklar için uygunluk, asansör düzenlemeleri ve kış sporları yapılan işletmelerde gerekli düzenlemeler şeklinde beş ana kategori altında sınıflandırılmıştır (Resmi Gazete, 2019).

Oteller*:* İnsanlar, seyahatleri sırasında dinlenme ve uyku ihtiyaçlarını karşılayabilecekleri tesislere ihtiyaç duyarlar. Ticari amaçlar, yeni yerler keşfetme, dini amaçlı seyahatler vb. nedenlerle yapılan bu yolculuklarda, temel ihtiyaç olan konaklama hizmeti sunan ticari işletmeler bulunmaktadır. Otel olarak ifade edilen bu işletmeler, konaklayan misafirlerin dinlenme ve yeme-içme ihtiyaçlarının yanı sıra psikolojik tatmin sağlayarak mal ve hizmet sunan işletmelerdir (Şener 2009: 10) Belirli standartlar çerçevesinde düzenlenen tesisler olan oteller, seyahat eden yabancıların birçok ihtiyacını karşılamak üzere tasarlanmıştır. Konaklama odalarının yanı sıra otellerde yeme-içme, eğlence ve alışveriş alanları da bulunmaktadır (Aksu, 2014; Akın, 2019). Halka açık restoranlar, toplantı salonları ve farklı oda tipleri gibi birçok hizmet, otellerin sunduğu olanaklar arasındadır. Oteller, 20'den 2000'e kadar değişen oda sayılarına sahip olabilirler. Araba kiralama hizmetleri, farklı eğlence alanları ve acenteler gibi diğer hizmetler de otellerin parçası olabilen uygulamalardır (Kalıpçı, 2021: 5-6).

Moteller: Motorlu araçlarla seyahat edenlerin konaklama, yeme-içme ve araç parkı ihtiyaçlarını karşılamak üzere inşa edilen en az on odalı konaklama tesislerine moteller denir. Bu tesisler, özellikle tur programından bağımsız olarak seyahat eden turistler tarafından tercih edilir ve genellikle yerleşim merkezleri dışında veya yakın çevrelerinde bulunur (Çolakoğlu, 2014: 62).

Tatil Köyleri: Doğal güzelliklerle çevrili, konaklama yanında spor ve eğlence gibi aktivitelerin de yapılabildiği en fazla üç katlı ve seksen odalı tesislerdir (Azaltun ve Kaya, 2010: 3). Tatil köylerinde, çocukların oynayabileceği alanlar, rekreasyon aktiviteleri için uygun mekanlar, spor ve eğlence faaliyetleri için alanlar ve açık havuzlar mevcuttur (Akgöz, 2009: 12).

Pansiyonlar: Pansiyonlar, en az beş odası olan ve konaklama hizmeti veren küçük işletmelerdir. Konaklayan misafirler, yeme-içme faaliyetlerini genellikle pansiyonun imkânları dahilinde kendileri karşılarlar. Genellikle turistik yerlerde ve kırsal kesimlerde bulunan pansiyonlarda, otellere göre daha az sayıda personel bulunur ve hizmetleri daha sınırlıdır (Kozak ve diğerleri, 2014: 57).

2.1.2. Tamamlayıcı konaklama işletmeleri

Bu başlık altında tamamlayıcı konaklama işletmeleri kapsamında değerlendirilen oberj, kamping, hostel, apart otel, termal tesis, butik otel ve bungalov gibi konaklama işletmelerinin bazılarının tanımlarına yer verilmiştir.

Kamping: Kampingler, motorlu araçların yaygınlaşması sonrasında ortaya çıkan ve genellikle doğal güzelliklere veya tarihi turistik yerlere yakın konumlarda bulunan en az on birimlik konaklama tesisleridir. Bu tesisler, müşterilerin kendi imkânlarıyla geceleme, yeme-içme, dinlenme, eğlence ve spor ihtiyaçlarını karşılayabilecekleri hizmetleri sunarlar ve genellikle karayolu güzergâhları üzerinde veya kent girişlerinde yer alırlar (Arda, 2015: 29).

Hostel: Hosteller, özellikle genç turistlerin ihtiyaçlarına cevap verebilecek şekilde tasarlanmış en az on odalı konaklama tesisleridir. Misafirlerin kendi ihtiyaçlarını karşılayabilecekleri veya isteğe bağlı olarak yeme-içme hizmeti alabilecekleri tesislerdir (Batman, 2008: 49).

Apart otel: Turizm sertifikasına sahip büyük tesisler, villa ya da apartman olarak inşa edilmiş, tatil köyü veya turizm kompleksleri içinde bağımsız olarak yer alan konaklama birimleridir (Kozak ve diğerleri, 2020). Konaklama yapacak kişilerin kendi yemeklerini hazırlayabilecekleri minimum 10 oda sayısına sahip binada, gerekli ekipmanlar mevcuttur

(Akın, 2019). “Müstakil apart otel” adı altında tesis bünyesinde bulunma zorunluluğu olmadan en az on üniteden oluşacak şekilde, bir müstakil olarak da düzenlenebilmektedirler (Hazar, 2010).

Termal Tesis/Kaplıca: Termal tesisler, genellikle sağlık amaçlı ziyaret edilen, termal kaynak üzerine inşa edilmiş işletmelerdir. Bu tesisler, mineralize su veya çamur gibi sağlık hizmetlerini sunmanın yanı sıra yeme-içme ve eğlence hizmetleri de sunarlar (Kozak ve diğerleri, 2020).

Butik Otel: Hizmet kalitesi ve kişiye özel hizmetlerin ön planda olduğu oteller, 20 ya da 30 odalı ve özenle tasarlanmış tesislerdir. Bu otellerin mimarisi, iç dekorasyonu ve kullanılan malzemeler diğer tesislerden daha farklıdır (Akın, 2019).

Oberjler: Yardımcı konaklama işletmesi olan oberj, dağ evi anlamına gelmektedir. Kırsal yörelerde kurulan en az tek yıldızlı otel niteliği taşıyan alanlarda konaklama imkânı sunan oberjler, doğa güzelliklerinden yararlanmak, avlanmak, kış sporları ve çim sporları gibi aktiviteler gerçekleştirmek isteyenler tarafından tercih edilir (Barutçugil, 1989: 54).

Bungalov: Bungalov, geniş bir verandayla çevrili küçük bir ev ya da kır evi olarak tanımlanır ve tek katlı ya da eğimli bir çatıya sahiptir. Çatı pencereleri ile süslenen bungalovlar, ikinci katları da barındırabilir. İlk kez 1869 yılında İngiltere'de inşa edilen bir ev bungalov olarak sınıflandırılmıştır (Powell ve Svendsen, 2004). Bungalov evleri genellikle sıcak ve güneşli iklimlerde inşa edilir ve isimleri de bu özelliklerine atfen verilmiştir. İlk bungalov evler geniş sundurmaları ve sarkıt/gölgelikleriyle dikkat çekerken, günümüzde Türkiye'de daha çok tatil evi olarak kullanılmaktadır. Bazı ülkelerde ise bungalov evleri daimî konut olarak kullanmaktadır (Yıldırgan, 2022: 61).

2.2. Konaklamada Hizmetin Önemi

Hizmet, müşteri veya tüketiciye sunulan fiziksel bir şekli olmayan soyut bir değer olarak ifade edilirken; hizmet kalitesi bir işletmenin müşteri beklentilerini karşılayabilme yeteneği olarak belirtilmektedir (Fettahlıoğlu, Polat ve Demir, 2016: 850). Hizmet sektöründe kalitenin tanımı ve ölçümünü yapmak oldukça zordur. Önemli olan, kalitenin tüketiciler tarafından belirlenmesidir. Kaliteli bir hizmet, tüketicilerin beklentilerini karşılayan ve

tatmin eden hizmettir. Tersine, tatmin etmeyen hizmetler ise kalitesiz olarak nitelendirilmektedir (Karahan, 2000: 116). Hizmet kalitesiyle ilgili en önemli faktör, müşteri algısındaki kalitedir. Bu nedenle, hizmet kalitesi, müşteri tarafından algılanan performans düzeyi veya müşterinin tatmin düzeyi ile ölçülür. Hizmet kalitesi, müşteri beklentileri ile gerçek hizmet performansının karşılaştırılması sonucu belirlenir.

Konaklama işletmelerinde hizmet kalitesi, müşteri memnuniyeti ve müşteri sadakati açısından oldukça önemlidir. Müşteriler, konaklama işletmelerinde aldıkları hizmet kalitesiyle ilgili algılarına göre işletmeleri tercih ederler ve tekrar ziyaret ederler. Bu nedenle, işletmelerin müşteri odaklı hizmet sunmaları gerekmektedir (Paraskevas ve Arendell, 2007). Müşterilerin beklentilerinin karşılanması ve hatta aşılması, işletmelerin müşteri memnuniyetini artırması açısından önemlidir. Konaklama işletmeleri, müşteri sadakatini sağlamak için müşterilerinin ihtiyaçlarına uygun hizmetler sunmalı, müşterilerinin deneyimlerini sürekli olarak iyileştirmeli ve işletme hakkındaki olumlu imajı sürdürmelidir (Bowen ve Chen, 2001).

Fiziksel ortamın ve çalışanların müşteri deneyiminde önemli bir rol oynadığı belirtilmektedir. Müşteriler, konaklama işletmelerinin fiziksel ortamındaki temizlik, düzen ve konfor gibi unsurları da dikkate alarak değerlendirirler. Ayrıca, çalışanların misafirperverliği, güler yüzlülüğü ve hızlı hizmet sunumu, müşteri memnuniyetini artırmada etkilidir (Kim H.B. ve Kim W.G., 2005). Otel işletmelerinin başarısı sadece iyi bir finansal ve fiziksel altyapıya sahip olmaktan değil, aynı zamanda etkili bir insan gücüne sahip olmaktan da geçmektedir. İşletmenin finansal ve fiziksel yapıları, onları değerlendiren ve anlam kazandıran işletme çalışanları tarafından yönetilmektedir. İşletmenin insan kaynakları, motive edici ve geliştirici bir yapıya sahip olduğunda, işletmenin hedeflerine daha kolay ulaşmasını sağlar. Bunun aksine, fiziksel ve finansal olanakları ne kadar mükemmel olursa olsun, insan kaynaklarını geliştiremeyen örgütler hedeflerine ulaşamazlar (Erdem, 2004).

Kısaca ifade etmek gerekirse konaklama işletmelerinde hizmetin kalitesinin işletme başarısı açısından oldukça önemlidir. Bu nedenle işletmelerin müşteri memnuniyeti sağlamak için müşteri odaklı yaklaşımlar benimsemeleri gerekmektedir.

2.3. Konaklama İşletmelerinde Müşteri Yorumlarının Önemi

Müşteri memnuniyeti, uzun süredir pazarlama uygulamaları ve akademik alanda araştırılan önemli bir kavramdır. Bu kavram, her işletmenin başarısı için kilit bir faktördür ve müşteri bağlılığı oluşturmak ve rekabet avantajı elde etmek için kritik bir rol oynamaktadır. Bu nedenle, konaklama işletmeleri gibi yoğun rekabet ortamında faaliyet gösteren işletmeler için, müşteri memnuniyeti ve memnuniyetsizliğine neden olan faktörlerin anlaşılması önemli bir zorunluluk olarak ifade edilmektedir (Ercan, 2019). Konaklama işletmelerinin müşterilerine sunduğu hizmetler, birçok farklı öğeden oluştuğundan misafir memnuniyetini etkileyen faktörler değişebilir. Misafir memnuniyeti konusu pratikte oldukça göreceli bir kavramdır ve müşteri yorumları incelenerek, yeni tüketici davranış biçimleri takip edilerek güncel kalınabilir (Arkadaş ve Ayyıldız, 2020).

Rekabet ortamı yüzyıllardan beri devam etse de teknolojinin gelişmesiyle birlikte sanal ortamda da varlığını sürdürmektedir. İnternet, en yaygın ve etkili iletişim kanalıdır ve insan hayatını derinden etkilemektedir. Bu nedenle, sosyal medya araçları popülerlik kazanmıştır ve pazarlama ve reklam faaliyetleri de bu platformlara taşınmıştır. Tüketiciler, çevrimiçi sitelerde otel, pansiyon, özel konaklama ve restoran gibi işletmeleri ziyaret ederken, deneyimlerini paylaştıkları bir alan olarak da görülmektedir (Kafa ve Demircioğlu, 2022).

Günümüzde turistik müşteriler, satın alma kararlarını verirken diğer insanların internet ve sosyal medya yorumlarına ve deneyimlerine büyük önem vermektedir. Bu tür yorumlar, turistik işletmelerin hizmetlerine yönelik bir hizmet kalitesi değerlendirme kaynağına dönüşmüştür ve potansiyel tüketiciler için önemli bir kriter haline gelmiştir. Bunun nedeninin müşterilerin herhangi bir turistik ürün ile ilgili, internet üzerinde bulunan bilgi kaynaklarından elde ettikleri bilgilerden ziyade, daha önce bu deneyimi yaşamış müşteri yorumlarına çok daha fazla güvenmeleri ile ilgili olduğu düşünülmektedir. Müşteriler, bu tür yorumları tarafsız ve dürüst olarak değerlendirmektedir (Nowacki ve Niezgoda, 2019). Müşteriler, bu yorumları okuyarak tesisin kendilerine ne kadar uygun olduğunu değerlendirmektedirler. Otel yöneticileri açısından da bu yorumların değerlendirilmesi, hizmet kalitesini ölçmek, müşterilerin hangi faktörlere daha çok önem verdiğini belirlemek ve müşterilerin duygularını anlama açısından oldukça önemlidir (Şahin, Çolakoğlu, Kılıçdere ve Atay, 2020).

İlgili alan yazın incelendiğinde müşteri yorumlarının konaklama işletmeleri açısından önemini ortaya koyan çalışmalar olduğu görülmektedir. Örneğin Kandampully ve Suhartanto (2003) tarafından yapılan çalışmada, müşteri yorumlarının konaklama işletmelerinde müşteri sadakati üzerindeki etkisi araştırılmıştır. Araştırmanın sonucunda, müşteri yorumlarının müşteri sadakati üzerinde olumlu bir etkiye sahip olduğu tespit edilmiştir. Litvin, Goldsmith ve Pan (2008) tarafından yapılan çalışmada, müşteri yorumlarının konaklama işletmeleri için önemli bir pazarlama aracı olduğu belirtilmektedir. İnternet ve sosyal medya gibi platformlarda yayınlanan müşteri yorumları, işletmelerin imajını etkilemekte ve potansiyel müşterilerin karar verme sürecine katkı sağlamaktadır.

Ye, Law ve Gu (2009) çalışmasında, online müşteri yorumlarının konaklama işletmelerinin satışlarına etkisi incelemiştir. Araştırmanın sonucunda, olumlu müşteri yorumlarının işletmelerin satışlarını artırdığı tespit edilmiştir. Xiang vd (2017) çalışmasında farklı online müşteri yorumu platformlarının analiz edilmesi sonucunda, müşteri yorumlarının işletmeler için önemli bir geri bildirim kaynağı olduğunu ve müşteri deneyimlerinin iyileştirilmesi için değerli bilgiler sağladığını belirlemiştir (Xiang, Du, Ma ve Fan, 2017).

Ercan (2019) tarafından yapılan bir çalışmada, müşteri yorumları incelendiğinde en sık bahsedilen memnuniyet unsurlarının sırasıyla personel, yiyecekler, diğer, misafir odaları, otel olanakları, hizmet kalitesi, konum ve içecekler olduğu belirlenmiştir. Ayrıca, müşteri memnuniyetsizliğiyle ilişkili ana unsurlar ise yiyecekler, personel, diğer, hizmet/ekipman eksiklikleri, temizlik ve içecekler olarak tespit edilmiştir. Şahin vd (2020) çalışmasında, Kuşadası'ndaki 5 yıldızlı otellerde konaklayan turistlerin en çok temizlik, personel ve yemek kalitesine önem verdiğini saptamıştır. Bu bulgular, turistlerin otel seçiminde bu unsurlara öncelik verdiğini göstermektedir ve otel işletmecileri için hizmet kalitesinin belirlenmesinde önemli bir faktördür. TripAdvisor gibi çevrimiçi yorum platformları, turistlerin otel seçimi sürecinde de etkili olabilmektedir.

Yılmaz (2020) tarafından yapılan araştırmaya göre İstanbul'daki otel müşteri yorumlarında, konum, ulaşım erişilebilirliği, yeme-içme konsepti, personel/hizmet kalitesi ve tesis temizliğinin otellere verilen yıldız puanları üzerinde etkili olduğu ortaya konmuştur. Benzer bir çalışma 2019 yılında, İstanbul'daki otel işletmeleri hakkındaki yorumları incelemek amacıyla gerçekleştirilmiştir. Ak ve Dinçer'in (2019) araştırmasında, İstanbul'daki beş yıldızlı otel işletmelerinin tüketici değerlendirme puanları konum, temizlik, hizmet, fiyat-

performans değeri ve genel performans kriterlerine göre incelenmiştir. Araştırmada, otel işletmelerinin tüketici inceleme puanları açısından konum, yorum sayısı, oda fiyatı ve çevresel faktörlere göre farklılıklar tespit edilmiştir. Avrupa yakasında bulunan otel işletmeleri, Asya yakasındaki işletmelere göre konumlarından dolayı daha yüksek puanlar almışlardır. Bu çalışmalar dışında müşterilerin otel tercihleri ve memnuniyetleri üzerine literatürde çok sayıda çalışma yer almaktadır (Budak ve Kılıç, 2021; Yılmaz, 2020; Güngör O., Güngör Y. ve Doğan, 2018; An, 2018; Uğurlu, Aslan ve İnce, 2017; Kızılırmak, Alrawadieh ve Aghayeva, 2015). Ayrıca, bazı literatür çalışmaları müşteri memnuniyetinin olumlu yorumlar ve yüksek derecelendirme puanlarıyla şekillendiğini göstermiştir (Mariani ve Borghi, 2018; Bagnera, 2017).

Müşteri yorumları arasında ulaşım olanakları ve konum (Glaveli, Manolitzas, Palamas, Liassidis ve Grigoroudis, 2022; Yılmaz, 2020; An, 2018), eğlence ve spor (Güngör ve diğerleri, 2018), hizmet ve temizlik (An, 2018; Bagnera, 2017), personel, restoran ve oda (Kuhzady ve Ghasemi, 2019; An, 2018; Bagnera, 2017) vb. ifadeler memnuniyeti önemli ölçüde etkilemektedir. Diğer taraftan bazı çalışmalarda, konumun müşteri memnuniyeti ve olumlu değerlendirmeler açısından en önemli kriter olduğu sonucuna ulaşılmıştır (Peres ve Paladini, 2021; Sann ve Lai, 2020; Alrawadieh ve Law, 2019; Ak ve Dinçer, 2019). Müşteriler, otellerin konumuyla ilgili yorum yaparken genellikle toplu taşıma, havaalanı, şehir merkezi, alışveriş merkezleri, restoranlar ve turistik yerlere yakınlığına odaklanmaktadırlar (Bagnera, 2017). Bu kaynaklar, müşteri yorumlarının konaklama işletmelerinin performansı üzerinde önemli bir etkiye sahip olduğunu ve işletmelerin müşteri sadakati, imajı ve satışlarına olumlu katkıda bulunduğunu göstermektedir. Bunun yanı sıra online müşteri yorumları gibi dijital platformlardaki yorumların işletmelerin pazarlama ve imaj stratejilerinde önemli bir yer tuttuğu da vurgulanmaktadır.

3. TÜRKİYE'DE DEĞİŞEN TURİZM VE PANDEMİNİN TURİZME ETKİSİ

3.1. Türkiye'de Turizm

Turizm, insanların günlük yaşamlarından koparak farklı amaçlarla (sosyalleşme, dinlenme, tatil yapma, farklı kültürleri tanıma, spor ve sağlık gibi) seyahat etmeleri ve geçici olarak konaklamaları şeklinde ifade edilebilir. Turizm faaliyetleri hem yurt içinde hem de yurt dışında gerçekleştirilebilmektedir (Uslu, 1990: 19). Turizm, insanların farklı amaçlarla kısa süreliğine bulundukları yerden ayrılarak yeni yerler keşfetmelerini, çeşitli aktivitelere katılmalarını, farklı kültürleri deneyimlemelerini ve günlük hayattan kaçış arzusunu tatmin etmelerini sağlayan bir sektördür. Hem gelişmiş hem de gelişmekte olan ülkeler tarafından önem verilen turizm sektörü, günümüzde de önemini korumaya devam etmektedir (Uluçeçen, 2011: 5; Yıldız, 2011: 54).

Turizm, ekonomik kalkınmaya katkısı, döviz kazandırması ve iş imkânı yaratması gibi pek çok alanda olumlu etkileri olan bir sektördür. Bu nedenle, dünya ekonomisi için önemli bir yer tutmaktadır ve gelir artışına önemli bir katkı sağlamaktadır (Bahar, 2006: 137-150). Turizm, çağdaş ekonomi içinde önemli bir sektör olarak kabul edilir çünkü arz ve talep unsurlarına dayanmaktadır (Göçen, Yirik ve Yılmaz, 2011: 493-509). Küresel istatistik çalışmalarında, turizmin dünya genelindeki istihdama katkısı %7,4 oranında ve toplam istihdamda %10'luk bir paya sahip olduğu belirtilmektedir (Bahar ve İlal, 2020: 125-139).

Turizm sektörü, dünya genelinde gelişmiş ve gelişmekte olan ülkelerin büyük ilgi gösterdiği hızla gelişen sektörlerden biridir. Turizm sektörü, ülkelerin turistik taleplerini belirleyen çeşitli faktörler tarafından etkilenir. Bu faktörler arasında, turist sayısı, yatak kapasitesi, sektöre yönelik teşvikler, seyahat acentelerinin sayısı, turizm çalışanları, ülkedeki suç ve terör unsurları ve hükümet politikaları gibi değişkenler yer almaktadır. Turist sayısı, bir ülkenin turizmdeki popülerliğini, ekonomik durumunu ve sosyal hareketliliğini en iyi ifade eden faktördür (Yıldız, 2011: 54).

Türkiye, güvenli ve uygun maliyetli turizm imajı sayesinde son 30 yılda büyük bir turist kitlesini ağırlamıştır. Coğrafik konumu da Türkiye'yi turizm açısından önemli bir ülke yapmaktadır (www.data.tuik.gov.tr). Türkiye, turizm açısından zengin bir ülkedir ve tarihi,

kültürel, doğal ve gastronomik zenginlikleri ile öne çıkar. Turizm sektörü, Türkiye ekonomisi için önemli bir gelir kaynağıdır ve bu nedenle Türkiye, turizm alanında önemli çalışmalar yürütmektedir.

Türkiye'de turizm sektörü, ülke ekonomisi için son derece kritik bir yere sahiptir ve pek çok kaynak bu önemi vurgulamaktadır. TÜİK verileri, 2021 yılında Türkiye'yi ziyaret eden turist sayısının 12,3 milyonu aştığını göstermektedir. Bu büyük hacimli turizm sektörü, Türkiye'nin milli gelirine önemli katkılar sağlamaktadır (Kambur ve Aksoy, 2022). Türkiye, her yıl milyonlarca ziyaretçiyi çeken popüler bir turizm merkezidir. Türkiye İstatistik Kurumu'nun açıkladığı bir rapora göre Türkiye'ye gelen yabancı ziyaretçi sayısı 2019 yılında bir önceki yıla göre %14,3 artarak 45,06 milyon olarak gerçekleşmiştir Raporda ayrıca Türkiye'de en çok ziyaret edilen şehrin İstanbul olduğu, ardından Antalya ve Edirne'nin geldiği belirtilmektedir (TUİK, 2019).

Turizm, Türkiye ekonomisi için önemli bir sektör, ülkenin GSYİH'sına önemli ölçüde katkıda bulunmak ve istihdam olanakları sağlamak. Dünya Seyahat ve Turizm Konseyi'ne göre seyahat ve turizmin Türkiye'nin GSYİH'sına toplam katkısı 2019 yılında %11,3 olmuş ve sektör 2,5 milyondan fazla kişiye istihdam sağlamıştır. Ancak COVID-19 salgını birçok ülkede olduğu gibi Türkiye'de de turizm sektörünü önemli ölçüde etkiledi. Seyahat kısıtlamaları ve virüsün yayılmasını kontrol altına almak için alınan diğer önlemler nedeniyle 2020 yılında Türkiye'ye gelen yabancı ziyaretçi sayısı bir önceki yıla göre %71,9 azalmıştır. Sonuç olarak, turizm Türkiye ekonomisi için önemli bir sektördür, ancak COVID-19 pandemisinden önemli ölçüde etkilenmiştir. Buna rağmen Türkiye, dünyanın dört bir yanından gelen turistler için popüler bir destinasyon olmaya devam etmektedir (Doğru ve Koçak, 2021).

3.2. Pandemi Dönemi Turizm

2019 Aralık ayında Çin'in Wuhan kentinde tespit edilen ve Dünya Sağlık Örgütü (WHO) tarafından 11 Şubat 2020'de pandemi olarak ilan edilen Koronavirüs (Covid-19), hızla dünya genelinde yayılmıştır. Türkiye'de ise ilk Covid-19 vakası 10 Mart 2020 tarihinde kaydedilmiş ve ilk ölüm ise 17 Mart 2020 tarihinde gerçekleşmiştir (Sağlık Bakanlığı, 2020). Salgın hastalıkların ortaya çıkması, dünya genelinde olduğu gibi Türkiye'de de önlemlerin alınmasını gerektirmiştir (Demir, Günaydın ve Demir, 2020: 82). Bu tedbirler, hastalığın

yayılmasını önlemek ve halk sağlığını korumak amacıyla alınmıştır. Ancak salgın hastalıkların ortaya çıkması, turistlerin seyahat planlarını da etkileyebilmekte ve hedef ülkeye olan güveni azaltabilmektedir (Çeti ve Ünlüönen, 2019: 109).

Covid-19 ile birlikte gelen kısıtlamalar, ailelerin gelir ve harcamalarını azaltmış ve turizm sektörünü olumsuz etkilemiştir. Özellikle hizmet sektöründeki işletmeler ve spor, eğlence ve kültürel etkinlikler gibi faaliyetler, bu kısıtlamalardan en çok etkilenen alanlar arasında yer almaktadır (Bahar ve İlal, 2020: 125-139). Turizm sektörü, Covid-19 salgınının dünya çapında pandemi ilan edilmesiyle ciddi şekilde etkilenmiştir. Dünya Turizm Örgütü'nün açıklamasına göre Covid-19 salgını dünya genelinde bir durgunluğa neden olmuş ve turizm sektörü tüm diğer sektörler arasında en çok etkilenen sektör olmuştur. Bu olumsuz etkinin, dünya turizmi üzerinde uzun süreli bir etkisi olacağı tahmin edilmektedir (UNWTO, 2020).

2020 Mayıs'ta "Birleşmiş Milletler Dünya Turizm Örgütü (UNWTO)" tarafından yayınlanan verilere göre COVID-19'un turizm sektörü üzerindeki etkisi, 2009 yılında yaşanan "Küresel Ekonomik Kriz"in etkisinden üç kat daha büyük olarak ifade edilmiştir. Ayrıca, uluslararası seyahatlerin bir yıl öncesine göre %98 azaldığı gözlemlenmiştir. 2020 yılının ocak-mayıs döneminde ise bir yıl öncesine göre %56 oranında bir düşüş meydana gelmiştir (TÜRSAB, 2021).

Dünya Turizm Seyahat Konseyi (WTTC) verilerinde ise 2020 yılı seyahat ve turizm sektörü zararının 4,5 trilyon doların üzerinde olduğu görülmektedir. Ayrıca 2019 yılı seyahat ve turizm sektörlerinin GSYİH katkısının %10,4 olduğu görülmekteyken, Covıd-19 salgınından kaynaklı kısıtlamalar dahilinde bu oranın 2020 yılında %5,5 oranına kadar düşmektedir. Ayrıca Dünyada 2019 yılı turizm sektörü istihdamının 334 milyon kişi olduğu, 2020 yılında ise 272 milyon istihdama düşerek %18,5 oranında bir daralma yaşadığı, milyon insanın da işsiz kaldığı bildirilmektedir (Özsarıoğlu, 2023: 31).

UNWTO tarafından yayınlanan 2021 Dünya Turizm Barometresi'ne göre Covid-19 salgını nedeniyle milletlerarası turizmde 2020 yılında %73 oranında bir düşüş yaşandığı görülmektedir. 2021 yılının ilk ayında ise kısıtlamaların artması sonucu bu oran %83'e yükselmiştir. Asya ve Pasifik bölgesi %96, Avrupa ve Afrika %85, Orta Doğu %84 ve Amerika kıtası %77 oranında uluslararası geçiş kaybı yaşamıştır. Bu durum turizm

sektörünün önümüzdeki dönemde toparlanması için ciddi zorluklar oluşturacaktır (UNWTO, 2021).

Pandemi döneminde Türkiye'nin turizm sektörü de önemli bir zarar görmüştür. Örneğin, Devlet Hava Meydanları İşletmesi (DHMİ) verilerine göre doğrudan transit yolcular da dahil olmak üzere toplam yolcu trafiği 2018 yılında 210.947.693 kişi iken 2019 yılında 208.911.338 kişiye düştü. Ancak pandemi sonrası 2020 yılında sadece 81.703.685 kişi seyahat etmiş ve 2021 yılında 128.565.706 kişi seyahat edebilmiştir (Ulaştırma ve Altyapı Bakanlığı Devlet Hava Meydanları İşletmesi Genel Müdürlüğü, 2022).

Turizm sektörü, sürekli gelişen ve değişen bir sektördür ve bu değişimler genellikle turistlerin beklentileri ve istekleri doğrultusunda şekillenirken bazen de turistlerin kontrolü dışındaki unsurların etkisi altında kalabilir. Örneğin salgın hastalıklar, terör olayları veya diplomatik ilişkiler gibi sebepler turizm sektörünü etkileyebilir. Covid 19 süreci de ekonomik yönden turizm sektörünü etkilediği gibi insanların turizm tercihlerini de etkilemiştir. Salgının etkisiyle insanlar klasik deniz, kum, güneş ve kalabalık ortamlardaki tatil anlayışı yerine daha bireysel hareket edebilecekleri, mümkün olduğunca az insanla temas kurabilecekleri daha sakin ortamları tercih etmeye başlamışlardır (Düzgün, 2021). Bu süreçte insanların bu ihtiyacına yönelik olarak, lüks kampçılık olarak adlandırılan 'glamping' trendi başlamıştır. Glamping, lüks ve doğayla iç içe, çevre dostu bir konaklama deneyimi sunmakta ve güvenlik veya sağlık sorunlarına neden olmadan konforlu bir tatil imkânı sağlamaktadır (Ergüven ve diğerleri, 2015).

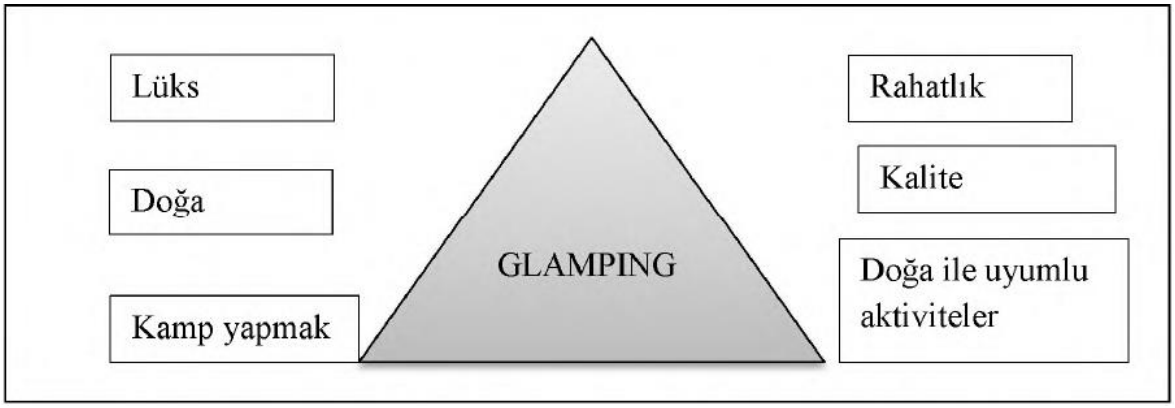
3.3. Glamping Tanımı ve Özellikleri

Glamping, lüks ve konforlu bir kamp deneyimi sunan bir turizm türüdür. Terim, "glamorous" (büyüleyici ve göz alıcı) ve "camping" (kamp yapma) kelimelerinin birleşiminden oluşmaktadır (Vres G. ve Vres T.D., 2015: 42; Pereira, 2013: 11). Glamping, doğaya özlem duyan ancak klasik kampçılıkta karşılaşılan sorunlardan kaçınmak isteyen, günlük ihtiyaçlarını rahatlıkla karşılamak ve kaliteli hizmet almak isteyen, aynı zamanda güvenlik konusunda endişe duymak istemeyen turistler için tasarlanmış yeni bir turizm türüdür. Bu turizm türü, şehir hayatından bunalan ve doğa ile iç içe olmak isteyen ama konforundan ödün vermek istemeyen birçok insan için cazip bir seçenek olabilir (Ergüven ve diğerleri, 2015).

Glamping, doğa turizmi kapsamında yer alan ve kamp yapma zahmeti olmadan doğada konaklama imkânı sunan bir turizm türüdür. Genellikle doğayla iç içe olmayı seven ancak kamp kurulumu için gerekli malzemeleri taşımak istemeyen insanlar tarafından tercih edilmektedir (Çelik, Bahar ve Tatar, 2017). Glamping, turistlere doğa ile bütünleşebilme fırsatı sunarken aynı zamanda lüks ve konforlu bir konaklama deneyimi yaşatır. Bu sayede turistler, doğada vakit geçirirken de rahat ve konforlu bir şekilde konaklama imkânına sahip olurlar (Aksöz, Önem ve Aydın, 2020). Turistler, bungalov, çadır, dome vb. glamping konaklama otellerinde konaklayarak, doğaya uygun konseptlerde lüks ve konforlu bir deneyim yaşayabilmektedirler (Yetgin Akgün ve Ünal, 2021). Türkiye'deki glamping işletmeleri ise yurt şeklindeki çadırlar ve bungalov tarzı evler gibi konaklama birimlerine sahiptir. Ancak her işletmenin iç tasarımı ve eşyaları farklıdır. Bazı işletmeler, yemek yeme alanları ve duş hizmetleri gibi ortak alanları sunarken diğerleri bu hizmetleri özel olarak sunmaktadır (Olcay ve Turhan, 2017).

Glamping, çoğu uzman tarafından genellikle lüks veya en azından yenilikçi kiralık konaklama birimlerinin, genellikle üst düzey hizmetle birleştirildiği ve çekici bir atmosfere gömülü olduğu kamp alanlarında bulunmaktadır. Bu nedenle, glamping, aynı anda doğa ve kamp deneyimi ile lüks ve konforu birleştirir (Walter, Comino ve Reimann, 2019: 314-315).

Glamping turizmi, çevreye saygılı turizm (sustainable tourism) olarak da adlandırılmaktadır. Bu turizm türü, doğal güzellikleri korumak ve bozulmamış çevreleri keşfetmek üzerine kuruludur. İşletmeler, çevreye olan bu duyarlılığı göz önünde bulundurarak, çevre dostu ürünleri tercih ederek, organik ve yerel ürünler sunarak, atıkları geri dönüştürerek ve kaynakları daha verimli kullanarak bu felsefeye uygun hareket etmektedirler (Çelik, Bahar ve Tatar, 2017). Glamping turizmi, doğa ile iç içe kamp yapmak isteyen ancak konforlu bir konaklama arayışında olan turistler için tasarlanmıştır. Bu turizm anlayışı, sıcak su, rahat bir yatak, lüks eşyalar gibi ev konforunu kamp alanında sunar. Ziyaretçilere, çadırların yanı sıra ağaç evler, Eskimo evleri (iglo), kabinler, karavanlar, bungalovlar, cam evler, safari çadırları, çiftlik evleri, kulübeler, lüks villalar, kızılderi çadırları, mağaralar, hobbit evleri ve özel adalar gibi farklı konseptlerde konaklama imkânları sunulmaktadır. Bu konaklama seçenekleri, doğanın güzelliğine sahip çıkan işletmeler tarafından tasarlanmakta ve sunulmaktadır (Boscoboinik ve Bourquard, 2011: 158; Sakacova, 2013). Glamping turizmi kavramsal olarak Şekil 3.1'de gösterilmektedir.



Şekil 3.1. Kavramsal olarak glamping turizmi (Çelik, Bahar ve Tatar, 2017: 1284).

Glampingin özelliklerini genel olarak şu şekilde ifade etmek mümkündür (Çakmaklı, 2019; Gloe, 2016; Filipe, Santos ve Barbosa, 2018; Glamping, 2021):

- Lüks Konaklama: Glamping, lüks bir konaklama deneyimini sunar. Konaklama genellikle iyi donanımlı bir çadır, karavan, yurt, kır evi, çiftlik evi veya ağaç evi şeklindedir. Konaklama birçok durumda banyo, yatak, klima, ısıtma, mini buzdolabı, ocak ve hatta bazı durumlarda bir jakuzi veya özel havuz gibi özelliklere sahip olabilir.
- Doğa ile Uyumlu: Glamping, doğayla uyumlu bir şekilde tasarlanmıştır. Konaklama, çevreye zarar vermeden, yürüyüş yollarına, doğal göller veya nehirler gibi doğal güzelliklere yakın bir şekilde inşa edilir. Aynı zamanda çevre dostu malzemeler kullanarak yapılan yapılar ekolojik açıdan sürdürülebilirdir.
- Yeme ve İçme: Glamping, genellikle tam pansiyon hizmeti sunar. Konaklamaya dahil olan yemekler genellikle yöresel lezzetleri içerir. Misafirler genellikle günlük kahvaltı, öğle yemeği ve akşam yemeği gibi öğünleri dahil olmak üzere konaklamaları boyunca tam bir yemek paketi alırlar. Bazı glamping tesisleri, restoran hizmetleri, piknik sepetleri veya barbekü alanları gibi seçenekler de sunar.
- Özel Deneyim: Glamping, özel bir deneyim sunar. Konaklama genellikle yalnızca birkaç kişiliktir ve doğal güzelliklerin tadını çıkarmak için gizli köşelere yerleştirilmiştir. Konaklama, geleneksel kamp alanlarından daha fazla gizlilik, özgürlük ve huzur sunar.
- Etkinlikler: Glamping tesisleri, konaklama süresince çeşitli etkinlikler sunar. Bu etkinlikler arasında balık tutma, kano, yürüyüş, bisiklet, at sırtında gezi, yoga, kuş gözlemleme, doğal yaşamı keşfetme ve daha fazlası bulunabilir.

- Lüks Hizmetler: Glamping, konforlu bir tatil deneyimi sunarken, aynı zamanda lüks hizmetleri de içerir. Tesisler genellikle spa, masaj, jakuzi, özel yüzme havuzu gibi lüks hizmetler sunar.
- Romantik Tatiller: Glamping, özellikle romantik tatiller için popülerdir.
- Doğa dostu: Glamping tesisleri, doğal kaynakları korumak ve çevreye minimum etki bırakmak için tasarlanmıştır. Bu tesisler genellikle geri dönüştürülmüş veya yenilenebilir malzemelerden yapılır ve atık yönetimi ve su tasarrufu gibi sürdürülebilir uygulamaları benimserler.
- Aktif tatil imkânı: Glamping genellikle doğa yürüyüşleri, bisiklet turları, dağ tırmanışı, su sporları ve diğer doğa aktiviteleri için uygun konumlandırılmıştır. Bu aktiviteler, konukların doğal çevreyle daha iyi bağlantı kurmasını ve keyifli bir tatil deneyimi yaşamasını sağlar.
- Konforlu yataklar: Glamping tesislerinde yataklar genellikle lüks yatak takımları ve rahat yataklar ile donatılmıştır. Bazı tesisler, güneş enerjisi veya ahşap soba gibi alternatif enerji kaynaklarından yararlanarak konukların sıcak ve konforlu kalmasını sağlar.
- Doğal lezzetler: Glamping tatilleri, yerel yemekler ve organik ürünlerin tadını çıkarmak için bir fırsattır. Bazı glamping tesisleri, konukların kendi organik sebzelerini yetiştirebilecekleri veya yemek pişirebilecekleri bahçeler sağlar.

3.3.1. Glamping'in tarihsel seyri

Glamping turizmi, tarihi geçmişiyle ve kapsamlı küresel yayılımıyla ilgi çekicidir. Kökeni Osmanlı İmparatorluğu'na dayanan bu turizm türü, 1900'lerin başlarında Afrika safarileri ile şekillenmiş ve sonrasında diğer kıtalara yayılmıştır (Korjula, 2017). Günümüzde ise Asya, Okyanusya, Kuzey ve Güney Amerika, Orta Doğu ve Avrupa gibi birçok yerde popüler hale gelmiştir. Bu turizm türü, lüks çadırlar, ağaç evler, Eskimo evleri, kabinler ve karavanlar gibi farklı konseptlerle farklılaşmıştır.

Glamping terimi, ilk olarak 2005 yılında İngiltere'de kullanılmıştır ve 2016 yılında Oxford sözlüğüne eklenmiştir (Olcay ve Turhan, 2017). 'Glamping' kelimesi yeni olmasına rağmen, aslında lüks kamp yaşamı fikri eski zamanlara dayanmaktadır. Örneğin, 16. yüzyılda İskoçya'da Kont Athol, Kral 5. James ve annesi için dağlık bölgelerde lüks bir konaklama

hazırlamıştır. Kont Athol, kurduğu bu çadırlara sarayının lüks eşyalarını yerleştirerek, konuklarına unutulmaz bir deneyim sunmuştur (Güvenol ve Sarıbaş, 2022).

1520 yılında İngiltere'nin VII. Henry'si ve Fransa'daki Francis arasında gerçekleşen 'Altın Bezi Alanı' diplomatik zirvesi, tarihteki en abartılı saray çadırı örneklerinden biridir. Bu alanda yaklaşık 2800 çadır kurulmuş ve şarap çeşmeleri dahi oluşturulmuştur.

Osmanlılar, aynı dönemlerde gösterişli saray çadırlarını askeri amaçlar için kullanmışlardır. İmparatorluğun zanaatkâr ekibi, bu imparatorluk (emperyal) çadırlarını kurmak ve korumak için orduyla birlikte seyahat etmiştir. Glamping turizmi Türkiye açısından tarihsel olarak Osmanlı İmparatorluğu dönemine kadar dayanabilir. Osmanlılar, padişahlar için abartılı ve lüks taşınabilir çadırlar kurarak, bu tarz konaklama imkânlarının öncülüğünü yapmışlardır. Glamping turizminin gelişiminde bir diğer önemli aşama ise 1990'ların başında Afrika'daki safari seyahatleri sırasında, zengin gezginlerin lüks yaşam tarzlarından vazgeçmeden çadırda konaklama istekleriyle ortaya çıkmıştır. Turizm sektörü, bu taleplere mükemmel bir şekilde cevap vererek, glamping konaklama imkânlarını sunmaya başlamıştır. Dolayısıyla glamping kavramının tamamen yeni bir fikir olmadığı açıkça söylenebilir (Sakacova, 2013: 1).

Thomas Hiram Holding liderliğindeki İngiliz yüksek sınıfının gösterişli üyeleri, 19. yüzyılın sonunda glamping'in öncüleriydi. Holding, tasarladığı portatif kamp ekipmanlarını ilgililere tanıtarak, onları kendisiyle iletişime geçmeleri için davet etti. Bu, Kamp ve Karavan Kulübü'nün kuruluşuna yol açtı ve günümüzde 300.000'den fazla üyesi bulunuyor. Kitaplar yazan Holding, Connemara'da "Bisiklet ve Kampçılık" adlı kitabıyla da tanınıyor (Bonifacic, Milohnic ve Cerovic, 2017). Glamping, modern teknolojinin imkânlarıyla geçmişin kampçılık anlayışını birleştirerek ortaya çıkan bir turizm çeşididir. Lüks kampçılık olarak da bilinen glamping, özellikle kitle turizmine veya deniz-kum-güneş turizmine alternatif olarak sunulan ve ekoturizm kapsamında özel ilgi turizmi olarak öne çıkan bir turistik ürün türüdür (Ergüven ve diğerleri, 2015: 256).

