



**TURİSTLERİN ÇEVRE DOSTU OTELLERLE İLGİLİ
MEMNUNİYETİNİ ETKİLEYEN FAKTÖRLERİN
MAKİNE ÖĞRENMESİ TEKNİKLERİ
KULLANILARAK ARAŞTIRILMASI**

Mahmud ALRAHHAL

Danışman: Doç. Dr. Ferhat BOZKURT

Yüksek Lisans Tezi

Bilgisayar Mühendisliği Ana Bilim Dalı

2023

(Her hakkı saklıdır.)

T.C.
ATATÜRK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**TURİSTLERİN ÇEVRE DOSTU OTELLERLE İLGİLİ MEMNUNİYETİNİ
ETKİLEYEN FAKTÖRLERİN MAKİNE ÖĞRENMESİ TEKNİKLERİ
KULLANILARAK ARAŞTIRILMASI**

(Investigating the Factors That Impact Tourists' Satisfaction with Environment-Friendly
Hotels Using Machine Learning Techniques)

YÜKSEK LİSANS TEZİ

Mahmud ALRAHHAL

Danışman: Doç. Dr. Ferhat BOZKURT

Erzurum
Mayıs, 2023



FEN BİLİMLERİ ENSTİTÜSÜ

Graduate School of Natural and
Applied Sciences

T.C.
ATATÜRK ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü Müdürlüğü
TEZ KABUL VE ONAY TUTANAĞI

**TURİSTLERİN ÇEVRE DOSTU OTELLERLE İLGİLİ MEMNUNİYETİNİ ETKİLEYEN
FAKTÖRLERİN MAKİNE ÖĞRENMESİ TEKNİKLERİ KULLANILARAK
ARAŞTIRILMASI**

Doç Dr. Ferhat BOZKURT danışmanlığında, Mahmud ALRAHHAL tarafından hazırlanan bu çalışma, 25/05/2023 tarihinde aşağıdaki jüri tarafından Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Bilim Dalı'nda Yüksek Lisans tezi olarak **oybirliği (3/0)** ile kabul edilmiştir.

Jüri Başkanı: Dr. Öğr. Üyesi Mete YAĞANOĞLU
Atatürk Üniversitesi

Aslı Islak İmzalıdır

Danışman: Doç. Dr. Ferhat BOZKURT
Atatürk Üniversitesi

Aslı Islak İmzalıdır

Jüri Üyesi: Dr. Öğr. Üyesi Işıl KARABEY AKSAKALLI
Erzurum Teknik Üniversitesi

Aslı Islak İmzalıdır

Enstitü Yönetim
Kurulunun/.../.... tarih
ve sayılı kararı.

Bu tezin Atatürk Üniversitesi Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili maddelerinde belirtilen şartları yerine getirdiğini onaylarım.

Prof.Dr. Saltuk Buğrahan CEYHUN

Enstitü Müdürü

Aslı Islak İmzalıdır

ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU

Yüksek Lisans Tezi olarak Doç. Dr. Ferhat BOZKURT danışmanlığında sunulan “TURİSTLERİN ÇEVRE DOSTU OTELLERLE İLGİLİ MEMNUNİYETİNİ ETKİLEYEN FAKTÖRLERİN MAKİNE ÖĞRENMESİ TEKNİKLERİ KULLANILARAK ARAŞTIRILMASI” başlıklı çalışmanın tarafımızdan bilimsel etik ilkelere uyularak yazıldığını, yararlanılan eserlerin kaynakçada gösterildiğini, Fen Bilimleri Enstitüsü tarafından belirlenmiş olan Turnitin Programı benzerlik oranlarının aşılmadığını ve aşağıdaki oranlarda olduğunu beyan ederiz.

Tez Bölümleri	Tezin Benzerlik Oranı (%)	Maksimum Oran (%)
Giriş	0	30
Kuramsal Temeller	15	30
Materyal ve Yöntem	7	35
Araştırma Bulguları ve Tartışma	2	20
Sonuçlar ve Öneriler	0	20
Tezin Geneli	7	25

Not: Yedi kelimeye kadar benzerlikler ile Başlık, Kaynakça, İçindekiler, Teşekkür, Dizin ve Ekler kısımları tarama dışı bırakılabilir. Yukarıdaki azami benzerlik oranları yanında tek bir kaynaktan olan benzerlik oranlarının %5'den büyük olmaması gerekir.

Sunulan bilgilerin doğru olduğunu, aksi halde doğacak hukuki sorumlulukları kabul ettiğimizi beyan ederiz.

Tez Yazarı (Öğrenci)	Tez Danışmanı
MAHMUD ALRAHHAL	Doç. Dr. Ferhat BOZKURT
25.5.2023	25.5.2023
İmza: Aslı Islak İmzalıdır	İmza: Aslı Islak İmzalıdır

* Tez ile ilgili YÖKTEZ’de yayınlamasına ilişkin bir engelleme var ise aşağıdaki alanı doldurunuz.

☐ Tezle ilgili patent başvurusu yapılması / patent alma sürecinin devam etmesi sebebiyle Enstitü Yönetim Kurulunun/....../.... tarih ve sayılı kararı ile teze erişim 2 (iki) yıl süreyle engellenmiştir.

☐ Enstitü Yönetim Kurulunun/....../.... tarih ve sayılı kararı ile teze erişim 6 (altı) ay süreyle engellenmiştir.

TEŞEKKÜR

Yüksek Lisans eğitimim süresinde bana zaman ayıran, bilgi ve deneyim birikimlerini esirgemeyen sayın hocam Doç. Dr. Ferhat BOZKURT danışmanıma şükranlarımı büyük mutlulukla sunmak istiyorum.

Ayrıca hayatım boyunca ve tezim süresince bana hem maddi hem de manevi olarak yardım eden annem, babam, ve tüm aileme teşekkür ederim.

Mahmud ALRAHHAL



ÖZET

YÜKSEK LİSANS TEZİ

TURİSTLERİN ÇEVRE DOSTU OTELLERLE İLGİLİ MEMNUNİYETİNİ ETKİLEYEN FAKTÖRLERİN MAKİNE ÖĞRENMESİ TEKNİKLERİ KULLANILARAK ARAŞTIRILMASI

Mahmud ALRAHHAL

Danışman: Doç. Dr. Ferhat BOZKURT

Amaç: Turizm endüstrisi sadece Türkiye ekonomisi için değil, aynı zamanda gezginler için de kritik öneme sahiptir. Oteller, özellikle çevreci oteller gibi turizm tesislerinin geliştirilmesi, hem çevreyi hem de Türkiye ekonomisini mutlaka olumlu yönde etkileyecektir. Son birkaç yılda, yeşil turizm artan bir ilgi görmektedir. Türkiye'de çevre dostu turizm bağlamında büyük sosyal veri analizi, çevre dostu otellerde seyahat edenlerin memnuniyetini etkileyen faktörleri anlamak için kritik öneme sahiptir. TripAdvisor gibi sosyal platformlarda sunulan sosyal verilerin çıkarılan boyutlarının değerlendirilmesi, otel yöneticilerinin ve karar vericilerin iyileştirmeleri gereken tesisleri bilmelerine yardımcı olacaktır. Bu çalışmanın amacı Türkiye'deki çevre dostu otellerdeki turistlerin deneyimlerini memnuniyetini etkileyen faktörleri makine öğrenmesi yöntemleriyle ortaya çıkarmaktır.

Yöntem: Çalışmanın amacını gerçekleştirmek için bilgiler, Tripadvisor web sitesinden tarama yöntemiyle alınmıştır. Kullanıcı tarafından oluşturulan içerikten memnuniyet yönlerini çıkarmak için makine öğrenimi teknikleri, özellikle Gizli Dirichlet Tahsisi (LDA) kullanılmıştır. Müşteri segmentasyonu için k-means kümeleme yöntemi kullanılmıştır. Son olarak Uzun Kısa Süreli Bellek (LSTM), Çift Yönlü Uzun Kısa Süreli Bellek (BiLSTM), ve Geçitli Tekrarlayan Birim (GRU) müşteri yorumlarının sınıflandırması için eğitildi ve karşılaştırıldı.

Bulgular: Bu çalışmada Türkiye'deki çevre dostu otellerdeki turistlerin deneyimlerini ve memnuniyetini etkileyen faktörlerin elde edilmesi için LDA konu modelleme tekniği kullanılmıştır. LDA ile 4 konu elde edilmiştir. Müşteri segmentasyon işlemi için K-means ile turistler 3 gruba ayrılmıştır. Yorum sınıflandırmada LSTM, BiLSTM, ve GRU kullanılmıştır. 96% doğruluk değeri ve 10% kayıp değeri ile en iyi performans gösteren BiLSTM modeli elde edilmiştir.

Sonuç: Sonuç olarak bu çalışmada turistlerin Türkiye'deki çevre dostu otellerle olan deneyimlerini etkileyen bir takım önemli faktörü ortaya koyulmuştur. Bu faktörlerin çıkarılması hem müşteriler için hem de işletme sahipleri için önemlidir ve ona göre işletme hizmetleri iyileştirilebilir. K-means kümeleme tekniği ile müşteriler 3 ayrı gruba (az memnun, orta memnun ve çok memnun) ayrılarak grup bazlı müşteri memnuniyeti araştırılmıştır. Son olarak LSTM, BiLSTM, ve GRU yorum sınıflandırma işlemi için karşılaştırılmıştır ve en yüksek doğruluk veren model elde edilmiştir.

Anahtar Kelimeler: Büyük Sosyal Veri Analizi, Çevrimiçi Müşteri İncelemeleri, Makine öğrenmesi, NLP, Konu modelleme

Mayıs 2023, 84 sayfa

ABSTRACT

MASTER'S THESIS

INVESTIGATING THE FACTORS THAT IMPACT TOURISTS' SATISFACTION WITH ENVIRONMENT-FRIENDLY HOTELS USING MACHINE LEARNING TECHNIQUES

Mahmud ALRAHHAL

Supervisor: Assoc. Prof. Dr Ferhat BOZKURT

Purpose: Tourism industry is critically essential for not only Turkish economy but also for travelers. Enhancing tourism facilities like hotels, particularly environment-friendly hotels will absolutely affect both environment in general and Türkiye's economy positively. Over the past few years, green tourism has drawn increasing focus. Big social data analysis in the context of environment-friendly tourism in Türkiye is critically important to comprehend the factors that impact travelers' satisfaction in eco-friendly hotels. Evaluating the extracted dimensions of social data presented on social platforms like TripAdvisor will help hotel managers and decision-makers to know the facilities that they should improve. The aim of this study is to reveal the factors that affect the satisfaction of tourists' experiences in environmentally friendly hotels in Turkey with machine learning methods.

Method: In order to fulfill the study's objective, the information was taken from the Tripadvisor website by means of a crawling method. To extract satisfaction aspects from user-generated content, machine learning techniques, particularly Latent Dirichlet Allocation (LDA), was used. The k-means clustering method was utilized for customer segmentation. Finally, Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (BiLSTM), and Gated Recurrent Unit (GRU) were trained and compared for customer review classification.

Findings: In this study, LDA subject modeling technique was used to obtain the factors affecting the satisfaction of the experiences of tourists in environmentally friendly hotels in Turkey. 4 subjects were obtained with LDA. For the customer segmentation process, the tourists were divided into 3 groups with K-means. LSTM, BiLSTM, and GRU were used for comment classification. The best performing BiLSTM model was obtained with 96% accuracy value and 10% loss value.

Results: As a result, in this study, a number of important factors that affect tourists' experience with eco-friendly hotels in Turkey are revealed. Extracting these factors is important for both customers and business owners, and business services can be improved accordingly. With the K-means clustering technique, customers were divided into 3 different groups (less satisfied, moderately satisfied and very satisfied) and group-based customer satisfaction was investigated. Finally, LSTM, BiLSTM, and GRU were compared for the interpretation classification process and the model with the highest accuracy was obtained.