Dr. Eike Wenzel, Almanya'daki Trend ve Gelecek Araştırmaları Enstitüsü'nün yöneticisi olarak glamping'in doğa sevgisiyle ilgili bir turizm anlayışı olduğunu belirtir. Wenzel'e göre glamping'in doğal hayatla ilgili ihtiyaçları karşıladığı için, "Landlust" adlı derginin gelişimiyle örtüşen bir süreçtir. Glamping, turizm sektöründeki standartlaşmaya ve kitle turizmine tepki olarak ortaya çıkmış ve doğa, lüks ve hazza olan ihtiyaçları birleştiren kamp

ürünleri sayesinde daha da popüler hale gelmiştir. Wenzel, bu nedenle glamping turizminin 2020 yılında turizm trendi olacağını öngörmektedir (Ergüven ve diğerleri, 2015: 258).

Tüketici tercihleri, kitle turizminin gerilemesi ve özel ilgi turizmi çeşitlerinin popülerleşmesi yönünde değişmektedir. Glamping benzeri alternatif konaklama seçeneklerinin tercih edilmesi, geleneksel otellerin ve her şey dahil konseptinin ilgi kaybına uğradığını göstermektedir. Covid-19 salgınıyla birlikte, insanların teması azaltmayı tercih ettiği izole tatillerin popülerleştiği ve örneğin "karavan turizmi"nin yeniden canlandığı görülmektedir.

3.3.2. Glamping konaklama türleri

Glamping turistlere çeşitli konaklama seçenekleri sunar. Ekolojik bölmeler, çadırlar, kabinler, doğal rustik evler, mağaralar, kubbeler ve karavanlar gibi farklı yerlerde konaklama imkânı sağlar. Glamping, henüz yeni bir turizm türü olmasına rağmen çeşitli türleri vardır ve bazıları aşağıdaki gibidir (Göktaş, Çetin ve Kızılırmak, 2017: 109-111):

- Ağaç Evleri: Ağaçların dallarında inşa edilen, dikey bir dünya yaratmak için gövde ve güçlü ağaçların şekillerini kullanan ahşap yapılar.
- Bungalovlar: : Tek katlı veya eğimli bir çatıya yerleştirilmiş ikinci bir kata sahip olan ve geniş verandalarla çevrili olabilen yapılardır.
- Otağılar: Genişleyebilen ahşap kafes kesitleriyle inşa edilmiş. Yüksek rüzgarlara ve düşük sıcaklıklara son derece dayanıklı olan, geniş ve silindirik tabanlı yapılar.
- Bungalovlar: Tek katlı veya eğimli bir çatıya yerleştirilmiş ikinci bir kata sahip olan ve geniş verandalarla çevrili olabilen yapılardır.
- Safari Çadırları: Alüminyum veya fiberglas direklerle desteklenen, büyük ve dikdörtgen yapılar. Zaman zaman platform veya sert zemin kaplamaları da kullanılabiliyor.
- Seyyar Karavanlar (Campervan): Tekerlekli, gezgin yapılardır. Mütevazi bir yatak odası, mutfak ve banyosu bulunmaktadır.
- Kızılderili Çadırları (Tipis): Küçük kapısı olan ve üçgen biçiminde dikilmiş ahşap direklerden yapılmış, dayanıklı kumaşlar ile kaplanmış yapılardır.
- Kubbe Çadırlar: Ahşap, demir, çelik veya yüksek dirençli PVC/polyester kumaştan inşa edilebilen, kubbe şeklinde yapılar.

- Eko-Kapsüller: Tamamen yerel, doğal ve geri dönüştürülebilir kaynaklardan oluşan ve rahat bir konaklama imkânı sunan yapılardır.
- Doğa Kulübeleri: Lüks bir otelin konforunu doğal peyzaj içinde sağlayabilen yapılar.
- Mağaralar: Yataklar, masa ve fenerler eklenerek konaklama tesisine dönüştürülmüş, doğal olarak oyulmuş mağaralar.
- Ahırlar: Tarımsal yapılar olarak inşa edilmiş, daha sonra restore edilerek turizme kazandırılan yapılar.
- Kulübeler: Saman, ahşap, kar, taş, çimen, palmiye yaprakları, buz, dallar, örtüler veya kumaştan inşa edilmiş yapılardır.
- Eskimo Evler: Inuitler tarafından yapılan bu evler, katı kar bloklarından oluşan kubbe şekilli yapılar olarak bilinir ve Grönland, Kanada ve Amerika Birleşik Devletleri'nin Arctic bölgelerinde sıklıkla kullanılır.
- Kabinler: Ahşap veya çadır şeklinde olabilen bu yapılar, otantik mimarileriyle öne çıkar ve tatilcilere doğal bir konaklama imkânı sunar.
- Vagonlar: Tren vagonlarının restore edilmesiyle yapılan bu evler, içerisinde mutfak ve banyo olanakları sunarak konforlu bir tatil geçirmek isteyenler için idealdir.
- Kır Evleri: Orta büyüklükte ahşap veya taş yapılardan oluşan bu evler, doğal bir çevrede konforlu bir konaklama sağlar.
- Çan Çadırlar: Tek bir merkez direğe sahip koni şeklindeki yapılarıyla dünyanın en popüler çadırlarından biridir.
- Villalar: Lüks bir donanıma sahip, doğal bir çevrede yer alan ahşap ve taş yapıların, çok katlı ve havuzlu tesislere dönüştürülmüş halleridir.
- Yüzen Evler: Su üzerinde yüzen veya su seviyesinin üstünde dikmelerle desteklenen yapılar olarak bilinirler.
- Çadırlı Kabinler: Ahşap veya ince beton iskeletlerin üst kısmının gerdirilmiş kumaşla kaplandığı bu yapılar, rahat bir tatil geçirmek isteyenler için idealdir.
- Kulübeler: Rahat odalar, şömineler, tuvalet banyo ve güneş enerjisi sistemleriyle donatılan, büyük pencerelerle çevredeki manzaraların izlenebildiği yapılar olarak bilinirler.
- Lüks Çadırlar: Ahşap zemini, rahat yatakları, şık dekorasyonu ve evin tüm olanaklarını sunan konaklama tesisleridir. Ortalama bir kamp çadırından çok daha üstündürler.

- Karavanlar: Doğayla iç içe olan bu yapılar, sabit bir noktada kalabileceği gibi, bir aracın arkasına bağlanabilen kabin tarzı tekerlekli araçlarda da bulunabilirler. Genellikle ahşaptan yapılmıştır.
- Yüksek Kabinler: Su ya da kara üzerinde konaklama imkânı sağlayan, standart bir kabin ile ağaç evi arasında bir karışım olan yükseltilmiş kabinlerdir.
- Adalar: Doğal yapılar konaklamaya uygun hale getirilerek küçük bir ada üzerinde oluşturulmuş yaşam alanlarıdır. Kiralanarak tatilcilerin yalnızca adayı keşfetmeleri amaçlanır.
- Kuleler: Yel değirmenleri, deniz fenerleri ve orta çağ kulelerinden esinlenerek inşa edilmiş konaklama tesisleridir.
- A-Frame Kabinler: Üçgen çatı yapısına sahip geniş açıdan bakıldığında A çerçeveyi andıran ahşap yapılardır. İçerideki eşsiz iç mimarisi ve lüks tasarımıyla bilinirler.

Glamping tesisleri, çeşitli konseptlerde tasarlanabilir ve bu farklılıklar, pazarlama faaliyetleri açısından oldukça önemlidir. Bu durum, işletmelerin kendilerini benzersiz bir şekilde konumlandırarak rekabet avantajı elde etmelerini sağlayabilir.

3.4. Türkiye’de Glamping Hizmeti Sunan İşletme Örnekleri

Glamping konaklama seçeneklerinin farklı türleri Türkiye'nin farklı bölgelerinde ve illerinde mevcuttur ve ülkemizin doğal güzellikleri bu turizm sektörünü desteklemektedir.

Kapadokya Ihlara Konakları: Güzelyurt adı verilen eski bir Rum kasabasında yer alır ve Kapadokya'nın önemli yerlerinden biri olan Ihlara Vadisi'ne sadece 14 km uzaklıktadır.

Club Amazon Bördiibet: Marmaris ile Datça arasındaki Bördübet bölgesinde yer almaktadır ve doğal bir ortamda çam ve günlük ağaçları arasında konumlanmıştır. Glamping turizmi için farklı konaklama seçenekleri sunan tesis, butik van, çingene arabası, taş oda, mini bungalov, küp oda, bonibon arabalar gibi farklı alternatiflere sahiptir. Tesis, plaja 500 metre mesafededir ve resimleri aşağıda görülebilir.

Assos Gargara Doğal Tatil Kampı: Assos ilçesi, Marmara bölgesinde yer alır ve doğayla iç içe sessiz ve huzurlu bir tatil yapmak isteyenler için ideal bir seçenektir. Kamp alanları temiz ve yemekleri oldukça lezzetlidir. Hem kamp yapılabilir hem de dinlenilebilir.

Compo Partakal Çıralı: Çıralı'nın doğal güzelliklerinde yer alan bir işletme olan bu tesis, Kemer ilçesi sınırları içerisinde yer almaktadır. Doğa içinde konforlu bir konaklama sunan Safari Çadırda Eco Glamping tesisinde, misafirler dağ manzaralı geniş çadırlarda konaklayabilirler. Ekolojik portakal bahçesi ile çevrili olan tesis, lüks bir kamp deneyimi yaşatmaktadır. Her çadır, özel bir mutfak ve vantilatör gibi olanaklar sunmaktadır. Çadırlarda özel banyo da mevcuttur. Tesise 1 km uzaklıktaki Çıralı Plajı doğal bir güzelliktir ve bahçede çok sayıda gölgelik alan vardır.

Livera Camping: Trabzon'un Maçka ilçesinde yer alan bir tesiste, doğal güzellikler ekoturizme katkı sağlıyor. Çevre dostu bir yaklaşımla, güneş ve rüzgâr enerjisi kullanarak enerji ihtiyaçlarını karşılıyorlar. Tesis, doğanın güzelliklerini misafirlerine huzurlu bir ortamda sunarak, kamp alanlarında doğaya ve insanlara saygılı bir konaklama deneyimi sunmaktadırlar.

Hızır Kamp: Balıkesir ilinin Edremit ilçesinde yer alan tesis, Kazdağları Milli Parkı ile sınırda ve Zeytinli Çayı'nın yanında konumlanmıştır. Misafirlerine serin sularında yüzme, muhteşem doğasıyla Kazdağları'nda yürüyüş gibi birçok farklı aktivite seçeneği sunmaktadır.

Elbaba Konaklama Tesisleri: Tunceli'nin Ovacık ilçesinde bulunan Elbaba Konaklama Tesisleri öncelikle bölge sakinleri tarafından tercih edilen bir aile işletmesidir.

Sapanca Bungalovları: Sakarya'nın Sapanca ilçesinde bulunan, genellikle gölün kenarında/ göl manzaralı olarak konumlanmışlardır. Misafilerine günbatımını doyasıya izleyebilecekleri, jakuzili veya özel havuzlu odalar, evcil hayvanıyla tatil yapma imkânı gibi birçok olanak sunmaktadır.

4. METİN MADENCİLİĞİ, DUYGU ANALİZİ VE MAKİNE ÖĞRENMESİ

4.1. Metin Madenciliği Kavramının Tanımı ve Diğer Alanlarla İlişkisi

Metin madenciliği, doğal dil işleme, veri madenciliği ve makine öğrenimi gibi farklı alanların kesişiminde yer alan bir araştırma alanıdır. Bu alanda, büyük miktarda metin verilerinin analizi, keşfi ve anlaşılması amaçlanmaktadır. Bu veriler, sosyal medya mesajları, blog yazıları, e-postalar, gazete makaleleri ve daha birçok kaynaktan elde edilebilir (Feldman, 2013). Metin madenciliği, büyük verilerin içerisinde yer alan yapılandırılmamış metinsel verilerden, anlamlı modeller, trendler, kalıplar ve kurallar keşfetmeye yönelik bir uygulama alanıdır. Bu alanda veri madenciliği teknikleri özel olarak kullanılır ve geniş bir veri yelpazesinden faydalı bilgi elde etmek amaçlanır. Metin madenciliği, yazılı metinlerin analizi ve yorumlanması sürecini içerir ve bu sayede işletmeler, araştırmacılar ve diğer kullanıcılar için anlamlı veriler elde etmek mümkün olur (He, Zha ve Li, 2013: 464).

Metin madenciliği, doğal dil işleme (NLP) alanında yer alan bir disiplindir ve bilgisayar programları kullanarak metinleri analiz etmek, özelliklerini çıkarmak ve anlamlı bilgiyi elde etmek için kullanılır. Bu süreçte, metin madenciliği teknikleri, metin verilerindeki yapısal olmayan bilgileri ortaya çıkarır. Metin madenciliği teknikleri, ayrıca makine öğrenimi ve yapay zeka teknikleri ile birlikte kullanılarak, büyük veri setlerindeki bilgiyi keşfetmek ve anlamak için de kullanılabilir (Pang ve Lee, 2008).

Metin madenciliği, dokümanlardan çabucak özetler ve anahtar kelimeler elde etmek için kullanılabilecek bir yöntemdir. Bu, doğru ve etkili kararlar almak için zaman ve emek açısından oldukça yararlı olabilir. Diğer yandan, metin madenciliği. Yüksek hacimli veri tabanlarından değerli bilgileri bulmayı amaçlayan bir disiplindir. Veri tabanındaki veriler genellikle programlanmış ve otomatik işleme için uygun bir yapıdadır. Ancak analiz edilecek metinler insanlar tarafından tamamen hata yapmaya açık, karmaşık bir yapıya sahiptir (Oğuz, 2009).

4.1.1. Metin madenciliği uygulama alanları

Metin madenciliği, doğal dil işleme, makine öğrenimi ve istatistiksel yöntemler gibi teknikleri kullanarak metinlerdeki bilgileri analiz etme sürecidir (Bhatia ve Jain, 2019). Bu teknoloji, metinlerin bulunduğu birçok farklı uygulama alanında kullanılmaktadır. Temel metin madenciliği uygulamalarının en çok kullanıldığı alanlar şunlardır (Bolasco, Canzonetti, Capo, Della Ratta-Rinaldi ve Singh, 2005: 31-50):

- Siyasi kurumlar, siyasi analistler, kamu yönetimi ve yasal belgeler.
- Telekomünikasyon, enerji ve diğer hizmet endüstrileri,
- Sağlık hizmetleri, ilaç ve araştırma şirketleri,
- Bankalar, sigorta ve finans piyasaları,
- İnternet ve bilgi teknolojisi sektörü,
- Medya ve yayıncılık

4.1.2. Metin madenciliği aşamaları

Metin madenciliği süreci, çeşitli kaynaklardan bir belge koleksiyonu ile başlar. Metin madenciliği aracı, belirli bir belgeyi alır ve biçim ve karakter kümelerini kontrol ederek onu önceden işler. Daha sonra belge bir metin analizi aşamasından geçer. Metin analizi, metinden yüksek kaliteli bilgi türetmek için anlamsal analizdir. Birçok metin çözümleme tekniği mevcuttur; organizasyonun amacına bağlı olarak tekniklerin kombinasyonları kullanılabilir. Bazen metin analiz teknikleri, bilgi çıkarılana kadar tekrarlanır. Ortaya çıkan bilgi, bir yönetim bilgi sistemine yerleştirilebilir ve bu sistemin kullanıcısı için bol miktarda bilgi elde edilebilir (Gaikwad, Chaugule ve Patil, 2014: 42-45).

Metin madenciliği aşamaları aşağıda açıklanmaktadır.

<u>Veri setinin oluşturulması</u>

Metin madenciliği, bir belge koleksiyonundan başlayarak, metinleri önceden işleyen ve daha sonra anlamsal analiz yaparak yüksek kaliteli bilgi üreten bir süreçtir. Metin analizi için birçok teknik mevcuttur ve bunlar organizasyonun amaçlarına bağlı olarak kombinasyon

halinde kullanılabilir. Teknikler, bazen bilgi çıkarılana kadar tekrarlanabilir. Elde edilen bilgi, yönetim bilgi sistemine aktarılabilir ve kullanıcılar tarafından bol miktarda bilgi elde edilebilir (Oğuzlar ve Kızılkaya, 2019: 24).

Metin madenciliği, sosyal medya platformlarının yanı sıra web sayfaları, e-postalar, bloglar, dijital ve fiziksel kütüphaneler, makaleler, haber metinleri, kurum dokümanları ve ses dosyalarının transkripsiyonları gibi birçok farklı kaynaktan yararlanabilir (Han, Kamber ve Pei, 2012: 596; Aksu, 2021: 19). İnternet ortamından metinlere - erişime açıksa - direkt olarak ulaşılabileceği gibi çeşitli programlar vasıtasıyla da ulaşmak mümkündür. Metin madenciliği çalışmalarında kullanılabilecek metinler, birçok farklı formatta olabilirler. Word, excel, csv, pdf ve düz metin dosyaları gibi formatlarda bulunabilirler. Ancak bu metinlerin hepsi aynı formata dönüştürülmelidir. Bu amaçla, toplanan tüm metinler bir derleme olarak adlandırılır (Ağdeniz, 2017: 80).

Doğal dil işleme

Metin madenciliği için kullanılan yöntemler, bilgisayar bilimi, makine öğrenimi ve hesaplamaya dayalı dilbilimi alanlarının etkileşimli çalışması sonucu doğal dil işleme adı verilen bir metotlar bütününü oluşturur (Adalı, 2012: 8-9). Doğal dil işleme, dijital ortamda metinleri işleyebilen yazılımların geliştirilmesi ile ilgilenmektedir. Ayrıca doğal dil işlemenin bir başka uygulama alanı da ses dalgalarının dijital formatta analiz edilmesidir (Altıntaş, 2014).

Doğal dil işleme, metin verilerindeki dil bilgisi, semantik, sözdizim ve morfolojik özellikleri anlamak için kullanılan bir dizi teknik ve algoritmadır. Bu teknikler, metin verilerindeki kelime sıklığı, kelime öbekleri, cümlelerin anlamı, duygusal tonu, konuları ve daha birçok özellikleri analiz etmek için kullanılır (Chowdhary, 2020: 603-649).

Metin verilerindeki anlamı, gramer yapısını ve yinelenen verileri tespit etmek için kullanılan teknikler bütünü olan doğal dil işleme, eğitim sistemlerinde de faydalıdır. Doğal dil işleme algoritmaları, giriş metinlerinden bilgi çıkarmak için kullanılır (Arslan, 2021: 33-34). Bilgisayarların anlayabileceği formata dönüştürülmek üzere insanların doğal dillerle iletişimini işleyen bir bilim dalı olan doğal dil işleme, çeşitli teknikler ve algoritmalar kullanmaktadır (Vosoughi, Deb ve Aral, 2018). Doğal dil işleme, ilk örneklerinden biri olan

makine çevirisi ile başladı ve doğal diller arasındaki çeviri işlemini ele almıştır. İlk dönemlerinde geleneksel yöntemler kullanılmış ancak makine öğrenmesinin kullanımı ile birlikte daha etkili sonuçlar elde edilmeye başlanmıştır. Makine öğrenmesinin doğal dil işleme alanındaki kullanımı, bu alanda yapılan çalışmalarda artışa yol açmış ve doğal dil işleme daha da tanınır hale gelmiştir (Taşkın, 2020: 25).

İşaretleme (Tokenization)

Metnin okunabilirliğini artırmak için, dizgeciklere ayırma işlemi yapılarak metinden okunamayan karakterler, noktalama işaretleri, tab karakterleri ve satır sonu karakterleri temizlenir (Tunalı, 2011: 13). Eğer web temelli bir veri işleniyorsa, XML ve HTML gibi uzantılar da dahil olmak üzere metindeki okunamayan karakterler, noktalama işaretleri, tablar ve satır sonu karakterleri temizlenir (Çelikyay, 2010: 60). Dizgecikler, metin içerisindeki anlamlı sözcükleri ifade eden yapılar olup, XML ve HTML uzantıları web dili için anlamlı olsa da metnin içeriğine dair herhangi bir bilgi barındırmazlar. Dizgeciklere ayırma işlemi, metin madenciliği için gereksiz verilerden arındırarak anlamlı sözcüklere erişmeyi amaçlar (Değer, 2017: 67).

Kök bulma (Stemming)

Kelimenin anlamını koruyan, ek almaksızın anlaşılabilen başlangıç hali kök olarak adlandırılır. Türkçe, sondan eklemeli bir dildir, bu nedenle kelime kökleri, çeşitli yapım ve çekim ekleri almaktadır. Bu sayede aynı kökten farklı kelimeler oluşabilmektedir. Kök bulma işlemi ile çoğul ekler (evler → ev) ve fiil çekim ekleri (okuyor → oku) ortadan kaldırılır. Ancak kök bulma sırasında iki farklı sorunla karşılaşılabilmektedir: çok fazla ek çıkarılarak farklı bir kelimeye ulaşılması veya yeterince ek çıkarılmadan asıl köke ulaşılamaması (Çelikyay, 2010: 61). İyi bir dilbilgisine sahip olmadan yapılan kök bulma işlemi, kelimenin özel yapısı nedeniyle meydana gelen harf düşmesi, harf yumuşaması, kaynaştırma harfi alma veya büyük ünlü uyumu küçük ünlü uyumu gibi süreçler nedeniyle ciddi hatalara neden olabilir. Bu yüzden kök bulma işlemini gerçekleştirirken son derece dikkatli olunması gerekmektedir (Kızılkaya, 2018: 9).

Durak kelimeleri çıkartma

Metin içindeki anlamı pek taşımayan edatlar, bağlaçlar ve zamirler gibi sıklıkla kullanılan durak kelimeler, genellikle metinden çıkarılmalıdır (Haltaş, Alkan ve Karabulut, 2015: 419). Durak kelimeler, anlamsal bir değeri olmayan fakat yazım dilinde sıkça kullanılan edatlar, bağlaçlar ve zamirler gibi kelimelerdir. Bu kelimelerin metinden atılması, anlamsal olarak bilgi taşımayan kelimelerin çıkarılması anlamına gelir. Bu şekilde, analiz edilecek kelime sayısı önemli ölçüde azalabilir. Türkçe'de sıkça kullanılan kelimeler arasında "ile", "çünkü" gibi kelimeler bulunmaktadır. Bu kelimelerin atılması bile, analiz edilen kelime sayısında önemli bir düşüşe neden olabilir ve yapılan analizin doğruluğunu artırabilir (Kızılkaya, 2018: 10).

Terim ağırlıklandırma

Metinlerdeki kelimelerin katkısı metne göre değişkenlik gösterir. Bazı kelimeler metinde sadece bir kez geçebilirken, bazıları defalarca tekrar edebilir. Bir kelimenin önemi, diğer tüm kelimelerin frekansına göre göreceli değeri ile belirlenir. Bir kelimenin sıklığı, bir dokümanın belirli bir kategoride sınıflandırılmasında etkili bir rol oynar. Eğer bir kelime birden çok dokümanda geçiyorsa, o kelimenin ayırt edici özelliği azalır. Ayrıca belirli terimler bir arada kullanıldıklarında genellikle aynı kategoriye atanırlar (Çoban ve Özyer, 2018).

Vektör oluşturma

Belgelerin tamamı, bir kelime vektörü olarak temsil edilir. Bu sayede, bilgisayar tarafından iki boyutlu bir dizi olarak işlenebilirler. Bu yöntem, bilgisayar işlemlerini oldukça kolaylaştırır. Bu metin, bir matrisin temel bileşenlerini açıklamaktadır. Bu matris A olarak adlandırılabilir ve her bir elemanı, D belgesinden geçen T kelimesinin önemini ifade eder. Bu rakam, kelimenin belgedeki frekansını yansıtır. Kelime vektörü, N-Gram kelime modeli ve boyut ağırlıklandırma, vektör oluşturmak için kullanılabilecek üç farklı alternatiftir (Güven, 2007: 46). Kelime vektörü, kelimelerin tekil olarak ele alınması ve sınıflarıyla birlikte gösterilmesiyle oluşan bir vektördür. Bu vektör sayesinde, verilerin en iyi şekilde temsil edilebilecek öznitelikler kümesi oluşturulur (Topaçan, 2016: 83).

N-Gram Kelime Modeli, belgedeki kelimelerin "n" adedinin bir arada kullanılarak kelime vektörü oluşturulmasını sağlar. Tek bir kelime yalnızca anlam bütünlüğünü sağlamazken, yanındaki kelimelerle birlikte ele alındığında daha doğru sonuçlar elde edilebilir. N sayısı, yanındaki kelime sayısını belirtir. N iki olarak ele alındığında kelimeler ikişerli gruplar halinde ele alınır, n üç olarak ele alındığında ise kelimeler üçerli gruplar halinde ele alınır. Pratikte, genellikle kelimeler tek, iki veya üçerli gruplar halinde ele alınır (Kutlugün ve Şirin, 2018: 1-4).

Boyut ağırlıklandırma, kelime vektörü oluşturulurken kelimelerin sıklıklarını baz alarak yapılan bir yöntemdir. Kelimelerin vektörde yer alacağı değerlerin belirlenmesi için geliştirilmiş iki yöntem, kelimenin varlığına göre 1 ya da 0 değerinin atanması ve kelimenin frekansının vektöre dahil edilmesidir (Kızılkaya, 2018: 12).

Özellik seçme

Veri ön işleme işleminden sonra özellik seçimine geçilir. Bu aşama, veri kümesindeki özelliklerin seçilmesi için bir süzme işlemidir. Veri kümesi, özellik seçiminden önce çok yüksek boyutlu bir yapıya sahiptir ve hem nitelikli hem de niteliksiz terimler içerebilir (Haltaş ve diğerleri, 2015: 420). Özellik seçme aşamasında, verilerin ön işleme aşamasından geçirilerek önemli kelimeler belirlenir ve ilişkisiz özellikler çıkarılır. Ayrıca az sayıda dokümanda yer alan özellikler ayıklanırken, çok sayıda dokümanda yer alan özellikler de azaltılır (Ergün, 2012: 51). Verilerin özellik seçimi aşamasından geçirilmesi sonrasında boyutlar küçültülmüş, niteliksiz terimler ayıklanmış ve metinler daha işlenebilir hale getirilmiştir. Bu süreç, sınıflandırma performansını artırdığı gibi zaman açısından da avantaj sağlar (Kızılkaya, 2018: 12).

4.1.3. Metin madenciliği metotları

Metin Madenciliği metotları, Bilgiye Erişim, Bilgi Çıkarımı, Web Madenciliği ve Kümeleme olmak üzere dört grupta incelenmektedir (Zohar, 2002).

Bilgiye erişim

Bilgiye erişim sistemleri, kullanıcılara bilgi sağlamak amacıyla geliştirilen sistemlerdir. Örneğin, arama motorları gibi popüler platformlar, bu tür sistemlere iyi bir örnektir. Kullanıcılar, belirli anahtar kelimeleri kullanarak istedikleri bilgiyi içeren belgelere erişebilirler. Bu sistemler, geniş metin veri tabanlarından tüm belgelere erişebilme yeteneğine sahiptir ve kullanıcının ihtiyaç duyduğu bilgiye en uygun sonuçları sunmak için belgeleri filtrelerler. Bu sayede, kullanıcılar hedefledikleri bilgiye daha hızlı ve etkili bir şekilde ulaşabilirler (Oğuzlar, 2011: 15).

Bilgiye erişim sistemleri, metin madenciliği yöntemlerine dayanır. Kullanıcıların ihtiyaç duydukları bilgiye erişim sağlamak için farklı teknikler kullanılır. Örneğin kelime bazlı arama, belirli anahtar kelimeleri içeren belgeleri bulmada yaygın olarak kullanılır. Bunun yanı sıra, metin sınıflandırma ve öznitelik çıkarma gibi teknikler de kullanıcıların aradıkları bilgiye daha hızlı erişmelerine yardımcı olur (Parlak, 2022).

Bilgi çıkarımı

Bilgi çıkarımı, yapılandırılmış bilgilerin otomatik olarak çıkarılmasını ifade eder. Bu bilgiler, varlıkların, varlıklar arasındaki ilişkilerin ve öz niteliklerin yapılandırılmamış kaynaklardan elde edilmesini içerir. Genellikle doğal dil işleme tekniklerini kullanarak bir metin üzerinde belirli kriterler hakkında bilgi edinmeyi amaçlar. Temel amaç, büyük miktarda veriyi otomatik olarak işleyen ve insan müdahalesini en aza indiren bir yazılım geliştirmektir. Bilgi çıkarımı genellikle yazılı metinler üzerinde gerçekleştirilir, ancak bu metinlerin kaynakları farklı alanlarda bulunabilir (Değer, 2017: 65). Örneğin, veritabanları, internet üzerindeki belgeler veya taranmış metinler, bilginin çıkarılacağı veri kaynakları olabilir. Bu yöntem, veri analitiği, bilgi yönetimi ve yapay zeka gibi alanlarda yaygın olarak kullanılmaktadır. Bilgi çıkarımı sayesinde, büyük miktarda veri içinde gizlenmiş olan önemli bilgileri keşfedebilir, ilişkileri anlayabilir ve daha bilinçli kararlar alınabilmektedir (Pasin, 2018).

Web madenciliği

Web madenciliği, kendi uygulama alanına sahip olan bir metin madenciliği alanıdır. Benzersiz yapısı ve muazzam veri hacmi, web belgelerini diğer metinlerden ayıran özelliklerdir. Web belgeleri genellikle köprülerle birbirine bağlı olan sayfalar şeklinde yapılandırılmış metinlerdir. Bu farklılık, web madenciliğinin zorlukları ve fırsatlarıyla birlikte gelir (Chakrabarti, 2002).

Web madenciliği, web üzerindeki bilgiyi keşfetmeyi, analiz etmeyi ve anlamayı amaçlar. İnternetin, sosyal medya platformlarının yükselişi gibi etkilerle günlük yaşamımızda daha da yaygınlaşmasıyla, web madenciliği de hızla gelişen bir alandır. Bu alanda yapılan çalışmalar, web üzerindeki veriyi çeşitli yöntemlerle analiz ederek önemli bilgilerin keşfedilmesini sağlar. Örneğin, web madenciliği teknikleri arama motorlarının etkinliğini artırmak, kullanıcı tercihlerini anlamak, e-ticaret trendlerini takip etmek gibi birçok alanda kullanılır (Miner, Delen, Fast, Hill, Elder ve Nisbet, 2012). Web madenciliği, büyük miktarda web verisi içindeki bilgiyi elde etme, desenleri tanıma, kullanıcı davranışını analiz etme gibi konuları ele alır. Veri madenciliği, makine öğrenimi, doğal dil işleme ve ağ analizi gibi disiplinlerden faydalanarak web madenciliği çalışmaları gerçekleştirilir. Bu sayede, web üzerindeki verilerin keşfedilmesi ve değerli bilgilerin elde edilmesi mümkün hale gelir (Saini ve Mohan, 2015).

Kümeleme

Kümeleme birimlerin ortak özelliklerini ortaya çıkarmayı ve bu gruplar hakkında genel tanımlamalar yapmayı sağlayan bir analiz yöntemidir. Kümeleme analizinin amacı, gruplandırılmamış verileri benzer özelliklere sahip olan gruplara ayırmak ve araştırmacıya yararlı, özetlenmiş bilgiler sunmaktır (Dinler, 2014). Sınıflandırma tekniğinde, sınıflar önceden tanımlanmıştır ve veriler bu sınıflara göre analiz edilir. Ancak kümeleme tekniğinde veriler, belirli bir sınıf olmadan incelenir. Başlangıçta, genellikle sınıflar bilinmez ve önceden hazır değildir. Kümeleme analizi sürecinde, sınıf sayısı esnek bir şekilde artırılabilir. Veriler, benzerliklerini en üst düzeye çıkarmak için gruplandırılır veya sınıflandırılır (Pasin, 2018). Kümeleme yöntemleri, çeşitli disiplinlerde kullanılır. Örneğin, pazarlama alanında müşteri segmentasyonunda, biyoinformatikte genetik verilerin analizinde, veri madenciliğinde büyük veri kümelerinin incelenmesinde ve sosyal ağ

analizinde kullanılır. Kümeleme analizi, verilerdeki yapıları ve ilişkileri keşfetmek, veri setlerini daha anlamlı ve kullanışlı bir şekilde düzenlemek ve karar verme süreçlerini desteklemek için önemli bir araçtır (Larose 2005: 16).

4.2. Duygu analizi

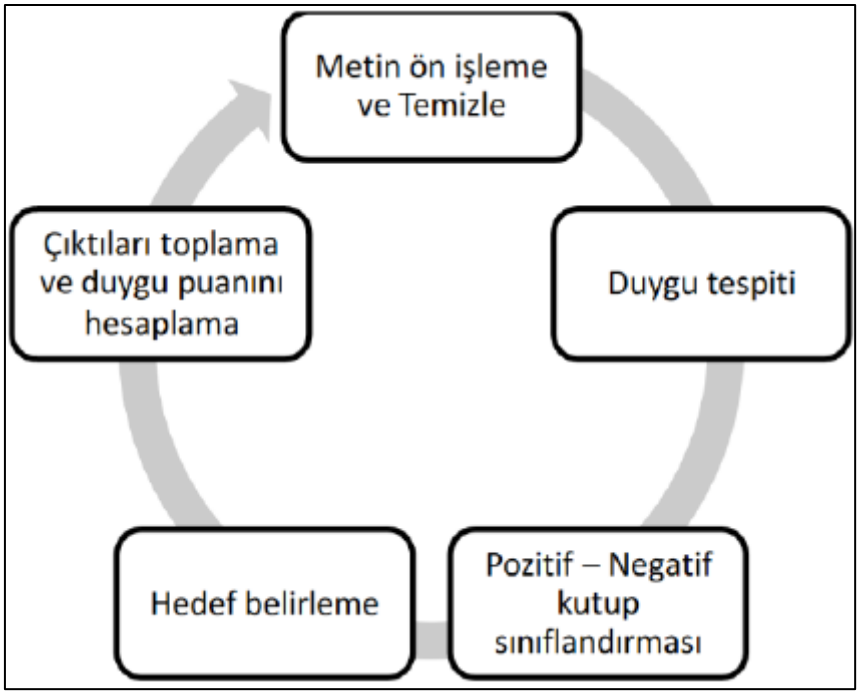
İnsanların duyguları, tarih boyunca bireylere ve toplumlara yön veren ve üzerinde önemli çalışmalar yapılan bir özellik olmuştur. Duygular, insanların yaşamları boyunca karşılaştıkları olayların veya nesnelerin anlamını gösteren en önemli unsurdur. İnsanların ilgi duydukları konu veya düşünceye karşı duygusal yaklaşımları, o konuyu baskın veya önemsiz hale getirebilir. Bu nedenle, insan duygularının analiz edilmesi bir ihtiyaç haline gelmiştir (Eröz, 2011).

Duygu analizi, insanların duygusal durumlarını anlamak, değerlendirmek ve yorumlamak için kullanılan bir disiplindir. Bu analiz, metinlerden, sosyal medya paylaşımlarından, müzikten veya diğer veri kaynaklarından toplanan veriler üzerinde yapılabilir. Duygu analizi, metin madenciliği, doğal dil işleme ve yapay zeka tekniklerini kullanarak duygusal ifadeleri tanımlamayı ve anlamlandırmayı hedefler (Medhat, Hassan ve Korashy, 2014). İnsanların görüşlerini, tutumlarını ve düşüncelerini içeren metinleri analiz etmek için kullanılan bir dizi teknikler kümesidir. Bu teknikler, metinlerdeki olumlu, olumsuz ve nötr duyguları tanımlamak, sınıflandırmak ve ölçmek için kullanılır. Müşterilerin marka, ürün, olaylar, şirket ve müşteri hizmetlerine ilişkin davranış ve yaklaşımlarını sınıflandırmaya ve incelemeye yardımcı olan bir süreç olan duygu analizi pazarlama, sosyal medya analizi, müşteri hizmetleri, sağlık hizmetleri, siyaset ve güvenlik birçok alanda kullanılmaktadır (Bird ve diğerleri, 2009).

Duygu analizi, veri madenciliği, web madenciliği ve sosyal medya analitiği gibi alanlarda yaygın olarak kullanılan bir metin madenciliği uygulamasıdır. Duygu analizi, insanların farklı olaylar, sorunlar, hizmetler veya diğer ilgi alanlarına yönelik ruh hallerini algılama, analiz etme ve değerlendirme sürecini ifade eden bir “Duygusal Bilgi İşleme” alanıdır (Yadollahi, Shahraki ve Zaiane, 2017). Duygu kavramı, davranışların değerlendirilmesinde önemli bir özelliği temsil eder. Bu bağlamda, metinler duygu analizinin temel veri kaynaklarını oluşturan duyguları barındırma ve iletişim işlevini yerine getirir (Chakraborty, Bhattacharyya, Bag ve Hassanien, 2019). Sosyal medya verileri gibi mesajlar, haberler ve

paylaşımlar, duygu ve düşüncelerin anlam bilimsel olarak ortaya çıkarılması amacıyla kullanılır. Bu süreçte, metinler duygusal kutuplara ayrılır ve duygusal kategoriler yerine pozitif ve negatif duygular şeklinde gruplandırılır. Bu analiz, metinlerde duyguların nasıl ifade edildiğini ve bu ifadelerde olumlu veya olumsuz durumların nasıl tespit edileceğini sağlar (Nasukawa ve Yi, 2003).

Cümlenin analizi sırasında oluşan sözdizimsel ağa bağlı olan duygu analizi, yalnızca olumlu veya olumsuz anlam ifade eden kelimeler veya kavramlarla sınırlı değildir. Duygu analizi, kullanıcının yazılı metninden duyguları otomatik olarak çıkarmak için yapılandırılmamış bilgileri işleyerek ve bilgi çıkarımı için bir model oluşturarak gerçekleştirilen bir işlemdir (Bird ve diğerleri, 2009). Duygu analizi, kişilerin ürünler, konular, olaylar veya bu başlıkların özellikleri hakkındaki değerlendirmelerini ve tutumlarını analiz eden bir kavramdır (Satapathy, Cambria ve Hussain, 2018). Delen vd. (2014), duygu analizi sürecinin akışını Şekil 4.1'de açıklamışlardır (Delen, Sharda ve Turban, 2014).



Şekil 4.1. Duygu analizi süreç akışı

Duygu Analizi, dijital ortamdaki tüm metin içeriklerinin makine öğrenmesi tabanlı veya istatistiksel yöntemlerle analiz edilerek duyguların sınıflandırılmasını içeren bir işlemdir (Poria, Hussain ve Cambria, 2018). Bu, yazılı metinlerin çeşitli bilgisayar yazılımları aracılığıyla analiz edilmesini ifade eder. Bu analizler, algoritmik veya istatistiksel

çözümlemelerle gerçekleştirilir. Duygu Analizi üzerinde çalışırken doküman seviyesi, cümle seviyesi ve varlık-görüş seviyesi olmak üzere üç farklı teknik seviyeden söz etmek mümkündür.

Genellikle doküman seviyesinde gerçekleştirilen duygu analizinde amaç, tüm metindeki duygunun pozitif veya negatif olduğuna dair bir fikir elde etmektir. Bu yaklaşımın avantajı, uygulanması kolay olmasıdır, ancak dezavantajı sadece genel bir duygu yönüne varılabilmesidir (Pang ve Lee, 2008). Cümle seviyesinde yapılan duygu analizinde, cümlelerin içerik taşıyıp taşımadığı negatif, pozitif veya nötr olarak belirlenir. Nötr ifadeler fikir beyanında dikkate alınmaz. Ancak bu teknikte, ima veya ironi içeren cümlelerin hassasiyetiyle ilgili bazı zorluklarla karşılaşılabilir (Satapathy ve diğerleri, 2018).

Metin yapılarına odaklanmadan sadece pozitif ya da negatif duyguların varlığı ve bir fikrin varlığı üzerinde duran bir analiz seviyesi olan Varlık-Görüş seviyesi, metnin yönünü anlamaya çalışmayı hedefler (Satapathy ve diğerleri, 2018). Metin verileri üzerinde yapılan duygu analizleri, çeşitli amaçlar doğrultusunda kullanılmaktadır. Bu amaçlar arasında görüş ve yorumların özetlenmesi, görüş ve yorumların ortaya çıkarılması, spam ve sahte yorumların tespiti, metin türü ve yazarının tespiti, alaycılık ve ironin tespit edilmesi, öznelliğin sınıflandırılması, duygu ve görüşlerin sınıflandırılması gibi hedefler yer almaktadır (Kaynar, Yıldız, Görmez ve Albayrak, 2016).

4.2.1. Duygu analizi amacı ve önemi

Duygu analizi günümüzde doğal dil işlemenin (NLP) en yaygın kullanım alanlarından biridir. Sosyal medya platformlarındaki kullanıcılar farklı konularda düşüncelerini paylaşırken, izledikleri filmler, okudukları kitaplar veya satın aldıkları ürünler hakkında yorumlar yaparak önemli miktarda veri üretirler. Bu verilerin analizi, uzmanların ve veri bilimcilerin ilgisini çekmektedir. Duygu analizinin amacı, bu verilerin kutupluluğunu ölçmek ve metinlerin duygu polarite değerlerini hesaplamaktır. Dolayısıyla duygu analizi, bir metin sınıflandırma yöntemi olarak kabul edilebilir (Chawai, 2019: 60). Duygu analizinin bir başka tanımı ise metinlerde yer alan öznel bilgileri ortaya çıkarmak veya yazarın belirli bir konuya yönelik duygusal tutumunu belirlemek için kullanılan bir tekniktir (Kumar ve Sebastian, 2012). Wilson, Wiebe, Bruce, Bell ve Martin (2004) ise öznelliği, bireylerin duygu, düşünce, değerlendirme ve inançlarının dil yoluyla ifadesi olarak tanımlamaktadır.

Duygu analizi veya duygu madenciliği, doğal dil işleme (NLP) alanında kullanılan bir alt alan olup metin verilerindeki insan duygularını, tutumlarını ve görüşlerini belirlemek ve analiz etmek amacını taşır. İnsan duyguları, sosyal, ekonomik, pazarlama, müşteri hizmetleri, sağlık hizmetleri, siyaset ve diğer birçok alanda büyük bir öneme sahiptir. Bu nedenle, da bu alanlardaki metin verilerini etkili bir şekilde analiz etmek ve duygu, tutum ve görüşleri belirlemek için önemli bir araç olarak kullanılmaktadır (Kumar ve Sebastian, 2012). Duygu analizi çalışmalarının önemi aşağıdaki gibi sıralanabilir (Değer, 2017):

- Marka yönetimini etkin bir şekilde yapmak için markaya bakış açısı kazandırabilir.
- Manuel olarak zor bir süreç olan analizi otomatik hale getirerek zaman ve çaba tasarrufu sağlar.
- Anlık, günlük, haftalık veya istenilen sıklıkta raporlar oluşturarak, markanın müşterilerini sürekli olarak dinleme imkânı sunar.
- Ürünler ve hizmetlerle ilgili fikir ve görüşleri tespit ederek, krizlere hızlı müdahale edebilme ve maliyetleri azaltma olanağı sağlar.
- Marka ve ürüne ilişkin duygusal durumları tespit ederek, sektörle ilgili anahtar kelimeleri belirleme ve müşteri düşüncelerini sınıflandırma yeteneği sağlar.
- Duygu analizi, şirketlerin kendi markalarının yanı sıra rakip şirketlerle ilgili içerikleri takip etme ve rekabette avantaj elde etme imkânı sunar.

İnternetin yaygınlaşmasıyla birlikte ticaret, yeni kavramlar ve stratejiler ile şekil değiştiriyor. Sosyal medya araçlarının artması, insanların düşüncelerini her yerde rahatlıkla paylaşabilmesine olanak sağlıyor. Tüketiciler, deneyimlerini markalara karşı olumlu ya da olumsuz bir şekilde paylaşarak üreticiler ve hizmet sağlayıcılar üzerinde etkilerini gösteriyorlar. Bu etkiler, firmaların rekabet dünyasında göz ardı edilemez bir öneme sahiptir. Twitter örneğinde, sadece İngilizce değil, Türkçe de dahil olmak üzere diğer dillerde de sosyal medya takibi manuel olarak yapılamayacak kadar hızlı bir şekilde büyüyor. Günde 175 milyon Tweet ve 1 milyon yeni hesap ile Twitter, bu trendi hızla sürdürmektedir (Köktürk ve Dirsehan, 2012: 7).

Firmalar, müşterilerinin düşüncelerini öğrenmek ve kendilerini geliştirmek, ürünlerini değerlendirmek amacıyla sürekli olarak ölçüm yapmaya çalışırlar. Bu ölçümler, anketler veya marka etkinlikleri gibi farklı faaliyetler aracılığıyla gerçekleştirilebilir, ancak temel

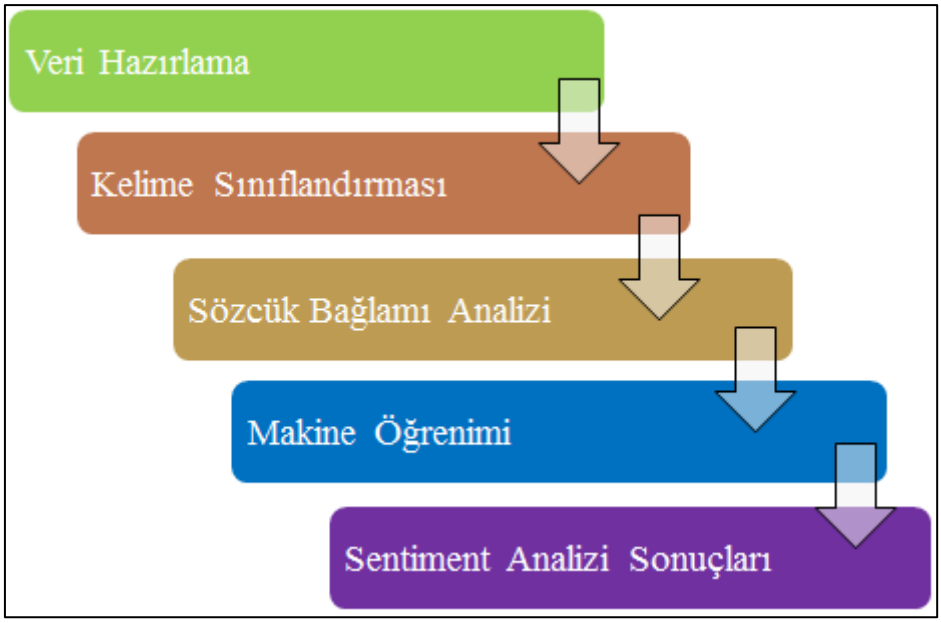
hedef her zaman müşterilere kulak vermektir. Ancak günümüzde, müşterilerin duygu ve düşüncelerini saniyeler içinde paylaştıkları sosyal medya platformları, firmaların tüm bu çabalarının ötesine geçmiştir. Sürekli içerik üreten bu platformlar, firmalar için olumlu veya olumsuz sonuçlar doğurabilir. Olumlu içerikler, potansiyel müşterilerin satın alma kararlarını olumlu yönde etkilerken, olumsuz içerikler potansiyel ve mevcut müşteri kaybına neden olabilir. Sosyal medya içerikleri bu kadar önemli ve bu kadar yaygınken, bunları anlamlı hale getirmek kesinlikle zor bir görevdir. Önceden manuel olarak takip edilen bu içerikler, sosyal medya araçlarının ve kullanıcıların içerik üretim hızının sürekli arttığı bu çağda ciddi bir zorluk haline gelmiştir (Seker, 2016)

Verinin hızlı bir şekilde bilgiye dönüşmesi, firmaların rekabetçi avantajını belirleyen en önemli faktör haline gelmiştir. Bu nedenle, sosyal medya analizi üzerine yapılan her çalışma son derece değerlidir. Sosyal medya paylaşımlarının analiz edilmesi sırasında, içeriklerin duygusal öğeler içerip içermediği ve duygunun yönünün belirlenmesi, duygu analizi çalışmalarının temel amacını oluşturur. Analiz sonuçlarına dayalı olarak, hızlı aksiyon almanın firmaların başarısını artırdığı görülmektedir. Özellikle müşterilerin Twitter gibi sosyal medya platformlarında paylaştıkları olumsuz problemler ve kötü deneyimler sonrasında, firmaların bu müşterilerle iletişime geçerek sorunları çözmesi ve özür dilemesi gibi adımların ardından müşterilerin olumlu içerikler paylaştığı gözlemlenmektedir. Bu tür örneklerden dolayı, firmalar duygu kutuplarına göre hızla müdahale ederek müşterilerin düşüncelerini değiştirebilmektedir (Polat ve Ağca, 2022). Bu nedenlerle, firmalar rekabetçi bir avantaj elde etmek veya rakipleriyle eşit şartlarda mücadele etmek için sürekli olarak kendileri hakkında internet ve sosyal medya üzerinde ne söylendiğini takip etmek zorundadır.

Otomatik duygu analizi, bu amaçla en kolay ve etkili yöntem olarak öne çıkmaktadır. Firmalar, kendileriyle ilgili yapılan olumsuz bir yorumu hızlı bir şekilde olumlu bir hale dönüştürebilmek için bu yöntemi kullanabilirler. Örneğin, bir müşteri uzun süredir çözümlenmeyen bir şikayetini ifade ediyorsa, firma hemen bir çözüm sunarak, müşterilere şikayetlere duyarsız kalmayan bir imaj sergileyebilir. Eğer o şikayetin sosyal medyada paylaşıldığından haberdar olunmamış ve çözülmemiş olsaydı, diğer potansiyel müşteriler de firma hakkında olumsuz etkilenebilir ve satın alma kararlarından firmayı çıkarabilirlerdi. Otomatik duygu analizi modelleri sayesinde, bu gibi durumlar firmaların lehine dönüştürülebilir (Değer, 2017: 11-12).

4.2.2. Duygu analizi kelime sınıflandırma aşamaları

Duygu analizi, metin verilerindeki kelimelerin duygusal anlamlarını belirlemek için kullanılan çeşitli teknik ve yöntemlerin bir dizi adımını içerir. Şekil 4.2'de, duygu analizi kelime sınıflandırmasının aşamaları gösterilmektedir:



Şekil 4.2. Duygu analizi kelime sınıflandırma şeması

Bu aşamalar şu şekilde özetlenebilir (Mohammad ve Turney, 2013; Medhat ve diğerleri, 2014; Gandomi ve Haider, 2015):

1. Veri Hazırlama: Metin verilerinin hazırlanması aşaması, verilerin toplanması, ön işleme ve normalizasyon adımlarını içerir. Bu adımda, özel karakterlerin çıkarılması, küçük harflere dönüştürülmesi ve kelime köklerinin belirlenmesi gibi işlemler gerçekleştirilir.
2. Kelime Sınıflandırması: Metin verilerindeki kelimelerin duygusal anlamlarını belirlemek için çeşitli tekniklerin kullanıldığı bir adımdır. Kelime tabanlı yöntemler, kelimelerin pozitif, negatif veya nötr olarak sınıflandırılmasına dayanır. Bu yöntemde, duygusal kelime listeleri kullanılır ve kelimeler bu listelere göre sınıflandırılır. Örneğin, "iyi" kelimesi pozitif olarak sınıflandırılırken, "kötü" kelimesi negatif olarak sınıflandırılabilir.
3. Sözcük Bağlamı Analizi: Bir kelimenin duygusal anlamını belirlemeden önce, o kelimenin çevresindeki diğer sözcüklerin ve cümlelerin anlamını dikkate alan bir aşamadır. Bu analiz, kelimenin duygusal anlamını belirlemek için önemlidir. Örneğin,

"güneşli gün" ifadesi, kelime tabanlı yöntemle ele alındığında pozitif olarak sınıflandırılabilirken, "yaralı köpek" ifadesi negatif olarak sınıflandırılabilir.

4. Makine Öğrenimi: Kelime sınıflandırma sürecini otomatikleştirmek için kullanılır. Bu yöntemde, bir eğitim veri seti kullanılır ve model, bu veri setindeki örneklerden duygusal anlamların sınıflandırılmasını öğrenir. Daha sonra, model, yeni metin verilerindeki kelimelerin duygusal anlamlarını belirlemek için kullanılır.
5. Sentiment Analizi Sonuçları: Sentiment analizi sonuçları, metin verilerindeki duygu, tutum ve görüşleri belirlemek için kelime sınıflandırma sonuçlarına dayanır. Bu sonuçlar, karar verme süreçlerinde ve stratejik planlama süreçlerinde çeşitli alanlarda kullanılabilir.

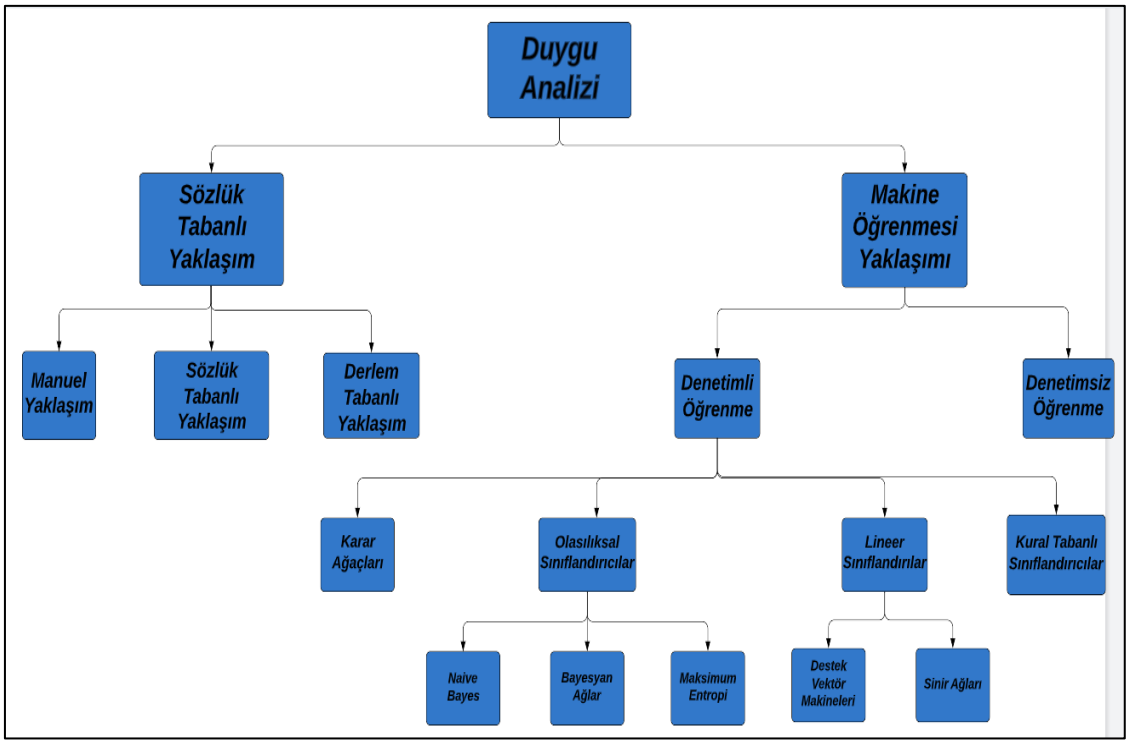
Duygu düzeylerinin belirlenmesi sürecinde, kelimeler ve cümleler için aşağıdaki formül kullanılarak hesaplamalar yapılır (Özkul, 2019):

$$\text{Duygu Değeri} = \frac{\text{Etiketlenmiş Pozitif Kelime} - \text{Etiketlenmiş Negatif Kelime}}{\text{Etiketlenmiş Toplam Kelime Sayısı}} \quad (4.1)$$

Bu hesaplama sonucunda, eğer duygu değeri sıfırdan büyükse (Duygu Değeri > 0), kelime pozitif olarak sınıflandırılır. Eğer duygu değeri sıfıra eşitse (Duygu Değeri = 0), kelime nötr olarak kabul edilir. Ayrıca duygu değeri sıfırdan küçükse (Duygu Değeri < 0), kelime negatif olarak belirlenir (Özkul, 2019).

4.2.3. Duygu analizinde kullanılan yöntemler

Duygu analizi alanında yapılan çalışmalar, temel olarak makine öğrenmesine dayalı yöntemler ve sözlüğe dayalı yöntemler olmak üzere iki kategoriye ayrılabilir (Medhat vd. 2014). Bazı kaynaklarda, bu iki yaklaşımın birlikte kullanıldığı hibrit yöntemler de ele alınmıştır. Uygulama alanına bağlı olarak duygu analizi çalışmalarında bu iki temel sınıf arasında seçim yapılır. Örneğin, makine öğrenmesi algoritmalarının basit yapısı ve istatistiksel yaklaşımların hızlı çalışması, sosyal medya gibi sürekli güncellenen veriler üzerinde gerçek zamanlı analizlerde tercih edilme nedenleri arasındadır (Seker, 2016). Bu nedenle çalışmada makine öğrenmesi yöntemi ele alınmıştır. Duygu analizi yöntemlerinin detaylı gösterimi Şekil 4.3'de gösterilmektedir.



Şekil 4.3. Duygu analizi yöntemleri (Kızılkaya, 2018)

Sözcük tabanlı yaklaşım

Duygu analizinde kullanılan sözlük tabanlı yaklaşımda, duygusal ifadeyi taşıyan kelimeler en önemli öğeler arasında yer alır. Bu kelimeler, pozitif veya negatif duyguları ifade eden tek bir kelime veya kelime grupları şeklinde olabilir. Örneğin, “iyi” veya “harika” gibi pozitif duygusal kelimeler, beğenilen bir durumu ifade etmek için kullanılırken, “kötü” veya “korkunç” gibi negatif duygusal kelimeler, beğenilmeyen bir durumu ifade etmek için kullanılır. Bu duygusal kelimeler ve kelime gruplarından oluşan liste, duygu sözlüğü olarak adlandırılır (Albayrak, 2018).

Duygu kelimelerinin derlenip duygu sözlüğü oluşturulması için kullanılan manuel yaklaşım, sözlük tabanlı yaklaşım ve derlem tabanlı yaklaşım olmak üzere üç temel yaklaşım söz konusudur (Liu, 2012: 90):

1. Manuel Yaklaşım: Bu yaklaşımda, uzmanların ve dilbilimcilerin deneyimlerine dayalı olarak duygu kelimeleri derlenir ve bir duygu sözlüğü oluşturulur. Uzmanlar, kelimelerin duygu ifadesini değerlendirir ve pozitif, negatif veya nötr olarak sınıflandırır. Bu yöntemde, insan becerisi ve bilgi birikimi temel alınır (Taboada, Brooke, Tofiloski, Voll

ve Stede, 2011).

2. Sözlük Tabanlı Yaklaşım: Bu yaklaşımda, önceden oluşturulmuş bir duygu sözlüğü kullanılır. Duygu sözlüğü, pozitif, negatif veya nötr duyguları ifade eden kelimelerin bir listesini içerir. Metin verileri analiz edilirken, bu sözlük kullanılarak kelimelerin duygu sınıflandırması yapılır. Örneğin, pozitif duygu kelimesi listede varsa, kelime pozitif olarak sınıflandırılır (Aksaya, 2021).
3. Derlem Tabanlı Yaklaşım: Bu yaklaşımda, büyük veri kümeleri veya derlem metinler kullanılarak otomatik bir şekilde duygu sözlüğü oluşturulur. Metinlerdeki kelimelerin frekansları, istatistiksel yöntemler veya makine öğrenimi teknikleri kullanılarak değerlendirilir. Örneğin, bir kelimenin sıklığı ve diğer kelimelerle olan ilişkisi, duygusal anlamını belirlemek için kullanılabilir (Elmas, 2019).

Bu üç yaklaşım, duygu analizi için farklı yöntemler sunar ve uygulanacak alana, veri kaynaklarına ve gereksinimlere bağlı olarak tercih edilebilir. Her bir yaklaşımın avantajları ve dezavantajları bulunmaktadır, bu nedenle araştırmacılar ve uygulayıcılar, hangi yöntemin ihtiyaçlarına en uygun olduğunu değerlendirmelidir.

Makine öğrenimi yaklaşımı

Makine öğrenmesi terimi ilk olarak, bilgisayarların deneyimlerden öğrenebileceği ve insan zekasını taklit edebileceği fikrini öne süren ve makine öğrenmesi alanında öncü bir isim olan Arthur L. Samuel tarafından 1959 yılında yayımlanan “Some Studies in Machine Learning Using the Game of Checkers" başlıklı makalede kullanılmıştır (Akyiğit ve Taşçı, 2021).

Makine öğrenmesi, sürekli gelişme gösteren bir hesaplama algoritması dalıdır ve çevredeki ortamdan öğrenerek insan zekasını taklit etmeyi amaçlar (El Naqa ve Murphy, 2015). Bu teknoloji, hesaplama maliyetlerini önemli ölçüde azaltır ve geliştirme döngüsünü kısaltır. Ayrıca makine öğrenmesini karmaşık veriler arasındaki ilişkileri tespit etmesi, işlemleri daha hızlı ve daha az hata payıyla tamamlaması ve daha ekonomik olması nedeniyle tercih edilmektedir. Konuşma tanıma, bilgisayarlı görü, otonom sürüşler, hastalık tanıma, kredi yönetimi, öneri sistemleri ve pazarlama gibi pek çok alanda yaygın olarak kullanılan makine öğrenmesi, teknolojik gelişmelere paralel olarak sürekli olarak gelişmektedir (Wei, Chu, Sun, Xu, Deng, Chen ve Lei, 2019).

Makine öğrenimi yaklaşımı, bilgisayar sistemlerinin verilerden öğrenme ve deneyim kazanma yeteneği üzerinde duran bir disiplindir. Bu yaklaşımda, duygu analizi gibi hedeflere yönelik olarak makine öğrenmesi algoritmaları ve özellik çıkarım yöntemleri kullanılır. Eğitim ve test verileri ayrı ayrı setlerde yer alır. Eğitim verisi, özellikleri ve çıktı değerleri önceden bilinen verilerden oluşur ve bazen "etiketli veri" olarak adlandırılır. Bu veriler, çeşitli sınıflandırma algoritmaları kullanılarak eğitilir. Test verisi ise kategorisi ve özellikleri belirsiz önceden bilinmeyen verileri içerir ve modelleme işlemi bu veri setleri üzerinde gerçekleştirilir. Bu sayede, sınıflandırma işlemi gerçekleştirilir (Amolik, Jivane, Bhandari ve Venkatesan, 2015).

Makine öğrenmesi temelli yöntemler; denetimli öğrenme ve denetimsiz öğrenme olmak üzere ikiye ayrılır. Bu yöntemler çeşitli alanlarda yaygın olarak kullanılan bir dizi teknik ve algoritmadan oluşur. İlgili yöntemler ve bu yöntemlerle ilgili kavramlar aşağıda açıklanmıştır.

Denetimli öğrenme

Denetimli Makine Öğrenmesi, etiketli verilerin eğitim kümesini temel alan bir öğrenme yöntemidir. Bu yaklaşımda, bir sınıflandırıcı, eğitim verileriyle eğitilir ve ardından test verileri üzerinde sınıflandırma performansı değerlendirilir. Denetimli öğrenme, makine öğrenmesi alanında en yaygın kullanılan yöntemlerden biridir. Denetimli öğrenme modellerinde, etiketli girdi verilerinden istenen çıktı verilerini elde etmek ve oluşan çıktıların hedef değerlere ne kadar yakın olduğu önemlidir. Oluşturulan denetimli öğrenme modeli, bir grup girdi değeri için onlara ait hedef değerleri kullanarak aralarındaki ilişkiyi öğrenir ve en yakın çıktıları üretme amacı taşır (Atalay ve Çelik, 2017). Denetimli öğrenme Şekil 4.3'te görüldüğü üzere; karar ağaçları, kural tabanlı sınıflandırıcılar, doğrusal sınıflandırıcılar ve olasılıksal sınıflandırıcılar olmak üzere dörde ayrılmaktadır.

Karar ağaçlarında; sınıflandırma yapmak için eldeki verilerden ağaç yapısı oluşturulur ve veri kümesinde yer alan kayıtlara oluşturulan bu yapı uygulanır. Ağaç yapısı; düğüm, dal ve yapraktan oluşmaktadır. Yapı kök düğümünden başlayarak yaprağa ulaşıncaya kadar devam etmektedir (Mermer, 2019).

Olasılıksal sınıflandırıcılar; olasılığa dayalı geliştirilmiş sınıflandırma tekniğidir, en sık kullanılan algoritmaları Naive Bayes, Bayesyan Ağlar ve Maksimum Entropi'dir.

Lineer sınıflandırıcılar; doğrusal bir fonksiyon tarafından yapılan sınıflandırma tekniğidir. Sinir Ağları ve Destek Vektör Makineleri lineer sınıflandırıcı algoritmalarıdır. Bu algoritmalardan Destek Vektör Makineleri iki farklı sınıf arasında iyi lineer ayrıştırıcılara karar vermede daha iyi sonuçlar vermektedir (Can ve Alataş, 2017).

Kural tabanlı sınıflandırıcılar; belirli bir kurala dayanarak veri setini sınıflandırır. Karar ağaçlarına benzer bir sistematiğe sahiptir fakat karar ağaçlarının farkı hiyerarşik yaklaşımla amacına ulaşmaya çalışmasıdır.

Bu çalışmada, uygulamalarda sıklıkla tercih edilen Naive Bayes, Destek Vektör Makineleri ve Rastgele Orman algoritmalarından yararlanılmıştır.

Naive bayes

Bayes teoreminden faydalanarak, belirli bir sınıfa ait öznitelik değerinin belirli bir etikete ait olma olasılığını tahmin etmek için kullanılan istatistiksel ve olasılık temelli bir sınıflandırma algoritmasıdır (Medhat ve diğerleri, 2014). Bu algoritma, basit yapısı, hesaplama etkinliği ve yüksek doğru sınıflandırma başarısı nedeniyle duygu analizinde yaygın olarak uygulanmaktadır (Onan, 2017). Bayes teoreminin temelinde; bir olayın gerçekleşme olasılığı, o olayın meydana gelme olasılığı ile o olayın oluşması durumunda diğer olayların meydana gelme olasılığı çarpımının toplamına bölünmesiyle hesaplanmaktadır. Hesaplamalar, sınıflar düzeyinde gerçekleştirilip, her bir sınıf için olasılık değeri hesaplanarak en yüksek olasılık değerine sahip sınıf, sınıflandırılması yapılacak olan dokümanın sınıfı olarak alınır (Kızılkaya, 2018). Bayes teoremi şu şekilde ifade edilir:

$$P(A|B) = (P(B|A) * P(A)) / P(B) \quad (4.2)$$

Burada:

P(A| B), B'nin verildiği durumda A'nın gerçekleşme olasılığını temsil eder.
P(B| A), A'nın verildiği durumda B'nin gerçekleşme olasılığını temsil eder.

P(A) ve P(B), sırasıyla A ve B'nin ayrı ayrı gerçekleşme olasılıklarını temsil eder.

Bayes Teoreminden hareketle Navie Bayes sınıflandırmaya geçmek için;

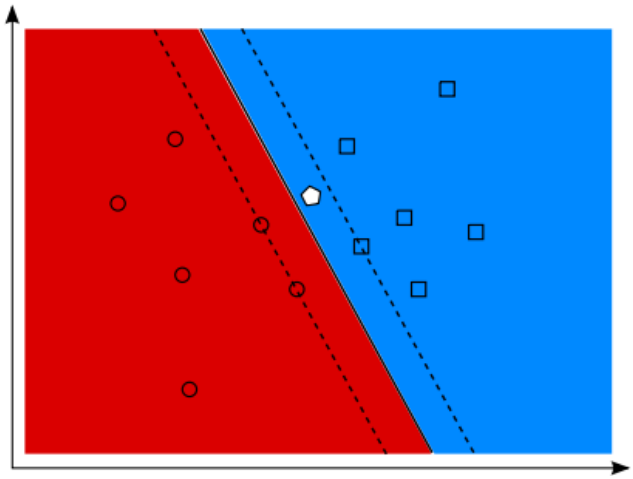
$$P(S_i|X) = \frac{P(X|S_i)\,P(S_i)}{P(X)} \tag{4.3}$$

- X olarak seçilen bir örneğin i duygu sınıfına (pozitif, negatif, nötr) ait olma ihtimali $P(Si\ |X)$,
- i duygu sınıfının henüz veri testine başlamadan öncesinde duygu kütüphanesinde yer alan duygu sınıfına ait ilk olasılığı $P(Si\ )$,
- Seçilen bir örneğin X olma önsel olasılığı $P(X)$,
- i duygu sınıfına ait bir örneğin X olma olasılığı ise $P(X|\ Si\ )$ olarak ifade edilmektedir.

En yüksek olasılığa sahip sınıfın bulunması amacıyla, ele alınan bir örnek için eğitim seti kullanılarak her bir duygu sınıfı için $P(Si\ |X)$ olasılıkları ayrı ayrı hesaplanır. Bu hesaplamalar sonucunda olasılık değeri en büyük hesaplanan duygu sınıfı ele alınan örneğin sınıfı olarak kabul edilir. Bu eşitlikte $P(X)$ tüm olası sınıflar için sabit olduğundan $P(Si\ |X)$ = $P(X|Si\ )\ P(Si\ )$ olasılığı için en büyük değer aranması yeterlidir (Şahinaslan, Ö., Dalyan, H. ve Şahinaslan, E., 2022).

Destek vektör makineleri

İki sınıfa ait verileri birbirinden optimum şekilde ayırma amacına yönelik olarak doğrusal bir hiper düzlemi tahmin etmeyi temel alan sınıflandırma algoritmasıdır. Hiper düzlem, iki sınıfı en iyi ayıran bir sınırdır ve sınıflandırmadaki hatayı en aza indirerek sınıfları birbirinden olabildiğince maksimum uzaklıkta sınıflara ayırmaktadır (Korkusuz, 2019; Can ve Alataş, 2017). Şekil 4.4'te ayırıcı bir hiper düzlem ve destek vektörleri gösterilmektedir.



Şekil 4.4. Destek vektör makineleri algoritması kesikli çizgiler ve düz çizgi (Korkusuz, 2019)

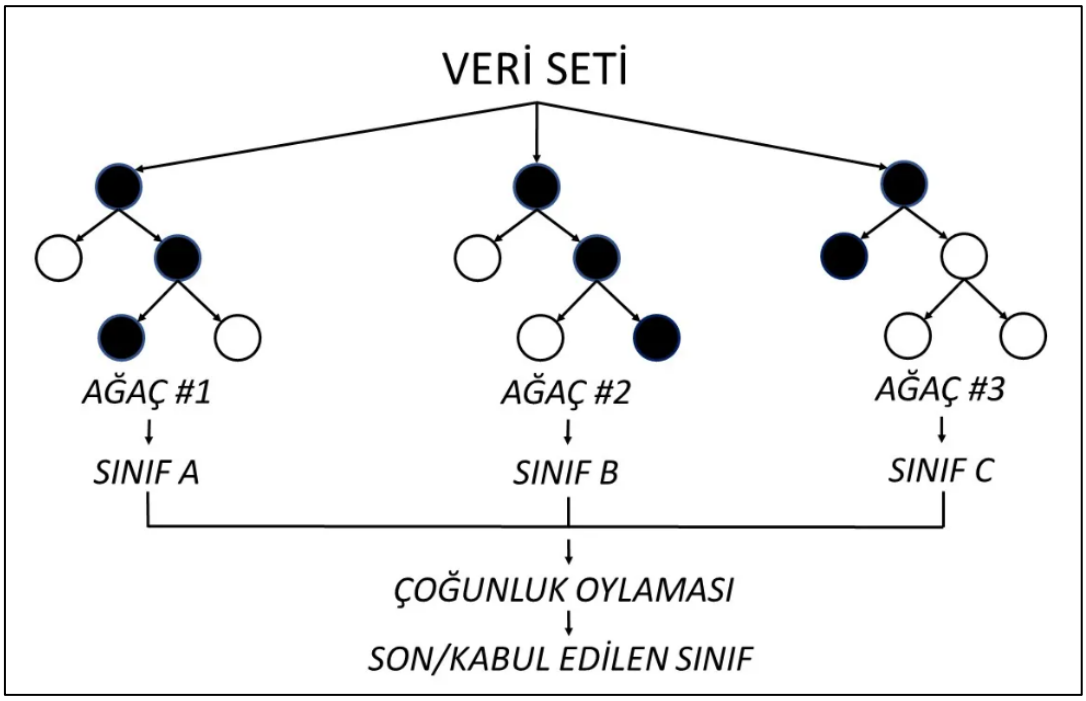
Şekil 4.4'te gösterilen kesik çizgiler iki sınıf arasındaki en büyük sınırı, düz çizgi ise ayırma hiper düzlemini ifade etmektedir. Yani ayırma çizgisinin sağındaki tüm nesnelerin "kare" olduğu ve ayırma çizgisinin solundaki tüm nesnelerin "yuvarlak" olduğu bir sınırı belirtir. Ayırma çizgisinin sağına gelen her yeni nesne kare olarak; soluna gelen her yeni nesne yuvarlak olarak sınıflandırılmaktadır.

Fakat veri setlerinde boyut artıkça (üç boyutlu veya çok boyutlu gibi) işlem yapmak zorlaştığı için kernel fonksiyonları kullanılır (Kaynar ve diğerleri, 2016). Kernel fonksiyonları sayesinde doğrusal olarak ayrıştırılamayan girdi uzayı daha yüksek boyutlu lineer olarak ayrıştırılabilen uzaya taşınarak, doğrusal olmayan verilerin sınıflandırılmasına imkan sağlanmaktadır. Literatürde en yaygın kernel fonksiyonları; doğrusal fonksiyon, polinomiyal fonksiyon, sigmoid fonksiyon, Gauss fonksiyon ve radyal tabanlı fonksiyonlardır (Serdarer ve Giray, 2020).

Destek Vektör Makineleri algoritması, hem doğrusal hem de doğrusal olmayan verilerin sınıflandırılmasında kullanılması, sınıflandırmadaki başarısı. Yüksek güvenilirlik oranı ve ezbere öğrenmeye karşı güçlü olması yönleriyle metin sınıflandırma ve duygu analizi çalışmalarında sıklıkla tercih edilmektedir (Onan, 2017; Karaöz, 2018).

Rastgele orman

Birden çok karar ağacının bir araya gelerek tahmin yapmasına dayalı bir sınıflandırma algoritmasıdır (Alqaraleh, 2020). Algoritmanın çalışması için ilk olarak hem öznitelikler hem de hedef değişkenler içeren veri kümesi seçilir. İkinci aşamada bir dizi karar ağacı eğitilir. Her bir ağaç, rastgele seçilen örneklemlerden ve rastgele seçilen özniteliklerden oluşan alt veri kümesiyle eğitilir. Her bir karar ağacı eğitildikten sonra, tahmin yapmak için tüm ağaçların sonuçlarını birleştirir. Birleştirme işleminde, her bir ağacın tahmini sınıfları oylanır ve en fazla oy alan sınıf genel tahmin olarak kabul edilir (Baid, Gupta ve Chaplot., 2017). Bir Rastgele Orman algoritması Şekil 4.5'teki gibi şematize edilebilir.



Şekil 4.5. Rastgele orman algoritması çalışma prensibi (Miraç Öztürk, 2023)

Karar ağaçları algoritması ile Rastgele Orman algoritması arasındaki temel farklar;

- Rastgele Orman algoritmasında kök düğümü bulma ve düğümleri bölme işlemlerinin rastgele seçiliyor olması,
- Rastgele seçime bağlı olarak overfitting (aşırı öğrenme) sorunun önüne geçiliyor olmasıdır (Kemaloğlu, Küçüksille ve Özgünsür, 2021).

Rastgele Orman algoritması, büyük veri setleri içerisinde yararlı fakat gizli durumdaki bilgileri ayrıştırabilmek için yaygın olarak kullanılan ve başarılı sonuçlar bir yöntemdir (Kumral, Topal, Ersoy, Çolak ve Yiğit, 2022).

Denetimsiz öğrenme

Denetimsiz öğrenme etiketlenmemiş veri setleri üzerinde yapılan bir öğrenme yaklaşımıdır. Bu yöntemde, veri setindeki yapı ve desenlerin otomatik olarak keşfedilmesi amaçlanır. Denetimsiz öğrenmede, veri setinde herhangi bir etiket veya doğru çıktı bilgisi bulunmaz. Model, veri setindeki girdiler arasındaki benzerlikleri, ilişkileri ve desenleri tanımlamak için kullanılır. Model, veri setindeki verileri analiz ederek gruplar, kümeler veya desenler oluşturabilir (Ahlemeyer-Stubbe ve Coleman, 2014: 17).

Denetimsiz öğrenmede yaygın olarak kullanılan teknikler arasında kümeleme (clustering), boyut azaltma (dimensionality reduction) ve dışlayıcı (anomaly detection) gibi yöntemler bulunur. Kümeleme, veri noktalarını benzerlik ölçütleri kullanarak farklı gruplara ayırır. Boyut azaltma, veri setindeki değişken sayısını azaltarak verinin özünü daha kolay anlaşılır hale getirir. Dışlayıcı, anormal veri noktalarını tespit eder ve ayırt edici özelliklerini belirler. Denetimsiz öğrenme, bazı basit işletmelerde karmaşık problemleri çözmek için kullanılsa da genellikle insanların ele alamayacakları problemleri çözmek için tercih edilir. Denetimsiz makine öğrenmesi algoritmalarının bazı örnekleri, k-ortalamalar kümeleme, temel bileşen analizi ve ilişkilendirme kurallarıdır (Kına, 2022).

Derin öğrenme

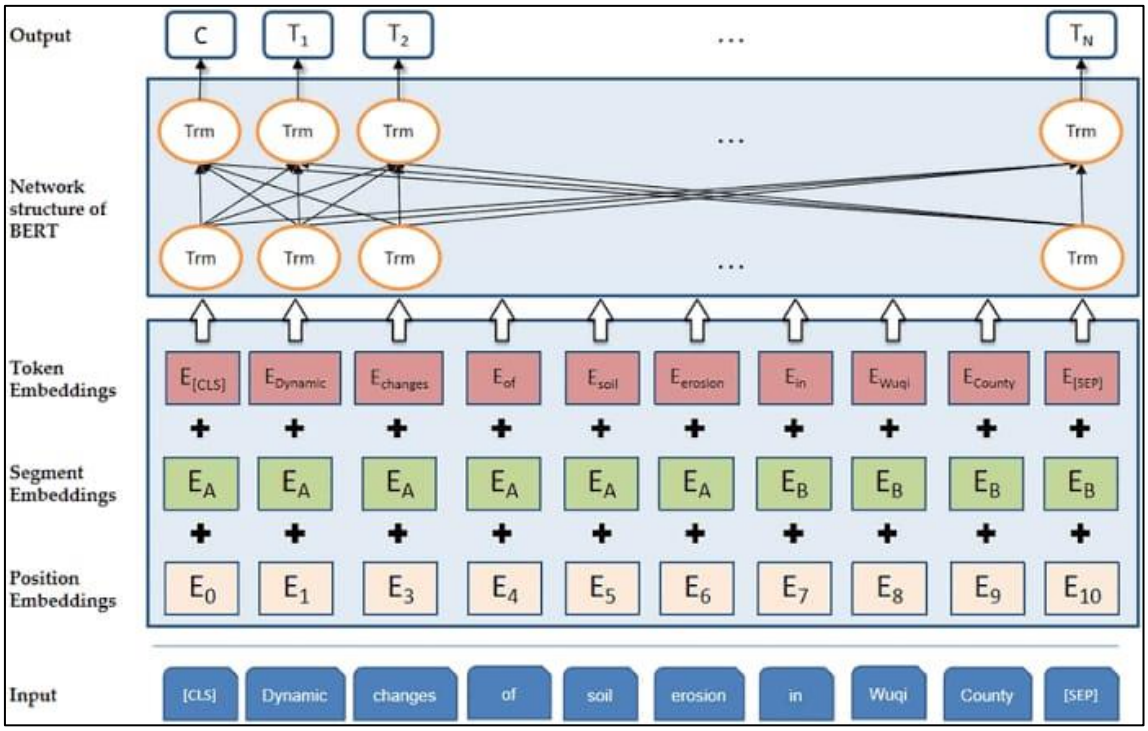
Derin öğrenme, yapay sinir ağlarından oluşan derin nöral ağlar kullanarak karmaşık modellerin öğrenilmesini sağlayan bir makine öğrenmesi alt dalıdır. Geleneksel makine öğrenmesi yöntemlerinden farklı olarak, derin öğrenme, ham veri üzerinde doğrudan öğrenme yapabilir ve veri içerisindeki karmaşık ilişkileri otomatik olarak çıkarabilir (Wani, Bhat, Afzal ve Khan, 2020). Derin öğrenme modelleri, çok katmanlı yapılara sahip olan derin nöral ağlardan oluşur. Bu katmanlar, veri işleme ve özellik çıkarımı için birbirine bağlı bir dizi yapay sinir hücresi veya nöron içerir. Her katman, önceki katmanın çıktılarını girdi olarak alır ve daha karmaşık özelliklerin temsil edildiği yeni çıktılar üretir.

Bu şekilde, veri üzerinde hiyerarşik bir temsil oluşturulur ve derin öğrenme modeli verinin içerdiği özelliklerin daha derin ve anlamlı düzeylerini keşfedebilir (Doğan ve Türkoğlu, 2019).

Derin öğrenme; görüntü tanıma, doğal dil işleme, konuşma tanıma, otomatik sürüş, robotik ve birçok diğer alanda başarılı sonuçlar elde etmiştir. Büyük veri setleri ve güçlü hesaplama kaynakları sayesinde derin öğrenme modelleri daha da geliştirilebilir ve karmaşık problemleri çözebilir hale gelmiştir. Ancak derin öğrenme modellerinin eğitimi, genellikle daha fazla veriye ve hesaplama gücüne ihtiyaç duyduğundan, uygun kaynaklara ve altyapılara sahip olmayı gerektirebilir (İnik ve Ülker, 2017). Çalışmada son yıllarda popüler olan ve sıkça kullanılan BERT (Bidirectional Encoder Representations from Transformers) derin öğrenme yönteminden yararlanılmıştır.

BERT

Ekim 2018'de Google tarafından geliştirilen, büyük miktarda metin üzerinde eğitilmiş, yapay zeka ve makine öğrenmesi yöntemlerini birlikte uygulayan doğal dilleri işleyebilmek için kullanılan transformer tabanlı bir derin öğrenme modelidir (Yılmaz, 2022). BERT, bir giriş dizisini belirteç (token) olarak işlemeden tüm diziyi tek seferde girdi olarak alan ve modeli önceden eğitmek için etiketlenmiş veriye ihtiyaç duymayan transformer modelini kullanır. Şekil 4.6'de BERT modelinin genel yapısı yer almaktadır.



Şekil 4.6. BERT modeli genel yapısı (Liu, Sun, Cui ve He, 2022)

Şekil 4.6'da yer alan en alt katmanda (input), bir örnek cümle girdi olarak yer almaktadır. Bu örnekte, [CLS], Dynamic, changes, of, soil, erosion, in, Wugi, County, [SEP] şeklinde bir giriş verisi kullanılmıştır. [CLS] ve [SEP] özel tokenları, cümlenin başlangıcını ve bitişini belirtmek için kullanılır. Position Embeddings katmanında, her bir belirtece giriş sırasını temsil eden pozisyon bilgileri eklenir. E0'dan E10'a kadar olan pozisyon gömülerini temsil eden bu bilgiler, her belirtecin cümle içindeki yerini belirlemek için kullanılır. Segment Embeddings olarak adlandırılan gömme katmanı, birden fazla cümlenin bir araya getirildiği durumlarda farklı cümlelerin ayrımını yapmak için kullanılır. Eğer tek bir cümle varsa, tüm belirteçler aynı segmente atanır. Fakat cümlelerin uzun olması gibi durumlarda istenilirse bir cümle farklı segmentlere gömülebilir. Token Embeddings katmanında, her bir belirtece karşılık gelen gömme vektörleri bulunur. Örneğin, [CLS] tokenına karşılık gelen gömme vektörü E[CLS], Dynamic tokenına karşılık gelen gömme vektörü E[Dynamic] şeklinde ifade edilir. Token gömme vektörleri, kelimenin anlamını ve temsilini yakalamak için kullanılır. Gömme katmanlarından sonra, gömülü belirteç vektörleri, BERT modelinin geri kalan aşamalarıyla birlikte ağ yapısında işlenir. Bu aşamalar, genellikle birçok dikkat mekanizması (attention mechanism) ve katman içi bağlantıları içeren bir Transformer mimarisi ile temsil edilir. Bu şekilde her bir belirteç, çevresindeki diğer belirteçlere bağlı olarak işlenir ve kapsamlı bir dil modeli oluşturulur. En üstteki çıktı katmanı, BERT

modelinin sonucunu temsil eder. Bu çıktı, genellikle bir dizi vektör veya olasılık dağılımı şeklinde ifade edilir ve belirli bir görevi gerçekleştirmek için kullanılır. Örneğin, sınıflandırma görevleri için çıktı, farklı sınıfların olasılıklarını içerebilir (Devlin, Chang, Lee ve Toutanova, 2019; Vaswani ve diğerleri, 2017).

Transformer modeli, girdileri dönüştürücüler olarak adlandıran bir dizi katmana aktararak çalışır. Her bir dönüştürücü, dikkat mekanizması ve ileri besleme ağı olmak üzere iki katmandan oluşur. Dikkat mekanizması; girdileri etkileşime sokarak, belirli bir kelimeyi anlamlandırmak için diğer kelimelerin önemini belirler. Benzer anlama sahip kelimelerin ilişkilendirilmesini sağlar ve daha sonra tüm verilerin anlamsal bir temsillemesi elde edilir. Bir kelimenin önceki bir kelime ile sıkı bir ilişkisi veya korelasyonu varsa o kelimeyi daha yüksek bir ağırlıkla işleyerek, modelin kelime sırasını anlamasında yardımcı olur. Dikkat mekanizması, kelime sırasındaki her bir kelimenin bir vektör temsili olarak ifade edildiği "kelime gömülmesi" (embedding) katmanından başlar. Bu vektörler, kelimenin özelliklerini ifade eder ve modelin kelime sırasındaki ilişkileri anlamasına yardımcı olur. Daha sonra, her bir kelime için bir “sorgu” (query), “anahtar” (key) ve “değer” (value) vektörü oluşturulur. Bu vektörler, dikkat mekanizmasında kullanılan matematiksel işlemlerde kullanılır. Sorgu vektörü, diğer kelime vektörlerine karşı "sorgu" yapar ve modelin hangi kelimelere dikkat etmesi gerektiğini belirler. Anahtar ve değer vektörleri ise, sorgu vektörüyle karşılaştırılarak her kelime için bir "dikkat ağırlığı" belirlemek için kullanılır. Son olarak; her kelime vektörü, diğer kelime vektörlerinin dikkat ağırlıklarına göre birleştirilerek "dikkat ağırlıklı" bir vektör oluşturulur. Bu, modelin kelime sırasındaki ilişkileri daha iyi anlamasına yardımcı olur ve daha doğru tahminler yapmasına olanak sağlar.

İleri-besleme ağının amacı, her bir kelime veya girdi öğesini daha yüksek boyutlu bir uzayda temsil etmektir. Bu, kelime sırası ve anlamlarının daha iyi anlaşılmasına yardımcı olur. Başka bir deyişle, bir cümlenin kendisinden sonra gelen cümle ile ilişkisine bakılarak cümlelerin birbirleri ile ilişkisini incelemektedir. Örneğin, bir kelimenin diğer kelime veya kelime öbekleriyle olan ilişkileri bu temsiller aracılığıyla anlaşılabilir (Ozan, Özdil, Taşar, Arslan ve Polat, 2021).

BERT, bir metni hem sağdan sola hem de soldan sağa işleyen “İki Yönlü Dil İşlem” özelliğini kullanarak metin içerisindeki öğeler arasındaki ilişkileri öğrenebilmektedir (Santiago ve Campos, 2022). Bu özelliklerinden dolayı diğer geleneksel dil işleme

modellerine göre daha hızlı çalışır ve daha iyi sonuçlar verir. Bu nedenle son yıllarda birçok doğal dil işleme (NLP) uygulamasında popüler bir seçenek haline gelmiştir.