Keywords: Big Social Data Analysis, Online Customers' Reviews, Machine learning, NLP, Topic modeling

May 2023, 84 pages

İÇİNDEKİLER

TEZ KABUL VE ONAY TUTANAĞI	i
ETİK BİLDİRİM VE İNTİHAL BEYAN FORMU	ii
TEŞEKKÜR	iii
ÖZET	iv
ABSTRACT	v
İÇİNDEKİLER.....	vi
TABLolar DİZİNİ.....	viii
ŞEKİLLER DİZİNİ.....	ix
KISALTMALAR VE SİMGELER DİZİNİ.....	x
GİRİŞ.....	1
KURUMSAL TEMELLER.....	4
Sosyal Verilerin Analizi.....	4
Çevrimiçi Kullanıcı Yorumları	4
Çevrimiçi Kullanıcı Yorumlarında Metin Madenciliği.....	5
Verideki Bilgi Keşfi	6
Veri Madenciliğinde Makine Öğrenmesi Yöntemleri.....	8
Denetimli (Supervised) Öğrenme	9
Destek vektör makinesi (Support vector machine)	9
Yapay sinir ağları (Artificial neural networks)	9
Naive bayes sınıflandırması (NBC)	10
Karar ağaçları (Decision Trees)	10
K-en yakın komşu (K-nearest neighbor).....	10
Tekrarlayan sinir ağları (Recurrent neural network RNN)	10
Uzun kısa süreli bellek (LSTM).....	11
Çift yönlü uzun kısa süreli bellek (BiLSTM)	11
Kapılı tekrarlayan hücre (Gated recurrent unit GRU).....	12
Denetimsiz (Unsupervised) Öğrenme	12
Hiyerarşik kümeleme (hierarchical clustering).....	13
K-ortalamalar kümeleme (K-means).....	13
Konu modelleme (Topic modelling).....	15
Çevrimiçi Kullanıcı Yorumları, Veri Madenciliği ve Müşteri Memnuniyeti ile İlgili Yapılan Bazı Çalışmalar	16
MATERYAL VE YÖNTEM	20

Veri Toplama	20
Metinsel veri toplama ve önişleme	21
Kriter derecelendirmeleri toplama	23
Veri Önişleme ve Temizleme	24
Konu Modelleme (LDA).....	26
Müşteri Segmentasyon	31
Yorum Sınıflandırma	33
Tekrarlayan sinir ağları (RNN)	33
Uzun kısa süreli bellek (LSTM).....	34
Çift yönlü uzun kısa süreli bellek (BiLSTM)	37
Kapılı tekrarlayan hücre (gated recurrent unit GRU)	39
Çapraz doğrulama	40
Sınıflandırma değerlendirme yöntemleri	41
ARAŞTIRMA BULGULARI	43
Veri Analizi ve Konu Modelleme	43
Müşteri Segmentasyonu	47
Yorum Sınıflandırma	48
LSTM model ile sınıflandırma.....	49
Çift yönlü LSTM model ile sınıflandırma	51
GRU model ile sınıflandırma	53
Sınıflandırma model performans değerlendirme	56
TARTIŞMA.....	59
SONUÇ ve ÖNERİLER	62
KAYNAKLAR.....	64
ÖZGEÇMİŞ.....	72

TABLÖLAR DİZİNİ

Tablo 1. Toplanan Kriter Derecelendirmelerin Bir Örneği	32
Tablo 2. Karışıklık Matrisi.....	41
Tablo 3. Konulardaki Kelimelerin Ağırlık Olasılığı.....	45
Tablo 4. Az memnun grubu istatistiksel bilgileri	47
Tablo 5. Orta memnun grubu istatistiksel bilgileri	47
Tablo 6. Çok memnun grubu istatistiksel bilgileri	47
Tablo 7. Müşteri Gruplarının Merkezleri (Centroids)	48
Tablo 8. LSTM Modeli Sonuçları.....	50
Tablo 9. BiLSTM modeli sonuçları	53
Tablo 10. GRU Modeli Sonuçları.....	55
Tablo 11. Negatif ve Pozitif Test Yorumların Sınıflandırılması Örnek Sonuç	57
Tablo 12. LSTM, BiLSTM, ve GRU model sonuçları	60

ŞEKİLLER DİZİNİ

Şekil 1. VBK süreç adımları.....	7
Şekil 2. K-means kümeleme algoritması akış diyagramı	15
Şekil 3. TripAdvisor sitesindeki bulunan çevrimiçi kullanıcı yorumlarının bir örneği	21
Şekil 4. Yeni elde edilen veri seti yorumlarının dağılımı	23
Şekil 5. TripAdvisor çevrimiçi kullanıcı yorumlarına ait kriter derecelendirmeleri.....	24
Şekil 6. TripAdvisor Çevrimiçi kullanıcı yorumlarına ait eksik girilmiş kriter derecelendirmeleri	25
Şekil 7. LDA üretici prosedürü (LDA generative procedure).....	28
Şekil 8. LDA örnek konu ve kelime dağılım temsili.....	29
Şekil 9. Araştırma metodu.....	30
Şekil 10. Elman hücre yapısı.....	33
Şekil 11. Tanh fonksiyonu	34
Şekil 12. Tekrarlayan sinir ağının çalışma prensibi	34
Şekil 13. LSTM birim yapısı.....	36
Şekil 14. Sigmoid fonksiyonu	37
Şekil 15. BiLSTM ağ yapısı	38
Şekil 16. GRU birim yapısı	40
Şekil 17. Çapraz doğrulama süreci.....	41
Şekil 18. Elde edilen başlıca konular	44
Şekil 19. Konu1’de oluşturulan kelime bulutu.....	44
Şekil 20. LSTM modeli eğitim ve doğrulama kayıp eğrisi	49
Şekil 21. LSTM modeli eğitim ve doğrulama doğruluk eğrisi	50
Şekil 22. LSTM modeli karışıklık matrisi.....	51
Şekil 23. Çift yönlü LSTM modeli eğitim ve doğrulama eğrisi.....	52
Şekil 24. Çift yönlü LSTM modeli eğitim ve doğrulama eğrisi.....	52
Şekil 25. Çift yönlü LSTM modeli karışıklık matrisi	53
Şekil 26. GRU modeli eğitim ve doğrulama eğrisi	54
Şekil 27. GRU modeli eğitim ve doğrulama eğrisi	55
Şekil 28. GRU modeli karışıklık matrisi.....	56

KISALTMALAR VE SİMGELER DİZİNİ

BiLSTIM	: Çift Yönlü Tekrarlayan Sinir Ağı (Bidirectional Recurrent Neural Network)
BK	: Bilgi Keşfi
BoW	: Kelime Torbası (Bag of Words)
CNN	: Evrişimli Sinir Ağları (Convolutional Neural Networks)
E-AAİ	: Elektronik Ağızdan Ağıza İletişim
GRU	: Geçitli Tekrarlayan Birim (Gated Recurrent Unit)
KKD	: Knowledge Discovery from Data
LDA	: Gizli Dirichlet Tahsis (Latent Dirichlet Allocation)
LSTM	: Uzun Kısa Süreli Bellek (Long Short-Term Memory)
ML	: Makine Öğrenmesi (Machine Learning)
NLP	: Doğal Dil İşleme (Natural Language Processing)
RNN	: Tekrarlayan Sinir Ağları (Recurrent neural network)
SEM	: Yapısal Denklem Modelleme (Structural Equation Modeling)
SEM-PLS	: Kısmi en küçük kareler yapısal eşitlik modellemesi (Partial Least Squares Structural Equation Modeling)
SOM	: Kendi Kendini Düzenleyen Haritalar (Self-Organizing Maps)
SVM	: Destek Vektör Makinesi (Support Vector machine)
TF-IDF	: Terim Frekansı-Ters Doküman Frekansı (Term Frequency-Inverse Document Frequency)
UNWTO	: Birleşmiş Milletler Dünya Turizm Örgütü
VBK	: Verideki Bilgi Keşfi

GİRİŞ

Türkiye’de turizm sektörü her geçen gün gelişmektedir. Türkiye, turistlerin tercihleri arasında önde gelen bir seçenektir. Türkiye’ye gidip dinlenmek turistlerin ilk tercihlerinde yer almaktadır. Dünyanın kalbinde Asya ve Avrupa arasında yer alan ve Afrika'ya da yakın olan Türkiye'nin coğrafi konumu nedeniyle pek çok gezgin tatilini Türkiye'de geçirmeyi tercih etmektedir. Bunun yanı sıra Türkiye’de turistlerin gezebileceği çok güzel yerler bulunmaktadır. Örnek olarak Uzun kıyılar, büyük ormanlar, ılıman hava koşulları, tarihi mekanlar (örneğin Kapadokya), kayak merkezleri (örneğin Erzurum) ve çevre dostu oteller; dünyanın dört bir yanından turistlerin ilgisini çekmektedir.

Turizm sektörü, hem Türkiye'de hem de küresel ölçekte, ülke ekonomisi ve doğa üzerindeki etkisiyle büyük bir öneme sahiptir. Milyonlarca turist turizm amaçlı etkinlikler gerçekleştirmek için her yıl Türkiye’ye giriş yapmaktadır. Sadece 2019 yılında, Türkiye'ye yaklaşık 34,5 milyar dolar gelirle toplam 51,7 milyon turist kaydedildi ve toplam turist sayısı açısından dünya altıncısı olmuştur (Tuna and Başdal 2021). Türkiye, çevre dostu otellere ev sahipliği yapması ve her bir bölgenin benzersiz doğal güzellikleri ve coğrafi yapısı sayesinde turistlere farklı aktivitelerle mutluluk seviyelerini yükseltme imkânı sunarak onları cezbediyor.

Türkiye’de yeşil otellerin bulunması ve turistlerin yapabilecekleri aktivitelerin çeşit sayısı az olmadığı için mesela yapabilecekleri aktivitelerin birkaç örneği: İstanbul boğazi çevresinde gezmek, İstanbul Eminönü’ndeki tarihsel yapıları izlemek, adalara gemi vasıtasıyla gitmek, Erzurum’da kayak yapmak, Kapadokya’da balon turu yapmak ve tarihsel mekânları gözlemlemek, Antalya hava sıcaklığından ve güneşinden yararlanmak ve Trabzon güzel doğasını keşfetmek gibi birçok aktivitede bulunabilirler. Turizm, doğayı gezmek ve farklı yerleri keşfetmek insanların ruh haline olumlu yönde etki etmektedir. Bu şekilde Turistler yaşam kalitelerini iyileştirmek için çoğu zaman Türkiye’yi tercih edebilmektedirler. Günümüzde insanlar, dinlenme, sağlık ve çevresel hizmetlerden yararlanmanın bir sonucu olarak yaşam standartlarını iyileştirdiği için doğayı keşfetmeye daha isteklidirler (Prihayati and Veriasa 2021).

Birçok kişi için turizm esastır ve giderek önem kazanmaktadır. Birleşmiş Milletler Dünya Turizm Örgütü (UNWTO), 2019'da 1,4 milyara yakın insanın seyahat ettiğini ortaya koyan bir araştırma yaptı (Streimikienė et al. 2021). İnsanların Turizm ve seyahat yapmadaki sayısı az olmadığı, Türkiye’de yeşil otellerin az olmadığı ve Türkiye’nin önemli bir konuma ve

çekici bir doğaya sahip olduğu için Türkiye çoğu zaman turistlerin seyahat hedeflerinden başında yer almaktadır.

Doğayı korumak ve turizmin doğaya verdiği olumsuz etkiyi azaltmak, turistlerin seyahatlerini sağlıklı bir biçimde yapmak, ve sağlıklı ortamlarda ve otellerde seyahatlerini geçirebilmek için yeşil turizmin önemi gittikçe artmaktadır. Turizm sektöründe yeşil turizm kavramı ortaya çıkmış ve artan bir ilgi görmüştür (Filimonau et al. 2022; Yeşiltaş, Gürlek, and Kenar 2022). Turizm alanında yeşil turizm turistlerin, otel yöneticilerinin ve karar verici yetkililerinin ilgisini çekmektedir. Çevresel sürdürülebilirliği ana pazarlama politikalarının kritik bir yönü olarak kullanan birçok kuruluşla birlikte, yeşil hizmetleri ve ürünleri benimsemek, büyüyen işletmelerde giderek daha fazla merkez noktası haline geldi (Chen, Hu, et al. 2022).

Çevre sorunlarının güçlü bir şekilde araştırılması ve bu tür sorunlarla başa çıkmak için çözümler aranması çoğu zaman turistlerin çevreyi kurtarmak için çevre dostu eylemler gerçekleştirmelerine yönelik güçlü şekilde teşvik eder (Han et al. 2018). Bu şekilde çevre dostu otellerin araştırılması, Advisor sitesinden alınan çevrimiçi kullanıcı yorumlarının analiz edilmesi ve yeşil otellere katkıda bulunulması turistlerin çevre dostu eylemler gerçekleştirmesine teşvik edebileceği için önemlidir. Ayrıca, sadece doğayı koruma, kullanıcıların yeşile oteller ile ilgili ilgilendikleri alanları keşfetme, yöneticilerin ve karar vericilere yardımcı olabileceği için yeşil sadece oteller iyileştirme konusunda değil, ayrıca doğayı korumada ve Türkiye ekonomisine katkıda bulunma açısından da önemlidir.