Aynı zamanda Türkçe'nin morfolojik yapısının zengin ama karmaşık olması, sondan eklemeli bir dil olması ve dil bilgisinden kaynaklanan zorluklar gibi metin sınıflama problemlerinin çözümünde sorun oluşturan etmenler BERT derin öğrenme modeli ile daha kolay çözülebilir hale gelmiştir (Özkan ve Kar, 2022).

Sınıflandırma algoritmalarının karşılaştırılmalarında kullanılan kriterler

Sınıflandırma algoritmalarının karşılaştırılması, doğru bir performans değerlendirmesi yapabilmek ve en uygun yöntemi seçebilmek için önemlidir. Bu karşılaştırmalar genellikle çeşitli kriterler üzerinden yapılır (Karakoyun ve Hacibeyoğlu, 2014). Sınıflandırma yöntemleri karşılaştırılırken, sıklıkla kullanılan bir yöntem karmaşıklık matrisi (confusion matrix) üzerinden çeşitli kriterler hesaplanarak performans değerlendirmesi yapılır. Karmaşıklık matrisi, gerçek sınıflar ile tahmin edilen sınıflar arasındaki ilişkiyi gösteren bir tablodur (Öncel ve Erduğan, 2015).

Çizelge 4.1. Karmaşıklık matrisi (Kızılkaya, 2018)

Gerçek Değerler / Tahmini Değerler	Pozitif	Negatif
Pozitif	Doğru Pozitif (TP)	Yanlış Pozitif (FP)
Negatif	Yanlış Pozitif (FP)	Doğru Pozitif (TP)

Karımaşıklık matrisinde yer alan değerler ve anlamları aşağıda listelenmektedir;

- True Positive (TP); pozitif olan aynı zamanda sınıflandırıcı tarafından da pozitif olarak sınıflandırılmış verileri temsil etmektedir.
- True Negative (TN); negatif olan aynı zamanda sınıflandırıcı tarafından da negatif olarak sınıflandırılmış verileri temsil etmektedir.
- False Negative (FN); gerçekte pozitif sınıfa ait olan bir verinin sınıflandırma sonucunda negatif sınıf olarak etiketlenmesidir.

- False Positive (FP); gerçekte negatif sınıfa ait olan bir verinin sınıflandırma sonucunda pozitif sınıf olarak etiketlenmesidir (Kaynar ve diğerleri, 2016; Alan ve Karabatak, 2020).

Karmaşıklık matrisi üzerinden hesaplanan kriterler şunlardır:

Kesinlik (Precision – P): Kesinlik (Precision), sınıflandırma algoritmalarının performansını değerlendirmek için kullanılan önemli bir kriterdir. Kesinlik, bir sınıflandırıcının bir sınıf için ne kadar doğru çalıştığını, yanlış pozitif sonuçların ne kadar az olduğunu ölçer. Özellikle dengesiz veri kümeleriyle çalışırken önemli bir ölçüttür. Kesinlik, aşağıdaki formülle hesaplanır:

$$Kesinlik = TP / (TP + FP) \quad (4.4)$$

Yüksek bir kesinlik değeri, sınıflandırıcının pozitif olarak etiketlenen örnekleri doğru bir şekilde ayırt etme yeteneğini gösterir. Düşük bir kesinlik değeri ise sınıflandırıcının yanlış pozitif sonuçlar ürettiğini veya pozitif olarak etiketlenen örnekleri doğru bir şekilde ayırt etmede zayıf olduğunu gösterir.

Duyarlılık (Recall – R): Duyarlılık (Recall), sınıflandırma algoritmalarının performansını değerlendirmede kullanılan bir kriterdir. Duyarlılık, bir sınıflandırıcının bir sınıfı ne kadar iyi tespit ettiğini ölçer. Yani gerçek pozitif sonuçların ne kadarını doğru bir şekilde sınıflandırabildiğini gösterir. Duyarlılık, aşağıdaki formülle hesaplanır:

$$Duyarlılık = TP / (TP + FN) \quad (4.5)$$

Duyarlılık, özellikle yanlış negatif sonuçların yüksek maliyetli olduğu durumlarda önemlidir. Örneğin, bir kanser teşhisinde yanlış negatif sonuçlar ciddi sonuçlara yol açabilir. Bu nedenle, duyarlılık hassasiyet gerektiren uygulamalarda kritik bir performans ölçütüdür.

Doğruluk (Accuracy – A): Doğruluk (Accuracy), sınıflandırma algoritmalarının performansını değerlendirmede kullanılan bir kriterdir. Doğruluk, bir sınıflandırıcının tüm örnekleri ne kadar doğru bir şekilde sınıflandırdığını ölçer. Doğruluk, aşağıdaki formülle hesaplanır:

$$Doğruluk = (TP + TN) / (TP + TN + FP + FN) \quad (4.6)$$

Yüksek bir doğruluk değeri, sınıflandırıcının genel olarak örnekleri doğru bir şekilde sınıflandırma yeteneğini gösterir. Ancak doğruluk, dengesiz veri kümeleri gibi durumlarda yanıltıcı olabilir. Örneğin, bir veri kümesinde pozitif örneklerin sayısı negatif örneklerden çok daha fazlaysa ve sınıflandırıcı tüm örnekleri pozitif olarak sınıflandırırsa. Yüksek bir doğruluk değeri elde edilebilir. Ancak bu durumda sınıflandırıcının gerçek performansını değerlendirmek için diğer kriterlere de bakmak önemlidir.

F-1 Ölçümü (F-1 Measure): Kesinlik (precision) ve duyarlılık (recall) değerlerinin harmonik ortalamasını göstermektedir.

$$F= (2* Duyarlılık* Kesinlik) / (Duyarlılık + Kesinlik) \quad (4.7)$$

ROC Eğrisi Altındaki Alan (AUROC): ROC (Receiver Operating Characteristics) eğrisi doğruluk ve duyarlılık metrikleri arasındaki dengeyi korumak için kullanılır. Basitçe doğru pozitiflerin yanlış pozitiflere olan kesri olarak ifade edilebilir (Düven, 2021).

ROC eğrisi altındaki alan AUROC (Area Under Receiver Operating Characteristics) büyüklüğü arttıkça, modellerin sınıfları ayırt etme yeteneği artar (Çelik, Dal ve Aydın, 2021).

Kappa İstatistiği: Kappa istatistiği, sınıflandırma algoritmalarının karşılaştırılmasında kullanılan bir başka kriterdir. Kappa istatistiği, sınıflandırma sonuçlarının rastgelelikten ne ölçüde daha iyi olduğunu değerlendirmek için kullanılır. Kappa istatistiği, sınıflandırma sonuçlarını gözlem ve tahmin olmak üzere iki farklı şekilde değerlendirir. Gözlem değerleri, gerçek etiketlenmiş verilere dayanırken, tahmin değerleri ise sınıflandırma algoritmasının tahminlerine dayanır. Kappa istatistiği, bu iki değeri karşılaştırarak anlaşmanın, yani gerçek ve tahmin edilen sınıflar arasındaki uyumun ölçüsünü sağlar (Kılıç, 2015). Kappa istatistiği aşağıdaki formülle hesaplanır:

$$k = \frac{Po-Pe}{1-Pe} \quad (4.8)$$

Kappa istatistiği formülünde, Po (gözlenen uyum oranı) gerçek gözlemciler arasındaki uyumu, Pe (tesadüfi olarak ortaya çıkan uyum oranı) ise rastgelelikten kaynaklanan uyumu temsil eder. k değeri, -1 ile +1 arasında değer alabilir. k değeri, 0'dan küçük olduğunda gözlemciler arasında bir uyum olmadığını gösterir. 0'dan büyük ancak +1'den çok küçük olan değerler, uyumun tesadüfi olduğunu düşündürebilir. k değeri +1'e yaklaştıkça, iki gözlemci arasındaki uyumun tesadüfi olarak oluşamayacağı ve mükemmel bir uyum olduğu, yani her iki gözlemcinin de homojen olduğu söylenebilir.

Model performans değerlendirme yöntemleri

Model performans değerlendirme yöntemleri, makine öğrenimi ve istatistiksel analiz gibi alanlarda kullanılan yöntemlerdir. Bu yöntemler, bir modelin ne kadar iyi çalıştığını, tahminlerinin ne kadar doğru olduğunu veya sınıflandırma işleminin ne kadar başarılı olduğunu ölçmek için kullanılır (Balaban ve Kartal, 2015). Modelin seçildiği veri setiyle sınırlı olmaksızın gelecekte ne kadar iyi çalışabileceğini tahmin etmek önemlidir. Bu nedenle, veri madenciliği alanında çeşitli model performans değerlendirme yöntemleri geliştirilmiştir.

Eğitim verileriyle sınırlı kalmayıp, genelleştirilebilirlik için çeşitli değerlendirme teknikleri kullanmak veri madenciliğinde kabul edilebilir bir yaklaşımdır (Kızılkaya, 2018).

Yaygın olarak kullanılan model performans değerlendirme yöntemleri şunlardır:

Holdout (Dışarıda Tutma): Holdout, model performansını değerlendirmek için kullanılan temel bir yöntemdir. Bu yöntemde, mevcut veri seti rastgele ikiye bölünür: bir kısım eğitim verisi olarak kullanılırken diğer kısım test verisi olarak kullanılır. Holdout yönteminin temel amacı, modelin eğitim verilerine nasıl uyum sağladığını değil, gerçek dünya verilerinde ne kadar iyi performans göstereceğini tahmin etmektir. Eğitim verileri, modelin parametrelerini öğrenmek ve optimize etmek için kullanılırken, test verileri, modelin tahmin yeteneğini objektif bir şekilde değerlendirmek için kullanılır (Boz, Canbazoğlu, Özen ve Gülseçen, 2018). Holdout yöntemi, veri setinin büyüklüğüne bağlı olarak farklı oranlarda uygulanabilir. Genellikle, veri setinin %70-%80'i eğitim verisi olarak kullanılırken, geriye kalan %20-%30'u test verisi olarak ayırılır. Ancak veri setinin boyutu ve özel gereksinimler göz önünde bulundurularak oranlar ayarlanabilir (Kartal, 2015).

Holdout yöntemi, basit ve hızlı bir değerlendirme yöntemi olmasının yanı sıra, modelin genelleştirme yeteneğini ölçmek için kullanışlıdır. Ancak veri setinin rastgele bölünmesiyle ilgili dikkat edilmesi gereken bazı noktalar vardır. Örneğin, veri setinde sınıf dengesizliği varsa, iyi bir temsil sağlamak için stratejik örnekleme veya tabakalı bölünme yapılabilir. Holdout yöntemi, model performansının bir kez ölçülmesini sağlar. Daha güvenilir sonuçlar elde etmek için, bu yöntemi tekrarlı olarak uygulayarak ve sonuçları ortalayarak performans tahminlerini iyileştirmek mümkündür (Oxford ve Daniel, 2001).

Tekrarlı Holdout: Tekrarlı Holdout, model performansını daha güvenilir bir şekilde değerlendirmek için kullanılan bir yöntemdir. Holdout yönteminin genelleştirilmiş bir versiyonudur ve birden fazla kez tekrarlanarak sonuçların ortalaması alınır. Tekrarlı Holdout yöntemi, veri setini rastgele ikiye böler ve her bir bölümü eğitim ve test verisi olarak kullanır. Ancak Holdout'tan farklı olarak, bu bölme işlemi tek seferle sınırlı değildir. Yani veri seti birden fazla kez rastgele bölünür ve her seferinde modelin performansı değerlendirilir (Akman, 2010).

Her bir bölme işlemi sırasında, model eğitim verileriyle eğitilir ve test verileriyle performansı değerlendirilir. Bu şekilde, birden fazla performans ölçümü elde edilir. Ardından, elde edilen performans ölçümleri ortalama alınarak modelin genel performansı tahmin edilir. Tekrarlı Holdout yöntemi, performans değerlendirmesinin daha güvenilir ve istatistiksel olarak anlamlı olmasını sağlar. Tek bir Holdout bölmesi yerine, birden fazla bölme işlemi gerçekleştirilerek, veri setinin farklı kombinasyonlarının kullanılması ve performansın ortalaması alınmasıyla daha sağlam sonuçlar elde edilir (Kızılkaya, 2018).

Tabakalı Örnekleme: Tabakalı Örnekleme, model performansının değerlendirilmesi için kullanılan bir yöntemdir. Bu yöntemde, veri seti farklı sınıflara veya özelliklere göre tabakalara ayrılır ve her tabaka eğitim ve test verisi olarak kullanılır. Tabakalı Örnekleme yöntemi, her sınıf veya özellik grubunun temsil edildiği daha dengeli bir performans değerlendirmesi sağlar. Veri setindeki her tabaka, örneklerin sınıf dağılımını yansıtacak şekilde oluşturulur. Bu, modelin tüm sınıfları eşit şekilde öğrenmesini ve performansını değerlendirmesini sağlar (Karagölge ve Peker, 2011). Örneğin, bir sınıflandırma probleminde veri setinde A, B ve C sınıfları bulunuyorsa, tabakalı örnekleme yöntemiyle veri seti her bir sınıf için ayrı ayrı tabakalara bölünür. Her tabakada bulunan örnekler, eğitim ve test verisi olarak kullanılır. Bu şekilde, her sınıfın doğru oranda temsil edildiği bir

performans değerlendirmesi gerçekleştirilir. Tabakalı Örnekleme yöntemi, dengesiz veri setlerinde özellikle etkilidir. Eğer veri setinde sınıflar arasında büyük bir dengesizlik varsa, yani bazı sınıflar diğerlerine göre çok daha fazla örneğe sahipse, tabakalı örnekleme yöntemi sınıflar arasındaki dengesizliği gidermek için kullanılabilir (Yener ve Abdulkadir, 2007).

Üçlü Ayırma: Üçlü Ayırma (Triple Split) yönteminde, veri seti eğitim verisi, doğrulama verisi ve test verisi olmak üzere üç ayrı parçaya bölünür. Üçlü Ayırma yönteminin amacı, modelin hem eğitim verilerine uyum sağlamasını hem de gerçek dünya verilerinde iyi performans göstermesini sağlamaktır. Veri setinin büyüklüğüne bağlı olarak, genellikle veri setinin %60-%70'i eğitim verisi olarak ayrılırken, geriye kalan kısmı eşit olarak doğrulama ve test verisi olarak ayrılır (Alan, 2020).

Eğitim verisi, modelin parametrelerini öğrenmek ve optimize etmek için kullanılır. Doğrulama verisi, modelin farklı hiperparametre ayarlamaları ve model seçimi için kullanılır. Modelin performansı doğrulama verisi üzerinde değerlendirilir ve hiperparametreler ayarlanır. Test verisi ise modelin nihai performansını objektif bir şekilde değerlendirmek için kullanılır. Test verisi, modelin daha önce görmediği verilere karşı nasıl performans göstereceğini tahmin etmek için kullanılır (Kartal, 2015).

K- Kat Çapraz Doğrulama: K-Kat Çapraz Doğrulama (K-fold Cross-Validation) yönteminde veri seti K eşit parçaya bölünür ve her bir parça sırayla test verisi olarak kullanılırken, geri kalanları eğitim verisi olarak kullanılır. Bu yöntem, veri setinin eğitim ve test verisi olarak ayrılmasını daha adil ve güvenilir bir şekilde yapmayı sağlar (Özer, 2020). Veri setinin farklı parçaları üzerinde K kez eğitim ve test işlemi yapılır ve sonuçlar ortalama alınarak genel bir performans tahmini elde edilir. modelin genelleştirme yeteneğini daha iyi değerlendirmek için kullanılan bu yöntem, veri setinin farklı parçaları üzerinde yapılan testler, modelin farklı verilere karşı nasıl performans gösterdiğini ortaya koyar (Fushiki, 2011).

K-Kat Çapraz Doğrulama yöntemi, veri setinin tamamının eğitim ve test verisi olarak kullanıldığı Holdout yöntemine göre daha verimli bir kullanım sağlar. K değeri, K-Kat Çapraz Doğrulama yönteminde belirleyici olan bir parametredir. Genellikle 5 veya 10 gibi yaygın değerler kullanılır, ancak veri setinin büyüklüğüne bağlı olarak değişebilir. Daha yüksek K değerleri daha fazla hesaplama gücü ve zaman gerektirebilir, ancak daha güvenilir bir performans tahmini sağlar (Kızılkaya, 2018).

Multiple Random Sampling (Çoklu Rassal Örnekleme): Multiple Random Sampling (Çoklu Rassal Örnekleme) yönteminde, veri seti rastgele örnekleme yöntemiyle birden fazla kez bölmeye tabi tutulur ve her bir bölme için ayrı eğitim ve test verisi oluşturulur. Çoklu Rassal Örnekleme yöntemi, veri setini birden fazla kez rastgele böler ve her bölme için ayrı bir eğitim ve test seti oluşturur. Bu yöntemde, her bölme bağımsız olarak bir modelin eğitimini ve testini gerçekleştirir. Ardından, elde edilen sonuçlar ortalama alınarak genel bir performans değerlendirmesi elde edilir (Tang, Han, Lin, Cui, Jia, Low ve Chen, 2007). Bu yöntem, veri setinin daha fazla varyasyonla değerlendirilmesini sağlar. Her bir rassal örnekleme, veri setindeki farklı veri kombinasyonlarını kullanarak modelin performansını test etme şansı verir. Bu sayede, modelin farklı veri dağılımları ve varyasyonlarına nasıl tepki verdiği daha iyi anlaşılabilir. Bu yöntemin dezavantajı ise diğer yöntemlere göre daha fazla hesaplama gücüne ve zaman kaynağına ihtiyaç duymasıdır. Her bir rassal örnekleme için ayrı bir eğitim ve test işlemi yapılması gerektiğinden, veri setinin büyüklüğüne bağlı olarak zaman ve kaynak tüketebilir (Kızılkaya, 2018).

Bootstrap Örnekleme: Bootstrap örnekleme yönteminde, veri setinden rastgele örneklemler çekilerek birden fazla örnek veri seti oluşturulur ve her bir örnek üzerinde ayrı eğitim ve test işlemleri gerçekleştirilir. Bootstrap örnekleme yöntemi, veri setindeki örneklerin tekrarlı olarak çekilmesiyle yeni veri setleri oluşturur. Bu sayede, veri setindeki gözlem noktalarının birden fazla kez kullanılması sağlanır. Her bir oluşturulan veri seti üzerinde ayrı bir model eğitilir ve performansı değerlendirilir. Bu yöntem, veri setinin daha etkin bir şekilde kullanılmasını sağlar ve modelin genelleştirme yeteneğinin daha iyi tahmin edilmesine yardımcı olur. Bootstrap örnekleme, veri setinin farklı örneklemeleri üzerinde yapılan deneyler sayesinde modelin çeşitli veri dağılımlarına ve varyasyonlara nasıl tepki verdiğini değerlendirmeye imkan tanır (Özdemir, 2011).

Bootstrap örnekleme yönteminin avantajı, veri setinin boyutuna bakılmaksızın uygulanabilir olmasıdır. Ayrıca veri setindeki her bir gözlem noktasının birden fazla örneklemede yer alması, daha güvenilir bir performans değerlendirmesi sağlar. Ancak bootstrap örnekleme yöntemi bazen veri setindeki çeşitliliği azaltabilir ve aşırı uyum (overfitting) riskini artırabilir. Bu nedenle, bootstrap örnekleme yöntemiyle elde edilen sonuçların dikkatli bir şekilde yorumlanması ve diğer değerlendirme yöntemleriyle karşılaştırılması önemlidir (Bala, 2013).

5. LİTERATÜR ARAŞTIRMASI

Bu bölümde, çalışmanın konusunu oluşturan duygu analizi yönteminin konaklama işletmeleri ve glamping turizm işletmeleri özelinde literatürdeki gelişmeleri iki ayrı başlık olarak incelenmiştir. İnceleme; "Google Scholar", "Ulusal Akademik Ağ ve Bilgi Merkezi (ULAKBİM)", "Yükseköğretim Kurulu Ulusal Tez Merkezi" ve "ScienceDirect" üzerinden 2013-2023 yılları arasında "sentiment analysis, duygu analizi, glamping turizm, makine öğrenmesi, BERT, otel yorumları analizi" anahtar kelimeleri kullanılarak gerçekleştirilmiştir.

5.1. Konaklama Sektörüne Yönelik Çalışmalar

Konaklama işletmeleriyle ilgili çevrimiçi platformlarda yer alan müşteri değerlendirmelerine yönelik yapılan çalışmalar incelendiğinde literatürde konuyu farklı yönlerden inceleyen çok sayıda çalışma olduğu görülmektedir.

Wang ve diğerleri (2010), tarafından yapılan çalışmada, TripAdvisor.com sitesinde yer alan otellere ilişkin 235.793 müşteri yorumları ve derecelendirmeleri incelenmiştir. Kullandıkları LRR (Gizli Derecelendirme Regresyonu) modelinin en iyi sonucu (%89,60) bulduğunu ifade etmişlerdir (Wang, Lu ve Zhai, 2010).

Katarzyna ve Tuchowski (2013), booking.com sitesinde yer alan Krokaw otelleri ile ilgili değerlendirmeleri Doğrusal Kelime Tabanlı, K-En Yakın Komşu ve Naive Bayes algoritmalarını kullanarak incelemişlerdir. Doğrusal Kelime Tabanlı algoritmanın en iyi sonucu; K-En Yakın Komşu algoritmasının en kötü sonucu verdiğini ifade etmişlerdir (Katarzyna ve Tuchowski, 2013).

Raut ve Londhe (2014), TripAdvisor.com sitesinde yer alan 1000 olumlu ve 1000 olumsuz otel yorumuna ilişkin duygu analizi gerçekleştirmişlerdir. Çalışmada makine öğrenmesi yöntemlerinden Naive Bayes, Destek Vektör Makineleri, Karar Ağaçları algoritmaları ile SentiWordNet yönteminin performanslarını karşılaştırmışlardır. Naive Bayes algoritmasının, %87 doğruluk değeriyle en başarılı algoritma olduğu sonucuna ulaşılmıştır (Raut ve Londhe, 2014).

Doğan (2017) çalışmasında, Aksaray ilinde yer alan 2-3-4 ve 5 yıldıza sahip otellerle ilgili booking.com internet sitesinde yer alan yorum ve puanlamaları incelemeyi amaçlamıştır. Yorumlara ve puanlamalara yönelik yapılan analizler sonucunda; misafirlerin en mutlu olduğu kriterler sırasıyla: otelin ve odaların temizliği, rahatlık, personel tutum ve davranışları, fiyat/fayda dengesi, olanaklar ve konum iken; misafirlerin daha az tatmin olduğu kriterler ise oda imkânları özellikle ücretsiz Wi-Fi ve gürültü konusu olmuştur (Doğan, 2017).

Perikos ve Hatzilygeroudis (2017), booking.com rezervasyon sitesinden toplamış oldukları 417 kullanıcı yorumu üzerinde makine öğrenmesi yöntemiyle duygu analizi gerçekleştirmişlerdir. Naive Bayes, Maksimum Entropi, Destek Vektör Makineleri ve Topluluk Öğrenmesi algoritmaları kullanmış ve en iyi sonucu %86,0 başarı performansına sahip olan Topluluk Öğrenmesi algoritması yardımıyla elde etmişlerdir (Perikos ve Hatzilygeroudis, 2017).

İnan (2019), Eskişehir'deki otel müşterilerinin tutumlarının analiz edilmesi üzerine çalışmıştır. Çalışma iki aşamada gerçekleştirilmiş olup; ilk aşamasında İstanbul'da hizmet veren 20 otele yönelik 708 yorum üzerinde denetimli makine öğrenmesi yöntemlerinden Lojistik Regresyon, k-en yakın komşu, Naive Bayes ve Destek Vektör Makineleri algoritmaları karşılaştırılmıştır. Lojistik regresyon yöntemi 0.92'lik bir oranla en başarılı yöntem olarak belirlenmiştir. İkinci aşamasında, en başarılı yöntem olarak belirlenen Lojistik Regresyon yöntemiyle Eskişehir'deki altı otele yönelik 107 yorumun duygu analizi yapılmış ve bu yorumların %68'inin olumlu olduğu belirlenmiştir (İnan, 2019).

Özen (2019), araştırmasında helalbooking.com sitesinde yer alan Antalya bölgesi otel işletmelerine yönelik müşteri şikâyetlerini metin madenciliği yöntemlerinden fikir madenciliği (bakış tabanlı duygu analizi) tekniğini kullanarak analiz etmiştir. Analiz sonuçlarına göre; müşterilerin rahatsız olmalarına neden olan en önemli konu başlıkları sessizlik ve otel çalışanlarının (personel) olduğu görülmüştür. Otellerdeki sessizliği-sakinliğini bozan en önemli unsurların yüksek ses ve müzik olduğu anlaşılmıştır. Ayrıca müşterilerin otel personeli ile iletişim sorunu yaşadığı ve bunun nedenin de otel personelinin yeterli düzeyde İngilizce konuşamamaları olduğu ortaya çıkmıştır (Özen, 2019).

Tengilimoğlu ve Öztürk (2019), booking.com üzerinde yer alan 5 yıldızlı Bakü otellerine ilişkin yorumlara olumlu-olumsuz ve hangi konuları içerdiğine göre kümeleme analizi uygulamıştır. Analiz sonucuna göre; olumlu yorumlarda sırasıyla konum, odalar ve manzara (%38) ile çalışanlar (%28) öne çıkarken olumsuz yorumlarda ise kahvaltı, çalışanlar (%63), odalar ve temizlik (%15) ile ilgili konular ön plana çıkmıştır (Tengilimoğlu ve Öztürk, 2019).

Çoban ve diğerleri (2019), Eskişehir'de faaliyet gösteren 10 adet şehir oteline yönelik TripAdvisor.com web sitesinde çeşitli dillerde yapılan 940 adet olumlu ve olumsuz yorumu incelemişler ve bu yorumların en çok hangi kategoriler altında toplandığını bulabilmek için içerik analizi kullanmışlardır. Araştırma sonucunda yorumların sırasıyla "insan kaynakları" temasında personelin güler yüzlü olması, "yiyecek ve içecek" temasında kahvaltı/yemeğin lezzetli olması, "kat hizmetleri" temasında ise odaların temiz olması konularında yoğunlaştığını tespit etmişlerdir (Çoban S., Çoban E. ve Yetgin, 2019).

Çınar (2019), TripAdvisor.com sitesindeki çeşitli otellere ait olumlu ve olumsuz 113 adet İngilizce yorum üzerinde duygu analizi gerçekleştirmiştir. Çalışmada makine öğrenmesi yöntemlerinden Gini, Naive Bayes ve C5.0 sınıflandırma algoritmaları kullanmıştır. Bu algoritmalar içerisinde Gini algoritmasının en yüksek doğruluk değerine sahip olduğunu tespit etmiştir (Çınar, 2019).

Büyükeke, Sökmen ve Gencer, (2020), tarafından yapılan çalışmada Antalya bölgesinde yer alan otellere yönelik Tripadvisor.com platformunda yapılan değerlendirmelerden elde edilen verilerle; güncel metin madenciliği yöntemleri kullanılarak otel işletmeleri açısından rekabetçi bir zekâ oluşturulmasını amaçlamışlardır. 39.800 yorum üzerinden duygu analizi gerçekleştirilmiştir. Çalışmada Lojistik Regresyon, Destek Vektör Makineleri ve Naive Bayes yöntemleri kullanılmıştır. Analiz sonucunda yorumlarının %80'inin olumlu, %20'sinin olumsuz olduğu bulunmuştur. Konu analizi sonucunda; Deneyim %26,70 ile birinci, Değer ve Eğlence %24,68 ile ikinci, Şikâyet %20,41 ile üçüncü sırada yer aldığı, diğer konuların; %16,15 ile Temel Hizmetler ve %12,06 ile Yapılacak Şeyler olduğu belirlenmiştir (Büyükeke ve diğerleri, 2020).

Ahmetoğlu ve Daş (2020), Türkçe otel yorumlarından elde edilen verileri Word2Vec ile eğiterek kelime vektörleri oluşturmuşlardır. Oluşturulan vektörler "yapay sinir ağının bir

çeşidi olan geçitli tekrarlayan birimler" ile bir sınıflandırma modeli geliştirilmiş, geliştirilen geniş kelime torbalarıyla eğitilmiş kelime vektörleri derin öğrenme yöntemiyle (RNN altındaki GRU) incelenmiştir. Elde edilen sonuçlara göre daha geniş kapsamlı kelime torbalarının sınıflandırma başarısını arttırdığı gözlemlenmiştir (Ahmetoğlu ve Daş, 2020).

Tuna, Kaynar ve Akdoğan, (2021), otel müşterilerinden gelen geri bildirimlerdeki gizli olan duyguların derecelendirme puanlarıyla örtüşme düzeyini araştırmışlardır. Araştırmada; Lojistik Regresyon, Rastgele Orman (RF), Karar Ağacı (CART), K-En Yakın Komşu (KNN), Destek Vektör Makineleri (SVM), Doğrusal Diskriminant Analizi (LDA) ve Naive Bayes (NB) olmak üzere yedi farklı makine öğrenme algoritması kullanılmıştır. Kullanılan algoritmaların başarı değerleri sırası ile Lojistik Regresyon (%87,99), Destek Vektör Makineleri (%86,84), Doğrusal Diskriminant Analizi (%86,24), Naive Bayes (%82,66), Rastgele Orman (%82,00), Karar Ağaçları (%76,92) ve K-En Yakın Komşu (%63,91) şeklinde olmuştur (Tuna ve diğerleri, 2021).

Gerenaz ve Yetkin (2021), Sapanca'da hizmet veren beş yıldızlı SPA otellerine yönelik Tripadvisor.com web sitesinde yapılan toplam 485 olumlu ve olumsuz yorumu incelenmişlerdir. Yorumlar içerik analizi yöntemi ile anlamlı bölümlere ayrılmış, dokuz ana tema ve 54 alt tema altında toplanmıştır. Araştırma sonucunda yorumların sırasıyla "insan kaynakları" kategorisinde ilgili olma, "SPA" kategorisinde SPA hizmetlerinin başarılı olması, "yiyecek ve içecek" kategorisinde yemeklerin lezzetli olması konularında yoğunlaştığı saptanmıştır (Gerenaz ve Yetgin, 2021).

Acar ve Uğur (2021), Ankara'da bulunan 13 adet Uluslararası zincir otelde konaklayan misafirlerin Tripadvisor.com sitesinde yer alan yorum ve görüşleri üzerine duygu analizi uygulamışlardır. Çalışma kapsamında ortaya çıkan duygu analizi sonuçlarına göre; misafirlerin otel işletmelerinin hizmetlerine ilişkin güçlü tonlama ile vurguladıkları duygular "güven" ve "keyif" olmuştur (Acar ve Uğur, 2021).

Tamer ve Ceylan (2021), çalışmalarında Uludağ'da kış turizmi faaliyetlerine katılan müşterilerin otel işletmelerine yönelik çevrimiçi yorumları memnuniyet ve şikâyet kategorilerine ayırarak içerik analizi yöntemiyle incelemişlerdir. Araştırma sonuçlarına göre müşteriler hem şikâyetlerinde hem de memnuniyetlerinde en çok personel ve restoran kategorilerinde yorumlarda bulunmuşlardır (Tamer ve Ceylan, 2021).

Polat ve Ağca (2022), TripAdvisor.com sitesinden elde edilen Türkçe ve İngilizce yorumlar üzerinde duygu analizi yöntemlerini karşılaştırmışlardır. Karşılaştırma için sözlük tabanlı analiz ve makine öğrenme yöntemlerinden Karar Ağaçları ve Rastgele Orman sınıflandırma algoritmaları kullanılmıştır. Sonuç olarak makine öğrenme yöntemlerinin sözlük tabanlı analize göre daha etkin sınıflandırma gerçekleştirdiği, makine öğrenme yöntemleri algoritmaları birbirleriyle karşılaştırıldığında ise Rastgele Orman algoritmasının başarı oranının Karar Ağaçları algoritmasından daha yüksek olduğu bulunmuştur (Polat ve Ağca, 2022).

Yılmaz (2022), tarafından Airbnb- İstanbul'da kalan kullanıcı yorumlarının analizi yapılmış ve kullanıcıları daha iyi yönlendireceği düşünülen yeni bir puanlama sistemi önerilmiştir. Yorum analizi için gözetimsiz öğrenme yöntemlerinden BERT, TextBlob ve VADER(Valence Aware Dictionary and Sentiment Reasoner) algoritmaları kullanılmıştır. Yorumlar farklı kategoriler ekseninde değerlendirilmiş ve nihayetinde her bir evin tekrar puanlanmasıyla müşterilere farklı bakış açıları kazandırılması amaçlanmıştır (Yılmaz, 2022).

Leelawat ve diğerleri (2022), tarafından yapılan çalışmada Covid-19 pandemi sonrası Tayland turizmi ile ilgili İngilizce tweetlerin Destek Vektör Makineleri, Rastgele Orman ve Karar Ağaçları algoritmaları kullanılarak duygu analizi yapılmıştır. En iyi sonuç Destek Vektör Makineleri algoritması ile elde edilmiştir. Ayrıca kelime analizi ile; "gece hayatı, turistik yerler ve yemek" olumlu duygular sınıfında, "siyasi ortam ve Covid-19 pandemisi" olumsuz duygular sınıfında öne çıkmıştır (Leelawat ve diğerleri, 2022).

Bu çalışmalardan elde edilen sonuçlara göre, konaklama işletmeleriyle ilgili çevrimiçi platformlardaki müşteri değerlendirmeleri üzerine farklı analiz yöntemleri kullanarak konuyla ilgili farklı perspektifler sunulmuştur. Müşteri değerlendirmelerinin, çeşitli makine öğrenmesi algoritmaları ve içerik analizi yöntemleriyle incelendiği görülmüştür. Makine öğrenmesi algoritmalarından; Gizli Derecelendirme Regresyonu, Doğrusal Kelime Tabanlı, Lojistik Regresyon, Naive Bayes, Destek Vektör Makineleri, Rastgele Orman algoritmalarının kullanıldığı, Naive Bayes, Destek Vektör Makineleri ve Lojistik Regresyon algoritmalarının genellikle yüksek doğruluk değerlerine sahip olduğu bulunmuştur. Ayrıca, müşterilerin en çok odaların temizliği, personel tutumu, fiyat/fayda dengesi gibi faktörlerde mutlu oldukları, ancak oda imkanları ve gürültü gibi konularda daha az tatmin oldukları

belirlenmiştir. Bunun yanı sıra, müşteri yorumlarının içerik analiziyle incelenmesi sonucunda, insan kaynakları, yiyecek ve içecek gibi temaların ön plana çıktığı görülmüştür.

5.2. Glamping Turizm Sektörüne Yönelik Çalışmalar

Glamping turizm sektörüne yönelik yapılan çalışmalar incelendiğinde sınırlı sayıda araştırma yapıldığı tespit edilmiştir. Filipe ve diğerleri (2018), tarafından yapılan çalışmada müşterilerin glamping'i seçme motivasyonlarını ve engellerini belirlemeyi amaçlanmıştır. 19-73 yaş arası toplam 54 katılımcıdan oluşan bir örneklem ile 7 odak grup görüşmesiyle veriler toplamış ve bağlamsal yorumlamalar için içerik analizi teknikleri kullanılmıştır. Glamping yapmak için en önemli motivasyon kaynağının doğa ile doğrudan temas etmek olduğu bulunurken; cazibe, konfor, mahremiyet ve farklı bir deneyim de tüketicilerin takdir ettiği önemli unsurlar olarak elde edilmiştir. Bilgi eksikliği, sınırlı teklif ve kampa göre özgün olmaması ise engeller olarak tespit edilmiştir (Filipe ve diğerleri, 2018).

Zengin ve diğerleri (2019), çalışmalarında Kastamonu ilinin doğa turizm potansiyelini alternatif turizm kapsamında değerlendirmek için nitel araştırma yöntemlerinden görüşme tekniğini kullanmışlardır. Verileri toplamak amacıyla Kastamonu turizminin yerel paydaşları ile yüz yüze görüşmeler yapmışlardır. Elde edilen veriler betimsel analize tabi tutularak yorumlanmaya çalışılmış, yapılan analiz sonucunda Kastamonu ilinin sahip olduğu sayısız doğal zenginliklere rağmen yeterince turistik talep yaratamadığı değerlendirilmiştir (Zengin, Koç ve Ulama., 2019).

Brochado A. ve Brochado F. (2020), çalışmalarında Portekiz'deki yedi glamping tesisinde konaklayan turistlerin TripAdvisor.com sitesinde yer alan yorumlarının, doğal dil metinlerinin denetimsiz nicel içerik analizlerini yürüten bir yazılım programı olan Leximancer kullanarak incelemişlerdir. Analiz sonucunda; en önemli tema "deneyim(%100, 309)" olmak üzere "otel" (%84, 237), "öğrenme" (%81, 229), "ev sahibi" (%79, 192), "kamp" (%65, 173) takip ediyor. , "doğa" (yüzde 63, 161), "yiyecek" (yüzde 51, 155), "içerikler" (yüzde 50, 142), "fark" (yüzde 48, 111), "eko(loji) " (%24, 80) ve "yoga" (%25, 60) 11 tema bulunmuştur (Brochado A. ve Brochado F., 2020).

Akgün ve Ünal (2021), Ölüdeniz, Marmaris ve İzmir Seferihisar'da faaliyet gösteren glamping işletmelerinde konaklayan turistlerin çeşitli dillerde (Türkçe, İngilizce, Almanca,

Rusça) 302 adet olumlu yorumunu içerik analiziyle değerlendirmiştir. Çalışma sonucunda; "yiyecek ve içecek olanakları" kategorisinde yiyecek ve içecek lezzeti, "otelin konumu" kategorisinde doğanın güzel olması, "otelin mimarisi ve otelin genel durumu" kategorisinde glamping işletmelerin konaklama tipi ve "çalışanların davranışları" kategorisinde çalışanın güler yüzlü olması konularının üzerinde sıklıkla durulduğu belirlenmiştir (Akgün ve Ünal, 2021).

Lu ve diğerleri (2021), glamping müşterilerinin ağızdan ağıza iletişimini etkileyen faktörleri belirlemeyi amaçlamışlardır. Bunun için 20 glampsite ait çevrimiçi incelemeleri gömülü teori tekniklerini kullanarak niteliksel analiz gerçekleştirmişlerdir. Analiz sonucunda ekoloji ve konaklama ortamının; müşterilerin gerginliğini atmasını, keyif almasını ve sosyal bağ kurmasını etkilediğini ve bu olumlu duygusal değerlerin ağızdan ağıza iletişimin esin kaynağı olduğu görülmüştür (Lu, Wang, Dai, Chen ve Feng, 2021).

Craig ve Karabaş (2021), tarafından yapılan çalışmada COVID-19'un glamping üzerindeki mevcut ve gelecekteki etkilerini ele almak için ABD ve Kanada'da 2926 aktif eğlence amaçlı seyahat eden kişiyle anket yapılmıştır. Karşılaştırma için katılımcılara COVID-19 sonrası glamping gezi planları ve otel/tatil köyü gezi planları sorulmuştur. Yapılan analiz sonuçlarına göre glamping gezi planlarının pandemi sırasında ve sonrasında geçerli bir eğlence seyahati alternatifi haline geldiği ve tercih edilme oranının iki katına çıktığı tespit edilmiştir (Craig ve Karabaş, 2021).

Düzgün (2021) çalışmasında Covid-19 pandemi süreci etkisiyle turistlerin kalabalık ortamlar yerine daha bireysel hareket edebildikleri için turistik amaçlı olarak tercih ettikleri glamping turizminin örnek bir işletme üzerinden değerlendirilmesini amaçlamıştır. Bu kapsamda Hypnos Sapanca GlampingveSpa işletmesi örnek olay kapsamında belirlenmiştir. Araştırma sonucunda, pandemi sonrası Türkiye'deki ve dünyadaki turist tercihlerine bağlı olarak glamping turizmine olan ilgi ortaya çıkarılmıştır (Düzgün, 2021).

Çepur ve Dalgıç (2022), Covid-19 pandemisiyle birlikte artan bir trend haline dönüşen glamping turizminin Isparta ili ve çevresindeki potansiyelinin ortaya konulmasını amaçlamıştır. Araştırma kapsamında, Isparta ilini tanıyan ve glamping konusunda bilgisi olan akademisyenlerle ve yerel yönetimlerin üst düzey yöneticileriyle görüşmeler gerçekleştirmiş olup veri toplamada görüşme tekniği kullanmıştır. Veriler betimsel analiz

yöntemiyle çözümlenmiştir. Verilerin analizi sonucunda, Isparta ili ve çevresinde Eğirdir ilçesinin göl kesiminin glamping tesisi için ön plana çıktığı, rekreatif faaliyet çeşitliliği, sulak alanların varlığı, erişebilirlik, tesisler, çekici faktörler, destekleyici hizmetler gibi hususlar dikkate alındığında en uygun alanın Eğirdir olduğu tespit edilmiştir (Çepur ve Dalgıç, 2022).

Glamping turizmi üzerine yapılan çalışmalarda genellikle içerik analizi yöntemi tercih edilmiştir. Bunun yanı sıra görüşme teknikleri, çevrimiçi incelemelerin analizi ve ankete dayalı yöntemler de kullanılmıştır. Pandemi döneminde glamping turizmine olan ilginin, kalabalık ortamlardan kaçınma ve doğa ile iç içe olma arayışı nedeniyle arttığı görülmüştür.

Bu çalışmaların, glamping turizmi üzerine genel eğilimleri ve glamping tercihlerini etkileyen faktörleri gösterdiği tespit edilmiştir.

Sonuç olarak; otellere ilişkin duygu analizi ile alakalı çalışmalarda genel olarak makine öğrenmesi yöntemlerinin kullanılarak birbirleriyle karşılaştırılmalarının yapıldığı, ancak bu konuda derin öğrenme yöntemlerinin kullanılmadığı görülmüştür. Ayrıca otellere ilişkin çevrimiçi yorumların değerlendirilmesine yönelik literatürde birçok çalışma yer alırken, COVID 19 salgınıyla birlikte popüler hale gelen bungalov işletmeleri özelinde herhangi bir çalışmaya rastlanmamıştır. İlgili akademik literatür incelendiğinde Sapanca Bölgesi'nde bulunan glamping turizm ve bungalov işletmelerine yönelik olarak makine öğrenmesi ve BERT yöntemleri kullanılarak gerçekleştirilen herhangi bir çalışmanın olmadığı görülmüştür. Bu nedenle çalışmanın ilgili literatüre ve sektöre önemli katkılarının olacağı öngörülmektedir.

6. UYGULAMA: COVİD-19 PANDEMİSİ SONRASI YENİ TREND BUNGALOV İŞLETMELERİNE İLİŞKİN ÇEVRİMİÇİ YORUMLARIN MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE DUYGU ANALİZİ

Çalışmanın uygulama aşamasını içeren altıncı bölümde; çalışmanın amacı, kapsamı, çalışmada kullanılan veri setinin hangi kriterlere göre seçildiği, hangi ön işlemlerden geçirildiği, analize uygun hale getirilen verilere makine öğrenmesi ve derin öğrenme algoritmaları uygulanarak elde edilen sonuçların bilgileri yer almaktadır. Aynı zamanda Koronavirüs salgınının turizm tercihleri üzerindeki etkisini değerlendirmek için veriler üzerinde gerçekleştirilen analizler bu bölümde sunulmaktadır.

6.1. Çalışmanın Amacı

Çalışmanın amacı, Sapanca bölgesinde yer alan bungalov işletmerine ait yorumların ve puanlamaların makine öğrenmesi ve derin öğrenme algoritmaları ile sınıflandırılması ve tahmin edilmesidir. Ek olarak Covid-19 pandemisinin bungalov işletmeleri talepleri üzerindeki etkisinin incelenmesi amaçlanmıştır. Yapılan literatür araştırmasında, bungalov ve glamping turizm işletmelerine yönelik olarak makine öğrenmesi yöntemiyle gerçekleştirilen herhangi bir çalışma olmadığı görülmüştür. Bu nedenle çalışma literatür ve sektöre katkı sağlaması açısından önem taşımaktadır.