Sosyal büyük veri analizleri ve çevrimiçi müşteri yorumları, Türkiye'de çevre dostu otellerde yer alan birçok özellik hakkında müşterilerin beklentilerini keşfetmek için çok önemlidir. Çevre dostu bir otelin sunduğu hizmet hakkında müşterilerin ilk beklentileri öncelikli olarak önemlidir. Müşterilerin ilk beklentileri ile ürünle ilgili gerçek deneyimleri arasındaki ölçüm, müşteri memnuniyetini tanımlayabilir (Yu et al. 2022). Çevrimiçi yorumlardan elde edilen özellikler, hükümetlere ve karar vericilere, müşterinin ilgilendiği özelliklerin neler olduğunu bilmelerinde yardımcı olacaktır. Yeşil turizm ve çevre dostu otel sitelerine ilişkin çevrimiçi yorumları değerlendirmek, hem Türkiye turizm sektörü hem de gezginlerin memnuniyetini artırmak açısından büyük önem taşımaktadır. Araştırmacılar, çevrimiçi yorumları değerlendiren birçok yaklaşım ve yöntem yürütmüştür. (Ekinci and Omurca 2017) Türk otel yorumları üzerinde LDA konu modelleme uyguladı. (Polat B 2021) Türkçe ürün yorumları verisi ile duygu analizi gerçekleştirmiştir (Mengutaycı Ü 2021). Otel müşteri yorumları yapay sinir ağları kullanılarak sınıflandırmış ve bu yorumları duygusal

içeriklerine göre analiz etmiştir. Bu çalışmada diğer çalışmalardan farklı olarak Türkiye’deki çevre dostu otellere yönelik çevrimiçi yorumları ML yöntemleri kullanılarak araştırıldı.

Çevrimiçi yorumları elde etme ve madenciliği yapma, son birkaç yılda doğal dil işlemede araştırmacıların odak noktası olmuştur (Afrizal, Rakhmawati, and Tjahyanto 2019). Bu, turistler arasında çevrimiçi yorumların artan popüleritesi ile açıklanmaktadır, çünkü %90’ı bu yorumları seyahat kararına ulaşmak ve seyahatlerini planlamak için kullanmaktadır (Godnov and Redek 2016). Sosyal büyük verileri analiz etmek, hem müşteriler hem de işletme sahipleri için temel olarak çok önemlidir. Büyük sosyal veriler Türkiye’deki yeşil otellerle müşteri memnuniyetinin değerlendirilmesinde büyük bir etkiye sahiptir. Türkiye’de yeşil turizme yönelik çevrimiçi yorumların analizi, karar vericilere ve otel sahiplerine müşterilerin ilgi duyduğu özellikleri öğrenmelerine yardımcı olacaktır. Çok sayıda yorum barındıran otel sitelerinde bulunan bu paylaşılan çevrimiçi yorumları kullanmak önemlidir. Çünkü bir müşteri olarak otel hakkında herhangi bir fikir yazabilir, o oteldeki işletme sahiplerinden veya çalışanlardan herhangi bir baskıya maruz kalmadan istediği gibi otel ile ilgili görüşünü paylaşabilir. Miktar veya büyüklük açısından değerlendirilen birçok ürünün aksine, oteller gibi turizm yerleri tecrübe ile değerlendirilir (Zibarzani et al. 2022).

Sosyal büyük veriler büyük sayıda sosyal medya sitelerinde paylaşılan verilerden oluşur (Nilashi, et al. 2021). Sosyal büyük veri analizi, literatürde, makul varsayımlar elde etmek ve pazar taleplerini tanımlamak amacı ile farklı gelişmiş yaklaşımlar ve yöntemler kullanılarak gerçekleştirilmiştir. İnternette paylaşılan sosyal büyük verilerin ve çevrimiçi kullanıcı yorumlarının artması nedeniyle doğal dil işleme (NLP), yalnızca makine öğrenimi (ML) bilim adamları için değil, piyasadaki karar vericiler için de vazgeçilmez hale geldi. Türkiye çevre dostu otellerin TripAdvisor sitelerinde bulunan kullanıcı yorumlarından yararlanıp kullanıcıların ilgilendikleri alanları keşfetmek için bu çalışma Türkiye’de turizm sektörüne katkıda bulunacaktır. Ayrıca doğayı korumak için çevre dostu olan otelleri tercih eden kullanıcıların yorumlarını analiz etmek ve ilgilendikleri alanları göz önünde bulundurup iyileştirmek Türkiye’deki çevre dostu turizmi desteklemeye yardımcı olacaktır. Bu çalışmanın temel amacı, Türkiye’deki turistlerin TripAdvisor çevre dostu oteller sitelerinde çevrimiçi portallarda yayınladıkları içeriğe dayalı olarak deneyimlerini keşfetmektir. Bu tez sonucunda TripAdvisor sitelerindeki kullanıcı yorumlarından edilen yeşil otel kullanıcı memnuniyet boyutlarının doğaya ve Türk turizm sektörüne katkıda bulunabileceği düşünülmektedir.

KURUMSAL TEMELLER

Sosyal Verilerin Analizi

İnsanlar tarafından ham verileri anlamak, içselleştirmek, yorumlamak ve yeni bilgiler edinmek için bu veriler üzerine gerçekleştirilen uygulamalar dizisine genel olarak veri analizi denir (Martin-Rodilla et al. 2018). Ayrıca genel olarak veri analizi çalışmaları yapısının tamamına veri analitiği şeklinde atıfta bulunur ve var olan bilgiler hakkında sonuçlar elde etmek amacıyla verilerin satırlarının araştırılması olarak tanımlanabilir (Martin-Rodilla et al. 2018). Veri analizi teknikleri genel olarak birçok alanda temel olarak kullanılmakta ve büyük bir öneme sahiptir. Sosyal verilerin analizinde müşterilerin görüşlerini değerlendirilmesinde (Nilashi, et al. 2021), tıp alanında (Emerson, Kara, and Glassey 2020; Tolba and Al-Makhadmeh 2021), ekonomi ve finans alanında (Guo et al. 2020), ve tarımda (Kamilaris, Kartakoullis, and Prenafeta-Boldú 2017) veri analizi uygulamalarının bilim dünyasında çok büyük bir önemi bulunmaktadır.

İnternet sitelerinde oluşturulan sosyal verinin büyüklüğünün artmasıyla birlikte araştırmacıların ve karar vericilerin ilgisini artırmaktadır. Müşterilerin satın alma kararları, çevrimiçi müşteri yorumlarından büyük ölçüde etkilenir (Huang et al. 2019). Müşterilerin otellerin çeşitli yönleri için çevrimiçi kullanıcı yorumlarında yer alan tercihleri, yalnızca müşterilerin rezervasyon kararlarını etkilemekle kalmaz, aynı zamanda karar vericilerin sahip oldukları otellerin hizmet kalitesini sürekli olarak iyileştirmelerine yardımcı olur (Bian et al. 2022). Sosyal medyada bir ürün ile ilgili paylaşılan veriler her türlü insanın istediği görüşü istediği gibi paylaşabildiği için, oluşan doğruluğuna ve açıklığına güvenebileceğimiz için bu görüşleri analiz etmek ve değerlendirmek oldukça önemlidir. Sosyal medya iletişimi ile ilgili kırsal yerlerde yaşamının uzaklık dezavantajlarını ve insanların arasındaki mesafenin aşırı uzaklığı üstesinden gelebildiği ve sosyal refahı teşvik ettiği için sağladığı ayrık faydaların giderek artan bir küresel tanınırlığa sahiptir (Tiwari et al 2019).

Çevrimiçi Kullanıcı Yorumları

Sosyal medya, ekonomi, ticaret, siyaset ve eğitim gibi birçok araştırma alanında kullanılabilecek önemli bir etkileşim kaynağı sağlar (Bozkurt et al. 2019). Çevrimiçi yorumların kullanılabilirliği ve güvenilirliği, kullanıcı görüşlerinin onaylanmasını ve müşterilerin çevrimiçi perakendecilere güvenme eğilimini teşvik eder (Shaheen et al. 2019). Çevrimiçi

kullanıcı yorumları, fiyatlandırma stratejisinin belirlenmesinde ve getirilerin artırılmasında önemli bir etkiye sahiptir (Tian and Zhang 2022). Çevrimiçi kullanıcı yorumları, müşterilerin yalnızca fikirlerini yaymakla kalmayıp aynı zamanda deneyimlerini tartışabilecekleri güçlü bir iletişim ve elektronik ağızdan ağıza iletişim (E-AAİ) türünü temsil eder ve otel pazarlamasının güçlü bir şeklidir (Nilashi, et al. 2021).

Çevrimiçi kullanıcı yorumlarının güvenilirliği, tüketiciler tarafından okunmakta olan belirli bir yoruma karşı tüketicilerin olan güvenini gösterir (Wang et al. 2022). Bundan dolayı kullanıcıların o görüş veya yoruma verdiği güven ürün alma kararını büyük ölçüde etkilemektedir. kullanıcı Görüşlerini göz önünde bulundurarak ve o görüşler doğrultusunda aksiyon ve gerekli tedbirleri almak işletme, veya yeşil otellerin gelirini büyük oranda etkileyebilmektedir. Kısa bir zaman diliminde ürün satışları üzerinde çevrimiçi kullanıcı yorumlarının ve derecelendirme puanlarının güçlü bir etkisi bulunmaktadır (Ma et al. 2022). COVID-19 salgını sırasında havayolu şirketleri için yapılan çevrimiçi kullanıcı görüşleri araştırması, geri ödeme stratejileri ve prosedürleriyle bağlantılı çeşitli sorunlar nedeniyle son derece olumsuz sonuçlar ortaya çıkardı. Böylece, çevrimiçi kullanıcı görüşleri, karar vericilere, havayollarının COVID-19 salgını ciddi etkileriyle nasıl başa çıkabileceği konusunda müşterilerin bakış açısından bir algı sağlar (Rita, Moro, and Cavalcanti 2022).

Çevrimiçi Kullanıcı Yorumlarında Metin Madenciliği

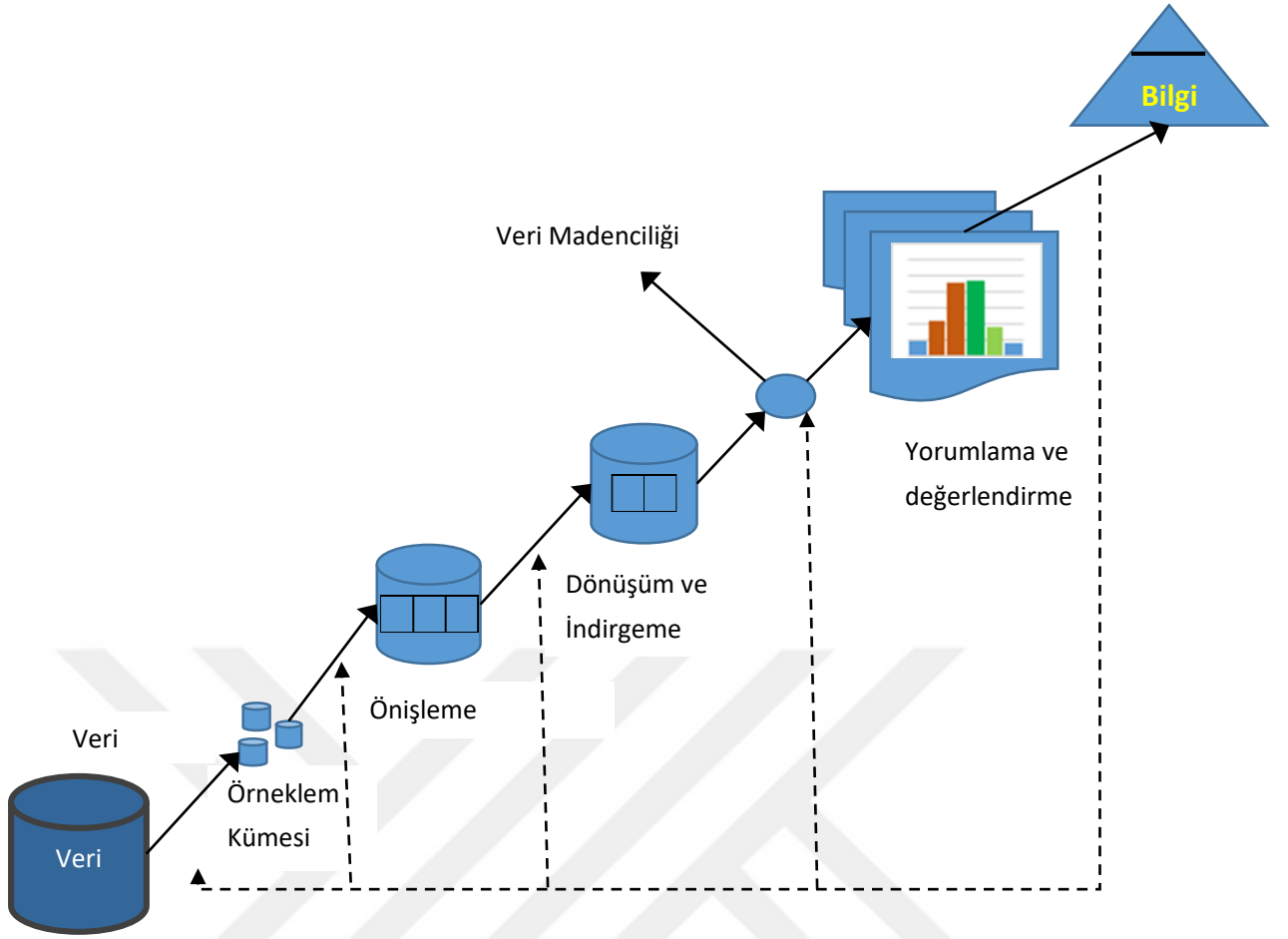
Büyük miktarlardaki veriden bilgi çıkarmak ya da madencilik yapma tekniğine veri madenciliği olarak işaret edilir (Krishnamoorthy and Karthikeyan 2022). Elde edilen verilerin belirlenmiş bir alanda bir hedefe erişmek için, o verilerin görülmeyen bağlantılarını keşfedilmesi operasyonu veri madenciliği olarak tanımlanabilir (Shalvi and DeClarıs 1998). Ayrıca Geleceğe dair tahminlerde bulunmak ve bu doğrultuda kararlar verebilmek konusunda yardımcı olabilir. Veri madenciliği terimi veya alanı genişliği nedeniyle veri tabanları, görselleştirme, bilgi toplama, makine öğrenmesi, istatistik, finans, market analizi gibi birden fazla bilim dalı veri madenciliğinden faydalanabilmektedir (Zhou 2003). Veri madenciliği veri setlerinde ve özellikle büyük veri setlerini oluşturan verilerden çıkarımda bulunmada ve verilerden istenilen bilgi ve görsel elde etmede önemli katkılar sağlamaktadır ve birçok alanda kullanılmaya devam edilmektedir. Büyük veri işlemek için veri madenciliği makine öğrenmesi ve istatistikten türetilmiş kavramları kullanır. Veri madenciliği, bilgi, iç görü ve veri anlayışı sağlar. Veri madenciliği, gelecekteki gözlemleri tahmin etmenin yanı sıra, verilerdeki temel ilişkiyi özetlemede etkilidir (Krishnamoorthy and Karthikeyan 2022). Büyük veri setlerinden saklı kalmış önemli bilgileri ortaya çıkarıp elde edilen veriler doğrultusunda karar vermede

destek sağlanmaktadır ve bu şekilde veri madenciliği popülaritesi büyük oranda yükselmeye devam etmektedir.

TripAdvisor gibi turizm sitelerinde verilerin ve kullanıcıların oluşturduğu içeriklerin gittikçe artması nedeniyle bu veri ve içerikler büyük bir öneme sahiptir. Böylece çevrimiçi kullanıcı yorumları ve içeriklerini oluşturan metinlerin madenciliği büyük bir önem taşımaktadır. TripAdvisor sitesi seyahat planlama ve rezervasyonda çevrimiçi deneyimi oluşturmaya yardımcı olmakla beraber, tüketicilerin gidecekleri hedefleri ve deneyimleri araştırmalarını, kullanıcı tarafından oluşturulan içeriği okumalarını ve katkıda bulunmalarını, turizm hedefleri ve işletmeleri kalite, fiyat ve müsaitlik temelinde karşılaştırmalarını ve rezervasyonları tamamlamalarını kolaylaştırmaya odaklar (SEC 2019). TripAdvisor siteleri gibi yeşil oteller hakkında kullanıcı içeriklerini, görüşlerini ve seyahat deneyimlerini oluşturan metinlerin madenciliği yeşil otellerin hizmetlerini ve turistlerin deneyimlerini iyileştirmek için oldukça önemlidir.

Verideki Bilgi Keşfi

Bilgilerin hacmi her gün artan bir şekilde büyümektedir. İnternet sitelerindeki sosyal verilerden, internet sitelerindeki verilerden, Sensörlerden, bilimsel verilerden, ver resim ve videolardan gelen veriler her geçen gün artmaktadır. Bundan dolayı bu verilerden anlamlı bilgiler ortaya çıkarabilecek bir sistemin geliştirilmesine ihtiyaç duyulur. Verideki bilgi keşfi veri setlerindeki bilgilerden faydalı bir çıkarım veya anlam keşif işlemine işaret edilir ve genellikle VBK (KKD: Knowledge Discovery from Data) kısaltması ile bilinir. Veri madenciliği, VBK süreci adımlarında bir adım olarak yer almaktadır. BK sürecinin bir adımı olan veri madenciliğinin görevi veriden anlamlı sonuç elde etmek için verilen veriye özel yapay zekâ modellerinin tasarlanarak geliştirilmesi, verilerin görselleştirilmesi, istatistik formüllerin uygulanması ve makine öğrenmesi modelleri geliştirilmesi gibi birçok alanın kesiştiği bir alan olup gelişmeye ve birçok araştırma alanında kullanılmaya devam etmekte olan bir tekniktir.



Şekil 1. VBK süreç adımları

VBK büyük ölçüde yinelemeli, etkileşimli ve Şekil 1’de görüldüğü gibi birden fazla adımdan oluşan bir prosestir (Imberman 2001).

1. Adım: Veriler bir veya birden fazla konumdan toplanır ve toplanan veriler merkezi bir konuma yerleştirilir. Bu merkezi konumlara bazen veri deposu ya da veri mart olarak adlandırılırlar. Bu aşamada farklı konumdan toplanan verinin, veri deposuna yerleştirilmeden önce veri içindeki varsa tutarsızlıkların giderilmesi için veri üzerinde düzenleme işlemi yapılabilir.
2. Adım: Hedef verinin oluşturulması: Veri toplanıp düzenlendiğinde, bu adımda üzerinde daha çok analiz gerçekleştirilecek veri seçim işlemi yapılır. Seçilmiş olan veri hedef veri setini oluşturur. Veri analisti, hedef veri setini oluştururken ilgi alanını, son kullanıcıların ihtiyaçlarını ve veri madenciliği işlemi ne olabileceğini anlaması önemlidir.
3. Adım: Veri önişleme ve temizleme: veri bazen özel amaçlı toplanır. Toplanan veriler bazen kayıp, bilinmeyen veya veri girişi hatalarından oluşan bir veri içerebilir. Önişleme ve temizleme prosesi sırasında gürültü kaldırılabilir, sapan veri silinir, kayıp veri alanları için özel işlemler yapılır.

4. Adım: Veri dönüşüm ve indirgeme: Bu adımda verileri temsil edebilecek yararlı özellikler bularak veri değişkenlerinin sayısının azaltılması amaçlı dönüşüm teknikleri kullanılır.
5. Adım: Veri madenciliği: Veri madenciliği VTBK’da bir adım oluşturur. Bu adımda ilgi gören örüntüleri ortaya çıkarmak için arama işlemi gerçekleştirilir. İstenilen örüntülerin ortaya çıkarılması için veri madenciliği işlemi, model temsili ve veri üzerinde hangi analizin gerçekleştirileceği bağlamında arama işlemi yapılır.
6. Adım: Veri madenciliği algoritmasının seçilmesi: Bu aşamada verilen veri üzerinde hangi veri madenciliği algoritmasının ve görevinin uygulanacağı kararı vermek oldukça önemlidir (yapay sinir ağları, lineer/lojistik regresyon, sınıflandırma, segmentasyon vb.).
7. Adım: Elde edilen örüntülerin yorumlanması: Veri madenciliği adımıyla elde edilen örüntüler uygun veya önemli örüntüler olmayabilir, bu şekilde önceki adımların herhangi birine dönülmesi ihtimali göz önünde bulundurularak veri madenciliği vasıtasıyla elde edilen örüntüler üzerinde yorumlama işlemleri yapılır.
8. Adım: Elde edilen bilgilerin raporlanması ve depolanması: İstenilen iyi örüntülerin ortaya çıkarılması ardından bu örüntüler son kullanıcılara rapor edilir. Raporlama işlemi tipleri değişebilir metin halinde, görsel olarak, grafik, diyagram vb. Bu şekilde rapor ilgili birime veya son kullanıcıya iletilir (Imberman 2001).

Veri Madenciliğinde Makine Öğrenmesi Yöntemleri

Makine öğrenmesi, bilgisayar sistemlerinin, veriden öğrenerek veri analitik sürecindeki performanslarını kademeli olarak geliştirmek için bilgisayar bilimi ve istatistiksel tekniklerin kullanıldığı süreci ifade eder (Shu X 2022). Makine öğrenmesi, bilgisayarların verilere dayalı bir şekilde nasıl öğrenebileceğini (veya performanslarını iyileştirebileceğini) ele alır (Han, Pei, and Kamber 2011). Makine öğrenmesi ana araştırma alanlarından biri, bilgisayar programlarının karmaşık biçimleri otomatik bir şekilde ayırt etmeyi öğrenmesi ve verilere dayalı zeki kararlar icra etmesidir (Han, Pei, and Kamber 2011).

Geliştirilen makine öğrenmesi (Machine Learning ML) modelinin başarısı aşağıdaki gibi faaliyetlere bağlıdır: Uygun özniteliklerin modele verilmesi, öznitelik mühendisliği ve örnek olarak birden fazla kaynaktan elde edilen verilerin özniteliklerini çıkarılması ve modele giriş olarak verilmesi işlemine öznitelik çıkarma adı verilir (Lepakshi V. 2022).

Makine öğrenmesi artan bir şekilde gelişmekte olan bir tekniktir. Veri madenciliği ile ilgili en fazla ilgi gören makine öğrenmesi teknikleri şu şekilde değerlendirilebilir:

Denetimli (Supervised) Öğrenme

Denetimli öğrenme temel olarak sınıflandırma ile aynı anlamı taşımaktadır. Sınıflandırma işleminde denetleme, keskin bir hedefe veya farklı bir adıyla etiketli veriye sahip olan öğrenme setleri üzerinde yapılır. Eğitim verileri üzerindeki tahminleri aşamalı olarak iyileştirir, bunları gerçek sonuç değişkenleriyle karşılaştırır ve denetimli geri bildirimle düzeltmeler yapılır (Shu X 2022).

Denetimli öğrenmede örnek olarak bir hastanın hasta olup olmadığını öğrenmek için geliştirilen model tarafından önce etiketli veri seti üzerinde öğrenme gerçekleştirilir ve test, doğrulama işlemleri yapıldıktan sonra etiketi olmayan ya da hastalık durumu bilinmeyen bir kişiden alınan verilerden hastanın hastalık durumu hakkında sonuç alınır. Bu işlem aynı şekilde metin madenciliği ve metin sınıflandırma işlemlerinde benzer şekilde uygulanır.

Denetimli öğrenme (Supervised Learning) yaygın olan yöntemleri:

- Destek vektör makinesi (Support Vector Machine)
- Yapay sinir ağları (Artificial Neural Networks)
- Naif Bayes sınıflandırıcısı (Naive Bayes Classifier)
- Karar ağaçları (Decision Trees)
- K-en yakın komşu (K-Nearest Neighbor)

Destek vektör makinesi (Support vector machine)

Destek vektör makinesi (Support Vector Machine) hem doğrusal hem de doğrusal olmayan verilerin sınıflandırılması için bir yöntemdir (Shu X 2022). Özet olarak bir Destek Vektör Makinesi (Support Vector machine SVM) sınıflandırıcının eğitimi, karar yüzeyi olarak (decision surface), pozitif eğitim örneklerini en büyük uzaklığını bularak negatif olanlardan ayıran bir hiperdüzlem bulmaya çalışır (Sun, Lim, and Liu 2009). SVM veri madenciliğinde bir çok alanda kullanılmıştır (Borkar et al. 2019; Chen et al. 2011).

Yapay sinir ağları (Artificial neural networks)

Yapay sinir ağları, insanların beyin hesaplama mantığını taklit ederek oluşturulmuştur. İnsandaki algı yapısı aslında bir sensörlere benzetilebilir (Göz, kulak gibi) ve insanların sinir ağları vasıtasıyla kendi kendine öğrenebiliyorlar. Bilgisayara da benzer bir şey uygulanabilirliğinden yapay sinir ağları icat edilmiştir. Bir sinir ağı, sınıflandırma için kullanıldığında, temel olarak, birimler arasında ağırlık sahibi bağlantılara sahip bir nöron benzer işlem birimlerinin bir koleksiyonudur (Shu X 2022). Sağlam bir model oluşturabilmek için modeli eğitilirken, eğitim dengesini korumak oldukça önemlidir. Çok fazla öğrenen

modellerde aşırı öğrenme (overfitting) durumu meydana gelir. Az öğrenme gören modeller ise yanlış tahmin verebilmektedir (Bozkurt 2012).