6.2. Çalışmanın Kapsamı

Yapılan çalışma Sapanca bölgesi ve Sakarya ili ile sınırlı tutularak bu bölgedeki bungalov işletmeleri dikkate alınmıştır. Bu bölgenin seçilme nedeni; pandemi süreciyle birlikte önem kazanan sosyal mesafe ve konaklayanlara bireysel hareket edebilme imkânı sağlaması, sahip olduğu atmosferik ortamıyla (süsleme, ışıklandırma, jakuzi, şömine vb.) sosyal medyada paylaşım ve tanıtımlarının artmasıyla buradaki işletmelerin daha popüler hale gelmesidir.

6.3. Veri Setinin Oluşturulması

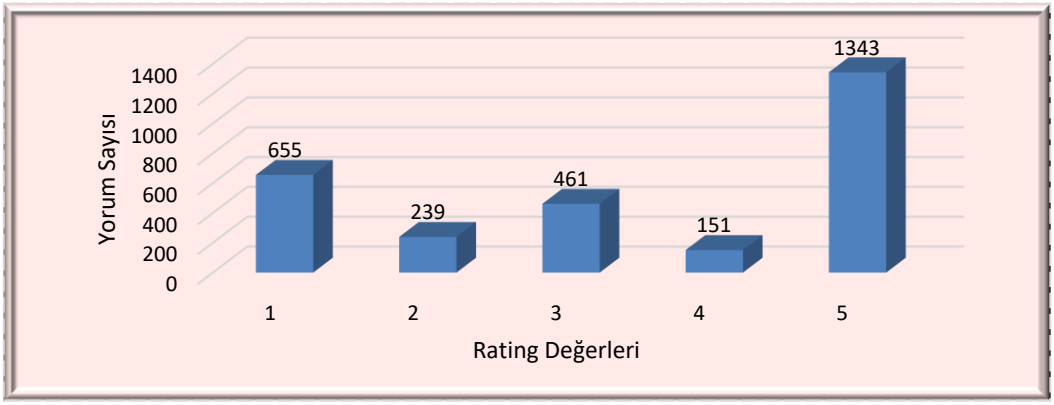
Çalışmaya temel teşkil eden veriler "Google Places" kullanarak, Sapanca bölgesindeki bungalovların sayfalarına bırakılan müşteri yorumlarından oluşmaktadır. Genel olarak konaklama işletmelerine ait yorumların değerlendirilmesine yönelik çalışmalarda Tripadvisor.com, Booking.com gibi çevrim içi rezervasyon sitelerinden yararlanılmıştır fakat Sapanca bölgesindeki bungalovlara ait yorumların belirtilen sitelerde yeterli sayıda olmadıklarından dolayı "Google Places" tercih edilmiştir.

"Google Places" Google tarafından uygulama geliştiricilere sunulan, konum tabanlı hizmetlerden en üst düzeyde yararlanabilmeyi sağlayan bir ara yüzdür. Sadece konum ve harita hizmeti sunmakla sınırlı değildir. Yer ayrıntıları, yer fotoğrafları ve otomatik yer tamamlama gibi çeşitli hizmetlerle kullanıcılarına kolaylıklar sağlamaktadır.

- Yer ayrıntıları: sorgulama yapılan konuma ait bilgilere detaylı olarak erişim sağlayabilme olanağı sunmaktadır.
- Yer fotoğrafları: konuma ait detaylı bilgilerin yanı sıra konum ve çevresinin fotoğraflarına erişim imkânı sunmaktadır.
- Otomatik yer tamamlama: kullanıcıların veri tabanına yeni yerler ekleyebilmesidir.

Çalışmaya dahil edilen veriler 03.02.2023 tarihine kadar 17 adet bungalov işletmesine yönelik yapılan yorumlardan elde edilmiştir. İlgili işletmelere yönelik toplam 5504 adet yorum ve değerlendirme puanı olmasına rağmen sadece puan verilmiş fakat herhangi bir metin ifadesi içermeyen 2690 değerlendirme Scrapy isimli Pyhton kütüphanesi ile otomatik olarak tespit edilmiş, veri setine dahil edilmemiştir. Çalışmada sonuçların daha anlaşılır olabilmesi için yorumların aynı dilde incelenmesi önemli bir konu olduğundan orijinal hali yabancı dillerde olan fakat "Google tarafından çevrildi" şeklinde Türkçeye çevirisi yapılan yorumlarla birlikte veri setinde Türkçe dili kullanılmıştır. Gerekli eleme işlemleri sonrasında veri seti toplamda 2849 yorumdan elde edilmiştir.

Yorumlar; yorum tarihi, yorum rating değeri ve yorum metnini içermektedir. Toplamda çalışmaya 1494 olumlu, 1355 olumsuz yorum dâhil edilmiştir. Şekil 6.1'de yorumların rating değerlerine göre dağılımı gösterilmektedir.



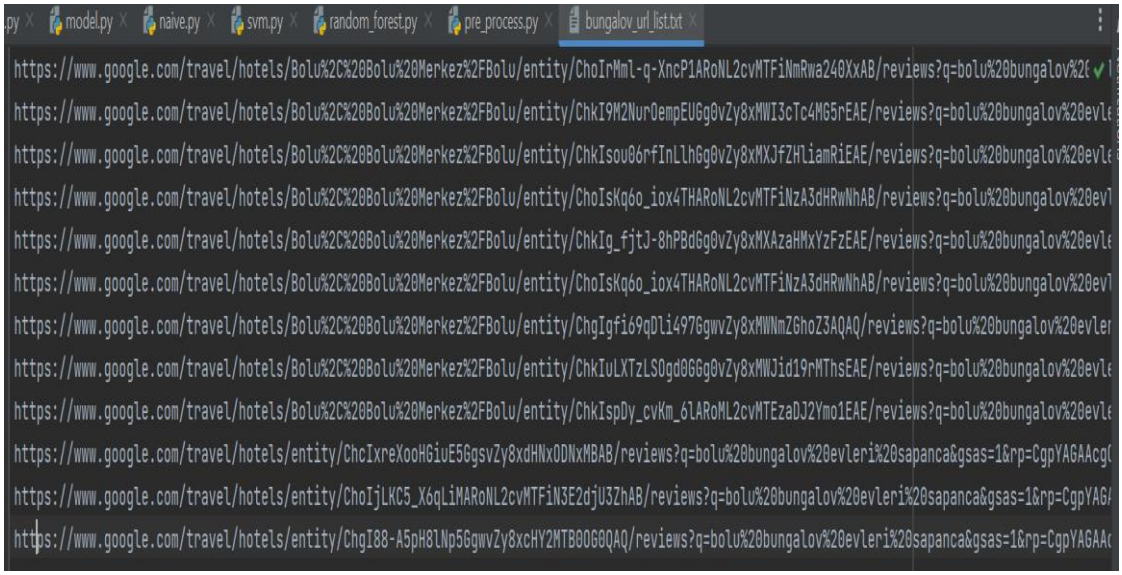
Şekil 6.1. Rating değerlerine göre yorumların dağılımı

Şekil 6.1' e göre yorumların genel dağılımı; rating değeri 4 ve 5 olan yorumlar yaklaşık %52, rating 1,2 ve 3 değerine sahip olan yorumlar ise yaklaşık %48 şeklindedir. Müşteri yorumlarının doğru bir şekilde değerlendirilebilmesi için negatif ve pozitif yorumların birbirlerine sayısal olarak üstünlük sağlamamasına dikkat edilmiştir. Bulgular bölümünde, makine öğrenmesi ve derin öğrenme yöntemleri kullanılarak sınıflandırılmış tüm yorumların duygusal kutupsallığı gösterilmiştir.

6.3.1. Çalışma verilerinin web ortamından çekilmesi

Çalışmaya konu olan Sapanca bölgesindeki bungalovlar "Google Places" platformunda belirlenip, işletmelerin platformdaki yorum sayfasına yönlendiren linkler bir listeye düz metin formatında satır satır eklenmiştir. Bu linklerdeki yorumların web kazıma yöntemi ile çekilmesinde açık kaynak kodlu "Python" programlama dili ile "Selenium" ve "Scrapy" kütüphanelerinden faydalanılmıştır.

Çalışma verilerinin seçileceği bungalovlara ait linkler için bir liste sayfası oluşturulmuş, bu listede yer alan linkler düz metin formatında otomatik olarak çağrılarak çalıştırılması için "Selenium" kütüphanesi kullanılmıştır. "Selenium" bir web tarayıcı otomasyon aracıdır. Bu araç sayesinde programatik ve algoritmik olarak "Chrome" ya da "Firefox" web tarayıcıları oluşturulabilir ve webde gezilebilir (https://selenium-python.readthedocs.io/) Yorumları çekilecek bungalovların "Google Places" sitesindeki linklerinin tutulduğu ve web kazıma aşamasında linklerin otomatik olarak kod dizinine aktarıldığı ekran görüntüsü Şekil 6.2'de gösterilmektedir.



Şekil 6.2. Bungalov işletmeleri linklerinin tutulduğu liste görüntüsü

Şekil 6.2'de gösterilen, çalıştırılan kodun bir bungalovun yorumunu bitirip diğer bungalovun yorumunu çekmek için her seferinde yeniden otomatik olarak yönlendirildiği bir dosya konumudur. Elde edilen HTML belgelerinde yorumları toplu olarak seçmeye yarayan CSS seçicisi "Scrapy" kütüphanesi yardımıyla oluşturulmuş ve yorumlar toplu olarak çekilmiştir. Bir işletmedeki yorumlar tarayıcı ekranı aşağıya kaydıkça geldiği için "Selenium" yardımıyla tarayıcıda periyodik olarak ekranı aşağıya kaydırma algoritması geliştirilmiştir. Bu sayede eski yorumların da çekilebilmesi sağlanmıştır.

Web kazıma, HTTP istekleri ile sunucudan alıcıya servis edilen HTML, JSON ve farklı türde belgelerden ve ortam dosyalarından istenilen verilerin çekilmesi için otomatizasyon yöntemidir (Mitra, Paul ve Singrodia, 2019). Bu çalışmada, web kazıma yöntemi "Google Places" platformunda yorumlara tekabül eden linklerden alınan HTML belgelerinden CSS seçicileri aracılığıyla yorumların, puanların ve tarihin çekilmesi için kullanılmıştır. Bungalov yorumlarını çekmede kullanılan kod EK-1'de yer almaktadır.

EK-1'de verilen kod satırlarının çalıştırılması sonucunda bungalov yorumları çekilmiştir. Her işletme için çekilen yorumlar, puanlar ve tarihler ayrı CSV dosyalarına kaydedilmiştir. Bir bungalova ait çekilen yorumların kaydedildiği CSV dosyası Şekil 6.3'te gösterilmektedir.

timestamp	rating	comment
üzerinde 3 hafta önce	5	2 gece dublex odada kaldık, manzarası, jakuzisi, şöminesi bir harika, tam bir romantik balayı konseptiydi (çocuksuz gidilmesini tavsiye ederim)
üzerinde 5 ay önce	3	Otelde bir gece konakladık.
üzerinde 3 ay önce	3	Otelde 2 gece konakladık. Konum olarak oldukça güzel. Havası huzurlu ve balkondan ağaç manzarası gözüküyor. Odaların temizliği gayet güzeldi. Oda içinde jakuzi güzel
üzerinde 5 ay önce	1	Hem süit hem dublex odasında 2 gece kaldık ve verdiğimiz paraya düşününce bizi hayal kırıklığına uğratan bir oteldi. Jakuzi olduğu için tutup jakuzisini pislikten kullanam
üzerinde 3 ay önce	2	13.11.2022-17.11.2022 tarihleri arasında 4 gece balayı için konakladık. Yorumlardan iyi bir otel olduğunu görmüştüm fakat umduğumu bulamadım.
üzerinde 5 ay önce	3	Ağustos ayında 2 gecelik bir konaklamamız oldu. Otel odalarının o fiyatı hakettiğini düşünmüyorum. Sırf odada jakuzi ve dağ manzarası var diye o kadar fiyat biraz abart
üzerinde 3 ay önce	3	18-20 kasım tarihleri arasında 2 gece konakladık. Otelin konumu, manzarası, odanın imkanları(şömine ve jakuzi) çok güzel. Fakat odamıza verdiğimiz siparişlerin tabakla
üzerinde 6 ay önce	5	Balayı oteli olarak mutlaka tercih edilmeli. Hem sunulan hizmet hem çalışanların ilgisi çok iyiydi. Oda manzarası, jakuzisi, sauna-hamam imkanı, kahvaltısı, atmosferi ve s
üzerinde 5 ay önce	3	Jakuzi, sauna, hamam var. Odalar çok güzel.
üzerinde 2 ay önce	1	Eşimin doğum günü nedeniyle tercih edip gittik ama Kesinlikle fiyatına göre otel değil. Biz 23/24 Aralık 2022 tarihinde dubleks odada kaldık. Jakuzi suyu odaya taşıyor. Yo
üzerinde 5 ay önce	5	Kesinlikle tavsiye edebileceğim bir hotel 🌲ormanın yanı başında etkili manzarası oda içerisinde ki kaliteli yapı, resepsiyon çalışanların ilgisi sizi etkileyen unsurlardan bir
üzerinde 5 ay önce	5	Ailemle 1 gece abantı gezmek için kaldığımız otel. Tek kelime ile muazzam geçti.Baştan sona Tüm çalışanlar çok nazik ve ilgiliydi. Her şey için teşekkürler. Kahvaltı ayrı ba
üzerinde 3 ay önce	4	Kahvaltı yapmak için geldik. Kahvaltı menüsü serpme olarak geliyor. Sıcaklar konusunda biraz yetersiz gibi geldi bana. Diğer yiyecekler konusunda hepsi kararında konul
üzerinde 4 ay önce	5	Aden suit odada kaldık, temiz ve ferah bir oda. Jakuzinin temizliği idare eder. Balkon mobilyaları çok rahat değil, daha konforlu dizayn edilebilir. Personeller güler yüzlü,
üzerinde bir ay önce	5	Genel olarak çok iyiydi.
üzerinde 2 ay önce	1	Kesinlikle gitmeyin diyecegim bir yer. Odalarin temizligi iyi degil. Internet hic cekmiyor. Oda da bi kettle bile koymamislar. Odalarda isiklandirma cok yetersiz.Herseyi par
üzerinde 3 ay önce	1	Rezervasyonumu yanlış alıp üstelik bütün parayı ödememe rağmen otele geldiğimde dündü rezervasyonunuz diyip hiç bir ilgi göstermeden küfür edip tehdit edip hakar
üzerinde 3 ay önce	5	Harika iki gün geçirdik eşimle beraber:) tertemiz bir otel :) o kadar temizdi ki ayakkabimizi çıkarıp mi girsek diye tereddüt ettik :) çok mutluyduk yalnız jakuzisinde biraz a
üzerinde bir ay önce	4	Odaları çok güzel, hizmet çok iyi bütün personel guleryuzlu hepsi 5 yıldız, ama odanın balkonu kirli ve tozluydu biz odaya giriş yaptigimizda ve jakuziden dışarı su sızıntısı
üzerinde bir ay önce	4	Sıcaklığıyla,odaların konforuyla kahvaltıları ile cok güzel misafir edildik.Odalarda su ısıtıcı bulunmaması biraz üzdü dogrusu. Fiyatı da biraz fazla bulduk o yüzden 4 puan
üzerinde 3 ay önce	4	Dublex odasında konakladım ilgi alaka ve kahvaltısı gayet güzeldi odalarda jakuzi ve şömine olması mükemmeldi hele ki sabah orman manzarası çık iyi ...

Şekil 6.3. Bir bungalov işletmesine ait yorumların CSV dosyasındaki örnek görüntüsü

Şekil 6.3'te verilen örnekteki gibi tüm bungalovlara ait yorumlar çekilmiş ve tek bir Excel dosyasında birleştirilmiştir. Böylelikle veriler ön hazırlık süreci için tümüyle hazır hale getirilmiştir.

6.3.2. Çalışma verilerinin ön hazırlık süreci

Yorumlar kelimelerden, özel karakterlerden ve yapısal bilgilerden oluşmaktadır. Yorumların program tarafından işlenebilmesi için sayısal hale dönüştürülmesi ve analize hazır hale getirilmesi gerekmektedir (Büyükeke ve diğerleri, 2020). Bu amaçla aşağıdaki aşamalar elde edilen verilere uygulanmıştır.

Çalışma verilerinin temizlenmesi

Çalışma verilerinin temizlenmesi sürecinde; yorumlar noktalama işaretleri, sayılar ve emojilerden arındırılmış, büyük harfler küçük harflere çevrilmiş, yorumların başındaki ve sonundaki boşluklar temizlenmiştir. "Google tarafından çevrildi" kalıbı kaldırılmış ve durak sözcükler (stopwords) NLTK kütüphanesi kullanılarak temizlenmiştir. Bu sürecin sonunda tamamı küçük harflerden oluşan noktalama işaretlemelerinin olmadığı metinler elde edilmiştir.

Çalışma verilerinin köklerine ayrılması ve etiketlenme süreci

Veri temizleme işleminden sonra, yapay zeka modellerinde kelimelerin kullanabilmesi için kelimelerin köklerine ayırılması gerekmektedir. Kelimelere gelen ekler ve kelimelerin çekimli halleri öznitelik uzayının (feature space) fazlasıyla büyümesine ve öğrenme sürecinin sağlıksız olmasına sebep olur. Kelimelerin köklerine ayrılmasında hece bölme (stemmatization) ve morfolojik kök (lemmatization) çıkarımı olmak üzere iki yaklaşım kullanılmaktadır (Tuna, 2020). Bu çalışmada morfolojik kök ayrımı yöntemi kullanılmıştır. Bu yöntemle kelimenin aday kökleri çıkarılır morfolojik anlam ayrıştırılması ile en uygun aday kök bulunur. Kök ayırılması için "StarlangSoftware" in "MorphologicalDisambiguation" ve "MorphologicalAnalysis" kütüphaneleri kullanılmıştır. Öncelikle yaklaşık ikiyüz bin kelimeden oluşan ağaç yapılı derlem kullanılarak kök öncellikli anlam ayrıştırıcısı eğitilmiştir. Kök öncelikli anlam ayrıştırıcı, morfolojik analiz sonucunda anlam ayrıştırılması için ekleri ve çekimlere ağırlık vermeden salt kökün bulunmasına öncelik verir. Bu sayede en uygun aday kök bulunur (Siddhartha, 2021). Morfolojik analiz için ise "Sonlu Durum Makinesi ile Morfolojik Analiz" yöntemi kullanılmıştır. Sonlu durum makineleri herhangi bir girdi dizisi için sınırlı sayıda ve aynı girdi dizisi için tek ve sınırlı bir duruma sahip olan soyut makinelerdir. Bu özelliği sayesinde önceden oluşturulmuş bir sözlükteki kelimelerin hecelerine ayrıştırılmasında kullanılabilir. "StarlangSoftware" in "MorphologicalAnalysis" kütüphanesi Türkçe dili için önceden oluşturulmuş Türkçe sözlük ve bu kelimelerin durum geçişlerini yöneten sonlu durum makinesine sahiptir. Verilerin temizlenmesi ve kelimelerin köklere ayrılması işlemleri için kullanılan kodlar EK-2'de gösterilmektedir. EK-2'de verilen kodların çalıştırılması sonucu analize hazır hale getirilmiş CSV dosyası örneği Şekil 6.4'te verilmektedir.

timestamp	rating	comment	lemmas
üzerinde 3 hafta önce	5	2 gece dut	['gece', 'dublex', 'oda', 'kal', 'manzara', 'jakuzi', 'şömine', 'bir', 'harika', 'tam', 'bir', 'romantik', 'balay', 'konsept', 'çocuksuz', 'git', 'tavsiye', 'et']
üzerinde 5 ay önce	3	Otelde bir	['otel', 'bir', 'gece', 'konakla']
üzerinde 3 ay önce	3	Otelde 2 g	['otel', 'gece', 'konakla', 'konum', 'ol', 'oldukça', 'güzel', 'hava', 'huzurlu', 've', 'balkon', 'ağaç', 'manzara', 'gözük', 'oda', 'temizlik', 'gayet', 'güzel', 'oda', 'içinde', 'ja
üzerinde 5 ay önce	1	Hem süit h	['hem', 'süit', 'hem', 'dublex', 'oda', 'gece', 'kal', 've', 'ver', 'para', 'düşün', 'biz', 'hayal', 'kırıklık', 'uğra', 'bir', 'otel', 'jakuzi', 'ol', 'için', 'tut', 'jakuzi', 'pislik', 'kullan', '
üzerinde 3 ay önce	2	13.11.202	['13', '11', '2022', '17', '11', '2022', 'tarih', 'ara', 'gece', 'balay', 'için', 'konakla', 'yorum', 'iyi', 'bir', 'otel', 'ol', 'gör', 'fakat', 'um', 'bul']
üzerinde 5 ay önce	3	Ağustos ay	['ağustos', 'ay', 'gecelik', 'bir', 'konakla', 'ol', 'otel', 'oda', 'fiyat', 'haket', 'düşün', 'sırf', 'oda', 'jakuzi', 've', 'dağ', 'manzara', 'var', 'diye', 'kadar', 'fiyat', 'biraz', 'abart
üzerinde 3 ay önce	3	18-20 kası	['18', '20', 'kasım', 'tarih', 'ara', 'gece', 'konakla', 'otel', 'konum', 'manzara', 'oda', 'imkan', 'şömine', 've', 'jakuzi', 'çok', 'güzel', 'fakat', 'oda', 'ver', 'sipariş', 'tabak',
üzerinde 6 ay önce	5	Balayı otel	['balay', 'otel', 'ol', 'mutlaka', 'tercih', 'et', 'hem', 'sun', 'hizmet', 'hem', 'çalış', 'ilgi', 'çok', 'iyi', 'oda', 'manzara', 'jakuzi', 'sauna', 'hamam', 'imkan', 'kahvaltı', 'atmo
üzerinde 5 ay önce	3	Jakuzi, sau	['jakuzi', 'sauna', 'hamam', 'var', 'oda', 'çok', 'güzel']
üzerinde 2 ay önce	1	Eşimin doğ	['eş', 'doğum', 'gün', 'nedeniyle', 'tercih', 'et', 'git', 'ama', 'kesinlikle', 'fiyat', 'göre', 'otel', 'değil', 'biz', '23', '24', 'aralık', '2022', 'tarih', 'dubleks', 'oda', 'kal', 'jakuzi',
üzerinde 5 ay önce	5	Kesinlikle	['kesinlikle', 'tavsiye', 'et', 'bir', 'hotel', 'orman', 'yan', 'başında', 'etki', 'manzara', 'oda', 'içeri', 'ki', 'kalite', 'yapı', 'resepsiyon', 'çalış', 'ilgi', 'siz', 'etkile', 'unsur', 'bir
üzerinde 5 ay önce	5	Ailemle 1	['aile', 'gece', 'abant', 'gez', 'için', 'kal', 'otel', 'tek', 'kelime', 'ile', 'muazzam', 'geç', 'baş', 'son', 'tüm', 'çalışanlar', 'çok', 'nazik', 've', 'ilgili', 'her', 'şey', 'için', 'teşekkü
üzerinde 3 ay önce	4	Kahvaltı yı	['kahvaltı', 'yap', 'için', 'gel', 'kahvaltı', 'menü', 'serpme', 'ol', 'gel', 'sıcak', 'konu', 'biraz', 'yetersiz', 'gibi', 'gel', 'bana', 'diğer', 'yiyecek', 'konu', 'hepsi', 'karar', 'kon',
üzerinde 4 ay önce	5	Aden suit	['aden', 'suit', 'oda', 'kal', 'temiz', 've', 'ferah', 'bir', 'oda', 'jakuzi', 'temizlik', 'idare', 'et', 'balkon', 'mobilya', 'çok', 'rahat', 'değil', 'daha', 'konforlu', 'dizayn', 'et', 'pe
üzerinde bir ay önce	5	Genel olar	['genel', 'ol', 'çok', 'iyi']
üzerinde 2 ay önce	1	Kesinlikle	['kesinlikle', 'git', 'de', 'bir', 'yer', 'oda', 'temizlik', 'iyi', 'değil', 'internet', 'hiç', 'çek', 'oda', 'da', 'bir', 'kettle', 'bile', 'koy', 'oda', 'ışıklandırma', 'çok', 'yetersiz', 'hersey
üzerinde 3 ay önce	1	Rezervasy	['rezervasyon', 'yanlış', 'al', 'üstelik', 'bütün', 'para', 'ödememe', 'rağmen', 'otel', 'gel', 'dün', 'rezervasyon', 'de', 'hiç', 'bir', 'ilgi', 'göster', 'küfür', 'et', 'tehdit', 'et', '
üzerinde 3 ay önce	5	Harika iki	['harika', 'iki', 'gün', 'geçir', 'eş', 'beraber', 'tertemiz', 'bir', 'otel', 'kadar', 'temiz', 'ki', 'ayakkabimizi', 'çık', 'mi', 'gir', 'diye', 'tereddüt', 'et', 'çok', 'mutlu', 'yalnız', 'jak
üzerinde bir ay önce	4	Odaları ço	['oda', 'çok', 'güzel', 'hizmet', 'çok', 'iyi', 'bütün', 'personel', 'guleryuzlu', 'hepsi', 'yıldız', 'ama', 'oda', 'balkon', 'kir', 've', 'tozlu', 'biz', 'oda', 'gir', 'yaptigimizda', 've'
üzerinde bir ay önce	4	Sıcaklığıyl	['sıcaklık', 'oda', 'konfor', 'kahvaltı', 'ile', 'çok', 'güzel', 'misafir', 'et', 'oda', 'su', 'ısıtıcı', 'bulunmama', 'biraz', 'üz', 'doğru', 'fiyat', 'da', 'biraz', 'fazla', 'bul', 'yüzden',
üzerinde 3 ay önce	4	Dublex od	['dublex', 'oda', 'konakla', 'ilgi', 'alaka', 've', 'kahvaltı', 'gayet', 'güzel', 'oda', 'jakuzi', 've', 'şömine', 'ol', 'mükemmel', 'hele', 'ki', 'sabah', 'orman', 'manzara', 'çık', 'i

Şekil 6.4. Temizleme ve köklerine ayırma işlemleri sonucu elde edilen CSV dosyası örnek görüntüsü

Şekil 6.4'ten de anlaşılabileceği gibi çözümlemenin temelini oluşturan kelimeler köklerine ayrıştırılmıştır. Yorumların etiketlenmesinde müşteri puanları esas alınmıştır. "Google Places" sitesinde müşteriler bungalovlara ilişkin değerlendirmelerini 1 ila 5 arasında yapmaktadırlar. Ön etiketleme işlemi gerçekleştirilen yorumlardan 3 puan ve altındakiler negatif; 4 puan ve üzerindekiler pozitif kabul edilerek ikili etiketlenmiştir.

Vektör uzay modelinin oluşturulması

Vektör uzay modelleme, herhangi bir metin veya metin parçasını vektör olarak gösterilebilmek ve bir uzay içerisinde modelleyebilmek için makine öğrenmesi ve doğal dil işleme alanlarında sıklıkla kullanılan bir yöntemdir. Bu matematiksel yöntem metin verilerinin filtrelenmesinde, bilgi getiriminde, indeksleme ve ilişki sıralamalarında kullanılarak; sayısal olmayan string (nümerik olmayan) bir veri yapısını, integer (nümerik) bir yapıya dönüştürmektedir (Tuna, 2020). Vektör uzay modelinin oluşturulmasında en çok tercih edilen yöntemlerden biri TF-IDF (Term Frequency — Inverse Document Frequency) ağırlıklandırmaları, bu çalışmada kullanılan vektör uzay modeli oluşturma yöntemidir.

TF-IDF metin madenciliği ve doğal dil işleme alanlarında kullanılan bir istatistiksel yöntemdir. Bir belgede veya metindeki her bir kelimenin önemini belirlemek için kullanılır. TF bir kelimenin metinde geçme sıklığını gösteren bir değerdir yani terim frekansıdır. IDF ise kullanıcı tarafından belirlenen kelimenin metinlerin kaç adedinde yer aldığını göstermektedir yani ters belge frekansıdır (Kaynar ve diğerleri, 2016).

Çalışmada, "sklearn.feature_extraction.text" kütüphanesinin "TfIdfVectorizer" sınıfından faydalanılmıştır. Bu sınıf sayesinde; yorum kökleri TF-IDF özelliklerinden oluşan matrislere çevirilmiş, aynı kütüphanede yer alan "CountVectorizer" sınıfı yardımıyla sayı vektörizasyonu yapılmış ve son olarak "TfIdfTransformer" ile "TF-IDF" matrisi hesaplanmıştır. Vektör uzay modelinin oluşturulması için kullanılan kod Şekil 6.5'de verilmektedir.

```
tf_wb = TfidfVectorizer(min_df=0.01,ngram_range=(1,2))
```

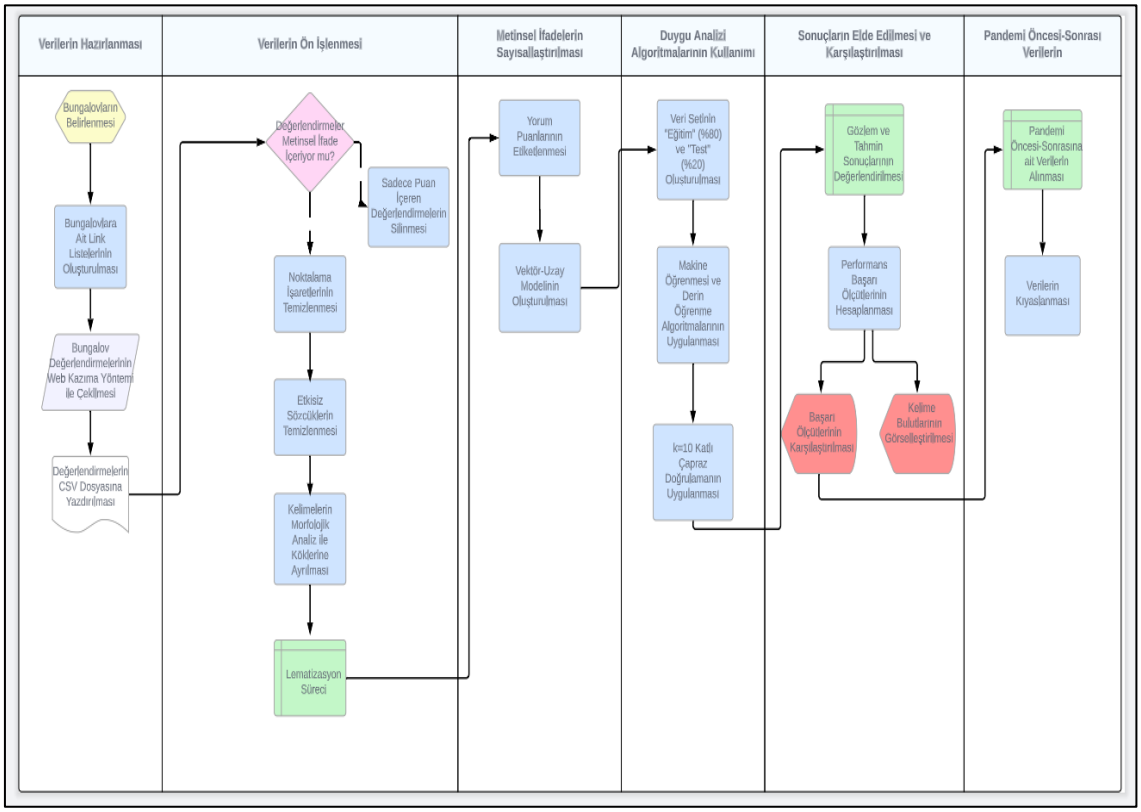
Şekil 6.5. Vektör uzay modeli için kullanılan kod

Şekil 6.5'deki kodda yer alan n-gram yöntemi, bir dizilim içindeki tekrar sayısını bulmaya yarayan, bir dizi verinin ardışık sıralı n elemanlı alt kümelerini bulunduran ve metin sınıflandırmada sıklıkla kullanılan bir yöntemdir (Sel, 2022). Çalışmada n-gram yönteminin uni-gram (tekli) ve bi-gram (ikili) metotları kullanılmıştır. Ayrıca, olumlu ve olumsuz yorumlarda sıkça tekrarlanan kelime dizilerinin bulunmasında da faydalı olacağı düşünülmüştür. Uni-gram metodunda her bir kelimenin metinde varlığı ya da yokluğu modellenirken; bi-gram metodunda birbiri ardına gelen ikili kelime gruplarının metindeki varlığı ya da yokluğu modellenebilmektedir (Onan, 2017). Örneğin; cümlemiz "Oda çok temizdi." olsun. "Uni-gram" kümesi {"oda", "çok", "temizdi"}, "bi-gram" kümesi ise {["oda", "çok"], ["oda", "temizdi"], ["çok", "temizdi"]} olur. Kelimeler arasındaki korelasyonları kaçırmamak ve vektör uzayının fazla büyümesinden kaynaklanabilecek öğrenme sorunları sebebiyle vektörizasyon sürecinde "uni-gram" ve "bi-gram" modelleri seçilmiştir. Doküman frenkası bütün dokümanın yüzde onundan az olan terimler "min_df" parametresi 0.01 seçilerek vektorizasyon işlemine alınmamıştır. Bu sayede az kullanılan kelime kökleri elenmiştir.

Bu işlemler sayesinde kelimelerin metin içindeki önemleri gösterilmiş olup, verilerin analize hazır hale gelmesi için gerçekleştirilmesi gereken ön hazırlık süreci tamamlanmıştır.

6.4. Verilerin Analizi ve Çalışma Modelinin Tanımlanması

Çalışmada kullanılan model temelde iki aşamadan oluşmaktadır: Birinci aşamada, "Google Places" sitesinden alınan müşteri yorumları makine öğrenmesi ve derin öğrenme kapsamında duygu analizleri ile incelenmiştir. İkinci aşamada ise Koronavirüs pandemisi öncesi-sonrası müşteri değerlendirmeleri analizlere tabi tutulmuş ve bu sayede pandeminin bungalov taleplerindeki değişimlerini belirlemek amaçlanmıştır. Araştırmanın modeli Şekil 6.6'da gösterilmiştir.



Şekil 6.6. Çalışmanın modeli

Çalışmanın modelini gösteren Şekil 6.6'da veri toplanmasından duygu analizinin gerçekleştirilmesine kadar olan verilerin ön işlenmesi, algoritmaların kullanılması, 10 katlı çapraz doğrulama uygulaması ve yöntemlerin nasıl bir sıralamayla analizde kullanıldığı süreçleri açıklanmıştır. Ayrıca çalışma sonucunda elde edilen bulgular da yer almaktadır.

6.4.1. Yorumların duygu analizi ile incelenmesi

Yorumların duygu analizi ile incelenmesi için ilk olarak pozitif ve negatif yorumlar polarite değerleri ile 0 (negatif); 1 (pozitif) olarak sayısallaştırılmış ardından veri seti eğitim ve test verisi olarak ayrılmıştır. Bu makine öğrenmesi ve derin öğrenme temelli duygu analizlerinin başlangıç prosedürdür. Ardından duygu analizinin diğer adımlarına geçilmiştir.

Nihai veri setinin eğitim ve test verisine ayrılması

Metin madenciliği çalışmalarında, veri setinin eğitim ve test veri seti olarak ayrıştırılması gerekmektedir. Çalışmada veri setinin %80'i eğitim, %20'si test veri seti olarak ayrılmıştır. %80-%20 seçiminde Scikit-learn tarafından önerilen bölme oranları esas alınmıştır (Ayan ve diğerileri, 2019). Veri setindeki dağılma ve parçalanmalardan ileri gelen hata ve gürültülerin asgari konuma getirilmesi amacıyla 10 katlamalı çapraz doğrulama metodundan (10-fold cross validation) yararlanılmıştır. Bu yöntemle analize konu olan veri seti 10 eşit parçaya bölünmüş, programın çalıştırılmasından sonra bu 10 eşit parçadan her biri her seferinde test verisi olarak seçilmek suretiyle sürekli değişen %20'lik bir kesim test veri seti olarak kullanılacak şekilde, eğitim veri seti (yani kalan %80'ıık kısım) 10 sefer çapraz doğrulama ile test edilmiştir.

Yorumların makine öğrenmesi algoritmaları ile analizi

Çalışmada duygu analizi için denetimli makine öğrenmesi yöntemi kullanılmıştır. Denetimli makine öğrenmesi yönteminin tercih edilme nedeni, yorumlardan elde edilen duyguların doğrulanması için konaklayanlar tarafından yorumlar yapılırken elde edilen rating değerlerinin kullanılmasıdır (Tuna ve diğerleri, 2021). Denetimli makine öğrenmesi yöntemlerinden; Naive Bayes, Destek Vektör Makineları ve Rastgele Orman algoritmalarından yararlanılmıştır. Söz konusu algoritmalar, literatürde çevrimiçi paylaşımları üzerinden yapılan veri analizlerinde daha fazla kullanılmaları, başarı oranlarının daha yüksek olması ve ikili sınıflandırmada kullanılmalarının yanı sıra çoklu sınıflandırma problemlerine de kolaylıkla uygulanabilmeleri nedeniyle seçilmiştir (Ekim ve İnner, 2021; Can ve Alataş, 2017). Belirtilen algoritmalar kullanılarak veri setleri eğitilmiş ve başarıları belirlenen ölçüm kriterleriyle karşılaştırılmıştır.

Naive Bayes Algoritması için; “nltk” kütüphanesinin “NaiveBayesClassifier” sınıfından faydalanılmıştır. Bu sınıf her kelime kökünün koşullu olasılıklarını göz önünde tutarak Naïve Bayes sınıflandırması yapar (NTLK, t.y). EK-3’te Naïve Bayes Algoritmasına ait kodlar sunulmaktadır. EK-3’te yer alan kodlar, 10 katlamalı çapraz doğrulama metodu ile eğitilip test edilmiştir ve sonuçları CSV formatında raporlanmıştır. Başlangıç modeli olarak kabul edildiği için sınıflandırma algoritması basit tutulmuştur.

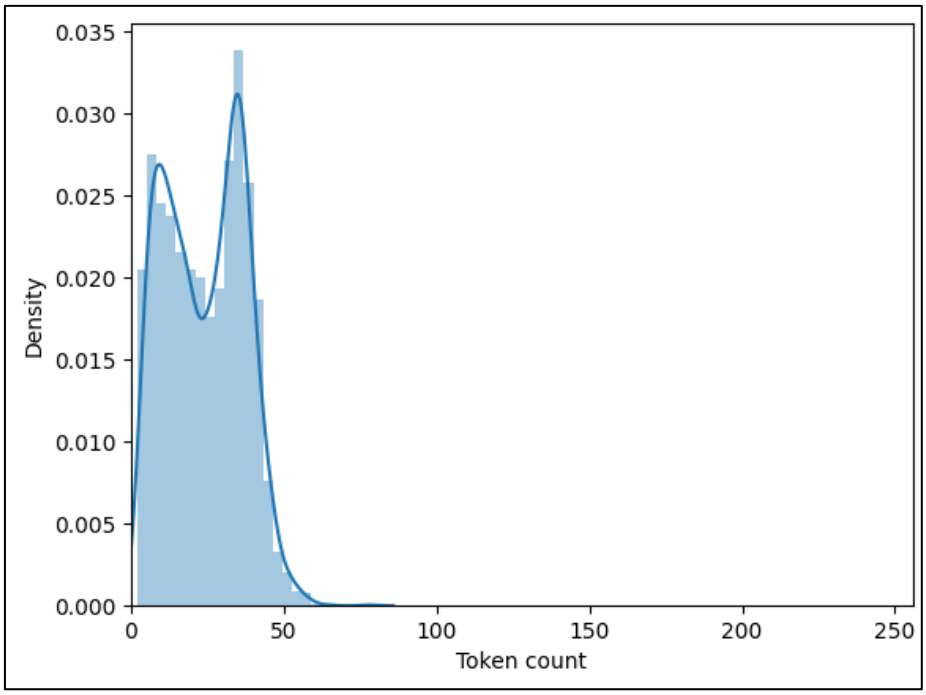
Destek Vektör Makineleri Algoritması için; “sklearn.svm” kütüphanesinin “SVC” sınıfı kullanılmıştır. “SVC” bir c-destek uzay sınıflandırması modelidir. Buradaki “c” değeri düzenlileştirme (regularization) parametresi olarak adlandırılır. Bu parametre değeri; büyüdükçe uzayda tanımlı hedef değerlerin yanlış sınıflandırmasına yol açabilir, küçüldükçe modelin aşırı öğrenmeye sebep olabilir (Tharwat, 2019). Bu yüzden c değerinin optimazsyonu için “sklearn.model_selection” kütüphanesinin “GridSearchCV” sınıfı kullanılmıştır. Bu sınıf k-katlamalı çapraz doğrulama tekniğinin üzerine modelde kullanılan hiperparametrelerin önceden belirlenmiş değerlerin kombinasyonları ile çapraz doğrulamasını yapar ve en uygun modelin seçilmesinde yardımcı olur (https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html). Burada “c” değeri için 0.1, 0.01 ve 0.001 değerleri 5 katlamalı çapraz doğrulama tekniği ile kullanılmıştır. Algoritmaya ait kodlar EK-4’te yer almaktadır. EK-4’te verilen kodlar, 10 farklı rastgele durum değerler için çalıştırılmış ve çapraz doğrulamadan çıkan metrikler raporlanmıştır.

Rastgele Orman Algoritması için; “sklearn.ensemble” kütüphanesinin “RandomForestClassifier” sınıfı kullanılmıştır. Destek Vektör Makinelerinde olduğu gibi hiperparametre optimizasyonu için “GridSearchCV” sınıfından yararlanılmıştır. Algoritmaya ait kodlar EK-5’te verilmektedir. EK-5’teki kodda yer alan; “n_estimators” parametresi karar ağaçlarının sayısını belirler. Çok sayıda karar ağacı, modelin aşırı öğrenme sorununu azaltarak daha iyi bir performans sergilemesini sağlar (Oshiro, Perez ve Baranauskas, 2012). “max_depth” parametresi ise, her bir karar ağacının derinliğini belirler. Karar ağaçlarının derinliği modelin eğitim verilerine uyma eğilimini artırır. Ayrıca “max_depth=None” olarak belirtilmesi, her bir karar ağacının olabildiğince derin olacak şekilde oluşturulmasını ve modelin esnek olmasını sağlar.

Yorumların derin öğrenme algoritması ile analizi

Konaklama sektörlerine ilişkin çevrimiçi yorumların duygu analizi ile incelenmesi çalışmalarında genel olarak makine öğrenmesi yöntemlerinin birbirleriyle karşılaştırılması ele alınmış olup derin öğrenme yöntemine rastlanmamıştır. Bu nedenle çalışmada, geçmiş araştırmalardan farklı olarak BERT derin öğrenme yöntemi kullanılmıştır.

Çalışmada, Turkish OSCAR corpus sözlüğü ile eğitilmiş olan MDZ Digital Library takımının açık kaynak olarak verdiği BERT-Base-Turkish-Uncased modeli kullanılmıştır. Vikipedi makalelerinden oluşturulan Turkish OSCAR corpus sözlüğü, 35 GB büyüklüğündedir ve 44 milyar küsür belirteçten (token) oluşmaktadır. Bu modelin tercih edilme sebebi; geçmiş çalışmalarda benzer sınıflandırma problemleri içerisinde en yüksek doğruluk oranlarına bu model ile ulaşılmış olmasından kaynaklanmaktadır. Model her kullanıcı yorumu için; yorum pozitif duyguya sahipse 1, negatif duyguya sahipse 0 çıktısını verecek şekilde etiketlenmiştir. Modele erişebilmek için, önceden eğitilmiş Transformer modellerine erişim için “from_pretrained” fonksiyonu sağlayan “transformer” kütüphanesi kullanılarak “dbmdz/bert-base-turkish-uncased” modeline erişilmiştir. BERT modeli için Transformers kütüphanesinden BertTokenizer sınıfı kullanılarak önceden eğitilmiş BERT modeli çekilmiştir. “BertTokenizer” sınıfı, BERT için önceden eğitilmiş bir kelime dağarcığını kullanarak metni belirteçlere (tokens) dönüştürür. Kelime kökleri belirteç haline getirildikten sonra modele beslenecek maksimum belirteç dizisi uzunluğu histogram grafiği kullanılarak belirlenmiştir. Histogram grafiği Şekil 6.7’de paylaşılmaktadır.