Naive bayes sınıflandırması (NBC)

Naive Bayes Sınıflandırması (NBC), sınıflandırmada Bayes teorisini uygulayan veri madenciliği tekniklerindeki algoritmalarından biridir (Luke 2015). Bayes sınıflandırıcıları istatistiksel sınıflandırıcılardır. Belirli bir özelliğin belirli bir sınıfa ait olma olasılığı gibi sınıf üyeliği olasılıklarını tahmin edebilirler (Shu X 2022).

Karar ağaçları (Decision Trees)

Karar ağaçları, düğümler ve dallardan oluşan bir bilgi temsil yapısına sahiptir, her bir dahili düğüm, bir veya daha fazla nitelik üzerinde bir sorgu ile ilişkilendirilir; her dal bir sorgu sonucunu temsil eder; ve her yaprak bir sınıf etiketi ile tasarlanır. Karar ağaçları diyagram görsel şekli desteğiyle SQL sorgusuna basitçe dönüştürme olanağı sağlar ve C 4.5, C5.0, C&RT ve CHAID karar ağaçlarında en yaygın olarak kullanılan yöntemlerdir (Bozkurt 2012).

K-en yakın komşu (K-nearest neighbor)

K-en yakın komşu, orijinal olarak Cover ve Hart tarafından önerilen bir sınıflandırma algoritmasıdır ve son yıllarda, parametrik olmayan bir regresyon yöntemi olarak yaygın yaygın bir şekilde kullanılmaktadır (Lin, Lin, and Gu 2022). K – en yakın komşu, çok çeşitli problemlere uygulanabilen olağanüstü bir sınıflandırma algoritmasıdır. Bu algoritma, test örneğine olan mesafeye göre komşuları tanımlamak için yeterli eğitim ve bu test örneğinin ait olduğu sınıfı belirlemek için yeterli test adımları gerektirir (Ghiasi, Ng, and Sheikh 2022).

Tekrarlayan sinir ağları (Recurrent neural network RNN)

RNN (Tekrarlayan Sinir Ağı), girdi verileri içindeki zaman bağımlılıklarını işlemek için tasarlanmış bir yapay sinir ağı türüdür. Bu zaman bağımlılıkları, önceki girdi verileriyle ilişkilendirilen bir dizi çıktı verisini oluşturur. RNN, diğer yapay sinir ağı türlerinden farklı olarak, girdi verilerindeki zaman bağımlılıklarını yakalar ve bu bilgiyi daha sonra kullanarak çıktıları üretir. RNN'nin bu özelliği, doğal dil işleme, konuşma tanıma, zaman serileri analizi ve müzik yapay zekası gibi uygulamalarda yaygın olarak kullanılmasını sağlamıştır (Khaldi et al. 2023; Chimmula and Zhang 2020).

RNN'nin temel yapısı, bir dizi hücre ve bağlantıdan oluşur. Her hücre, bir girdi verisini ve bir önceki hücreden gelen çıktıyı alır. Daha sonra, bu girdi verisi ve önceki çıktı, hücrenin

içindeki bir işlemle birleştirilir ve bir sonraki hücreye geçirilir. Bu işlem, ağı her bir hücresinde tekrarlanır ve son çıktı elde edilene kadar devam eder.

RNN'nin en önemli avantajlarından biri, NLP gibi metin verilerini işlemek için kullanılabilecek olmasıdır. Metin verileri, her kelimenin birbirine bağlı olduğu bir zaman serisi olarak düşünülebilir. Bu nedenle, RNN, metin verilerini analiz etmek için özellikle etkilidir.

RNN, LSTM ve GRU gibi türevleri ile birlikte, derin öğrenme alanında önemli bir rol oynamaktadır ve birçok uygulama için kullanılmaktadır.

Uzun kısa süreli bellek (LSTM)

LSTM kısa ve uzun vadeli belleği yönetmek için tasarlanmış bir tür yapay sinir ağıdır. LSTM, diğer geleneksel yapay sinir ağı türlerinden farklı olarak, zaman serisi verileri ve NLP gibi uzun vadeli bağımlılıkları işlemek için özellikle etkilidir. LSTM'nin ana amacı, geleneksel yapay sinir ağlarında görülen "gradientsizlik" problemini çözmektir. Bu probleme, ağı eğitim sürecinde, girdi verilerinin uzun vadeli bağımlılıkları hakkında yeterli bilgiye sahip olmadığından kaynaklanır. LSTM, bu probleme çözüm olarak hafıza hücreleri ve kapı mekanizmaları kullanır (Liang and Niu 2022).

Hafıza hücreleri, ağı uzun vadeli bilgiyi depolamasına ve yeniden kullanmasına olanak tanır. Kapı mekanizmaları ise, hafıza hücrelerinin ne kadar bilgi depolayacağına karar verir. Bu kapılar, hafıza hücreleri tarafından depolanan bilginin, girdi verileriyle birlikte nasıl güncelleneceğine karar verir.

LSTM, dil işleme, metin tahmini, konuşma tanıma ve zaman serileri analizi gibi birçok uygulama için kullanılabilir. LSTM'nin kullanımı, derin öğrenme alanında büyük bir ilgi uyandırmıştır ve halen aktif bir araştırma konusu olarak kabul edilmektedir.

Çift yönlü uzun kısa süreli bellek (BiLSTM)

BiLSTM, "Bidirectional Long Short-Term Memory" kelimelerinin kısaltmasıdır ve doğal dil işlemede NLP sıklıkla kullanılan bir derin öğrenme modelidir. LSTM, uzun vadeli bağımlılıkları işleyebilme yeteneği ile öne çıkan bir RNN çeşididir. BiLSTM, iki yönlü (ileri ve geri) LSTM'lerin birleşimidir ve hem önceki hem de sonraki girdilere dayalı tahminler yapabilen bir modeldir.

BiLSTM, çeşitli NLP görevleri için kullanılabilir, örneğin doğal dil anlama (NLU), metin sınıflandırma, dil modelleri ve dil çevirisi. BiLSTM, metnin her bir kelimesini önceki ve sonraki bağlamları ile birlikte işleyerek, kelimenin anlamına daha iyi bir şekilde erişebilir. Bu, özellikle, doğal dildeki anlamsal bağlamları yakalamak için önemlidir.

BiLSTM, eğitim verileriyle öğrenilir ve daha sonra yeni girdiler için tahminler yapmak için kullanılabilir. Derin öğrenme modellerinin gücü, veri büyüklüğü arttıkça daha iyi performans göstermeleridir, bu nedenle BiLSTM gibi modeller, büyük miktarda veri kullanılabilirse NLP problemlerinin çözümü için özellikle yararlıdır.

Kapılı tekrarlayan hücre (Gated recurrent unit GRU)

GRU, "Gated Recurrent Unit" kelimelerinin kısaltmasıdır ve NLP'de kullanılan bir derin öğrenme modelidir. GRU, uzun vadeli bağımlılıkları işleyebilme yeteneği ile öne çıkan bir RNN çeşididir.

GRU, LSTM'den farklı olarak, daha az parametre ve daha az hesaplama ile benzer bir performans sergiler. GRU, LSTM'de bulunan ayrı kapıların bir kısmını bir araya getirerek ve bazı sınırlamalar ekleyerek bir basitleştirme sağlar, GRU'nun LSTM kadar karmaşık olmamasına rağmen, benzer şekilde uzun vadeli bağımlılıkları işleyebilmesini sağlar (Jung, Lee, and Tani 2018). GRU, NLP'deki çeşitli görevler için kullanılabilir, örneğin doğal dil anlama (NLU), metin sınıflandırma, dil modelleri ve dil çevirisi. GRU, özellikle daha küçük veri setleri için uygundur, ancak büyük veri setleriyle de kullanılabilir.

GRU'nun LSTM'den daha hızlı eğitim süreleri ve daha az hesaplama maliyeti vardır, ancak performans açısından LSTM'ye göre bazı durumlarda geri kalabilir. Bu nedenle, kullanılacak modelin özelliklerine ve kullanım amacına bağlı olarak, LSTM veya GRU kullanmak daha uygun olabilir.

Denetimsiz (Unsupervised) Öğrenme

Kümeleme, belirli bir benzerlik ölçüsü kullanarak denetimsiz bir veri gruplandırma yöntemidir (García and González 2004). Kümeleme algoritmaları, etiketlenmemiş özellik vektörlerini kümeler veya "doğal gruplar" halinde düzenlemeye çalışır, öyle ki bir küme içindeki numuneler, farklı kümelere ait numunelerden "daha benzer" olur (García and González 2004). Kümeleme önemli ve oldukça kullanışlı bir araçtır ve veri madenciliğinde birçok konuda uygulanmıştır (Sato et al. 2019 et al 2015; Pliakos and Vens 2018).

Denetimsiz öğrenmenin amacı, araştırmacıların yalnızca girdi verilerine sahip olduğu ve karşılık gelen çıktılarına veya sonuç değişkenlerine sahip olmadığı durumlarda temel yapıyı verilerden modellemektir. Sürece denetimsiz öğrenme denir, çünkü tahmin edilecek açık sonuçların yokluğunda denetim veya öğretim gerçekleşmez. Algoritmalar, verilerden geri bildirim almadan, verilerdeki ilginç yapıları keşfetmeyi kendi kendilerine öğrenirler (Shu X 2022). Denetimsiz öğrenmenin uygulandığı alanların bir örneği: Müşterilerin bir mağazadaki harcamalarına göre gruplara ayrılmasıdır.

Denetimsiz öğrenme (Unsupervised Learning) yaygın olan yöntemleri:

- Hiyerarşik kümeleme (Hierarchical Clustering)
- K-ortalamlar kümeleme (K-means)
- Konu modelleme (Topic Modelling)

Hiyerarşik kümeleme (hierarchical clustering)

Hiyerarşik kümeleme, verileri bazı benzerlik ölçütlerine dayalı olarak gruplara ayırmaya yönelik denetimsiz bir öğrenme algoritmasıdır; etiketlenmemiş özellik vektörlerini kümeler halinde gruplandırmanın bir yolunu bulmayı amaçlar, böylece bir küme içindeki örnekler birbirine benzer ancak diğer kümedekilerden farklı olur (Zhong et al 2022). Hiyerarşik kümeleme, her bir gözlemi kendi içinde ayrı bir küme olarak ele alarak başlar ve daha sonra bu gözlemleri daha büyük bir kümede gruplandırır (Zhong et al 2022; Wu, Peng, et al. 2021). En yakın iki kümeyi belirleyerek ve en benzer iki kümeyi tüm kümeler bir araya gelinceye kadar birleştirerek, sonunda hiyerarşik bir kümeleme ağacı oluşturularak böyle bir işlemi tekrar tekrar gerçekleştirir (Zhong et al 2022).

K-ortalamlar kümeleme (K-means)

Kümeleme, makine öğreniminde denetimsiz öğrenme algoritmasına aittir (Zhang, Li, et al. 2022). K-ortalamlar kümeleme, birden fazla tekrarlama (iteration) yoluyla en uygun küme merkezini elde eden bir algoritmadır ayrıca veri madenciliği, makine öğrenimi, bilgi alma vb. alanlarda çok sayıda uygulamaya sahip bir kümeleme analizi algoritmasıdır (Zhang, Li, et al. 2022). K-ortalamlar kümelemesinde her grup, "ortalama" (mean) veya "merkez" (centroid) olarak adlandırılan grubun merkezi bir miktarı ile tanımlanır ve değerler, en yakın merkeze sahip gruplardan birine dağıtılır (Ida et al 2022).

Bir veri kümesi verildiğinde, kümeleme algoritması, verilerin özelliklerine göre veri noktalarını farklı gruplara eşler. K-ortalama kümeleme algoritması, verileri birden fazla yinelemeden (iteration) sonra k kategorisine böler. K-ortalama kümeleme yöntemi iki adımdan oluşur (Zhang, Li, et al. 2022):

1. Küme merkezini başlatılması

Kümeleme koşullarına göre uygun bir başlatma yöntemi seçilir. En yaygın yöntem, ilk küme merkezi olarak veri noktalarından rastgele k nokta seçmektir (Zhang, Li, et al. 2022).

2. Lloyd'un Yinelemesi (Lloyd's Iteration)

Yinelemenin ilk adımında, her bir veri noktasından küme merkezine olan mesafeyi hesaplanır ve ardından bu veri noktasını en yakın küme merkezine atılır. İkinci adımında ise,

küme merkezlerini güncelleme yöntemi, her kümedeki veri noktalarına karşılık gelen öznitelik değerlerinin ortalama değerini hesaplanır. Yinelemenin son adımında, iki yinelemenin küme merkezleri arasındaki mesafe hesaplanır. Mesafe değeri belirli bir eşikten küçükse iterasyon sonlandırılır, aksi halde iterasyona devam edilir (Zhang, Li, et al. 2022).

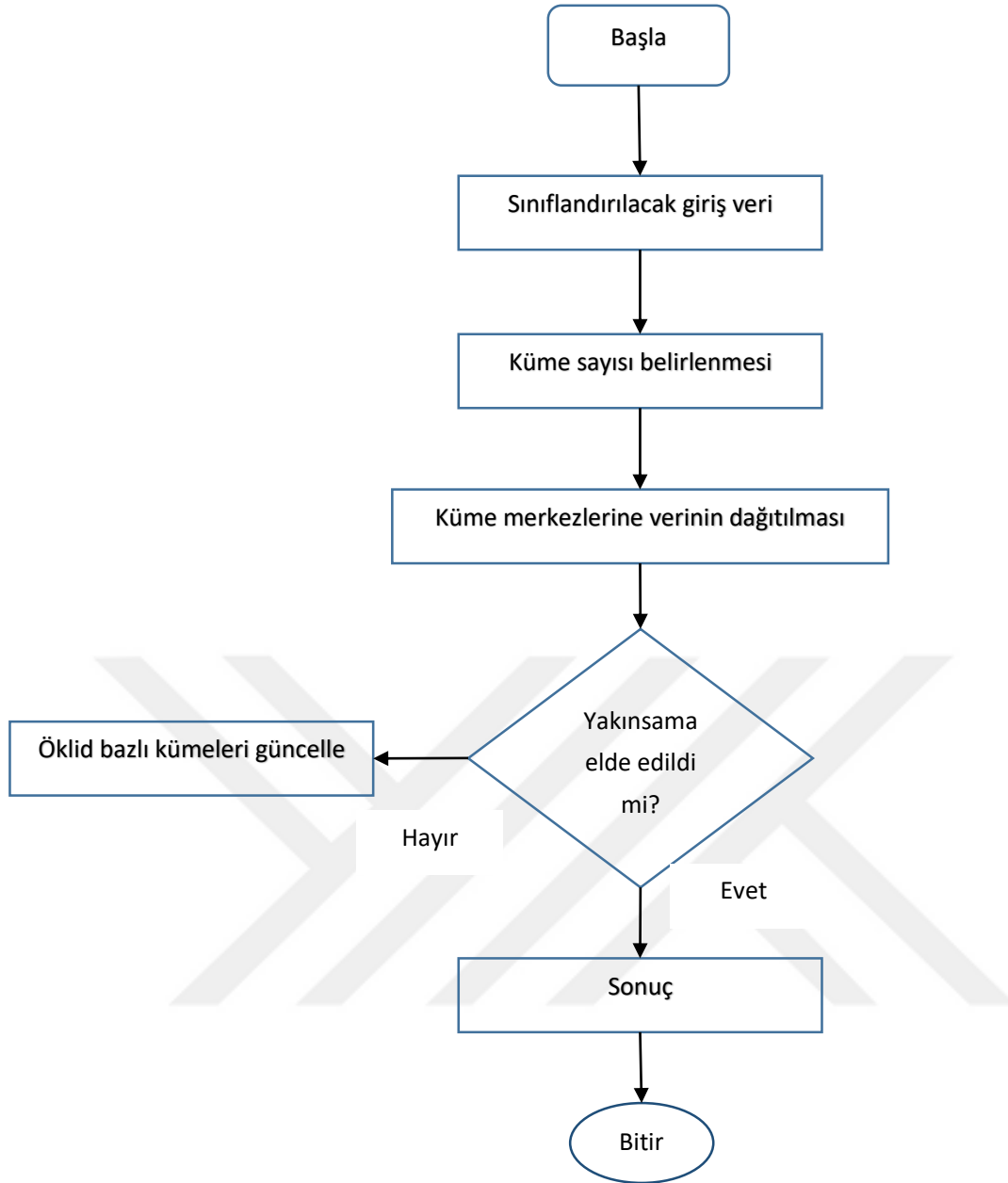
K-ortalamlar algoritması:

Giriş:

- $D = (d_1, d_2, d_3, \dots, d_i, \dots, d_n)$: n veri noktasının kümesi
- k: Küme sayısı

Çıkış:

- K sayısı kadar küme
1. Başlangıç küme merkezleri olarak rastgele D'den K nesne seçilir.
 2. Döngüye gir:
 - a. Her bir veri noktası d_i için:
 - i. d_i 'nin en benzer olduğu kümenin merkezini bul ve d_i 'yi bu kümeye yeniden ata.
 - b. Her bir küme için:
 - i. Kümedeki nesnelerin ortalama değerini hesapla ve küme merkezini güncelle.
 - c. Eğer küme atamalarında bir değişiklik olmadıysa döngüden çık.
 3. Algoritma tamamlandı.(Mohamed and Celik 2022).



Şekil 2. K-means kümeleme algoritması akış diyagramı (Mohamed and Celik 2022)

Konu modelleme (Topic modelling)

Veri madenciliği ve metin madenciliği denildiğinde konu modelleme akla gelen ilk yöntemlerden biridir. Konu modelleme metin madenciliğinde vazgeçilmez ve oldukça yaygındır (Ahadh et al 2021; Alghamdi and Alfalqi 2015), Biyoloji ve sağlık ile ilgili veri madenciliğinde konu modelleme (Narasimhulu and Abarna 2022; Albahri et al. 2020; van Altena et al. 2016), (recommender systems) tavsiye sistemlerinde (Zhang and Zhang 2022) ve Turizm alanında özellikle otellerin ve lokantaların müşteri yorumlarının değerlendirilmesinde ve madenciliğinde ayrıca konu modelleme tekniklerinden yaygın bir şekilde faydalanılmaktadır (Gao et al. 2022; Hu et al. 2019; Srivastava and Kumar 2021; Zibarzani et al. 2022). İnternet ortamında paylaşılan metinleri alanında önemli bir araştırma alanı olan konu modelleme, büyük

miktarda metin içeriğindeki gizli konuları tanımlayabilen önemli ve güçlü bir araçtır (Wu et al. 2023).

Birçok alanda konu modelleme tekniklerinden geniş bir şekilde faydalandığından konu modellemenin bilim ve araştırma alanlarında büyük bir önemi bulunmaktadır. Olasılıksal Gizli Semantik İndeksleme (PLSI) ve LDA modelleri, metin belgelerinin veya tweet'lerin gizli konularının türetilmesinde yaygın olarak kullanılmaktadır (Narasimhulu and Abarna 2022). LDA konu modeli, büyük ölçekli verileri işlemek için verimli olasılıksal çıkarım algoritmaları kullanan işbirlikçi filtreleme, metin sınıflandırma ve belge modelleme için denetimsiz bir makine öğrenimi modelidir (Wu et al. 2023).

Çevrimiçi Kullanıcı Yorumları, Veri Madenciliği ve Müşteri Memnuniyeti ile İlgili Yapılan Bazı Çalışmalar

(Zibarzani et al. 2022) tarafından yapılan çalışmada COVID-19 krizi dönemi içerisinde lokantalarda müşterilerin memnuniyeti değerlendirmek için çevrimiçi kullanıcı yorumları kullanarak ve lokantaların aldıkları COVID-19 tedbirlerini müşteri üzerindeki etkisini değerlendirerek iki aşamalı bir metod geliştirildi. Müşterilerin tercihlerini kümelemek için kümeleme, denetimli öğrenme ve (LVQ) kullanarak bir hibrit bir yöntem geliştirildi. Son olarak lokantaların kalite faktörleri ve müşteri memnuniyeti arasındaki ilişkiyi incelemek için (Partial Least Squares PLS) tekniği kullanıldı

(Massive Open Online Courses MOOCs) Kurs veren bir online sitede eğitim gören öğrencilerin aldıkları eğitimden memnuniyet seviyelerini değerlendirmek için (Nilashi, et al. 2022) tarafından bir çalışma yapıldı. Bu çalışma, öğrencilerin (MOOCs) ile memnuniyet seviyesini iyileştirebilecek faktörleri araştırmak adına hem makine öğrenmesi hem de anket tabanlı metodolojileri bir araya getiren yeni bir eğitimsel veri madenciliği yaklaşımı sunarak öğrenim görenlerin (MOOCs)'den memnuniyet seviyelerini ortaya çıkarmayı hedeflemektedir. Bu çalışma amacına ulaşmak için öğrenci tarafından oluşturulan içeriğin analiz edilmesi için LDA kullanıldı. Veri segmentasyonu için (self-organizing maps SOM) kullanıldı. Uyarlanabilir sinirsel bulanık çıkarım sistemi tabanlı (ANFIS) belirlenen faktörlerden öğrencilerin memnuniyetini tahmin etmek için kullanıldı. Araştırma modelinin validasyon işlemi yapmak ve modelin güvenilirliğini ispat etmek için (PLS-SEM) kullanıldı.

Çevrimiçi kullanıcı yorumları müşterilerinin beklenti ve deneyimlerinin değerlendirmesi açısından zengin bir içeriğe sahiptir (Zhang et al. 2021). Buna dayanarak (Zhang et al. 2021) tarafından yapılan çalışmada metin madenciliği metodu ile fonksiyonel değer, hedonic değer, epistemic değer ve sosyal ilişki değeri algılanan değeri ölçüm işaretçileri oluşturuldu. Airbnb platformundan toplanan veri sayısı P2P konaklama tüketicilerinin

algılanan değerinin tekrar satın alma niyeti üzerindeki etkisini ölçmek, çevrimiçi incelemeleri analiz etmek için Metin Madenciliği, Sentiment Analytics yöntemleri, LDA konu modeli ve makine öğrenmesi kullanıldı. LDA ile konuların bulunması ardından elde edilen konuların algılanan değerler ile eşleştirildi. Makine öğrenmesi tekniği kullanarak konular sınıflandırıldı ve (PLS-SEM) analizi gerçekleştirildi.

COVID-19 pandemi sürecinde malezya'daki oteller için müşterilerin memnuniyetini ortaya çıkarmayı hedefleyen (Nilashi, et al. 2021) bu çalışmada ayrıca COVID-19 süreci içerisinde servis kalitesinin otel performans kriterini ve sonucunda müşteri memnuniyeti üzerindeki etkisi olup olmadığını araştırıldı. Çalışma amacına ulaşmak için makine öğrenmesi teknikleri kullanarak yeni bir metot geliştirildi. Metin madenciliği, kümeleme ve tahmin öğrenme teknikleri kullanarak metot geliştirildi. Müşteri tercihini belirlemek için ve büyük veri analizi için LDA kullanıldı. Beklenti-Maksimizasyon (EM) kümeleme için kullanıldı. (ANFIS) memnuniyet seviyesini tahmini için kullanıldı. Sonuçlar COVID-19 sürecinde servis kalitesi otel performans kriteri ve bunun sonucunda müşteri memnuniyeti üzerinde etkisi olduğunu göstermiştir. Ayrıca, müşterilerin her zaman daha iyi performansa sahip otelleri aramalarının yanı sıra COVID-19 pandemisinde ilgili hizmetlerin kalitesine de ilgi gösterdiklerini ortaya çıkardı.

Bu çalışma (Nilashi, et al. 2022) COVID-19 sürecinde kullanıcıların internet ortamında oluşturduğu içerikleri kullanarak gezginlerin memnuniyet analizini yapmak için iki yöntem kombinasyonundan oluşan yeni bir yöntem sundu. Ayrıca metinsel veri analizi için LDA kullanıldı. Data segmentasyonu için K-Ortalama kümeleme yöntemi kullanıldı. Kayıp veri yerlerini doldurmak için boyut indirgeme yaklaşımı kullanıldı. Gezginlerin memnuniyet seviyesini tahmin etmek için bulanık mantık (fuzzy) bazlı bir yöntem kullanıldı. Validasyon işlemi için Sonrasında Yapısal Denklem Modelleme (SEM) kullanıldı. Sonuç olarak COVID-19 pandemi sırasında hizmet kalitesi otelin verdiği hizmet ile kullanıcı memnuniyeti arasındaki ilişkiyi önemli bir şekilde etkilediğini doğrulamaktadır.