Şekil 6.7. Belirteç dizisi uzunluğu histogram grafiği

Şekil 6.7'de maksimum belirteç uzunluğunun 80 olduğu görülmeketedir. 80'den az belirtece sahip yorumlara 80'e tamamlanana kadar PAD belirteçleri eklenir ve bu eklenenler "0" olarak temsil edilmektedir. Veri seti sınıfı, "torch" kütüphanesinin "Dataset" sınıfından oluşturulmuştur. Bu sınıf; yorumlardaki kelime köklerini, yorumların duygu sınıfını, belirteç dönüştürme sınıfını ve Şekil 6.7' de verilen maksimum belirteç uzunluğunu argüman olarak almıştır. Bu sınıftan oluşturulan veri kümesi nesnesi; eğitim ve test sırasında belirteç dizilerini, dikkat mekanizmasında kullanılan kelime pozisyonlarını ve hedef duyguları sağlamak için kullanılır.

Veri kümesinin eğitim, doğrulama ve test süreçleri sırasında modele yüklenebilmesi için "torch" kütüphanesinden türetilen "veri yükleyici (data loader)" sınıfı oluşturulmuştur. Eğitim, doğrulama ve test veri testleri ayırmak için "sklearn" kütüphanesinin "train_test_split" metodu kullanılmıştır. Veri setinin %80'i eğitim, %10'u validasyon ve %10'u test için ayırılmıştır. Son olarak eğitim sırasında kullanılacak kümelerin büyüklüğü, eğitim sırasında daha iyi tutarlılık yakalamak için mini-batch standartlarından (32,64,128,256) daha küçük tutularak 16 olarak belirlenmiştir. Çalışmada pozitif-negatif duygulara sahip yorumlar analiz edildiği için sınıf sayısı iki olacak şekilde "torch" kütüphanesinin "Module" sınıfından oluşturulmuştur. Modelde aşırı öğrenmeyi önlemek için bir "drop-out" katmanı eklenmiştir. Son olarak, BERT modelinin çıktılarından sınıflandırma

yapmak için bir lineer katman oluşturulmuştur. Kodlara ilişkin veriler EK-6'da yer almaktadır.

Eğitim tur sayısı (epoch) 3 olarak belirlenmiştir. Optimizasyon için "Adam Optimizer" kullanılmıştır. "Adam Optimizer, derin sinir ağlarını eğitmek için özel olarak dizayn edilmiş, uyarlanabilir bir öğrenme hızı (learning rate) optimizasyon algoritmasıdır ve bu algoritma sayesinde daha az aşırı öğrenme problemi yaşanmaktadır" (Özkan ve Kar, 2022). Öğrenme oranı, derin öğrenme kütüphanelerinin tavsiyesi olan 0.001 değeri olarak seçilmiştir. Kayıp fonksiyonu için ikili sınıflandırma problemlerinde standard olan "Binary Cross-Entropy Loss Function" seçilmiştir. Her döngüde eğitim için çağıralacak "train_epoch" metodu yazılmıştır.

Öğrenme metodları tanımlandıktan sonra model döngüsü çalıştırılmıştır. Her döngüde eğitim ve doğrulama doğruluk değerleri saklanmış ve en iyi validasyon doğruluğuna sahip modelin durum parametleri "torch" kütüphanesinin "save" metodu kullanılarak kaydedilmiştir.

Algoritmalara ait kodlar çalıştırıldıktan sonra karmaşıklık matrisleri ve bu matrislerden elde edilen başarı performansları Bulgular bölümünde detaylıca yer almaktadır.

6.5. Bulgular

Sınıflandırma algoritmalarının performans karşılaştırılması için; doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F-1 puanı ve ROC Eğrisi Altındaki Alan (AUC) kriterleri kullanılmaktadır (Polat ve Ağca, 2022). Tahminlerin doğruluğu hakkında bilgi veren, tahmin edilen ve gerçek değerlerle elde edilen karmaşıklık matrisi (confusion matrix) ilgili kriterlerin bulunmasında kullanılır (Tuna, 2020). Çalışmada gerçekleştirilen makine öğrenmesi ve derin öğrenme algoritmaları sonucunda elde edilen her bir algoritmaya ait karmaşıklık matrisleri Çizelge 6.1'da gösterilmektedir. Algoritmalara ait karmaşıklık matrislerinin program çıktıları sırasıyla; EK-7, EK-8, EK-9 ve EK-10'da yer almaktadır.

Çizelge 6.1. Her algoritma için elde edilen karmaşıklık matrisleri

Naive Bayes	Pozitif	Negatif	Toplam
Pozitif	262	41	303
Negatif	19	251	270
Toplam	281	292	573
Destek Vektör Makineleri	Pozitif	Negatif	Toplam
Pozitif	274	26	300
Negatif	38	235	273
Toplam	312	261	573
Rastgele Orman	Pozitif	Negatif	Toplam
Pozitif	267	32	299
Negatif	43	231	274
Toplam	310	263	573
BERT	Pozitif	Negatif	Toplam
Pozitif	142	12	154
Negatif	18	114	132
Toplam	160	126	286

Çizelge 6.1'de TP ve TN değerlerinin yüksek, FN ve FP değerlerinin düşük olduğu dolayısıyla sınıflandırıcının tahmin yeteneğinin gerçek değerler ile uyumlu olduğu gözlemlenmiştir.

Yukarıda yer alan karmaşıklık matrisindeki veriler kullanılarak; doğruluk (accuracy), kesinlik (precision), duyarlılık (recall), F-1 puanı ve ROC eğrisi altında kalan alan(AUROC) kriter değerleri hesaplanmıştır. Bu değerler Çizelge 6.2'de verilmektedir**.**

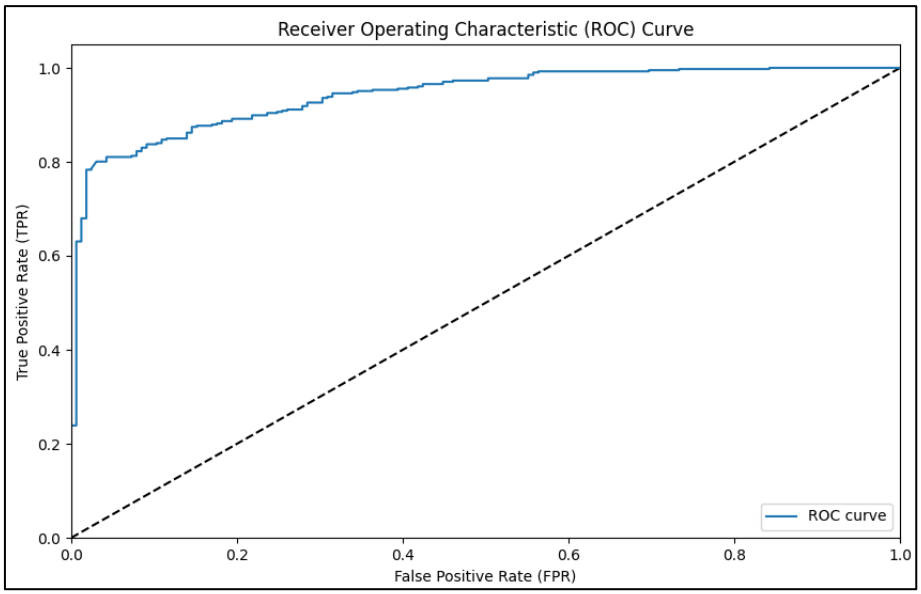
Çizelge 6.2. Duygu analizi sonucunda elde edilen performans değerleri

	Doğruluk (Accuracy)	Kesinlik (Precision)	Duyarlılık (Recall)	F-1 Ölçümü(F Measure)	ROC eğrisi altındaki alan (Area under the curve)
Naive Bayes	0,8954	0,9321	0,8653	0,8972	0,8974
Destek Vektör Makineleri	0,8882	0,8779	0,91308	0,8950	0,94606
Rastgele Orman	0,8693	0,8612	0,8939	0,87708	0,9366
BERT	0,8951	0,8875	0,9220	0,9044	0,9800

Duygu analizinde kullanılan algoritmalardan elde edilen sonuçların performans değerleri Çizelge 6.2'de verilmiştir. Daha önce bahsedilen beş kritere göre (doğruluk, kesinlik, duyarlılık, F-1 puanı ve ROC eğrisi altında kalan alan) hesaplanan performans değerleri, birbirine yakın ve %85'in üzerinde çıkmıştır. Buradan hareketle, mevcut çalışmada uygulanan algoritmaların duygu analizini hesaplamada başarılı olduğu düşünülebilmektedir.

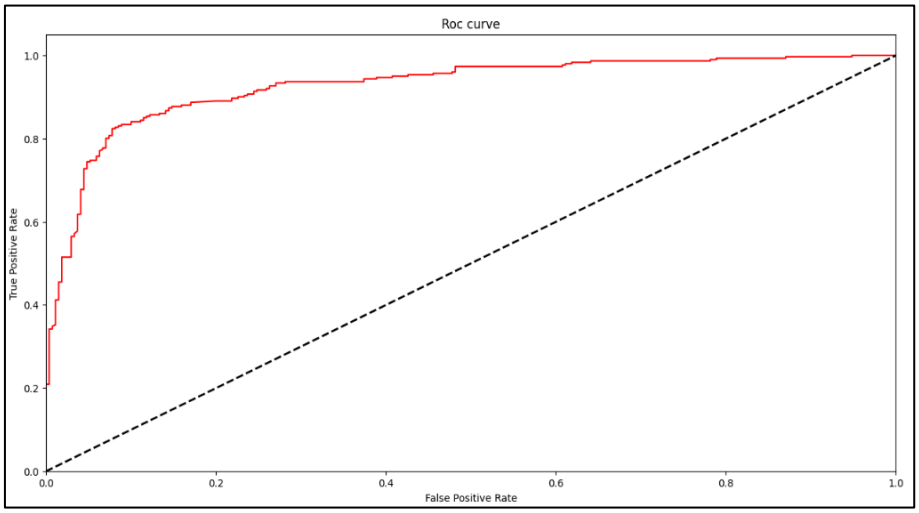
Algoritmaların sınıflandırma başarı karşılaştırmak için Çizelge 6.2'deki değerler incelendiğinde;

- Doğruluk oranı yüksekliği açısından Naive Bayes algoritmasının diğer algoritmalara göre üstün olduğu fakat BERT derin öğrenme algoritmasının doğruluk oranının da Naive Bayes yöntemine yakın olduğu görülmektedir.
- Kesinlik kriteri açısından Naive Bayes algoritmasının, duyarlılık ve kesinlik-duyarlılık kriterlerinin optimizasyonu (F1 ölçümü) kriterleri açısından BERT algoritmasının daha başarılı sonuçlar verdiği görülmektedir.
- Algoritmaların sınıflandırma performanslarının başka bir kritere göre karşılaştırılması için ROC eğrileri kullanılmış ve ROC eğrisi altında kalan alanda (AUROC) da en iyi sonucu BERT algoritması vermiştir. Ayrıca ilgili algoritmalara ait elde edilen ROC eğrileri sırasıyla, Naive Bayes; Şekil 6.8, Destek Vektör Makineleri; Şekil 6.9, Rastgele Orman; Şekil 6.10 ve BERT; Şekil 6.11'de yer almaktadır.



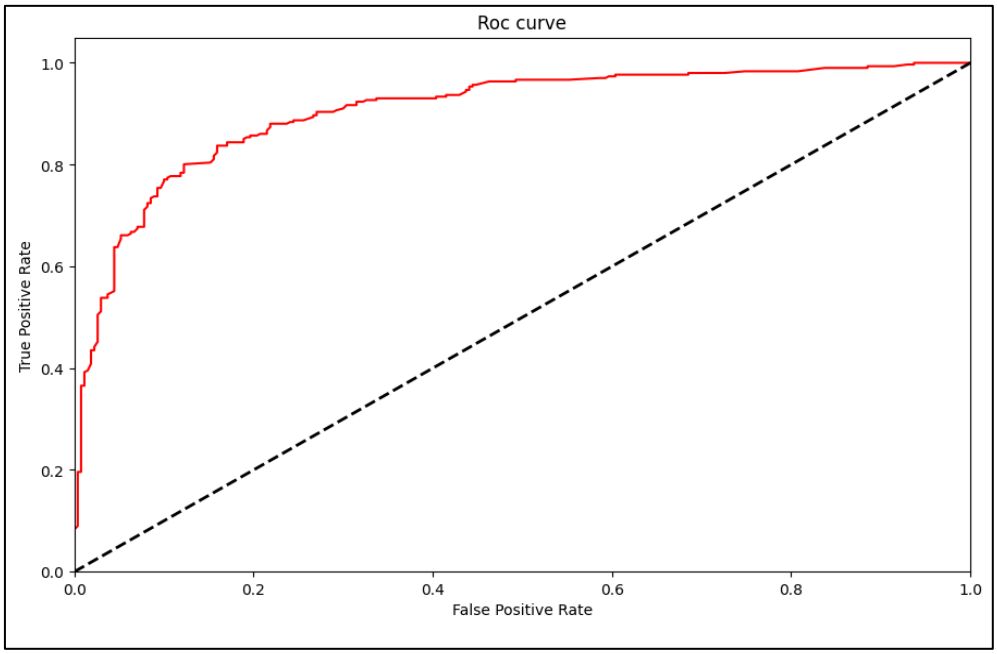
Şekil 6.8. Naive Bayes algoritmasına ait ROC eğrisi

ROC eğrisi, duyarlılık (True Positive Rate (Doğru Pozitif Oranı) - TPR) ile özgüllük (1 - False Positive Rate (Yanlış Pozitif Oranı) - FPR) arasındaki ilişkiyi göstermektedir. TPR, doğru pozitifleri (gerçek pozitiflerin toplam sayısı) toplam pozitiflerin (gerçek pozitiflerin ve yanlış negatiflerin toplam sayısı) oranı olarak hesaplarken; FPR ise yanlış pozitifleri (gerçek negatiflerin ve yanlış pozitiflerin toplam sayısı) toplam negatiflerin (gerçek negatiflerin ve yanlış pozitiflerin toplam sayısı) oranı olarak hesaplamaktadır. Eğri, sol üst köşeye ne kadar yakınsa, modelin o kadar iyi performans gösterdiğini göstermektedir (İlhan, Yıldırak ve Aladağ, 2023). AUROC değeri 0 ile 1 arasında bir değer alır. 1'e ne kadar yakınsa, modelin o kadar iyi bir performansa (mükemmel sınıflandırıcı eğrisi) sahip olduğunu söyleyebiliriz. Şekil 6.8'de algoritmanın dik yükselen bir eğriye sahip olduğu görülmektedir. Bu da söz konusu algoritmanın düşük bir hata oranıyla yüksek bir duyarlılık ve özgüllük sağladığını göstermektedir.



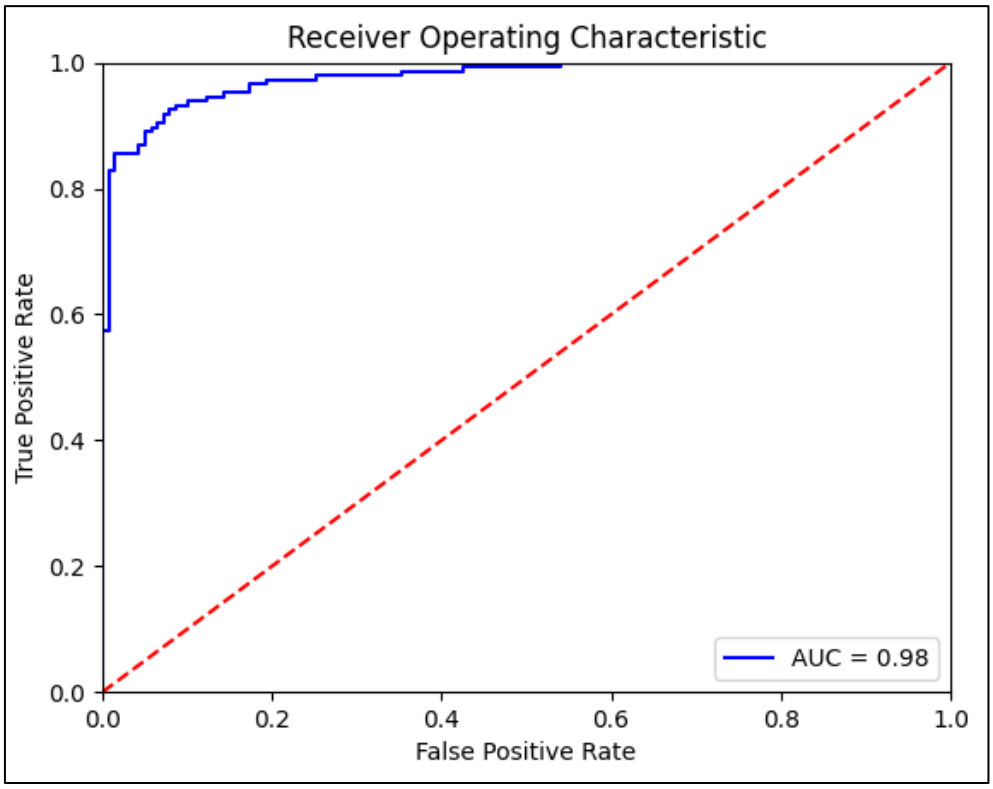
Şekil 6.9. Destek vektör Makineleri algoritmasına ait ROC eğrisi

Şekil 6.9'da Şekil 6.8'e göre eğri altında kalan alan arttığı ve daha yüksek duyarlılığa sahip olduğu görülmektedir.



Şekil 6.10. Rastgele orman algoritmasına ait ROC eğrisi

Şekil 6.10 da Şekil 6.9'a benzer bir eğriye sahiptir. İki algoritmaya ait AUROC değerleri de birbirine çok yakındır.



Şekil 6.11. BERT algoritmasına ait ROC eğrisi

Şekil 6.11'de BERT algoritmasının diğer algoritmalara kıyasla en yüksek duyarlılık ve özgüllüğe sahip olduğu, ROC eğrisinin mükemmel sınıflandırıcı eğrisine yakın olduğu görülmektedir.

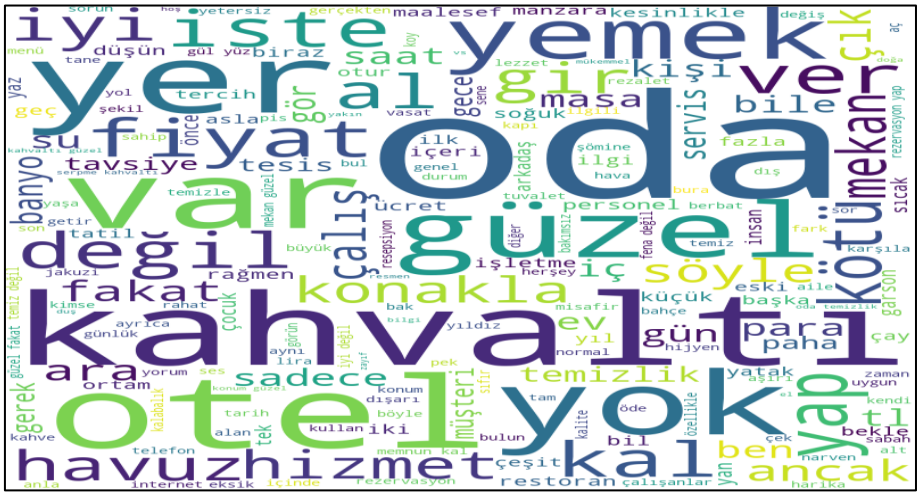
Algoritmaların sınıflandırma başarısı ROC eğrilerinin altında kalan alanın büyüklüğüne göre belirlenmiştir. Bütün eğriler rastgele seçim çizgisinin epey üzerinde olup sınıflandırma uyumunda aşırı öğrenme veya eksik öğrenme davranışı görülmemiştir. Elde edilen başarı oranları doğrultusunda, bungalov işletme sahipleri müşterilerden gelen geribildirimleri değerlendirmek için bütün yorumları tek tek okumak yerine Naive Bayes ve BERT algoritmalarını kullanarak zamandan, iş gücünden ve maliyetten tasarruf sağlayıp daha doğru bir değerlendirme yapabileceklerdir. Ayrıca söz konusu algoritmaları kullanılması, manuel değerlendirmede yorumların gözden kaçması ve yorumlarda gizli olan esas duyguların açığa çıkarılamaması gibi dezavantajları da ortadan kaldıracaktır.

Buraya kadar olan analizler, müşterilerin aldıkları bungalov işletmeleri hizmetlerine ilişkin yorumlarının duygu durumlarını ortaya koymada metinlerin duygu analizi ile incelenmesini içermektedir. Müşterilerin çalışmaya dahil edilen işletmelerde en beğendikleri ve beğenmedikleri yönlerini tespit etmek için kelime bulutları çıkarılmıştır. "Kelime bulutları, incelenen dokümanlar dahilindeki benzersiz kelimelerin frekansına bağlı olarak kompozisyonun genel temasına ait anahtar sözcüklerin görselleştirilmesi olarak tanımlanmaktadır" (Dikkaya, 2022). Olumlu yorumlara ait kelime bulutları Şekil 6.12'de; olumsuz yorumlara ait kelime bulutları Şekil 6.13'de yer almaktadır.



Şekil 6.12. Olumlu yorumlara ait kelime bulutu

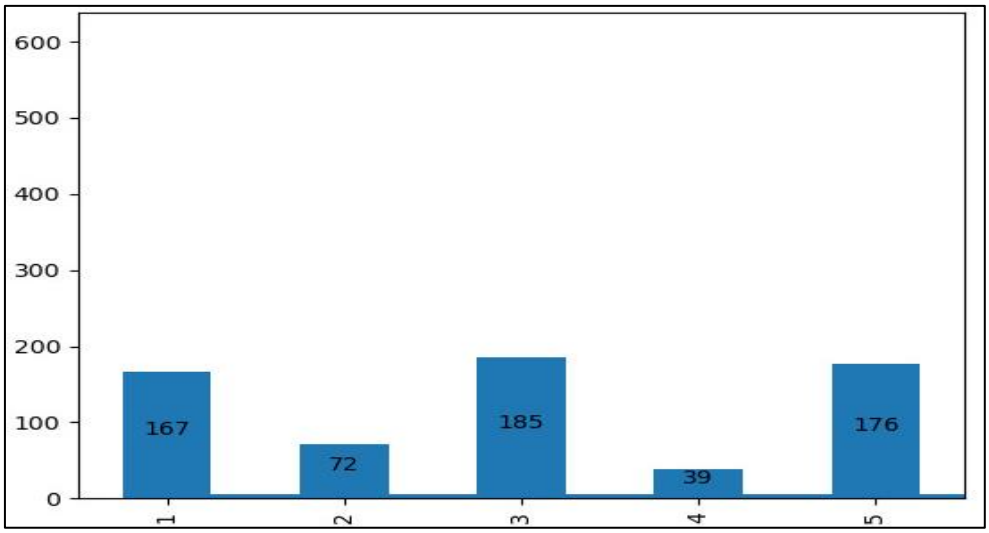
Şekil 6.12'de olumlu yorumlar için oluşturulan kelime bulutunda "kahvaltı", "yemek", "temiz, temizlik", "güler yüz, ilgili", "hizmet", ortam", "konum" , "manzara" ve "mekan" gibi kelimelerin ön plana çıktığı görülmektedir. Bu kelimeler doğrultusunda bungalovlarda konaklayanların; bungalovların temizliğinden, personelin ilgi-alakası ve güler yüzlülüğünden, bungalovların göl –doğa manzaralı konumlarından, başta kahvaltı olmak üzere yemeklerinden ve sıcak-huzurlu ortamından memnun kaldıkları değerlendirilmektedir.



Şekil 6.13. Olumsuz yorumlara ait kelime bulutu

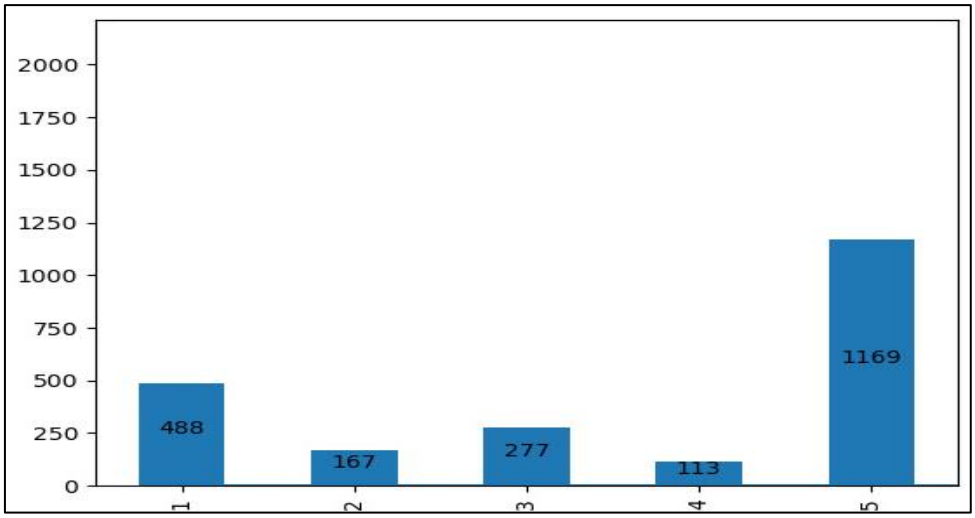
Şekil 6.13'de olumsuz yorumlar için oluşturulan kelime bulutunda en çok göze çarpan kelimelerin "fiyat", "ücret", "para", "pahalı", "ödeme", "küçük", "bahçe", "havuz", "banyo" ve "oda" olduğu görünmektedir. Buradan hareketle konaklayanların, fiyat-performans ilişkisi ve bungalovların iç-dış mekân boyutları konularında olumsuz görüş bildirdiği görülmektedir. Ayrıca olumlu yorumlarda yer alan "kahvaltı", "hizmet" ve "yemek" kelimeleri olumsuz yorumlarda da dikkat çekmektedir.

Çalışmada Covid-19 pandemisinin bu işletmeler üzerindeki etkisini tespit edebilmek için, çekilen yorumlar 2020 yılı esas alınarak pandemi öncesi ve sonrası olarak iki gruba ayrılmıştır. Şekil 6.14 ve Şekil 6.15'te bu gruplara ait yorum sayıları rating değerleriyle birlikte gösterilmektedir.



Şekil 6.14. Pandemi öncesi yorum sayıları

Şekil 6.14'de yer alan grafikte pandemi öncesi toplam yorum sayısı 639'dur. Yorumların yaklaşık %66'sı 1-2 ve 3 rating değerlerine sahiptir.



Şekil 6.15. Pandemi sonrası yorum sayıları

Pandemi sonrası toplam yorum sayısı 2214 olup rating değeri 4 ve 5 olan yorumlar yaklaşık %58'dir. Rating değerlerindeki dağılımın, pandemi öncesinde daha dengede iken pandemi sonrası dengesiz bir şekilde en yüksek rating olan 5'e doğru kaydığı görülmektedir. Dolayısıyla pandemi öncesi olumsuz yorumlar daha fazla iken pandemi sonrası olumlu yorum sayısının artışı dikkat çekmektedir.

7. SONUÇ VE ÖNERİLER

Bu çalışmada Google Places üzerinden elde edilen bungalov müşteri yorumlarını içeren bir veri setine makine öğrenmesi algoritmalarından (Naive Bayes, Destek Vektör Makineleri ve Rastgele Orman) ve derin öğrenme algoritmasından (BERT) yararlanılarak duygu analizi yöntemi uygulanmıştır.

Çalışmada ilk olarak yorum rating değerlerinin duygusal kutupsallığı (pozitif ve negatif) çıkarılmıştır. Buradan çıkarılan kutupsallık etiketleri sınıflandırma tahmini için (Naive Bayes, Destek Vektör Makineleri, Rastgele Orman ve BERT) kullanılmıştır. Kullanılan algoritmaların başarı performansları sırasıyla; (doğruluk kriterine göre) Naive Bayes (%89,54), BERT (%89,51) Destek Vektör Makineleri (%88,82) ve Rastgele Orman (%86,93) şeklinde elde edilmiştir. Çalışmada kullanılan algoritmaların ortalama başarısının %88,7 hesaplanması ile bu algoritmaların benzer çalışmalarda uygulanabileceği anlaşılmıştır. Literatür kısmında, konaklama sektörü içerisindeki otel işletmelerine yönelik müşteri yorumlarının makine öğrenmesi algoritmalarının birbirleriyle karşılaştırılması yapılarak analizlerinin gerçekleştirildiği gözlemlenmiştir. Bu çalışmada diğerlerinden farklı olarak bungalov işletmelerine yönelik müşteri yorumlarının, makine öğrenmesi algoritmasıyla birlikte derin öğrenme algoritması kullanılarak analizi gerçekleştirilmiştir. . Analiz sonucunda, geleneksel makine öğrenmesi yöntemleri ile derin öğrenme yönteminin birbirine yakın başarı oranları elde etmesi, sınıflandırmada derin öğrenme algoritmalarının da başarı ile uygulanabileceğini göstermiştir. Bu da araştırmacılara, BERT ve benzeri derin öğrenme yöntemlerini kullanarak metin verilerini analiz etmeleri ve daha iyi sonuçlar elde etmeleri için bir fırsat sunmaktadır.

Yorumlara yönelik yapılan analizler sonucunda, konaklayanlar tarafından en fazla olumlu yorumun “odaların temiz olması”, “personelin güler yüzlü, samimi tutum ve davranışları”, “kahvaltı” ve “bungalovların göl- orman manzaralı konumu” hususlarında yapıldığı; en fazla ifade edilen olumsuz yorumun ise “fiyatlandırma ” problemi üzerine yapıldığı gözlenmiştir. Olumlu yorumlarda değinilen “kahvaltı” ve “temizlik” hususları olumsuz yorumlarda da saptanmıştır. Bungalov işletmecilerinin, özellikle temizlik, fiyat-performans, işletme çalışanları ve işletmenin konumu hususlarında daha titiz davranarak hizmet kalitesini, müşteri memnuniyetini ve bungalovlara olan talebi artırabilecekleri değerlendirilmektedir.

Çalışma kapsamında sadece Türkiye'de Sapanca bölgesinde yer alan bungalovlara ait Türkçe yorumlar incelenmiştir. Bu çalışmanın bungalov işletmeleri için bir başlangıç noktası olarak kullanılabileceği ve diğer işletmeler için de faydalı olacağı düşünülmektedir.

Yapılan çalışmada ayrıca Covid-19 pandemisinin bungalov işletmelerine olan etkisi de incelenmiştir. Yorumlar 2020 yılı öncesi ve sonrası olarak iki gruba ayrılmış olup pandemi sonrası yorum sayısının pandemi öncesine göre yaklaşık 2,5 kat arttığı tespit edilmiştir. Aynı zamanda yorumların rating değerlerinde de pozitife doğru bir yönelim olduğu görülmüştür. Bu veriler ışığında, insanlara bireysel tatil imkânı sağlayan, farklı deneyimler yaşayabilecekleri ve otel konforuna sahip bungalov işletmelerine olan talebin pandemiyle birlikte arttığı değerlendirilmektedir. Bu durumun Türkiye'de bungalov işletmelerinin bilinirliğine, başka bölgelerde yayılmasına, doğayla iç içe olmasından dolayı sürdürülebilir turizm anlayışına önemli katkı sağlayacağı düşünülmektedir.

İleriki araştırmalarda veri seti genişletilerek, rating eşik değeri ile ikili sınıfa çevirme yerine farklı yöntemler araştırılıp geliştirilebilir. Geleneksel makine öğrenme algoritmaları halen duygu analizi problemlerinde kabul gören tekniklerden olsa da veri setindeki duygu ve kelime frekansları dağılımları daha düzenli dağılıma sahip ve cümle içindeki kelimelerin birbirlerine bağımlılıkları arttığında farklı derin öğrenme metodları tercih edilebilir. Ayrıca derin öğrenme metodlarının dinamik öğrenme yapısı sayesinde öğrenme turları arttırılıp, model üzerinde de ince ayar yapıldıktan sonra daha yüksek doğruluk elde edilebilir.

İleriki çalışmalarda diğer dillerdeki yorumların incelenmesi yabancı turistlerin olumlu-olumsuz gördüğü hususları anlamak açısından faydalı olacaktır. Ayrıca Türkiye'de farklı bölgelerde yer alan bungalov işletmelerine yönelik araştırmalar yapılarak sonuçlar karşılaştırılabilir ve Türkiye'deki bungalov işletmeleri hakkında genel bir tablo çıkarılabilir. Makine öğrenmesi algoritmaları ile farklı derin öğrenme algoritmaları kullanılarak analizler gerçekleştirilebilir.

KAYNAKLAR

Acar, A. ve Uğur, İ. (2021). Uluslararası Zincir Otellere Yönelik Tripadvisor Yorumlarının Duygu Analizi Yöntemi ile Değerlendirilmesi: Ankara Örneği. *Türk Turizm Araştırmaları Dergisi*, 5(3), 1803-1814

Adalı, E. (2012). Doğal Dil İşleme. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5(2), 1-19.

Ağdeniz, Ş. (2017). *Finansal Raporların Analizinde Metin Madenciliğinin Kullanımı: Borsa İstanbul Şirketlerinin Kurumsal Yönetim Niteliklerinin Tahmini*. Doktora Tezi, Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Enstitüsü, Eskişehir, 32-38.

Ahlemeyer-Stubbe, A. and Coleman, S. (2014). *A practical guide to data mining for business and ındustry*. West Sussex: John Wiley and Sons, 17.

Ahmetoğlu, H. ve Daş, R. (2020). Türkçe Otel Yorumlarıyla Egitilen Kelime Vektörü Modellerinin Duygu Analizi ile İncelenmesi, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 24(2), 455- 463

Ak, S. and Dinçer, M. Z. (2019). Investigation of Consumer Reviews on Social Media for Five-Star Hotel Enterprises in Istanbul: A Case of TripAdvisor. *Journal of Tourismology*, 5(2), 171-183.

Akgöz, E. (2009). *Kriz ortamında turizm işletmelerinin rekabet aracı olarak itibar yönetimi.* Modern Developmental Trends and Turkish World Conference, İstanbul, 158- 180.

Akgün, D. and Ünal, İ. (2021). Glamping İşletmelerine İlişkin Tripadvisor'daki Çevrim içi Yorumların İçerik Analizi ile Değerlendirilmesi. *Journal of Gastronomy, Hospitality and Travel (Online)*, 4(2), 199-214.

Akın, Ö. (2019). *Otel işletmeciliği*, Sağlık, E. & Karaman, A. (Edt). Otel İşletmeciliği, Erzurum: Atatürk Üniversitesi Açıköğretim Fakültesi Yayını, 20-22.

Akman, M. (2010). *Veri Madenciliğine Genel Bakış ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama*. Yüksek Lisans Tezi, Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Ankara, 21-36.

Aksaya, H. (2021). *Sözlük Tabanlı Yaklaşım ve Makine Öğrenmesi ile Türkçe Doğal Dili İşleme: Eğitim Kurumlarında Duygu Analizi*. Doktora Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 38-42.

Aksöz, E. O., Önem, B., Aydın, B.(2020). Kamp deneyimi yaşayan bireylerin glamping'e ilişkin algılarının belirlenmesi. Gastroia: *Journal of Gastronomy and Travel Research*, 4(2), 185-196.

Aksu, M. (2014). *Genel turizm*. İstanbul: Lisans Yayıncılık, 90.

Aksu, M. Ç. (2021). *Turizm için sosyal medya temelli mekansal karar destek sisteminin geliştirilmesi ve uygulanması: Artvin ili örneği*, Doktora Tezi, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü, Erzurum, 85-93.

Akyiğit, H. E. ve Taşcı, T. (2021). Sigortacılık sektöründe makine öğrenmesi ile müşteri kaybı analizi. *Tasarım Mimarlık ve Mühendislik Dergisi*, 2(1), 66-79.

Alan, A. (2020). *Makine öğrenmesi sınıflandırma yöntemlerinde performans metrikleri ile test tekniklerinin farklı veri setleri üzerinde değerlendirilmesi*. Yüksek Lisans Tezi, Fırat Üniversitesi Fen Bilimleri Enstitüsü, Elazığ, 16-22.

Alan, A. ve Karatabak, M. (2020), Veri Seti - Sınıflandırma İlişkisinde Performansa Etki Eden Faktörlerin Değerlendirilmesi, *Fırat Üniversitesi Mühendislik Bilimleri Dergisi* 32(2), 531-540.

Albayrak, A. (2018). *Duygu analizinde farklı vektör temsil yöntemleri ve sınıflayıcıların karşılaştırılması*. Yüksek Lisans Tezi, Sivas Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü, Sivas, 22-36.

Alqaraleh, S. (2020). Turkish Sentiment Analysis System via Ensemble Learning, *European Journal of Science and Technology Special Issue*, 122-129

Alrawadieh, Z. and Law, R. (2019). Determinants of Hotel Guests Satisfaction from the Perspective of Online Hotel Reviewers. *International Journal of Culture Tourism and Hospitality Research,* 13(1), 84-97.

Altıntaş, M. (2014). *Kullanıcı destek sistemlerinde yardım biletlerinin otomatik sınıflandırılması.* Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 103-105.

Amolik, A., Jivane, N., Bhandari, M andVenkatesan, M. (2015). Twitter sentiment analysis of movie reviews using machine learning techniques. *International Journal of Engineering and Technology*, 7(6), 2038-2044.

An, T. (2018). *Guest Satisfaction in New Zealand Luxury Hotels: A Study of Reviews from TripAdvisor*, Master Thesis, Auckland University of Technology, New Zeland, 45-66.

Arda, U. K. (2015). *Enez yöresinin turizm işletmeleri açısından gelişme perspektifleri*. Yüksek Lisans Tezi, Trakya Üniversitesi Sosyal Bilimler Enstitüsü, Edirne, 185-196.

Arkadaş, A. R. ve Ayyıldız, T. (2020). Uludağ'da faaliyet gösteren 4 ve 5 yıldızlı konaklama işletmelerinin çevrimiçi tüketici yorumlarının incelenmesi: TripAdvisor.com üzerinden içerik analizi. *OPUS International Journal of Society Researches*, 16(30), 2633-2658.

Arpacı, Ö., Uğurlu, K. ve Batman, O. (2015). Helal Konseptli Otel Müşterilerine Yönelik Yapılan Müşteri Şikayetleri Üzerine Bir Araştırma. *Bartın Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 6(11), 181-198.

Arslan, Ö. (2021). *Evrişimsel sinir ağları ve metin benzerliği kullanılarak fatura görüntülerinde sınıflandırma.* Yüksek Lisans Tezi, Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü, Konya, 32-38.

Atalay, M. ve Çelik, Ç. (2017). Büyük veri analizinde yapay zekâ ve makine öğrenmesi uygulamaları. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155-172.

Ayan, B., Kuyumcu, B. and Ciylan, B. (2019). Twitter Üzerindeki İslamofobik Twitlerin Duygu Analizi ile Tespiti. *Gazi University Journal of Science Part C*, 7(2): 495-502.

Azaltun, M. ve Kaya, E. (2010). *Muhasebe uygulamaları: konaklama işletmelerinde.* Ankara: Detay Yayıncılık, 3.

Bagnera, S. (2017). A (Diamond) Cut Above the Rest: Improving Hotel Operations Based on TripAdvisor Rating Attributes. *Boston Hospitality Review*, 5(3), 1-16.

Bahar, O. (2006). Turizm Sektörünün Türkiye'nin Ekonomik Büyümesi Üzerindeki Etkisi: VAR Analizi Yaklasımı. *Yönetim ve Ekonomi*, 137-150.

Bahar, O. ve İlal, N.Ç. (2020). Coronavirüsün (Covid-19) Turizm Sektörü Üzerindeki Ekonomik Etkileri. *International Journal of Social Sciences and Education Research*, 125-139.

Baid, P., Gupta, A. and Chaplot, N. (2017). Sentiment Analysis of Movie Reviews using Machine Learning Techniques, *International Journal of Computer Applications*, 179(7), 45-49.

Baker, S., Huyton, J. and Bradley P. (2000). Principles of Hotel Front Office Operations. *London: Second Edition*. 1,325.

Bala, S. (2013). *Bootstrap örnekleme yönteminin x-r süreç kontrol diyagramlarında kullanımı,* Yüksek Lisans Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 22-32.

Balaban, M. E. ve Kartal, E. (2015). *Veri madenciliği ve makine öğrenmesi temel algoritmaları ve r dili ile uygulamaları.* İstanbul: Çağlayan Kitabevi, 43.

Barutçugil, İ. (1989). *Turizm işletmeciliği.* İstanbul: Beta Basım Yayın, 54.

Batman, O. (2008). *Otel işletmelerinin yönetimi*. İstanbul: Değişim Yayınları, 49.

Bhatia, R. and Jain, A. (2019). *Text mining and its applications in business.* Big Data Analytics for Business In. Singapore: Springer, 49-67.

Bird, S., Wiles, J. L., Okalik, L., Kilabuk, J. and Egeland, G. M. (2009). Methodological consideration of story telling in qualitative research involving ındigenous peoples. *Global Health Promotion*, *16*(4), 16-26.

Bolasco, S., Canzonetti, A., Capo, F. M., Della Ratta-Rinaldi, F. and Singh, B.K. (2005). Understanding text mining: A pragmatic approach. In *Knowledge Mining*, Singapore: Springer, 31-50.

Bonifacic, J. C., Milohnic, I. and Cerovic, Z. (2017). Glamping – creative accommodation in camping resorts: Insights and opportunities. *Tourism in Southern and Eastern Europe*, 4, 101-114.

Boscoboinik, A. and Bourquard, E. (2011). Glamping and rural imagery. H. Horáková, & A. Boscoboinik (Ed.). *From Production To Consumption: Transformation of Rural Communities.* Zurich: LIT Verlag, 158.

Bowen, J. T. and Chen, S. L. (2001). The relationship between customer loyalty and customer satisfaction. *International Journal of Contemporary Hospitality Management,* 13(5), 213-217.

Boz, M., Canbazoğlu, E., Özen, Z. ve Gülseçen, S. (2018). Otel rezervasyon iptallerinin makine öğrenmesi yöntemleri ile tahmin edilmesi. *Veri Bilimi,* 1(1), 7-14.

Brochado, A. and Brochado, F. (2020). What makes a glamping experience great?. *Journal of Hospitality and Tourism Technology,* 10(1), 26-32.

Brooker, E. and Joppe, M. (2013), Trends in camping and outdoor hospitality—An international review, *Journal of Outdoor Recreation and Tourism,* 3(4), 1–6

Budak, İ. ve Kılıç, G. (2021). Muhafazakâr Otellerin Çevrimiçi Yorumlar ile Değerlendirilmesi. *İnternet Uygulamalar ve Yönetimi*, 12(1), 76-86.

Büyükeke A., Sökmen A. ve Gencer C. (2020). Metin Madenciliği ve Duygu Analizi Yöntemleri ile Sosyal Medya Verilerinden Rekabetçi Avantaj Elde Etme: Turizm Sektöründe Bir Araştırma. *Journal of Tourism and Gastronomy Studies*, 2020, 8 (1), 322-335

Can, U. ve Alataş, B. (2017). Duygu Analizi ve Fikir Madenciliği Algoritmalarının İncelenmesi. *International Journal of Pure and Applied Sciences,* 3(1): 75-111.

Cenni, I. and Goethals, P. (2017). Negative Hotel Reviews on TripAdvisor: A Cross-Linguistic Analysis. *Discourse, Context ve Media*, 16, 22-30.

Chakrabarti, S. (2002). *Mining the web: Analysis of hypertext and semi-structured data*. San Francisco: Morgan Kaufmann, 45-62.

Chakraborty, K., Bhattacharyya, S., Bag, R. and Hassanien, A.A. (2019). Sentiment analysis on a set of movie reviews using deep learning techniques. In *Social Network Analytics: Computational Research Methods and Techniques.* New York: Academic Press,127-147.

Chawai, A.İ.B. (2019): *Türkçe metinlerde sözlük tabanlı yaklaşımla duygu analizi.* Yüksek Lisans Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 85-92.

Chowdhary, K. (2020). Natural language processing. *Fundamentals of artificial intelligence*. New York: Springer Publishing, 603-649.