Bu araştırma, büyük sosyal verilerin analizi ile müşteri memnuniyetinin ortaya çıkarılmasını araştırmış (Nilashi, et al. 2021) ve gıda kalitesinin otel performans kriterleri ile seyahat edenlerin memnuniyeti arasındaki ilişkiyi etkileyip etkilemediğini ortaya çıkarmak için büyük sosyal veri analizi gerçekleştirmiştir. Ayrıca, turistlerin temizlik, konum, odalar, hizmet, uyku kalitesi ve değer gibi turist memnuniyetini etkileyebilecek çeşitli kriterlere yönelik algılarını araştırmayı amaçlamıştır. Bu kriterler, COVID-19 krizi sırasında TripAdvisor portalında yayınlanan yorumlara dayalı olarak ve metin madenciliği yaklaşımı kullanılarak araştırılmıştır. Bu çalışmanın amaçlarına yönelik olarak iki aşamalı bir metodoloji

geliştirilmiştir. Gezginlerin tercihlerini daha iyi tespit etmek için denetimli bir öğrenme tekniği ile küme analizi yapıldı. Bunu yapmak için LVQ tekniğini kullanıldı. Metodolojimizin bir sonraki adımında, LVQ tarafından oluşturulan her kümede CART yaklaşımı gerçekleştirildi. Metodolojinin bir sonraki aşamasında, kısmi en küçük kareler yapısal eşitlik modellemesi (SEM-PLS) yoluyla performans kriterleri ve memnuniyet arasındaki ilişkileri analiz etmek için varsayımsal model geliştirildi. Bulgular, çevre dostu otellerin performans kriterleri ile memnuniyet arasında pozitif bir ilişki olduğunu göstermiştir.

Bu çalışma (Xie et al. 2022) COVID-19 pandemi sürecinde Çinde insanların COVID-19 kısıtlamalarına uymaları yardımcı olmak için taze tarım ürünlerinin online olarak satıldığı sitelerin kullanıcı yorumlarında negatif yorumların artması nedeniyle o sitelerdeki çevrimiçi kullanıcı yorumları ele alınmıştır. Pycharm yazılımı yardımıyla taze tarım ürünlerinin satıldığı çevrimiçi platformlardan değerlendirme verisi olarak toplam 34,546 parça toplanmıştır. Toplanan veriler temizlenmiş ve durak kelimeleri metinden çıkartılmıştır. Son olarak, LDA modelinin sonuçları, tüketicilerin çevrimiçi taze tarım ürünleri satın alırken tercihlerini etkileyen faktörleri göstermektedir. Bu çalışma, COVID-19'un yayılması bağlamında sakinlerin çevrimiçi alışveriş tercihlerinin tipik özelliklerini belirlemekle kalmayıp, aynı zamanda yerel yönetimlere, belediyelere yapılan yaygın şikayetlerin önlenmesi için sakinlerin sosyal seviyede olumsuz duygularını yatıştırmak için pragmatik öneriler de sunmaktadır.

Çevrimiçi kullanıcı yorumlarının konaklama sektöründeki satın alma niyetine etkisini araştırılan bu çalışmada (Sim et al 2021) Güney Kore'deki 1256 konaklama yerinden 400.000'den fazla çevrimiçi incelemeyi kullanarak yorum içeriklerindeki konuları keşfetmek için LDA kullanılmıştır. Evrişimli sinir ağları (CNNs) yorumların değerliliğini belirtmek için kullanılmıştır ve mekânsal probit modelleri, inceleme içeriğinin ve değerliliğinin rezervasyon niyeti üzerindeki etkisini belirlemek için kullanıldı. Bir konaklama yerinin ambiyansı, değeri, hizmeti, resepsiyon hizmetleri, erişilebilirliği, çevredeki konum ve oda kapasitesi hakkındaki olumlu eleştirilerin rezervasyon niyetlerini önemli ölçüde artırdığı, hizmet, ön büro ve yakın çevredeki olumsuz yorumlar ise daha düşük rezervasyon oranı sonuçlandığı bulundu.

Bu çalışma (Xu et al. 2019) duygu analizinde Bi-LSTM kullanımını, Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), geleneksel LSTM ve Naïve Bayes (NB) gibi son teknoloji yöntemlerle karşılaştırdı. Elde edilen sonuçlar, Bi-LSTM'nin geleneksel tekniklerden daha iyi performans gösterdiğini, bağlam bilgisinin daha iyi yakalanmasını sağladığını ve diğer yöntemlere göre daha yüksek kesinlik, hatırlama ve F1'e sahip olduğunu göstermiştir.(Wu, He, et al. 2021) araştırmalarında Bi-LSTM birimlerinin gradyan

problemlerini etkili bir şekilde çözebildiğini ve bağlamsal semantik bilgileri iyi bir şekilde yakalayabildiğini göstermiştir.

Bu çalışma (Jelodar et al. 2020) COVID-19 ile ilgili sorunları kamuoyu görüşlerinden ortaya çıkarmak için konu modellemeyi kullanarak çevrimiçi Reddit tartışma forumundan COVID-19 tartışmalarının otomatik olarak çıkarıldığını bildirdi. Ayrıca, COVID-19 duyarlılık sınıflandırması için LSTM tekrarlayan sinir ağının nasıl kullanılacağını da araştırdı. Geliştirilen model ile %81.15 doğruluk elde edildi.

Bu çalışma (Nemes and Kiss 2021) Twitter'daki duygu ve tezahürleri (hashtag'ler, yorumlar, tweet'ler ve gönderiler) esas olarak "covid" ve "coronavirus" anahtar kelimelerine göre analiz etti. Duyguları sınıflandırmak için RNN'yi kullanarak çeşitli tweet'lerin duygusal doğasını analiz etmek için bir model geliştirildi. Tweet'leri çeşitli duygusal güç sınıflarına ayrıldı (zayıf olumlu/olumsuz, çok olumlu/olumsuz). Karşılaştırmalar esas olarak TextBlob (üçüncü taraf bir duyarlılık analizcisi) ile yapıldı ve tekrarlayan sinir ağı modelinin bu tür metin sınıflandırmasında iyi performans ve tahmin sağladığını gösterildi.

Bu çalışma (Mansoor et al 2020) LSTM ve ANN gibi çeşitli makine öğrenimi modellerini kullanarak bir duyarlılık sınıflandırması ve Coronavirüs tweet'lerinin analizini gerçekleştirdi. Yapılan çalışmalarında, farklı ülkelerdeki insanların duygularının zaman içinde nasıl değiştiğine odaklandı. Çevrimiçi öğrenme ve evden çalışma (WFH) ile ilgili tweet'ler kullanılarak zaman içinde duygu değişimi gözlemlendi. Deneysel sonuçlara göre, koronavirüs tweets veri setinde LSTM kullanarak %84,5 ve YSA kullanarak %76 doğruluk gözlemlendi. Pandemi boyunca korku duygusunun güven duygusundan önemli ölçüde daha yüksek olduğunu da gözlemlendi. Ülkelerle ilgili olarak, daha fazla oranda olumlu duyguya sahip ülkelerin Bangladeş, Pakistan, Mali ve Güney Afrika olduğunu görüldü. Avustralya, Hindistan, ABD, Türkiye, İngiltere ve Brezilya ise olumsuz duyguların yüksek olduğu ülkeler arasında yer almaktadır. Hesaplanan duygu puanı Tayland, Vietnam ve Polonya'nın en yüksek korku puanına sahip olduğunu, en yüksek güven puanlarının ise Umman, Suriye ve Kazakistan'da olduğunu gösterildi.

MATERYAL VE YÖNTEM

Yapılan çalışmada kullanılan materyal ve yöntem ele alınmıştır. Öncelikle veri toplama hakkında detaylar ele alınmıştır, veri toplamada kullanılan yöntem, programlama dili ve kütüphane gibi işlemler ele alındı. Veri temizleme ve önışleme hakkında detaylar açıklandı. Veri toplama ve önışleme işlemlerinden sonra veri analizi, konu modelleme, elde edilen konuların analizi ve piyasa üzerindeki faydalı etkisi ve müşteri segmentasyon çalışmaları detaylandırıldı. Şekil 9’da görüldüğü gibi veri seti crawler yardımıyla TripAdvisor sitesinden toplandı. Elde edilen veri önışleme aşamasından geçirildi. Müşterilerin memnuniyetini keşfetmek için LDA konu modelleme ile 4 konu elde edildi ve k-means ile müşteriler (az memnun, orta memnun, ve çok memnun) olarak 3 gruba ayrıldı. Son olarak yorum sınıflandırma aşamasında LSTM, BiLSTM ve GRU modelleri eğitildi ve karşılaştırıldı.

Veri Toplama

Bu araştırma için veri setini elde etmek için TripAdvisor’dan yararlanılmıştır. TripAdvisor sitesinden veri seti toplamada, Türkiye’deki 7 çevre dostu otel sitelerindeki kullanıcı yorumlarından ve kriter derecelendirmelerinden yararlanılmıştır. Toplam 17314 yorum elde edilmiştir. Veri toplama için TripAdvisor seçildi çünkü kullanıcıların yorumlarını açık bir yorum formuna göre göndermelerine izin veriyor (Nilashi, et al. 2022). Bu, kullanıcıların kapalı yorum formlarını benimseyen diğer portallardan ziyade hem olumlu hem de olumsuz ifadelerle yorumlarda çok daha yüksek boyutlar sunmasını sağlar (Nilashi, et al. 2022). TripAdvisor platformunda, yüksek oranlı ve olumlu yorumlara sahip oteller, daha uzun süre kalan turistler tarafından, yetersiz veya kötü yorumlara sahip otellere göre daha fazla talep görmektedir (Nilashi, et al. 2022).

TripAdvisor sitesinden veri toplama işlemi iki aşamadan oluşmaktadır. Birinci aşama kullanıcıların üretmiş olduğu metinsel yorumları toplamak. İkinci aşama ise kullanıcıların her bir otele ait 5 üzerinden yapılan derecelendirmelerde kriter derecelendirmelerini toplama. Bu şekilde iki tarayıcı (crawler) geliştirildi. Birincisi her bir çevre dostu otele ait metinsel yorumları toplar ve ikincisi ise seçilen her bir çevre dostu otele kriter derecelendirmeleri toplar.

Kriter derecelendirmeleri çok sayıda kullanıcı tarafından girilmiyor ve boş bırakılıyor. Bu sebepten dolayı kriter derecelendirmeleri toplama durumu için geliştirilen crawler sadece kriter derecelendirmeleri giren kullanıcıların kriter derecelendirmeleri toplanmıştır. En az bir derecelendirme girilmiş olanlar Crawler tarafından işleme alınmıştır. Boş kalan yani boş veri

oluşturan (missing values) kriter derecelendirmeler ile ilgili detaylı açıklama veri ön işleme ve temizleme bölümünde verilmiştir.

Metinsel veri toplama ve ön işleme

Müşterilerin çevrimiçi yorumları, TripAdvisor platformunda sunulan ve Türkiye'de bulunan farklı çevre dostu otel web sitelerinden toplanmıştır. TripAdvisor Yeşil Liderler programı, otellerin yeşil uygulamalara yönelik davranışlarını dikkate alır ve bunları 4 düzeye göre sıralar: Bronz, Gümüş, Altın veya Platin. Bu düzeyler özellikle mülkün TripAdvisor sitesindeki listesinde gösterilir. Daha fazla çevreci eylem sergileyen tesisler, daha yüksek TripAdvisor Eko Liderler seviyeleri (GreenLeaders level) elde edebilir (UNEP 2013). TripAdvisor sitesinde yer alan otelleri taramak için tarama tekniği (crawler) kullanıldı. Selenium kütüphanesi, TripAdvisor web sitesinde bulunan 7 yeşil otelden gelen çevrimiçi yorumları taramak için kullanıldı. Web sürücüsü Selenium kütüphanesinden import edildi ve tarama işlemi için google chrome kullanıldı.

Tripadvisor otel sitelerinde bulunan yorumlar, sayfa başına yaklaşık 10 yorum olacak şekilde dağıtılır. Tarama tekniğinde, bu sayfalarda gezinmek ve data-reviewid XPath'lerine göre çevrimiçi kullanıcı yorumlarının gövdesini almak ve ardından XPath'lerine göre çevrimiçi kullanıcı yorumları almak için bir döngü kullanılmıştır. Döngü işleminde Selenium kullanılmış olup, her yinelemede diğer sayfalara gitmek için sayfanın sonraki düğmesine tıklanır ve tarayıcıya bir sonraki URL adresi verilir.

Tripadvisor'da bulunan yeşil otelin son sayfasına ulaşana kadar operasyon sayfalarca devam eder. Şekil 3 tarayıcı (crawler) aracılığıyla toplanan metne dayalı çevrimiçi kullanıcı yorumlarının bir örneğini gösterir.



Michael W wrote a review Oct 15
2 contributions

Customer rating

Great hotel

"We had a wonderful stay. All the staff were friendly, helpful and kind. The morning buffet was diverse, tasty and generous. The room was comfortable and very clean. The hotel location was fabulous. It is close to all the main sites."