Craig, C. and Karabaş, İ. (2021). Glamping after the coronovirus pandemic. *Tourism and Hospitality Reseacrh,* 21(2), 251-256.

Çakmaklı, E. (2019). Glamping, Glamping Tipleri ve Özellikleri. Ondokuz Mayıs Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 8(2), 53-70.

Çelik, E., Dal, D. ve Aydın, T. (2021). Duygu Analizi İçin Veri Madenciliği Sınıflandırma Algoritmalarının Karşılaştırılması. *Avrupa Bilim ve Teknoloji Dergisi*, (27), 880-889.

Çelik, N., Bahar, O. ve Tatar, S. (2017). Kırsal Kalkınmada Glamping Turizmin Rolü: Club Amazon Bördübet Örneği. *Journal of International Social Research*, 10 (51), 1282-1287.

Çelikyay, E.K. (2010). *Metin madenciliği yöntemiyle Türkçede en sık kullanılan ve birbirini takip eden harflerin analizi ve birliktelik kuralları.* Yüksek Lisans Tezi, Beykent Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 112-115.

Çepur N. and Dalgıç A. (2022). Investigatıon of The Glamping Potential in Isparta and Its Surroundings. *Journal of Applied Tourism Research*, 3(1), 1-12.

Çeti, B. ve Ünlüönen, K. (2019). Salgın Hastalıklar Sebebiyle Oluşan Krizlerin Turizm Sektörü Üzerindeki Etkisinin Değerlendirilmesi. AHBVÜ Turizm Fakültesi Dergisi, 22 (2), 109-128.

Çınar, A. (2012). *Veri madenciliği ve makine öğrenimi temel kavramlar, algoritmalar, uygulamalar* içinde (105-140). İstanbul: Çağlayan Kitabevi, 92-103.

Çoban S., Çoban E. ve Yetgin D. (2019). Şehir Turizminde Faaliyet Gösteren Otellere Yönelik E-Yorumların İncelenmesi: Eskişehir İli Örneği. *Türk Turizm Araştırmaları Dergisi*, 3(3), 542-558.

Çoban, Ö. ve Özyer, G. T. (2018). Twitter duygu analizinde terim ağırlıklandırma yönteminin etkisi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 24(2), 283-291.

Çolakoğlu, O. (2014). Turistik ürün kavramına genel yaklaşım. A. Timur (Ed.). *Turistik Ürün Politikası*. Ankara: Detay Yayıncılık, 47-69.

Değer, N.S. (2017). *Sosyal medya mesajlarında veri madenciliği ile duygu analizi*. Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 17-22.

Delen D., Sharda R. and Turban E. (2014). *Business ıntelligence and analytics: System for decision support*. UK: Pearson Education Limited, 23.

Demir, M., Günaydın, Y. ve Demir, Ş. Ş. (2020). Koronavirüs (Covid-19) salgınının Türkiye'de turizm üzerindeki öncülleri etkileri ve sonuçlarının değerlendirilmesi. *International Journal of Social Sciences and Education Research*, 6(1), 80-107.

Devlin, J., Chang, M., Lee, K. and Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.* Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

Dikkaya, F. (2022), *Sosyal Medyada Gıda Sektörünün Netnografi ve Metin Madeciliği Yöntemi ile İncelenmesi: Torku Markasının Twitter Analizi*. Yüksek Lisans Tezi, Aksaray Üniversitesi Sosyal Bilimler Enstitüsü, Aksaray, 52-55.

Dinler, M. (2014). *Kümeleme analizi yöntemlerinin hayvancılık verilerinde karşılaştırmalı olarak incelenmesi.* Yüksek Lisans Tezi, Bingöl Üniversitesi Sosyal Bilimler Enstitüsü, Bingöl, 62-68.

Doğan, F. ve Türkoğlu, İ. (2019). Derin öğrenme algoritmalarının yaprak sınıflandırma başarımlarının karşılaştırılması. *Sakarya University Journal of Computer and Information Sciences,* 1(1), 10-21.

Doğan, S. (2017). Otellere Yönelik Yapılan Puanlama ve Çevrimiçi Yorumların Değerlendirilmesine İlişkin Aksaray İlinde Bir Araştırma. *Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* 9 (3) 27-36.

Doğru, E. ve Koçak, E. (2021). Covid-19 Salgınının Turistik Hizmet Sağlayan Konaklama İşletmelerine Etkisi (Fethiye Örneği*). Doğu Coğrafya Dergisi*, 26(45), 106-124.

Düven, B. (2021). *Yapay Öğrenme Yöntemleri ile Türkçe Metinlerde Duygu Analizi*. Yüksek Lisans Tezi, Maltepe Üniversitesi Lisansüstü Eğitim Enstitüsü, İstanbul, 85-92.

Düzgün, E. (2021). Pandemi sonrası yeni turist tercihi: glamping turizmi. *Ordu Üniversitesi Sosyal Bilimler Enstitüsü Sosyal Bilimler Araştırmaları Dergisi,* 11(1), 145-158.

Ekim, H.E. ve İnner, A.B. (2021). Duygu Analizi ve Fikir Madenciliği Uygulamları Üzerine Literatür Taraması, *KSÜ Mühendislik Bilimleri Dergisi*, 24(2), 93-114.

El Naqa, I. and Murphy, M. J. (2015). *What is machine learning? Machine learning in radiation oncology: theory and applications*. New York: Springer International Publishing, 3-11.

Elmas, S.Ş. (2019). *Sosyal medya mesajlarının veri madenciliği yöntemi ile duygu analizi (Sivas ili örneği).* Yüksek Lisans Tezi, Sivas Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü, Sivas, 102-105.

Ercan, F. (2019). Sosyal medyada otel işletmelerine yönelik yorumların müşteri memnuniyeti ve memnuniyetsizliği açısından analizi. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 22(2), 552-571.

Erdem, B. (2004). Otel İşletmelerinde İnsan Kaynakları Planlamasının Yeri ve Önemi. *Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,* 7 (11), 35-54.

Ergün, K. (2012). *Metin madenciliği yöntemleri ile ürün yorumlarının otomatik değerlendirilmesi*. Yüksek Lisans Tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya, 45-62.

Ergüven, M. H., Yılmaz, A. ve Kutlu, D. (2015). Turistik Ürün Çeşitlendirme Bağlamında Hibrit Turizm: Glamping Örneği, *International Journal of Social Science*, 41(2), 256 257.

Eröz, S. S. (2011). *Duygusal zekâ ve iletişim arasındaki ilişki: bir uygulama*. Doktora Tezi, Uludağ Üniversitesi Sosyal Bilimler Enstitüsü, Bursa, 12-15.

Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56(4), 82-89.

Fettahlıoğlu, Ö. O., Polat, M. ve Demir, S. (2016). Hizmet kalitesinin servqual analizi ile ölçümü: Kahramanmaraş ilinde bir uygulama. *Ordu Üniversitesi Sosyal Bilimler Enstitüsü Sosyal Bilimler Araştırmaları Dergisi,* 6(3), 849-860.

Filipe, S., Santos, N. and Barbosa, F. (2018). Glamping: A new niche in the tourist industry. *Tourism ve Management Studies,* 14(1), 11-20.

Fushiki, T. (2011). Estimation of prediction error by using k-fold cross-validation. In *Statistics and Computing.* New York: Springer, 138-142.

Gaikwad, S. V., Chaugule, A. and Patil, P. (2014). Text mining methods and techniques. *International Journal of Computer Applications,* 85(17), 42-45.

Gandomi, A. and Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management,* 35(2), 137-144.

Gerenaz, H. ve Yetgin, D. (2021). SPA Otel İşletmelerinin Çevrimiçi Müşteri Değerlendirmeleri: Tripadvisor Örneği. *İşletme Araştırmaları Dergisi,* 13(3), 2800-2818

Glaveli, N., Manolitzas, P., Palamas, S., Liassidis, C. and Grigoroudis, E. (2022). Breaking the Curse: Extracting Strategic Directions for Hotel Industry in Wuhan from TripAdvisor Quantitative Rating. *Anatolia,* 2, 275-279.

Gloe, T. (2016). Luxury camping–A new way to go on holiday: An analysis of the glamping market. *Journal of Tourism and Hospitality Management,* 4(1), 1-11.

Göçen, S., Yirik, Ş. ve Yılmaz, Y. (2011). Türkiye'de Krizler ve Krizlerin Turizm Sektörüne Etkileri. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 493-509.

Göktaş, L. S., Çetin, G. ve Kızılırmak, İ. (2017). Özel ilgi turizmi olarak glamping turizminin mevcut durumu: bir örnek olay analizi araştırması. *4. Disiplinlerarası Turizm Araştırmaları Kongresi Bildiri Kitabı*, İstanbul: Anatolia Turizm Akademisi, 107, 120.

Güngör, O., Güngör, M. and Doğan, S. (2018). Comparative Study About 4 and 5 Star Hotels in Antalya and Barcelona According to Tourist Ratings at Holidaycheck Web Site. *Journal of Travel and Hospitality Management,* 15(1), 74-88.

Güven, A. (2007). *Türkçe belgelerin anlam tabanlı yöntemlerle madenciliği*. Yüksek Lisans Tezi, Yıldız Teknik Ünivesitesi Fen Bilimleri Enstitüsü, İstanbul, 85-93.

Güvenol, B. and Sarıbaş, S. K. (2022). Konaklama sektöründe sürdürülebilirlik: Glamping işletmeleri örneği (Sustainability). *Journal of Tourism and Gastronomy Studies*, 10(3), 2369-2393.

Haltaş, A., Alkan, A. ve Karabulut, M. (2015). Metin sınıflandırmada sezgisel arama algoritmalarının performans analizi. *Journal of the Faculty of Engineering and Architecture of Gazi University,* 30(3), 419-425.

Han, J., Kamber, M. and Pei, J. (2012). *Data mining: concepts and techniques.* Massachusetts: Elsevier, 596.

Hazar, A. (2010). *Genel turizm.* Ankara: Nobel Yayın Dağıtım, 97.

He, W., Zha, S. and Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza ındustry. *International Journal of Information Management*, 33(3), 464-472.

İlhan, Z., Yıldırak, K. ve Aladağ, Ç.H. (2023). *Rastgele Orman Yöntemine Genel Bir Bakış, Teori ve Uygulamalarla Nicel Karar Verme*. Ankara: İKSAD, 56-66.

İnan, H.E. (2019), *Otel Yorumlarının Sınıflandırılmasında Makine Öğrenmesi Algoritmalarının Karşılaştırılması ve Eskişehir Otellerine Yönelik Yorumların Duygu Analizi*. 20. Ulusal Turizm Kongresi, Eskişehir, 45-52.

İnik Ö. ve Ülker E. (2017). Derin öğrenme ve görüntü analizinde kullanılan derin öğrenme modelleri. *Gaziosmanpaşa Bilimsel Araştırma Dergisi*, 6(3), 85-104.

İnternet: 101ayn (2023). *Marmara bölgesinde 10 büyüleyici kamp alanı.* Web: https: //101ayn.com/marmara- bolgesinde-10-buyuleyici-kamp-alani/ Son Erişim Tarihi: 14.05.2023.

İnternet: Agoda (2023). *Agoda,* www.agoda.com. Son Erişim Tarihi: 14.05.2023

İnternet: Clubamazon (2023). Web: https: //clubamazon.com.tr/ Son Erişim Tarihi: 9.05.2023.

İnternet: Glamping (2021). *What is glamping?* Web: https://glamping.com/what-is-glamping/, Son Erişim Tarihi: 06/03/2023.

İnternet: Hızır Kamp (2023). Web: http: //www.hizirkamp.com/index.html Son Erişim Tarihi: 10.05.2023.

İnternet: Kamp Yeri (2023). El Baba Camping. Web: https: //kampyeri.org/kamp-alani/el-baba-camping Son Erişim Tarihi: 8.05.2023.

İnternet: Kültür Bakanlığı (2023). Web: https://teftis.ktb.gov.tr/, Son Erişim Tarihi: 16/04/2023.

İnternet: Livera Camping (2023). Web: http: //www.liveracamping.com/ Son Erişim Tarihi: 10 2.05.2023

İnternet: Miraç Öztürk (2023). Web: http://miracozturk.com/python-ile-siniflandirma-analizleri-rastgele-orman-random-forest-algoritmasi, Son Erişim Tarihi: 16/02/2023.

İnternet: NLTK (2023). *Naivebayes,* Web: https://www.nltk.org/_modules/nltk/classify/naivebayes.html, Son Erişim Tarihi: 16/04/2023.

İnternet: NTV (2023). www.ntv.com.tr Son Erişim Tarihi: 17.05.2023

İnternet: QWE-Wiki (2023). Web: https: //tr.qwe.wiki/wiki/Field_of_the_Cloth_of_Gold Son Erişim Tarihi: 14.05.2023.

İnternet: Scikitlearn (2023a). Web: https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html Son Erişim Tarihi: 11.05.2023.

İnternet: Scikitlearn (2023b). Web: https://scikitlearn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html Son Erişim Tarihi: 11.05.2023.

İnternet: Selenium (2023). Web: https://selenium-python.readthedocs.io/ Son Erişim Tarihi: 13.05.2023.

İnternet: T.C. Sağlık Bakanlığı. (2020). *Türkiye Günlük Koronavirüs Tablosu*. Web: https://covid19.saglik.gov.tr/ / Son erişim tarihi: 13.05.2023.

İnternet: T.C. Ulaştırma ve Altyapı Bakanlığı Devlet Hava Meydanları İşletmesi Genel Müdürlüğü. (2022). *Devlet Hava Meydanları İşletmesi Genel Müdürlüğü* Web: https://www.dhmi.gov.tr/Sayfalar/Istatistikler.aspx, Son Erişim Tarihi: 03/05/2022.

İnternet: TUİK (2019). *Turizm istatistikleri*. Web: https://data.tuik.gov.tr/Bulten/Index?p=Turizm-Istatistikleri-IV.Ceyrek:-Ekim-Aralik-ve-Yillik,-2019-33669, Son Erişim Tarihi: 12.05.2023.

İnternet: Turkish Culture (2023). Web: www.turkishculture.org, Son Erişim Tarihi: 19.05.2023

İnternet: TÜRSAB (2021). *Turizmin ekonomideki yeri*. Web: https://www.tursab.org.tr/istatistikler/turizmin-ekonomideki-yeri, Son Erişim Tarihi: 17/03/2022.

İnternet: UNWTO (2023). Web: https://www.unwto.org/impact-assessment-of-the-covid-19-outbreak-on-internationaltourism, Son Erişim Tarihi: 13/03/2023.

İnternet: Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I. (2017). *Attention Is All You Need*, Web: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf, Son Erişim Tarihi: 16/03/2023.

Kafa, N. ve Demircioğlu, E. (2022). Gökçeada'daki konaklama işletmelerine yönelik çevrimiçi müşteri değerlendirmelerinin incelenmesi: Tripadvisor örneği. *Journal of Recreation and Tourism Research*, 9(2), 52-69.

Kale, B. (2018). *Veri Madenciliği Sınıflandırma Algoritmaları ile e-posta Önemliliğinin Belirlenmesi*. Yüksek Lisans Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana, 112-115.

Kalıpçı, M. B. (2021). *Konaklama işletmeleri satın alma müdürlerinin teknoloji kabul ve hizmet inovasyonu algılarının değerlendirilmesinde öğrenen örgütün aracılık rolü: Antalya örneği*. Doktora Tezi, Alanya Alaaddin Keykubat Üniversitesi Lisansüstü Eğitim Enstitüsü, Alanya, 185-192.

Kambur, E. ve Aksoy, M. (2022). 1998-2018 Döneminde Türkiye'de turizme yapılan yatırım ile ekonomik büyüme ilişkisinin granger nedensellik ve Toda-Yamamoto yöntemleri ile analizi. *Fırat Üniversitesi Sosyal Bilimler Dergisi,* 32(3), 1099-1112.

Kandampully, J. and Suhartanto, D. (2003). Customer loyalty in the hotel industry: the role of customer satisfaction and image. *International Journal of Contemporary Hospitality Management*, 15(2), 94-104.

Karagölge, C. ve Peker, K. (2011). Tarim ekonomisi araştırmalarında tabakalı örnekleme yönteminin kullanılması/using stratified sampling methods in the resource of agricultural economics. *Atatürk Üniversitesi Ziraat Fakültesi Dergisi,* 33(3), 313-316.

Karahan, K. (2000). *Hizmet pazarlaması.* İstanbul: Beta Yayınevi, 116.

Karakoyun, M. ve Hacibeyoğlu, M. (2014). Biyomedikal veri kümeleri ile makine öğrenmesi siniflandirma algoritmalarinin istatistiksel olarak karşilaştirilmasi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, 16(48), 30-42.

Karaöz, B. (2018). *Büyük veri ve işletme analitiği: Sosyal medya ve duygu analizi ile bir öngörü modeli*. Doktora Tezi, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 45-52.

Kartal E. (2015). *Sınıflandırmaya dayalı makine öğrenmesi teknikleri ve kardiyolojik risk değerlendirmesine ilişkin bir uygulama*. Yüksek Lisans Tezi, İstanbul Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 85-93.

Katarzyna, W. and Tuchowski, J. (2013). *Sentiment analysis of opinions about hotels extracted from the Internet*, Knowledge, Economy, Society. Global and Regional Challenges of the 21st Century Economy, Piteşti, 12-18.

Kaygısız A. (2021). Covid-19 sonrası türk turizm sektöründe oluşabilecek fırsatlar ve riskler. *Bartın Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 12, 23-35.

Kaynar, O., Yıldız, M., Görmez, Y. ve Albayrak, A. (2016). *Makine Öğrenmesi Yöntemleri ile Duygu Analizi*, International Artificial Intelligence and Data ProCessing Symposium (IDAP'16), Malatya, 1-18.

Kemaloğlu, N., Küçüksille, E.U. and Özgünsür M.E. (2021). Turkish Sentiment Analysis on Social Media. *Sakarya University Journal of Science* 25(3), 629-638

Kılıç, S. (2015). Kappa testi. *Journal of Mood Disorders*, 5, 142–144.

Kına, E. (2022). *Makine öğrenmesi algoritmaları kullanılarak twıtter mobil oyun verilerinde duygu analizi*. Doktora Tezi, Van Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü, Van, 45-63.

Kıvılcım, B. (2020). Covıd-19 (yeni koronavirüs) salgınının turizm sektörüne muhtemel etkileri. *Uluslararası Batı Karadeniz Sosyal ve Beşeri Bilimler Dergisi*, 4(1), 17-27.

Kızılırmak, İ., Alrawadieh, Z. ve Aghayeva, S. (2015). Doğu Karadeniz Bölgesindeki Otel Misafirlerinin Seyahat E-Yorumlarına Yönelik Bir İçerik Çözümlemesi. *Gümüşhane Üniversitesi Sosyal Bilimler Elektronik Dergisi*, 6(14), 110-121.

Kızılkaya, Y. M. (2018). *Duygu Analizi ve Sosyal Medya Alanında Uygulama*. Yüksek Lisans Tezi, Uludağ Üniversitesi Sosyal Bilimler Enstitüsü, Bursa, 17-23.

Kim, H. B. and Kim, W. G. (2005). The relationship between customer satisfaction and loyalty in the Korean lodging industry. *Journal of Hospitality and Tourism Research*, 29(4), 448-469.

Korjula, E. (2017). *Glamping-liiketoiminta Suomessa,* Bachelor's Thesis, Saimaa University of Applied Sciences Faculty of Tourism and Hospitality, Lappeenranta,52-66.

Korkusuz, R. (2019). *Futbola İlişkin Twitter Paylaşımlarının Duygu Analizi*. Yüksek Lisans Tezi, Trakya Üniversitesi Fen Bilimleri Enstitüsü, Tekirdağ, 45-52.

Kozak, N., Kozak M.A. ve Kozak, M. (2014). *Genel turizm, ilkeler, kavramlar*. Ankara: Detay Yayıncılık, 50-57.

Kozak, N., Kozak, M.A. ve Kozak, M. (2020*). Genel Turizm İlkeler ve İnkılaplar*. Ankara: Detay Yayıncılık, 18-23.

Köktürk, M.S. ve Dirsehan, T. (2012). *Veri madenciliği ile pazarlama etkileşimi*. Ankara: Nobel Akademik Yayıncılık, 7-10.

Kuhzady, S. and Ghasemi, V. (2019). Factors Influencing Customers' Satisfaction and Dissatisfaction with Hotels: A Text-Mining Approach. *Tourism Analysis*, 24(1), 69-79.

Kumar A. and Paul A. (2016) *Mastering Text Mining with R: Master text-taming techniques and build effective text-processing applications with R*. Birmingham: Packt Publishing, 45-56.

Kumar A. and Sebastian T. M. (2012). Sentiment analysis: a perspective on its past present and future. *Intelligent Systems and Applications*, 10, 1-14.

Kumral, C. D., Topal, A., Ersoy, M., Çolak, R. ve Yiğit, T. (2022). Random Forest Algoritmasının FPGA Üzerinde Gerçekleştirilerek Performans Analizinin Yapılması, *El-Cezerî Fen ve Mühendislik Dergisi*, 9(4), 1315-1327.

Kutlugün, M. A. and Şirin, Y. (2018). *Turkish meaningful text generation with class based n-gram model*. 2018 26th Signal Processing and Communications Applications Conference, New York, 1-4.

Larose, D.T. (2005). *Discovering knowledge in data an ıntroduction to data mining*. New York: Wiley, 16.

Leelawat, N., Jariyapongpaiboon, S., Promjun, A., Boonyarak, S., Saengtabtim, K., Laosunthara, A., Yudha, A.K. and Tang, J. (2022). Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning, *Heliyon* 8(10), 10894.

Litvin, S. W., Goldsmith, R. E. and Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458-468.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures On Human Language Technologies*, 5(1), 1-167.

Liu, Y., Sun, J., Cui, J. and He, H. (2022). Deep learning-based methods for natural hazard named entity recognition, *Scientific Reports* 12(1), 459.

Lu, J., Wang, X., Dai, Z., Chen, G. and Feng, Y. (2021). Antecedents of customer WOM in glamping: The critical role of original ecology, *International Journal of Hospitality Management*, 95, 102919

Mariani, M. M. and Borghi, M. (2018). Effects of the Booking.com Rating System: Bringing Hotel Class into the Picture. *Tourism Management*, 66(1), 47-52

Medhat, W., Hassan, A. and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113.

Mermer, A. (2019). *Veri Madenciği Yaklaşımı ile Proje Başvuru Yapma Eğiliminin Tahmini. Yüksek* Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 18-22.

Miner, G. Delen, D., Fast, A., Hill, T., Elder, J. and Nisbet, B. (2012). *Practical text mining and statistical analysis for non-structured text data applications*, Waltham, MA: Academic Press, 32-35.

Mitra, A., Paul, S. and Singrodia V. (2019). *A Review on Web Scrapping and its Applications*, International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, 1-22.

Mohammad, S. M. and Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational Intelligence*, 29(3), 436-465.

Nasukawa, T. and Yi, J. (2003). *Sentiment analysis: capturing favorability using natural language processing*, Proceedings of the 2nd International Conference on Knowledge Capture, Sanibel Island, 23.

Nowacki, M. and Niezgoda, A. (2019). Comparison of Poznań, Wrocław and Bratislava image attributes in the reviews published on tripadvisor. *Studia Periegetica*, 26, 63-77.

Oğuz, B. (2009). *Metin madenciliği teknikleri kullanılarak kulak burun boğaz hasta bilgi formlarının analizi.* Yüksek Lisans Tezi, Akdeniz Üniversitesi Sağlık Bilimleri Enstitüsü, Antalya, 22-36.

Oğuzlar, A. (2011). *Temel metin madenciliği*. Bursa: Dora Yayınları, 15.

Oğuzlar, A. ve Kızılkaya, Y. M. (2019). *Metin madenciliğinde duygu analizi: r uygulamalı*. Bursa: Dora Basım-Yayım, 24

Okuyucu, A. ve Akgiş, Ö. (2016). Türkiye'de Konaklama Sektörünün Yapısal ve Mekansal Değişimi: 1990-2013. *Türkiye Sosyal Araştırmalar Dergisi*, 20(1), 249-269.

Olcay, A. ve Turhan, U. (2017). Türkiye'de glamping hizmeti veren işletmelerin sahip ve yöneticilerinin yeni turizm trendi glamping ile ilgili görüşleri (Muğla ve Antalya örneği). *Gaziantep University Journal of Social Sciences,* 16(4), 980-996.

Onan, A. (2017). Sentiment Analysis on Twitter Messages Based on Machine Learning Methods. *Yönetim Bilişim Sistemleri Dergisi*, 3(2), 1-14.

Oshiro, T.M., Perez, P.S. and Baranauskas, J.A. (2012). *How Many Trees in a Random Forest?*. 8th International Conference on Machine Learning and Data Mining in Pattern Recognition. Toronto, 16-35.

Oxford, R. M. and Daniel, L. G. (2001). Basic cross-validation: using the holdout method to assess the generalizability of results. *Research in the Schools*, 8(1), 83-89.

Ozan, Ş., Özdil, U., Taşar, D.E., Arslan, B. ve Polat, G. (2021). BERT Modeli'nin Sınıflandırma Doğruluğunun Sıfır-Atış Öğrenmesi ile Artırılması. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 14(2), 99.

Öncel, S. Y. ve Erduğan, F. (2015). Kontenjans tablolarının analizinde log-lineer modellerin kullanımı ve sigara bağımlılığı üzerine bir uygulama. *Sakarya Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 19(2), 222-235.

Özdemir, A. (2011). *Doğrusal olmayan regresyonda asimptotik yöntemle bootstrap örneklemesi*. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 22-36.

Özen, A. (2019). *Müşteri Şikâyetlerinde Fikir Madenciliği: Antalya Helal Oteller Üzerine Bir Araştırma*. 2 nd Internatıonal Halal Tourism Congress, Antalya, 16-23.

Özer, İ. (2020). Uzun kısa dönem bellek ağlarını kullanarak erken aşama diyabet tahmini. *Mühendislik Bilimleri ve Araştırmaları Dergisi*, 2(2), 50-57.

Özkan, M. ve Kar, G. (2022). Türkçe Dilinde Yazılan Bilimsel Metinlerin Derin Öğrenme Tekniği Uygulanarak Çoklu Sınıflandırılması, *Mühendislik Bilimleri ve Tasarım Dergisi,* 10(2), 504-519.

Özkul, M. (2019). *Metin madenciliği yöntemi ile akıllı ulaşım sistemleri algısının duygu analizi ile ölçülmesi*. Yüksek Lisans Tezi, Bandırma Onyedi Eylül Üniversitesi Sosyal Bilimler Enstitüsü, Balıkesir, 22-45.

Özsarıoğlu, F. (2023). *Covid-19'un Türkiye Turizmine Etkilerinin Değerlendirilmesi (Analizi)*. Yüksek Lisans Tezi. Aydın Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü, Aydın, 36-45.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval, 2*(1-2), 1-135.

Paraskevas, A. and Arendell, B. (2007). Perceptions of service quality in the hospitality industry. *Journal of Hospitality ve Leisure Marketing*, 15(3-4), 49-66.

Parlak, İ. (2022). *Türkiye'deki Suriyeli sığınmacılar algısı: Türkçe twıtter verileri ile duygu analizi.* Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi Lisansüstü Eğitim Enstitüsü, Samsun, 17-20.

Pasin, E. (2018). *Investıgatıon of text mınıng methods on Turkısh text*. Yüksek Lisans Tezi, Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü, İzmir, 32-36.

Pereira, C. M. M. (2013). *Gamplinexp: Glamping experience*, Phd Thesis, ISCTE Business School Instituto Universitario de Lisboa, 11.

Peres, C. K. and Paladini, E. P. (2021). Exploring the Attributes of Hotel Service Quality in Florianópolis-SC, Brazil: An Analysis of Tripadvisor Reviews. *Cogent Business ve Management*, 8(1), 1-19.

Perikos, I. and Hatzilygeroudis, I. (2017), *Aspect based sentiment analysis in social media with classifier ensembles*, IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, 273-278.

Polat, H. ve Ağca, Y. (2022). Tripadvisor Kullanıcılarının Türkçe ve İngilizce Yorumları Kapsamında Duygu Analizi Yöntemlerinin Karşılaştırmalı Analizi, *Abant Sosyal Bilimler Dergisi*, 22(2): 901-916.

Poria, S., Hussain, A. and Cambria, E. (2018). *Multimodal sentiment analysis*. 1st Edition, New York: Springer International Publishing, 214.

Powell, J. and Svendsen L. (2004). *Bungalow details: Exterior*. London: Gibbs Smith Publisher, 36.

Raut, V.B. and Londhe, D.D. (2014). *Opinion Mining and Summarization of Hotel Reviews*, International Conference on Computational Intelligence and Communication Networks, Bhopal, 32-36.

Resmi Gazete (2019). *Turizm tesislerinin niteliklerine ilişkin yönetmelik*. Tarih: 01.06.2019, Sayı: 30791.

Saini, S. and Mohan Pandey, H. (2015). Review on web content mining techniques. *International Journal of Computer Applications*, 118(18), 33–36.

Sakáčová, K. L. (2013). *Glamping-nature served on silver platter*. Aalborg: Aalborg University, 1-10.

Sann, R. and Lai, P.-C. (2020). Do expectations towards thai hospitality differ? The views of english vs chinese speaking travelers. *International Journal of Culture Tourism and Hospitality Research*, 15(1), 43–58.

Santiago, M. and Campos, L.M. (2022). Discovering a tourism destination with social media data: BERTbased sentiment analysis. *Journal of Hospitality and Tourism Technology*, 13(5), 907-921.

Satapathy, R., Cambria, E. and Hussain, A. (2018). *Sentiment analysis in the bio-medical domain*. New York: Springer, 11-22.

Seker, S. E. (2016). Duygu analizi (sentimental analysis). *YBS Ansiklopedi*, 3(3), 21-36.

Sel, A. (2022). Pandemi Sürecinde Toplum Görüşünün Duygu Analizi Yöntemiyle İncelenmesi: Türkiye Örneği. *Beykoz Akademi Dergisi*, 10(2), 134-154

Serdarer, B. ve Giray, S. (2020). Destek Vektör Makineleri Yardımıyla İmalat Sanayisinde Mali Başarısızlık Tahminlerinin Teknoloji Yoğunluğuna Göre İncelenmesi. *Osmaniye Korkut Ata Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(2), 36-54.

Siddhartha, B.S. (2021). An Interpretation of Lemmatization and Stemming in Natural Language ProCessing. *Shanghai Ligong Daxue Xuebao/Journal of University of Shanghai for Science and Technology*, 22(10), 350-357.

Şahin, H., Çolakoğlu, O. E., Kılıçdere, S. ve Atay, H. (2020). Kuşadası'nda bulunan 5 yıldızlı otel işletmelerine yönelik yapılan sosyal medya yorumlarının incelenmesi. *Tourism and Recreation*, 2(1), 103-110.

Şahinaslan, Ö., Dalyan, H. ve Şahinaslan, E. (2022). Naive Bayes Sınıflandırıcısı Kullanılarak YouTube Verileri Üzerinden Çok Dilli Duygu Analizi. *Bilişim Teknolojileri Dergisi*, 15(2), 221-229.

Şener, B. (2009). *Konaklama işletmelerinde önbüro yönetimi*. Ankara: Detay Yayıncılık, 8.

Taboada, M., Brooke, J., Tofiloski, M., Kimberly, V. and Manfred, S. (2011). Lexicon-based methods for sentiment analysis. *Comput Linguist*, 37(2), 267–307.

Tamer, E.T. ve Ceylan, U. (2021). Kış Turizmi Kpasamında Faaliyet Gösteren Otellere Yönelik Çevrimiçi Yorumların İncelenmesi: Uludağ Örneği, *Journal of Gastronomy Hospitality and Travel (joghat)*, 4(1),62-74

Tang, Z. Q., Han, L. Y., Lin, H. H., Cui, J., Jia, J., Low, B. C. and Chen, Y. Z. (2007). Derivation of stable microarray cancer-differentiating signatures using consensus scoring of multiple random sampling and gene-ranking consistency evaluation. *Cancer Research*, 67(20), 9996-10003.

Taşkın, S.G. (2020*). Derin öğrenme algoritmaları ile Türkçe dilinde sahte haber tespiti.* Doktora Tezi, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü, Isparta, 17-22.

Tengilimoğlu, E. and Öztürk, Y. (2019). *Clusterıng of Text Mınıng Methods and Onlıne Comments: Case of Baku Hotels*. Conference: 5Th International Congress of Social Science Skopje, Macedonia, 16-22.

Tharwat, A. (2019). Parameter investigation of support vector machine classifier with kernel functions, *Knowledge and Information Systems*, 61(3), 1269-1302.

Topaçan Ü. (2016). *Sosyal medya paylaşımlarında duygu analizi: makine öğrenimi yaklaşımı üzerine bir araştırma*. Doktora Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 17-36.

Tuna M.F. (2020), *Çevrimiçi yorum ve şikâyetlerin otel işletmeleri üzerinden duygu analizi ile incelenmesi*. Doktora Tezi, Erciyes Üniversitesi Sosyal Bilimler Enstitüsü, Kayseri, 102-105.

Tuna, M.F., Kaynar, O. and Akdoğan, M.Ş. (2021). Otellere İlişkin Çevrimiçi Geribildirimlerin Makine Öğrenmesi Yöntemleriyle Duygu Analizi. *İşletme Araştırmaları Dergisi*, 13(3), 2232-2241

Tunalı, V. (2011). *Metin madenciliği için iyileştirilmiş bir kümeleme yapısının tasarımı ve uygulaması*. Doktora Tezi, Marmara Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 17-22.

Tuominen, P. (2011). *The Influence of TripAdvisor Consumer-Generated Travel Reviews on Hotel Performance*. 19th Annual Frontiers in Service Conference, Sweden, 17-18.

Uğurlu, K., Aslan, E. and İnce, Y. (2017). *Akdeniz bölgesinde faaliyet gösteren helâl konseptli otel işletmelerinin sosyal medya araçlarından olan facebook sayfalarının kullanım düzeylerinin incelenmesi.* 1. International Halal Tourism Congress, Kastamonu, 452-465.

Uluçeçen, T. H. (2011). *Özel ilgi turizmi: kapsamı, çeşitleri ve Türkiye'de uygulanabilirliği.* Uzmanlık Tezi, Kültür ve Turizm Bakanlığı Dış İlişkiler ve Avrupa Birliği Koordinasyon Daire Başkanlığı, Ankara, 17-36.

Uslu, O. (1990). *Turizm ve çevresel etkileri*, Turizm ve Çevre Konferansı, Ankara, 19-47.

Ünal, A. ve Bayar, S.B. (2021). Turizm işletmeleri. M. Bozkurt (Ed.). *Turizm İşletmelerinde İtibar Yönetimi.* Çanakkale: Paradigma Akademi, 71-110.

Vosoughi, S., Deb, R. and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.

Vres, G. and Vres, T. D. (2015). Garden Vıllage Bled Glampıng Kot Inovatıvna Revıtalızacıja Degradırane Krajıne/Garden Vıllage Bled Glampıng as an Innovatıve Revıtalızatıon of Degraded Landscape. *Arhitektura, Raziskave*, 3(2), 40.

Walter, M., Comino, S. and Reimann, T. (2019). Glamping: Camping, *Luxus, Nachhaltigkeit", in Conrady, R., Ruetz, D. and Aeber-hard,* M. (Eds.). Grundlagen und neue Perspektiven des Luxustourismus, Sprin-ger Fachmedien Wiesbaden, Wiesbaden, 313–323.

Wang, H., Lu, Y. and Zhai, C. (2010). *Latent aspect rating analysis on review text data: a rating regression approach,* KDD '10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington783-792.

Wani, M. A., Bhat, F. A., Afzal, S. and Khan, A. I. (2020). *Advances in deep learning*. Singapore: Springer, 23.

Wei, J., Chu, X., Sun, X. Y., Xu, K., Deng, H. X., Chen, J. and Lei, M. (2019). Machine learning in materials science. *InfoMat*, 1(3), 338-358.

Wilson, T., Wiebe, J. and Hwa, R. (2004). *Just how mad are you? Finding strong and weak opinion clauses.* Proceedings of Association for the Advancement of Artificial Intelligence, California, 761–769.

World Tourism Organization (2020). *Impact assessment of the covıd-19 outbreak on ınternatıonal tourısm*, *UNWTO*, Madrid, 1-8.

World Tourism Organization (2021). Tourist numbers down 83% but confidence slowly rising, 2 june 2021 report, *UNWTO*, Madrid, 1-2.

Xiang, Z., Du, Q., Ma, Y. and Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51-65.

Yadollahi, A., Shahraki A. G. ve Zaiane O. (2017). Görüşten duygu madenciliğine metin duyarlılığı analizinin mevcut durumu. *Association for Computing Machinery Computing Surveys*. 50, 2.

Ye, Q., Law, R. and Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1), 180-182.

Yener, Ö. ve Abdulkadir, G. (2007). Sosyal ve eğitim bilimleri araştırmalarında evren-örneklem sorunu. *Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi*, (15), 394-422.

Yıldırgan, M.S. (2022). *Rekreasyon Alanlarının ve Faaliyetlerinin Tamamlayıcı Konaklama İşletmeleri Üzerindeki Rolü: Sapanca Örneği*. Yüksek Lisans Tezi. Sakarya Uygulamalı Bilimler Üniversitesi Lisansüstü Eğitim Enstitüsü, Sakarya, 17-22.

Yıldız, Z. (2011). Turizm sektörünün gelişimi ve istihdam üzerindeki etkisi. *Süleyman Demirel Üniversitesi Vizyoner Dergisi*, 3(5), 54-71.

Yılmaz, E. S. (2020). The Effects on Consumer Behavior of Hotel Related Comments on the Tripadvisor Website: An Istanbul Case. *Advances in Hospitality and Tourism Research*, 8(1), 1-29.

Yılmaz, M. (2022). *Konaklama İşletmelerinde Helal Konsept Uygulamalarının Gıda Atığı ve Gıda İsrafı Üzerindeki Etkisi*. Doktora Tezi. Ankara Hacı Bayram Veli Üniversitesi Lisansüstü Eğitim Enstitüsü, Ankara, 26-28.

Yılmaz, S. (2022). *Konaklama Sektöründe Paylaşım Ekonomisi Uygulamalarının Kullanıcı Odaklı Değerlendirilmesi*. Yüksek Lisans Tezi, Sakarya Üniversitesi İşletme Enstitüsü, Sakarya, 65-82.

Zengin, B., Koç, D. ve Ulama, Ş. (2019). Kastamonu İlinin Doğa Turizmi Potansiyelinin Alternatif Turizm Kapsamında Değerlendirilmesi. *Safran Kültür ve Turizm Araştırmaları Dergisi,* 2(2): 251-274

Zohar, E. (2002). *Introduction to text mining. Supercomputing. Automated learning group national center for supercomputing applications.* Ullinois: University of Illinois, 32.

EKLER

EK-1. Bungalov yorumlarını çekmede kullanılan kodlar

```
from selenium import webdriver
from selenium.webdriver import ActionChains
from selenium.webdriver.common.by import By
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
import time
from scrapy import Selector
import pandas as pd
from selenium.webdriver.common.keys import Keys
from pre_process import lemmatize_word
browser = webdriver.Chrome()
with open("bungalov_url_list.txt") as file:
    urls = [line.rstrip() for line in file]
for b, url in enumerate(urls):
    browser.get(url)
    # 10 saniye bekle
    wait = WebDriverWait(browser, 10)
    df = pd.DataFrame()
    all_comments = []
    # Her scrollda 10 yorum geliyor. 10 * 30 = 300 yorum alacak sekilde sayfada asagiya in
    headlime = Selector(text=browser.page_source)
    last_height = browser.execute_script("return document.body.scrollHeight")
    for i in range(300):
        df_row = {"timestamp":[], "rating":[], "comment":[]}

        actions = ActionChains(browser)
        time.sleep(0.5)
        actions.send_keys(Keys.PAGE_DOWN)
        actions.send_keys(Keys.PAGE_UP)
        actions.send_keys(Keys.PAGE_DOWN)
            #actions.send_keys(Keys.PAGE_DOWN)
        actions.perform()
        comment_div_selector = Selector(text=browser.page_source)
        #Yorumlari gez
        for sel in comment_div_selector.css(".Svr5cf"):
            # Sabit zaman ve reyting CSS secicisi
            metadata = sel.css(".aAs4ib ::text").getall()
            # Sabit yorum CSS secicisi
            comment = sel.css(".K7oBsc span::text").get()
            if comment is None or comment in all_comments:
                continue
            all_comments.append(comment)
            timestamp = metadata[-2]
            rating = metadata[-1].split("/")[0]
```

EK-1. (devam) Bungalov yorumlarını çekmede kullanılan kodlar

```
            # Verileri pandasa ekle
            df_row["timestamp"] = timestamp
            df_row["rating"] = rating
            df_row["comment"] = comment
            df_row["lemmas"] = lemmatize_word(comment)
            df = df.append(df_row,ignore_index=True)
    # Pandastan yorumları yazdır.
    df.to_csv(f"yorumlar_{b}.csv",index=False)
browser.close()
```

EK-2. Verilerin temizlenmesi ve morfolojik analiz için kullanılan kodlar

```
import pandas as pd
import re
from Corpus.Sentence import Sentence
from DisambiguationCorpus.DisambiguationCorpus import DisambiguationCorpus
from MorphologicalAnalysis.FsmMorphologicalAnalyzer import
FsmMorphologicalAnalyzer
from MorphologicalDisambiguation.RootFirstDisambiguation import
RootFirstDisambiguation
morphologicalDisambiguator = RootFirstDisambiguation()
corpus = DisambiguationCorpus("penntreebank.txt")
morphologicalDisambiguator.train(corpus)
morphologicalDisambiguator.saveModel()
import ast
def remove_google_words(cell):
    tokens = ast.literal_eval(cell)
    filter_set = {"google", "tarafından", "çevir"}
    C = [x for x in tokens if x not in filter_set]
    return C
def process_csv():
    df = pd.read_csv("yorumlar_0.csv")
    df["comment"] = df["comment"].str.replace(r"\(Google tarafından çevrildi\)","")
    df["comment"] = df.comment.str.strip()
    df = df[df.comment.notnull()]
    return df
fsm = FsmMorphologicalAnalyzer()
def lemmatize_word(x):
    x.replace(r"\(Google tarafından çevrildi\)", "")
    x.strip()
    text = re.sub('\W+', ' ', str(x))
    text = text.replace('I', 'ı')  # lower() yapıldığı zaman I harfi i olarak çevirildiğinden dolayı
replace ile düzeltildi.
    text = text.replace('İ', 'i')
    text = text.lower()
    #text = " ".join([j for j in text.split() if j not in stop_word_list])
    text = " ".join([i for i in text.split() if len(i) > 1])
    #text = spellChecker(text.split())
    sentence = Sentence(text)
    fsmParseList = fsm.robustMorphologicalAnalysis(sentence)
    """
    print("All parses\n")
    print("-------------------------\n")
    for i in range(len(fsmParseList)):
        print(fsmParseList[i])
    """
```

EK-2. (devam) Verilerin temizlenmesi ve morfolojik analiz için kullanılan kodlar

```
    candidateParses = morphologicalDisambiguator.disambiguate(fsmParseList)
    """
    print("Parses after disambiguation\n")
    print("-------------------------\n")
    """
    roots = []
    for candidateRoot in candidateParses:
        roots.append(candidateRoot.root.name)
    return roots
```

EK-3. Naïve Bayes algoritmasına ait kodlar

```
import pandas as pd
import ast
import nltk
from matplotlib import pyplot as plt
from nltk import FreqDist
from sklearn.model_selection import train_test_split
from wordcloud import WordCloud

df = pd.DataFrame()
stop_word_listesi = nltk.corpus.stopwords.words('turkish')
stop_word_listesi.extend(["bir","kadar","sonra","kere","mi","ye","te","ta","nun","daki","nı
n","ten"])

def stopword_temizle(tokens):
    filtered_tokenler = [token for token in tokens if token not in stop_word_listesi]  # stop
word'lerden temizlenir.
    return filtered_tokenler

def tokenlari_isle(hucre):
    # hucredeki stringi listeye cevir
    tokenlar = ast.literal_eval(hucre)
    tokenlar = stopword_temizle(tokenlar)
    filtre_seti = {"google", "tarafından", "çevir"}
    C = [x for x in tokenlar if x not in filtre_seti]
    return C

import glob
for f in glob.glob("./hepsi/*.csv"):
    df = df.append(pd.read_csv(f), ignore_index=True)


df["lemmas"] = df["lemmas"].apply(lambda x: tokenlari_isle(x))

def rating_donustur(x):
    if x <= 3:
        return 0
    if x > 3:
        return 1

df["sentiment"] = df["rating"].apply(lambda x: rating_donustur(x))

def kelimeleri_getir(token_listesi):
    for tokenlar in token_listesi:
        for token in tokenlar:
```

EK-3. (devam) Naïve Bayes algoritmasına ait kodlar

```
 yield token


pozitif_token_listesi = df[df.rating > 3]['lemmas'].to_numpy()
negatif_token_listesi = df[df.rating <= 3]['lemmas'].to_numpy()
butun_token_listesi = df["lemmas"].to_numpy()

butun_kelimeler = kelimeleri_getir(butun_token_listesi)

pozitif_kelimeler = kelimeleri_getir(pozitif_token_listesi)

negatif_kelimeler = kelimeleri_getir(negatif_token_listesi)

frekans_dagilimi_hepsi = FreqDist(butun_kelimeler)

frekans_dagilimi_pozitif = FreqDist(pozitif_kelimeler)

frekans_dagilimi_negatif = FreqDist(negatif_kelimeler)

print("butun kelimeler")
print(frekans_dagilimi_hepsi.most_common(100))


print("pozitif kelimeler")
print(frekans_dagilimi_pozitif.most_common(100))


print("negatif kelimeler")
print(frekans_dagilimi_negatif.most_common(100))

def model_icin_kelimeleri_olustur(token_listesi):
        for yorum in token_listesi:
            yield dict([token, True] for token in yorum)

pozitif_tokenlar = model_icin_kelimeleri_olustur(pozitif_token_listesi)
negatif_tokenlar = model_icin_kelimeleri_olustur(negatif_token_listesi)

butun_wc_kelimeler = ""
pozitif_wc_kelimeler = ""
negatif_wc_kelimeler = ""

for token in butun_token_listesi:
    butun_wc_kelimeler += " ".join(token) + " "
```

EK-3. (devam) Naïve Bayes algoritmasına ait kodlar

```
for token in pozitif_token_listesi:
    pozitif_wc_kelimeler += " ".join(token) + " "


for token in negatif_token_listesi:
    negatif_wc_kelimeler += " ".join(token) + " "


positif_dataset = [(kelime, "Positive")
                   for kelime in pozitif_tokenlar][:]

negatif_dataset = [(kelime, "Negative")
                   for kelime in negatif_tokenlar]

print(f"Pozitif {len(positif_dataset)}")
print(f"Negatif {len(negatif_dataset)}")
dataset = positif_dataset + negatif_dataset


for i,cloud in enumerate([butun_wc_kelimeler, pozitif_wc_kelimeler,
negatif_wc_kelimeler]):

    title = ""
    file_name = ""
    if i == 0:
        title = "All Words Cloud"
        file_name = "all_words.png"
    elif i == 1:
        title = "Positive Words Cloud"
        file_name = "pos_words.png"
    else:
        title = "Negative Words Cloud"
        file_name = "neg_words.png"
    plt.title(title)
    plt.figure(figsize=(8, 8), facecolor=None)
    wc = WordCloud(width=800, height=800,
            background_color='white',
            min_font_size=10).generate(cloud)
    plt.imshow(wc)
    plt.axis("off")
    plt.tight_layout(pad=0)
    plt.savefig(file_name)

train_data = dataset[:2250]
```

EK-3. (devam) Naïve Bayes algoritmasına ait kodlar

```
test_data = dataset[2250:]

from nltk import classify
from nltk import NaiveBayesClassifier

from sklearn.metrics import auc, accuracy_score, f1_score, precision_score, recall_score, roc_curve

calisma_no = 1

def sonucları_raporla(y, pred):
    global calisma_no
    fpr, tpr, _ = roc_curve(y_true, pred, pos_label=1)
    auc_score = auc(fpr, tpr)
    acc = accuracy_score(y, pred)
    f1 = f1_score(y, pred)
    prec = precision_score(y, pred)
    rec = recall_score(y, pred)
    result = {'auc': auc_score, 'f1': f1, 'acc': acc, 'precision': prec, 'recall': rec}
    from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
    import matplotlib.pyplot as plt
    cm = confusion_matrix(y, pred)
    disp = ConfusionMatrixDisplay(confusion_matrix=cm)
    disp.plot()
    plt.savefig(f"naive_{calisma_no}.png")
    calisma_no += 1
    return result


def etiketten_numaraya(label):
    if label == 'Positive':
        return 1
    else:
        return 0

rep_df = pd.DataFrame()
calisma_sayilari = [1,17,31,97,841,14345,95234,341235,3764083,9999999]
metrikler = []
for sayi in calisma_sayilari:
    train, test = train_test_split(dataset, test_size=0.2, random_state=sayi, shuffle=True)

    y_true, y_score = [], []

    print("Model ogreniyor")
```

EK-3. (devam) Naïve Bayes algoritmasına ait kodlar

```
  classifier = NaiveBayesClassifier.train(train)

    for i, (feats, label_true) in enumerate(test):
        label_predicted = classifier.classify(feats)
        y_true.append(etiketten_numaraya(label_true))
        y_score.append(etiketten_numaraya(label_predicted))

    metrikler.append(sonucları_raporla(y_true,y_score))
    print("Model test ediliyor")

    print("Dogruluk degeri :", classify.accuracy(classifier, test_data))

    print(classifier.show_most_informative_features(50))

rep_df = pd.concat([rep_df, pd.DataFrame.from_dict(metrikler)])
rep_df.to_csv("naive_metrics.csv",index=False)
```

EK-4. Destek Vektör Makineleri algoritmasına ait kodlar

```
import pandas as pd
import ast
import nltk
from matplotlib import pyplot as plt
from nltk import FreqDist
from sklearn.metrics import roc_auc_score, accuracy_score, f1_score, precision_score, recall_score, roc_curve
from sklearn.model_selection import train_test_split, StratifiedKFold, GridSearchCV
from sklearn.pipeline import make_pipeline
from sklearn.svm import SVC
from wordcloud import WordCloud

df = pd.DataFrame()
stop_word_listesi = nltk.corpus.stopwords.words('turkish')
stop_word_listesi.extend(["bir","kadar","sonra","kere","mi","ye","te","ta","nun","daki","nın","ten"])

def stopword_temizle(tokens):
    filtered_tokenler = [token for token in tokens if token not in stop_word_listesi]  # stop word'lerden temizlenir.
    return filtered_tokenler

def tokenlari_isle(hucre):
    # hucredeki stringi listeye cevir
    tokenlar = ast.literal_eval(hucre)
    tokenlar = stopword_temizle(tokenlar)
    filtre_seti = {"google", "tarafından", "çevir"}
    C = [x for x in tokenlar if x not in filtre_seti]
    return C

import glob
for f in glob.glob("./hepsi/*.csv"):
    df = df.append(pd.read_csv(f), ignore_index=True)

df["lemmas"] = df["lemmas"].apply(lambda x: tokenlari_isle(x))

def rating_donustur(x):
    if x <= 3:
        return 0
    if x > 3:
        return 1

df["sentiment"] = df["rating"].apply(lambda x: rating_donustur(x))
```

EK-4. (devam) Destek Vektör Makineleri algoritmasına ait kodlar

```
def kelimeleri_getir(token_listesi):
    for tokenlar in token_listesi:
        for token in tokenlar:
            yield token


pozitif_token_listesi = df[df.rating >3]['lemmas'].to_numpy()
negatif_token_listesi = df[df.rating <= 3]['lemmas'].to_numpy()
butun_token_listesi = df["lemmas"].to_numpy()

butun_kelimeler = kelimeleri_getir(butun_token_listesi)

pozitif_kelimeler = kelimeleri_getir(pozitif_token_listesi)

negatif_kelimeler = kelimeleri_getir(negatif_token_listesi)

frekans_dagilimi_hepsi = FreqDist(butun_kelimeler)

frekans_dagilimi_pozitif = FreqDist(pozitif_kelimeler)

frekans_dagilimi_negatif = FreqDist(negatif_kelimeler)

print("butun kelimeler")
print(frekans_dagilimi_hepsi.most_common(100))


print("pozitif kelimeler")
print(frekans_dagilimi_pozitif.most_common(100))


print("negatif kelimeler")
print(frekans_dagilimi_negatif.most_common(100))

def model_icin_kelimeleri_olustur(token_listesi):
        for yorum in token_listesi:
            yield dict([token, True] for token in yorum)

pozitif_tokenlar = model_icin_kelimeleri_olustur(pozitif_token_listesi)
negatif_tokenlar = model_icin_kelimeleri_olustur(negatif_token_listesi)

# Kelime bulutlarını olustur
butun_wc_kelimeler = ""
pozitif_wc_kelimeler = ""
negatif_wc_kelimeler = ""
```

EK-4. (devam) Destek Vektör Makineleri algoritmasına ait kodlar

```
for token in butun_token_listesi:
    butun_wc_kelimeler += " ".join(token) + " "


for token in pozitif_token_listesi:
    pozitif_wc_kelimeler += " ".join(token) + " "


for token in negatif_token_listesi:
    negatif_wc_kelimeler += " ".join(token) + " "


positif_dataset = [(kelime, "Positive")
                   for kelime in pozitif_tokenlar][:]

negatif_dataset = [(kelime, "Negative")
                   for kelime in negatif_tokenlar]

dataset = positif_dataset + negatif_dataset


for i,cloud in enumerate([butun_wc_kelimeler, pozitif_wc_kelimeler,
negatif_wc_kelimeler]):
    # kelime bulutlarini dosyaya yaz
    title = ""
    file_name = ""
    if i == 0:
        title = "All Words Cloud"
        file_name = "all_words.png"
    elif i == 1:
        title = "Positive Words Cloud"
        file_name = "pos_words.png"
    else:
        title = "Negative Words Cloud"
        file_name = "neg_words.png"
    plt.title(title)
    plt.figure(figsize=(8, 8), facecolor=None)
    wc = WordCloud(width=800, height=800,
            background_color='white',
            min_font_size=10).generate(cloud)
    plt.imshow(wc)
    plt.axis("off")
    plt.tight_layout(pad=0)
    plt.savefig(file_name)
```

EK-4. (devam) Destek Vektör Makineleri algoritmasına ait kodlar

```
train_data = dataset[:2250]
test_data = dataset[2250:]

from sklearn.feature_extraction.text import CountVectorizer

def bos_fonksiyon(doc):
   return doc

vectorizer = CountVectorizer(
       tokenizer=bos_fonksiyon,
       preprocessor=bos_fonksiyon,
   )

import numpy as np

np.random.seed(65)

pipeline_svm = make_pipeline(vectorizer,
                     SVC(probability=True, kernel="linear", class_weight="balanced"))

kfolds = StratifiedKFold(n_splits=5, shuffle=True, random_state=1)

grid_svm = GridSearchCV(pipeline_svm,
                param_grid = {'svc__C': [0.01, 0.1, 1]},
                cv = kfolds,
                scoring="roc_auc",
                verbose=1,
                n_jobs=-1)

rep_df = pd.DataFrame()
calisma_sayilari = [1,17,31,97,841,14345,95234,341235,3764083,9999999]
metrics = []
calisma_no = 1
for sayi in calisma_sayilari:
   # random_state = 2
   train, test = train_test_split(df, test_size=0.2, random_state=sayi,shuffle=True)
   X_train = train['lemmas'].values
   X_test = test['lemmas'].values
   y_train = train['sentiment']
   y_test = test['sentiment']


   grid_svm.fit(X_train, y_train)
   print(grid_svm.score(X_test, y_test))
```

EK-4. (devam) Destek Vektör Makineleri algoritmasına ait kodlar

```
    print(grid_svm.best_params_)
    print(grid_svm.best_score_)


    def sonuclari_raporla(model, X, y):
        global calisma_no
        pred_proba = model.predict_proba(X)[:, 1]
        pred = model.predict(X)

        auc = roc_auc_score(y, pred_proba)
        acc = accuracy_score(y, pred)
        f1 = f1_score(y, pred)
        prec = precision_score(y, pred)
        rec = recall_score(y, pred)
        result = {'auc': auc, 'f1': f1, 'acc': acc, 'precision': prec, 'recall': rec}
        from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
        import matplotlib.pyplot as plt
        cm = confusion_matrix(y_test, pred, labels=model.classes_)
        disp = ConfusionMatrixDisplay(confusion_matrix=cm,
                                      display_labels=model.classes_)
        disp.plot()

        plt.savefig(f"svm_{calisma_no}.png")
        calisma_no += 1

        return result

    metrics.append(sonuclari_raporla(grid_svm.best_estimator_, X_test, y_test))

rep_df = pd.concat([rep_df, pd.DataFrame.from_dict(metrics)])
rep_df.to_csv("svm_metrics.csv",index=False)
def get_roc_curve(model, X, y):
    pred_proba = model.predict_proba(X)[:, 1]
    fpr, tpr, _ = roc_curve(y, pred_proba)
    return fpr, tpr
```

EK-5. Rastgele Orman algoritmasına ait kodlar

```
import numpy as np
import pandas as pd
import ast
import nltk
from matplotlib import pyplot as plt
from nltk import FreqDist
from sklearn.metrics import roc_auc_score, accuracy_score, f1_score, precision_score, recall_score, roc_curve
from sklearn.model_selection import train_test_split, StratifiedKFold, GridSearchCV
# kelime bulutlari icin
from sklearn.pipeline import make_pipeline
from sklearn.svm import SVC
from wordcloud import WordCloud

df = pd.DataFrame()
stop_word_listesi = nltk.corpus.stopwords.words('turkish')
stop_word_listesi.extend(["bir","kadar","sonra","kere","mi","ye","te","ta","nun","daki","nın","ten"])

def stopword_temizle(tokens):
    filtered_tokenler = [token for token in tokens if token not in stop_word_listesi]  # stop word'lerden temizlenir.
    return filtered_tokenler

def tokenlari_isle(hucre):
    tokenlar = ast.literal_eval(hucre)
    tokenlar = stopword_temizle(tokenlar)
    filtre_seti = {"google", "tarafından", "çevir"}
    C = [x for x in tokenlar if x not in filtre_seti]
    return C

import glob
for f in glob.glob("./hepsi/*.csv"):
    df = df.append(pd.read_csv(f), ignore_index=True)

df["lemmas"] = df["lemmas"].apply(lambda x: tokenlari_isle(x))

def rating_donustur(x):
    if x <= 3:
        return 0
    if x > 3:
        return 1

df["sentiment"] = df["rating"].apply(lambda x: rating_donustur(x))
```

EK-5. (devam) Rastgele Orman algoritmasına ait kodlar

```
def kelimeleri_getir(token_listesi):
    for tokenlar in token_listesi:
        for token in tokenlar:
            yield token


pozitif_token_listesi = df[df.rating >3]['lemmas'].to_numpy()
negatif_token_listesi = df[df.rating <= 3]['lemmas'].to_numpy()
butun_token_listesi = df["lemmas"].to_numpy()

butun_kelimeler = kelimeleri_getir(butun_token_listesi)

pozitif_kelimeler = kelimeleri_getir(pozitif_token_listesi)

negatif_kelimeler = kelimeleri_getir(negatif_token_listesi)

frekans_dagilimi_hepsi = FreqDist(butun_kelimeler)

frekans_dagilimi_pozitif = FreqDist(pozitif_kelimeler)

frekans_dagilimi_negatif = FreqDist(negatif_kelimeler)

print("butun kelimeler")
print(frekans_dagilimi_hepsi.most_common(100))


print("pozitif kelimeler")
print(frekans_dagilimi_pozitif.most_common(100))


print("negatif kelimeler")
print(frekans_dagilimi_negatif.most_common(100))

def model_icin_kelimeleri_olustur(token_listesi):
        for yorum in token_listesi:
            yield dict([token, True] for token in yorum)

pozitif_tokenlar = model_icin_kelimeleri_olustur(pozitif_token_listesi)
negatif_tokenlar = model_icin_kelimeleri_olustur(negatif_token_listesi)

butun_wc_kelimeler = ""
pozitif_wc_kelimeler = ""
negatif_wc_kelimeler = ""
```

EK-5. (devam) Rastgele Orman algoritmasına ait kodlar

```
for token in butun_token_listesi:
   butun_wc_kelimeler += " ".join(token) + " "


for token in pozitif_token_listesi:
   pozitif_wc_kelimeler += " ".join(token) + " "


for token in negatif_token_listesi:
   negatif_wc_kelimeler += " ".join(token) + " "


positif_dataset = [(kelime, "Positive")
                   for kelime in pozitif_tokenlar][:]

negatif_dataset = [(kelime, "Negative")
                   for kelime in negatif_tokenlar]

dataset = positif_dataset + negatif_dataset


for i,cloud in enumerate([butun_wc_kelimeler, pozitif_wc_kelimeler,
negatif_wc_kelimeler]):

   title = ""
   file_name = ""
   if i == 0:
      title = "All Words Cloud"
      file_name = "all_words.png"
   elif i == 1:
      title = "Positive Words Cloud"
      file_name = "pos_words.png"
   else:
      title = "Negative Words Cloud"
      file_name = "neg_words.png"
   plt.title(title)
   plt.figure(figsize=(8, 8), facecolor=None)
   wc = WordCloud(width=800, height=800,
           background_color='white',
           min_font_size=10).generate(cloud)
   plt.imshow(wc)
   plt.axis("off")
   plt.tight_layout(pad=0)
   plt.savefig(file_name)
```

EK-5. (devam) Rastgele Orman algoritmasına ait kodlar

```
train_data = dataset[:2250]
test_data = dataset[2250:]

from sklearn.feature_extraction.text import CountVectorizer

def bos_fonksiyon(doc):
    return doc

vectorizer = CountVectorizer(
        tokenizer=bos_fonksiyon,
        preprocessor=bos_fonksiyon,
    )
np.random.seed(65)

calisma_no = 1
def sonuclari_raporla(model, X, y):
    global calisma_no
    pred_proba = model.predict_proba(X)[:, 1]
    pred = model.predict(X)

    auc = roc_auc_score(y, pred_proba)
    acc = accuracy_score(y, pred)
    f1 = f1_score(y, pred)
    prec = precision_score(y, pred)
    rec = recall_score(y, pred)
    result = {'auc': auc, 'f1': f1, 'acc': acc, 'precision': prec, 'recall': rec}
    from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay
    import matplotlib.pyplot as plt
    cm = confusion_matrix(y_test, pred, labels=model.classes_)
    disp = ConfusionMatrixDisplay(confusion_matrix=cm,
                                  display_labels=model.classes_)
    disp.plot()

    plt.savefig(f"randomforest_{calisma_no}.png")
    calisma_no += 1

    return result

from sklearn.ensemble import RandomForestClassifier


from sklearn.model_selection import GridSearchCV

rf = RandomForestClassifier()
```

EK-5. (devam) Rastgele Orman algoritmasına ait kodlar

```
parameters = {
   'randomforestclassifier__n_estimators': [5, 50, 100],
   'randomforestclassifier__max_depth': [2, 10, 20, None]
}

pipeline_rf = make_pipeline(vectorizer,
                        RandomForestClassifier(class_weight="balanced"))

kfolds = StratifiedKFold(n_splits=5, shuffle=True, random_state=1)

grid_rf = GridSearchCV(pipeline_rf,
                    param_grid=parameters,
                    cv=kfolds,
                    verbose=1,
                    n_jobs=-1)

rep_df = pd.DataFrame()
calisma_sayilari = [1,17,31,97,841,14345,95234,341235,3764083,9999999]
metrikler = []
for sayi in calisma_sayilari:
   train, test = train_test_split(df, test_size=0.2, random_state=sayi, shuffle=True)
   X_train = train['lemmas'].values
   X_test = test['lemmas'].values
   y_train = train['sentiment']
   y_test = test['sentiment']



   grid_rf.fit(X_train, y_train)

   metrikler.append(sonuclari_raporla(grid_rf.best_estimator_, X_test, y_test))

rep_df = pd.concat([rep_df, pd.DataFrame.from_dict(metrikler)])
rep_df.to_csv("rf_metrics.csv",index=False)
```

EK-6. BERT algoritmasına ait kodlar

```
import seaborn as sns
import torch
from torch import nn
from torch.utils.data import Dataset, DataLoader
from transformers import BertTokenizer, TFBertForSequenceClassification, BertModel, AdamW, \
    get_linear_schedule_with_warmup
from transformers import InputExample, InputFeatures
from collections import Counter, defaultdict
import pandas as pd
import ast
import nltk
import sklearn.pipeline
from matplotlib import pyplot as plt
from nltk import FreqDist, word_tokenize
from sklearn.metrics import roc_auc_score, accuracy_score, f1_score, precision_score, recall_score, roc_curve, \
    classification_report, confusion_matrix
from sklearn.model_selection import train_test_split, StratifiedKFold, GridSearchCV
from sklearn.svm import SVC
from wordcloud import WordCloud
import numpy as np
from sklearn.feature_extraction.text import TfidfVectorizer
import tensorflow as tf
df = pd.DataFrame()
stop_word_listesi = nltk.corpus.stopwords.words('turkish')
stop_word_listesi.extend(["bir","kadar","sonra","kere","mi","ye","te","ta","nun","daki","nın","ten"])

def stopword_temizle(text):
    global stop_word_listesi
    removed = []
    stop_words = stop_word_listesi
    tokens = word_tokenize(text)
    for i in range(len(tokens)):
        if tokens[i] not in stop_words:
            removed.append(tokens[i])
    return " ".join(removed)

def tokenlari_isle(hucre):

    token_str = hucre.replace("'","")
    token_str = token_str.replace(",","")
    token_str = token_str.replace("[","")
```

EK-6. (devam) BERT algoritmasına ait kodlar

```
    token_str = token_str.replace("]","")
    #tokenlar = ast.literal_eval(hucre)
    tokenlar = stopword_temizle(token_str)
    #filtre_seti = {"google", "tarafından", "çevir"}
    #C = [x for x in tokenlar if x not in filtre_seti]
    return tokenlar



import glob

for f in glob.glob("./hepsi/*.csv"):
    df = df.append(pd.read_csv(f), ignore_index=True)


df["lemmas"] = df["lemmas"].apply(lambda x: tokenlari_isle(x))

def rating_donustur(x):
    if x <= 3:
        return 0
    if x > 3:
        return 1

df["sentiment"] = df["rating"].apply(lambda x: rating_donustur(x))

def kelimeleri_getir(token_listesi):
    for tokenlar in token_listesi:
        for token in tokenlar:
            yield token


pozitif_token_listesi = df[df.rating >3]['lemmas'].to_numpy()
negatif_token_listesi = df[df.rating <= 3]['lemmas'].to_numpy()
def dummy_tokenizer(a):
    return a
tf_wb = TfidfVectorizer()

tokenizer = BertTokenizer.from_pretrained("dbmdz/bert-base-turkish-uncased")

token_lens = []

for txt in df.lemmas:
    tokens = tokenizer.encode(txt, max_length=512)
    token_lens.append(len(tokens))
```

EK-6. (devam) BERT algoritmasına ait kodlar

```
sns.distplot(token_lens)
plt.xlim([0, 256]);
plt.xlabel('Token count')
plt.show()

MAX_LEN = 80


class BungalovDataset(Dataset):

  def __init__(self, reviews, targets, tokenizer, max_len):
    self.reviews = reviews
    self.targets = targets
    self.tokenizer = tokenizer
    self.max_len = max_len

  def __len__(self):
    return len(self.reviews)

  def __getitem__(self, item):
    review = str(self.reviews[item])
    target = self.targets[item]


    encoding = self.tokenizer.encode_plus(
      review,
      add_special_tokens=True,
      max_length=self.max_len,
      return_token_type_ids=False,
      pad_to_max_length=True,
      return_attention_mask=True,
      return_tensors='pt',
    )

    return {
      'review_text': review,
      'input_ids': encoding['input_ids'].flatten(),
      'attention_mask': encoding['attention_mask'].flatten(),
      'targets': torch.tensor(target, dtype=torch.long)
    }

df_train, df_test = train_test_split(df, test_size=0.2, random_state=31)
df_val, df_test = train_test_split(df_test, test_size=0.5, random_state=31)
```

EK-6. (devam) BERT algoritmasına ait kodlar

```
print(df_train.shape, df_val.shape, df_test.shape)


def create_data_loader(df, tokenizer, max_len, batch_size):
  ds = BungalovDataset(
    reviews=df.lemmas.to_numpy(),
    targets=df.sentiment.to_numpy(),
    tokenizer=tokenizer,
    max_len=max_len
  )

  return DataLoader(
    ds,
    batch_size=batch_size,
    num_workers=0
  )

BATCH_SIZE = 16
train_data_loader = create_data_loader(df_train, tokenizer, MAX_LEN, BATCH_SIZE)
val_data_loader = create_data_loader(df_val, tokenizer, MAX_LEN, BATCH_SIZE)
test_data_loader = create_data_loader(df_test, tokenizer, MAX_LEN, BATCH_SIZE)

data = next(iter(train_data_loader))
print(data.keys())

print(data['input_ids'].shape)
print(data['attention_mask'].shape)
print(data['targets'].shape)


class SentimentClassifier(nn.Module):

  def __init__(self, n_classes):
    super(SentimentClassifier, self).__init__()
    self.bert = BertModel.from_pretrained("dbmdz/bert-base-turkish-uncased")
    self.drop = nn.Dropout(p=0.3)
    self.out = nn.Linear(self.bert.config.hidden_size, n_classes)

  def forward(self, input_ids, attention_mask):
    _, pooled_output = self.bert(
      input_ids=input_ids,
      attention_mask=attention_mask,
      return_dict=False
    )
```

EK-6. (devam) BERT algoritmasına ait kodlar

```
        output = self.drop(pooled_output)
        return self.out(output)

device = torch.device("cuda" if torch.cuda.is_available() else "cpu")
class_names = ['negative', 'positive']
model = SentimentClassifier(len(class_names))


EPOCHS = 3

optimizer = AdamW(model.parameters(), lr=2e-5, correct_bias=False)

total_steps = len(train_data_loader) * EPOCHS

scheduler = get_linear_schedule_with_warmup(
  optimizer,
  num_warmup_steps=0,
  num_training_steps=total_steps
)

loss_fn = nn.CrossEntropyLoss().to(device)


def train_epoch(model, data_loader, loss_fn, optimizer, device, scheduler, n_examples):
  model = model.train()
  losses = []
  correct_predictions = 0

  for d in data_loader:
    input_ids = d["input_ids"].to(device)
    attention_mask = d["attention_mask"].to(device)
    targets = d["targets"].to(device)

    outputs = model(
      input_ids=input_ids,
      attention_mask=attention_mask
    )

    _, preds = torch.max(outputs, dim=1)
    loss = loss_fn(outputs, targets)
    correct_predictions += torch.sum(preds == targets)
    losses.append(loss.item())

    loss.backward()
```

EK-6. (devam) BERT algoritmasına ait kodlar

```
        nn.utils.clip_grad_norm_(model.parameters(), max_norm=1.0)
        optimizer.step()
        scheduler.step()
        optimizer.zero_grad()

    return correct_predictions.double() / n_examples, np.mean(losses)


def eval_model(model, data_loader, loss_fn, device, n_examples):
    model = model.eval()

    losses = []
    correct_predictions = 0

    with torch.no_grad():
        for d in data_loader:
            input_ids = d["input_ids"].to(device)
            attention_mask = d["attention_mask"].to(device)
            targets = d["targets"].to(device)

            outputs = model(
                input_ids=input_ids,
                attention_mask=attention_mask
            )

            _, preds = torch.max(outputs, dim=1)
            loss = loss_fn(outputs, targets)

            correct_predictions += torch.sum(preds == targets)
            losses.append(loss.item())

    return correct_predictions.double() / n_examples, np.mean(losses)


history = defaultdict(list)
best_accuracy = 0
for epoch in range(EPOCHS):


    print(f"Epoch {epoch + 1}/{EPOCHS}")
    print("-" * 10)

    train_acc, train_loss = train_epoch(
        model,
```

EK-6. (devam) BERT algoritmasına ait kodlar

```
      train_data_loader,
      loss_fn,
      optimizer,
      device,
      scheduler,
      len(df_train)
   )

   print(f"Train loss {train_loss} accuracy {train_acc}")

   val_acc, val_loss = eval_model(
      model,
      val_data_loader,
      loss_fn,
      device,
      len(df_val)
   )

   print(f"Val   loss {val_loss} accuracy {val_acc}")
   print()

   history['train_acc'].append(train_acc)
   history['train_loss'].append(train_loss)
   history['val_acc'].append(val_acc)
   history['val_loss'].append(val_loss)

   if val_acc > best_accuracy:
      torch.save(model.state_dict(), 'best_model_state.bin')
      best_accuracy = val_acc


plt.plot(history['train_acc'], label='train accuracy')
plt.plot(history['val_acc'], label='validation accuracy')

# Graph chars
plt.title('Training history')
plt.ylabel('Accuracy')
plt.xlabel('Epoch')
plt.legend()
plt.ylim([0, 1]);

plt.show()

test_acc, _ = eval_model(
```

EK-6. (devam) BERT algoritmasına ait kodlar

```
  model,
  test_data_loader,
  loss_fn,
  device,
  len(df_test)
)

test_acc.item()

def get_predictions(model, data_loader):
    model = model.eval()

    review_texts = []
    predictions = []
    prediction_probs = []
    real_values = []

    with torch.no_grad():
        for d in data_loader:
            texts = d["review_text"]
            input_ids = d["input_ids"].to(device)
            attention_mask = d["attention_mask"].to(device)
            targets = d["targets"].to(device)


            outputs = model(
                input_ids=input_ids,
                attention_mask=attention_mask
            )
            _, preds = torch.max(outputs, dim=1)

            review_texts.extend(texts)
            predictions.extend(preds)
            prediction_probs.extend(outputs)
            real_values.extend(targets)

    predictions = torch.stack(predictions).cpu()
    prediction_probs = torch.stack(prediction_probs).cpu()
    real_values = torch.stack(real_values).cpu()

    return review_texts, predictions, prediction_probs, real_values
```

EK-6. (devam) BERT algoritmasına ait kodlar

```
y_review_texts, y_pred, y_pred_probs, y_test = get_predictions(
    model,
    test_data_loader
)

print(classification_report(y_test, y_pred, target_names=class_names))
model.load_state_dict(torch.load('best_model_state.bin'))


import torch.nn.functional as F


def bert_predict(model, test_dataloader):
    model.eval()

    all_logits = []

    for batch in test_dataloader:

        b_input_ids = batch["input_ids"].to(device)
        b_attn_mask = batch["attention_mask"].to(device)
        with torch.no_grad():
            logits = model(b_input_ids, b_attn_mask)
        all_logits.append(logits)

    all_logits = torch.cat(all_logits, dim=0)

    probs = F.softmax(all_logits, dim=1).cpu().numpy()

    return probs

from sklearn.metrics import accuracy_score, roc_curve, auc


def evaluate_roc (probs, y_true):
    preds = probs[:, 1]
    fpr, tpr, threshold = roc_curve(y_true, preds)
    roc_auc = auc(fpr, tpr)
    print(f'AUC: {roc_auc:.4f}')

    y_pred = np.where(preds >= 0.5, 1, 0)
    accuracy = accuracy_score(y_true, y_pred)
    print(f'Accuracy: {accuracy * 100:.2f}%')
```

EK-6. (devam) BERT algoritmasına ait kodlar

```
    plt.title('Receiver Operating Characteristic')
    plt.plot(fpr, tpr, 'b', label='AUC = %0.2f' %roc_auc)
    plt.legend(loc='lower right')
    plt.plot([0, 1], [0, 1], 'r--')
    plt.xlim([0, 1])
    plt.ylim([0, 1])
    plt.ylabel('True Positive Rate')
    plt.xlabel('False Positive Rate')
    plt.show()

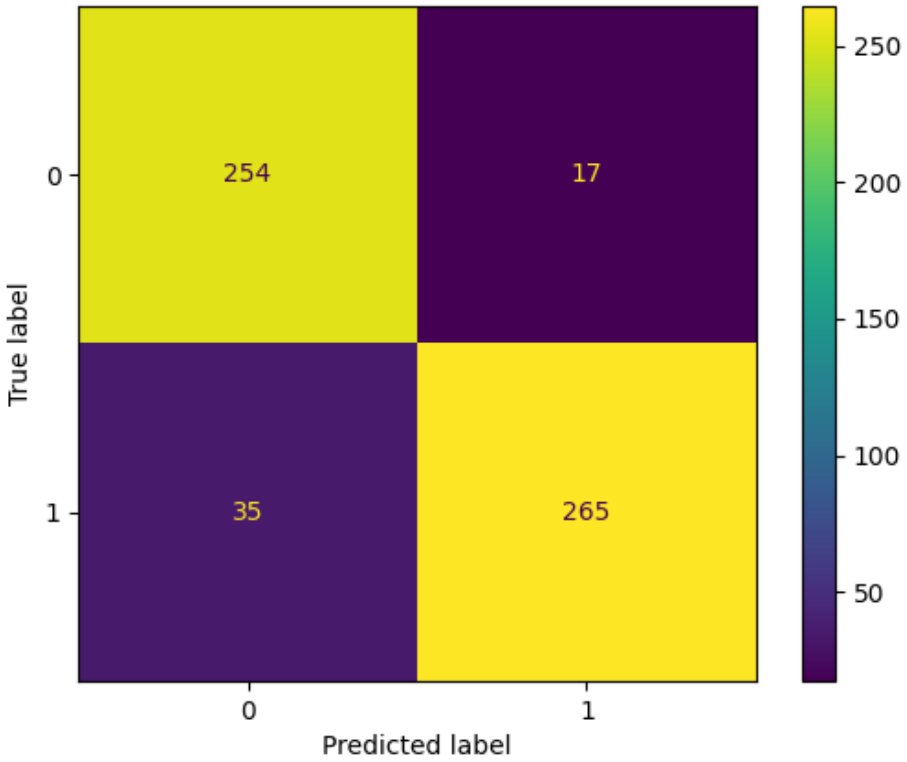
probs = bert_predict(model, val_data_loader)

evaluate_roc(probs, df_val.sentiment.to_numpy())

def show_confusion_matrix(confusion_matrix):
    hmap = sns.heatmap(confusion_matrix, annot=True, fmt="d", cmap="Blues")
    hmap.yaxis.set_ticklabels(hmap.yaxis.get_ticklabels(), rotation=0, ha='right')
    hmap.xaxis.set_ticklabels(hmap.xaxis.get_ticklabels(), rotation=30, ha='right')
    plt.ylabel('True sentiment')
    plt.xlabel('Predicted sentiment')
    plt.show()

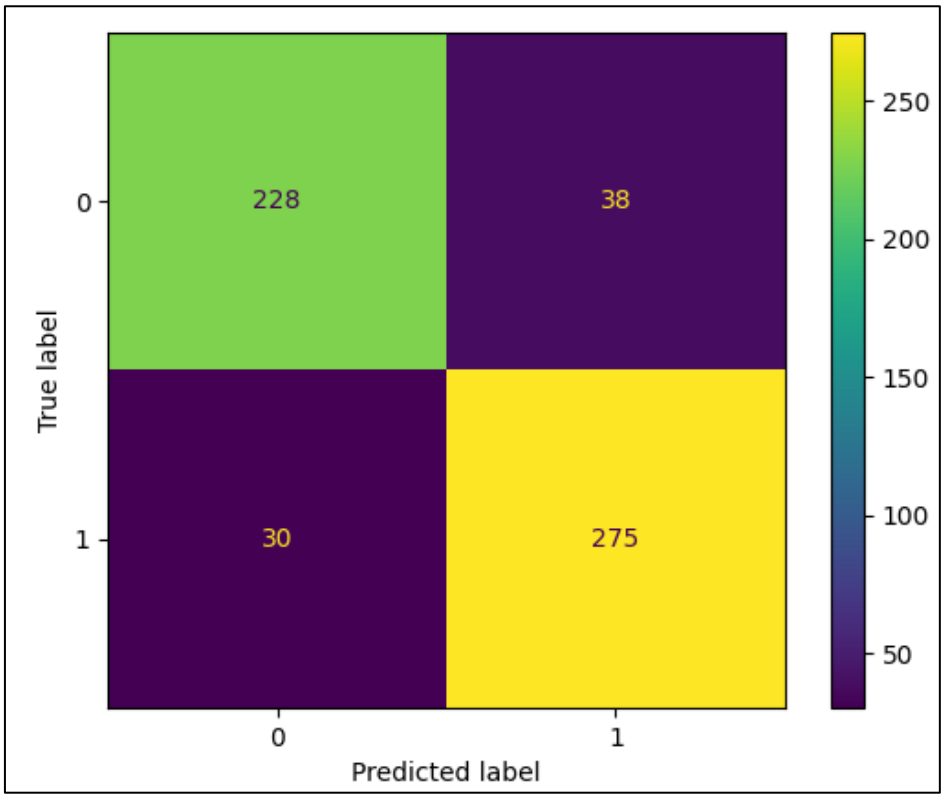
cm = confusion_matrix(y_test, y_pred)
df_cm = pd.DataFrame(cm, index=class_names, columns=class_names)
show_confusion_matrix(df_cm)
```

EK-7. Naive Bayes algoritmasına ait karmaşıklık matrisi program çıktısı örneği



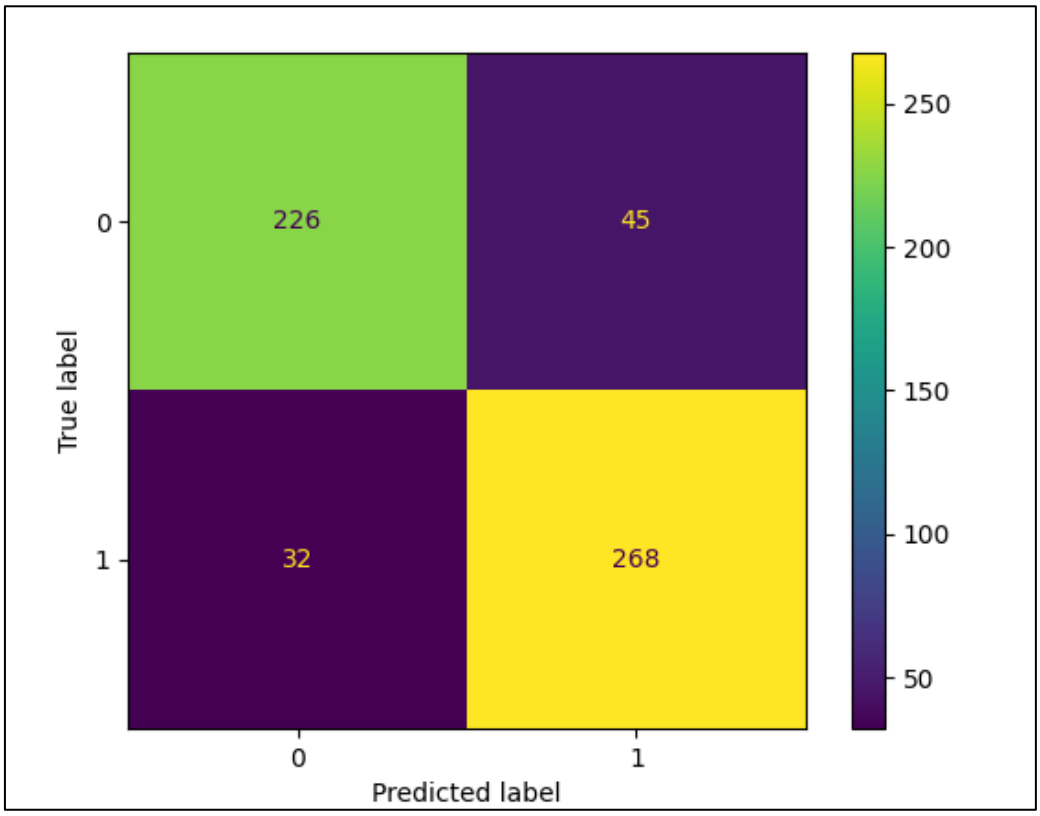
Şekil 7.1. Naive Bayes algoritmasına ait karmaşıklık matrisi program çıktısı örneği

EK-8. Destek Vektör Makineleri algoritmasına ait karmaşıklık matrisi program çıktısı örneği



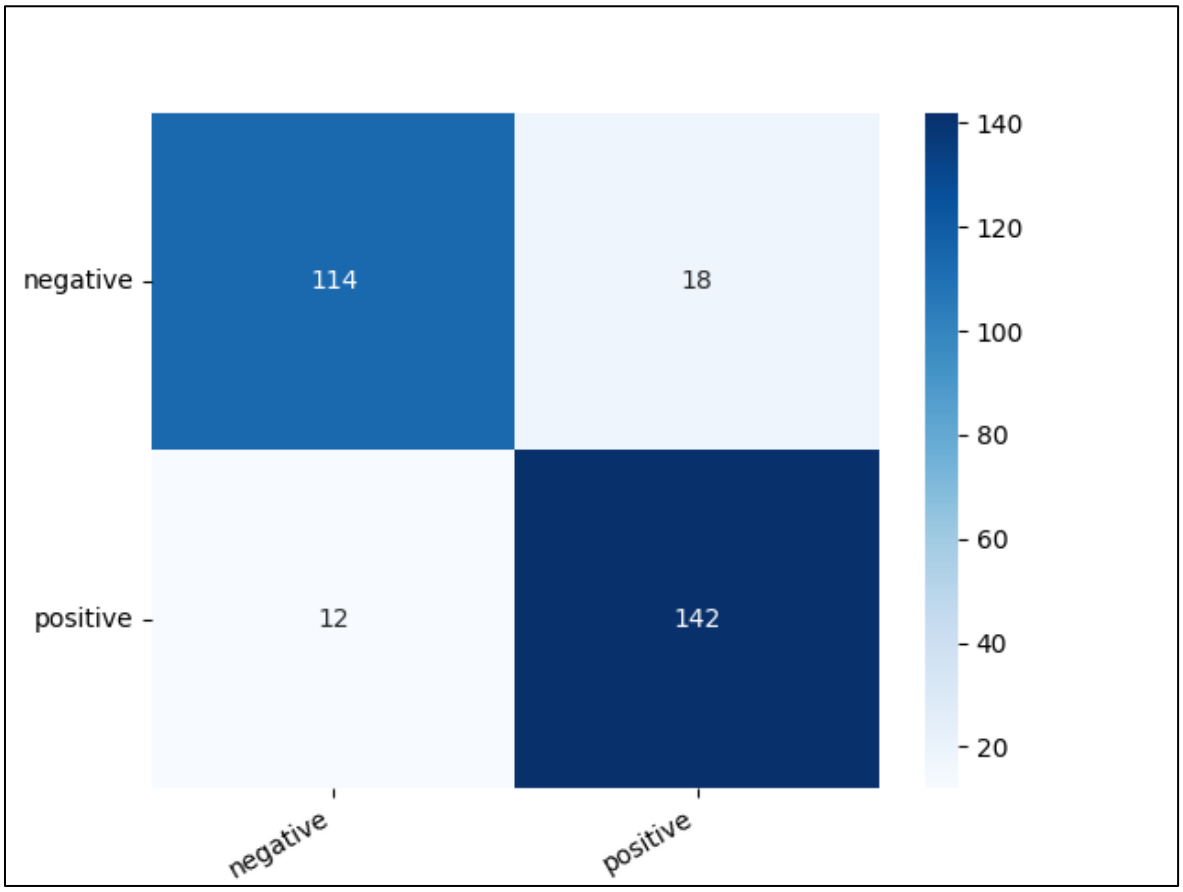
Şekil 8.1. Destek Vektör Makineleri algoritmasına ait karmaşıklık matrisi program çıktısı örneği

EK-9. Rastgele Orman algoritmasına ait karmaşıklık matrisi program çıktısı örneği



Şekil 9.1. Destek Vektör Makineleri algoritmasına ait karmaşıklık matrisi program çıktısı örneği

EK-10. BERT algoritmasına ait karmaşıklık matrisi program çıktısı örneği



Şekil 10.1. Destek Vektör Makineleri algoritmasına ait karmaşıklık matrisi program çıktısı örneği



Gazili olmak ayrıcalıktır