Read less ▲

Text-based reviews

Date of stay: October 2022

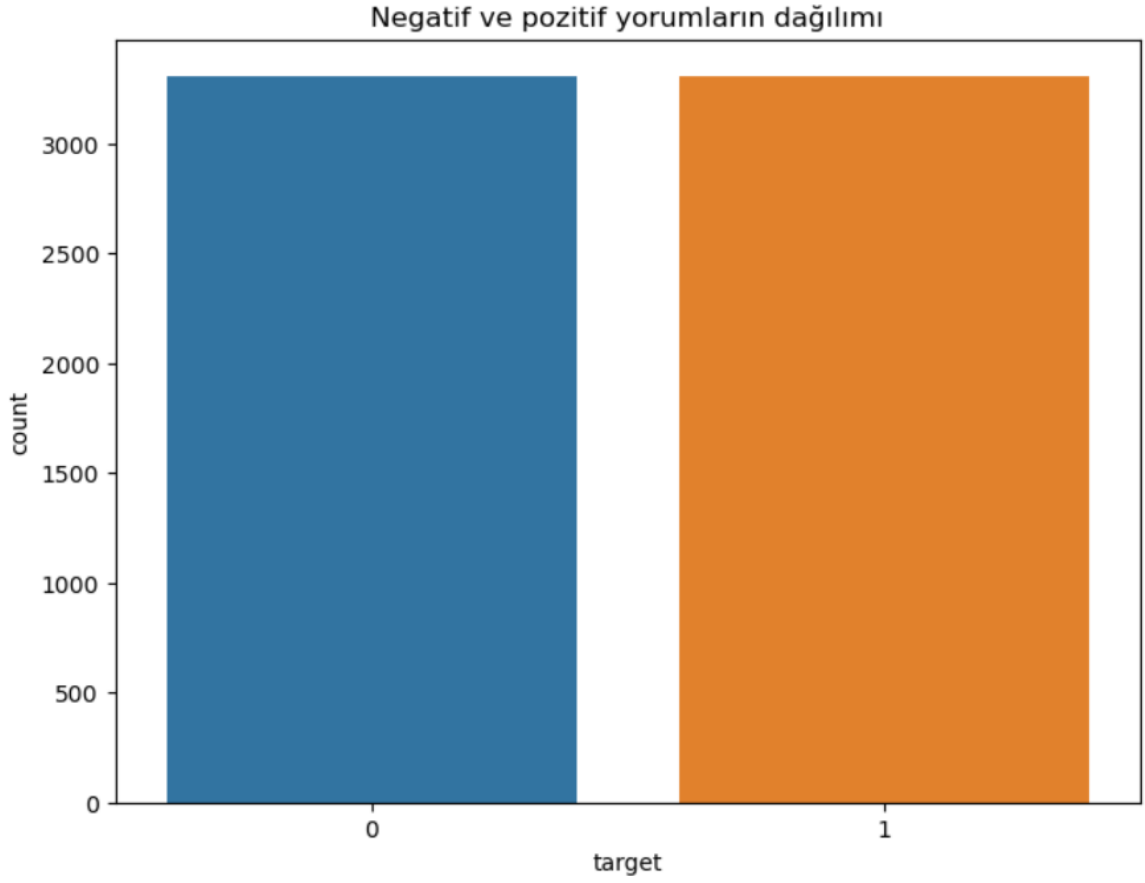
This review is the subjective opinion of a Tripadvisor member and not of Tripadvisor LLC. Tripadvisor performs checks on reviews.

Helpful Share

Şekil 3. Tripadvisor sitesindeki bulunan çevrimiçi kullanıcı yorumlarının bir örneği

Tarayıcı (crawler), araştırmayı amaçladığımız otellerle ilgili müşterilerin çevrimiçi yorumlarını toplamak için oluşturuldu. Türkiye'de bulunan 7 farklı otelden 17314 online yorum toplandı. Toplanan yorumların dili İngilizcedir. Verilerdeki yorumların kriter değerlendirmesi 5 üzerinden değerlendirilir ve değerlendirme 3 ve üzeri ise yorumların genel olarak pozitif olduğu görüldü. Bu çalışmaya benzer bir şekilde (Polat B 2021) yorumlardaki kriter değerlendirmesi 3'ten az olduğunda, genel olarak yorumların negatif olduğu görüldü. Metin sınıflandırma aşamasına gelindiğinde, veri seti içindeki yorumlar 951 negatif ve 16363 pozitif yorum olarak dengesiz olduğu görüldü. Bu nedenle sınıflandırma aşaması için özgün yeni dengeli bir veri seti oluşturuldu. Bu aşamada sadece negatif veriler toplanarak 3306 negatif veri elde edildi ve önceki dengesiz olan veri setinden 3305 pozitif yorum eklendi ve karıştırıldı. Böylelikle, TripAdvisor sitesinde ve Türkiye'deki çevre dostu otellerden müşteri yorumları kullanılarak 3305'i pozitif ve 3306'sı negatif olmak üzere toplamda 6611 yorum içeren yeni dengeli bir veri seti oluşturuldu. TripAdvisor sitesinde ve Türkiye'deki yeşil otel müşteri yorumlarından oluşturulan yeni veri seti yorum duygu dağılımı Şekil 4'de gösterilmiştir. Şekil 4'de gösterildiği gibi 3305 pozitif yorum ve 3306 negatif yorum olarak toplam 6611 yorum elde edildi.

Makine öğrenmesi modelinin metinsel veriyi anlayabilmesi için metinsel verinin sayısal hale getirilmesi gerekiyor. TensorFlow ve Keras içindeki Tokenizer API yardımıyla önce noktalama silme ve büyük harfleri küçük harflere çevirme gibi veri ön işleme yapıldı. Sonrasında cümleler kelimelere ayrıldı ve kelimeler sayısal hale getirildi. İkinci aşamada ise Tokenizer nesnesinde bulunan `texts_to_sequences` yardımıyla kelimeler dizi haline getirildi. Son aşamada ise diziler aynı uzunluğa sahip olabilmeleri için budama işlemi uygulandı.



Şekil 4. Yeni elde edilen veri seti yorumlarının dağılımı

Kriter derecelendirmeleri toplama

TripAdvisor sitesinde kullanıcılar sadece yorumlar vasıtasıyla oteli değerlendiriyor değiller aynı zamanda kullanıcılar bir grup performans faktörleri üzerinden oteli değerlendirebilirler. Şekil 5’de görüldüğü gibi kullanıcılar sadece metinsel olarak değil aynı zamanda performans faktörleri olarak uyku kalitesi, değer (maliyet-fayda), hizmet, konum, odalar ve temizlik olmak üzere altı kriter üzerinden değerlendirme yapabiliyorlar. Bu çalışmada performans faktörleri olarak altı kriter ele alınmıştır.



Xeres25 wrote a review May 2010
Birmingham, England • 51 contributions • 78 helpful votes

Genel derecelendirme

A nice hotel

"This was the first time I stayed in an all inclusive and I was pleasantly surprised by this hotel. The hotel was very good and clean. Rooms were well decorated and the other facilities were of a high quality. I would suggest that Irons are provided in the rooms, other than that the rooms were great. There is a good all inclusive entertainment programme which caters for the whole variety of the family.

The only issue we had staying at this hotel was the service of a few staff in the hotel and at the bar really let the side down. You will see some staff really work hard and will go out of the way to help, but it was noticeable that a certain number of staff lacked enthusiasm for their job or indeed showed a lack of interest in providing a good service."

[Read less](#)

Date of stay: May 2010

Trip type: Traveled with family

Value

Location

Service

Rooms

Cleanliness

Sleep Quality

Kriter derecelendirmeler

This review is the subjective opinion of a Tripadvisor member and not of Tripadvisor LLC. Tripadvisor performs checks on reviews.

Şekil 5. TripAdvisor çevrimiçi kullanıcı yorumlarına ait kriter derecelendirmeleri

Veri Önişleme ve Temizleme

Veriyi bir ML modelinde kullanılmadan önce verinin temizlenmesi ve önişleme işlemlerinin uygulanması model başarısını büyük ölçüde etkiler (Frye et al 2021). Bu nedenden dolayı verinin temizlenmesi ve önişleme işlemlerinin veri seti üzerinde uygulanması varsayılan model başarısı için oldukça önemlidir (Frye et al 2021).

Tüm oteller için toplanan veriler tek bir veri seti içerisinde birleştirilmiştir. Toplanan veriler gereksiz kelimelerden veya e-postalar ve yeni satır sembolleri gibi karakterlerden temizlenmiştir. Çok tekrar edilen kelimeler veya herhangi bir anlam taşımayan kelimeler durak kelimeler listesine eklenerek ayrıca metinlerden çıkartılmıştır.

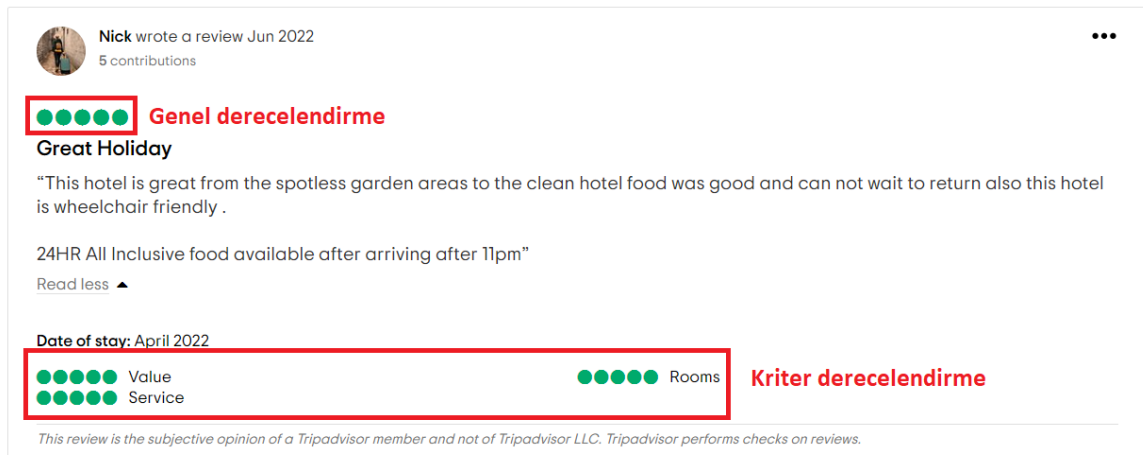
Durak kelimeler (stop words) listesine istenilmeyen ve anlamsız kelimeler “otel isimler, hotel” gibi kelimeler durak kelimeler listesine eklendikten sonra, durak kelimeleri corpus’ten çıkartılmıştır. Durak kelimelere ait birkaç örnek “the, a, is, are” gibi konu modelleme işlemlerinde katkıda bulunmayan kelimelerden oluşmaktadır.

Durak kelimeler silindikten sonra kelime torbası (bag of words BoW) uygulandı. Bu aşamada metinlerdeki tüm farklı kelimelerin bir sözlük oluşturacak şekilde toplanır. Bu sözlük, metindeki her kelimenin benzersiz bir numarasıyla eşleştirilir. Her bir metin, kelime sözlüğündeki her bir kelimenin metindeki varlığına bağlı olarak bir öznitelik vektörüyle temsil edilir (Bahmanyar et al 2015). Örneğin, bir metindeki bir kelimenin varlığı, o kelimenin sözlükteki numarasının indeksiyle temsil edilir. Bu durumda, öznitelik vektörü, metindeki tüm kelimelerin varlık veya yokluğunu gösteren bir dizi olur. Bu şekilde veri Terim Frekansı-Ters

Doküman Frekansı (Term Frequency-Inverse Document Frequency TF-IDF) ve ondan sonra topic modelleme işlemi uygulanması için veri hazırlanır.

Kelime torbası oluşturulduktan sonra, TF-IDF işlemi uygulanmıştır. Bu yöntem, bir belgedeki bir kelimenin önemini belirlemek için kullanılan istatistiksel bir ölçüdür. Bu yöntemde, bir kelimenin belirli bir belgedeki frekansı (TF) ile tüm belgelerdeki o kelimenin yaygınlığı (IDF) arasındaki çarpım hesaplanır. TF (Term Frequency), bir belgedeki bir kelimenin kaç kez geçtiğini ifade eder. Genellikle, bir kelimenin frekansı, belgedeki toplam kelime sayısına bölünerek normalize edilir. Böylece, bir kelimenin bir belgedeki önemi, o belgedeki sıklığıyla doğru orantılı olur. IDF (Inverse Document Frequency), bir kelimenin tüm belgelerdeki yaygınlığını ölçer. Kelimenin tüm belgelerde geçme sıklığı hesaplanır ve bu değer tersi alınarak elde edilir. IDF, bir kelimenin genel olarak ne kadar yaygın veya nadir olduğunu belirlemeye yardımcı olur. Nadir kelimelerin daha yüksek bir IDF değeri vardır, bu da onları daha fazla ağırlıklandırır. TF-IDF, bir belgedeki bir kelimenin önemini belirlemek için TF ve IDF'nin birleşimini kullanır (Chen, Wu, et al. 2022). Bir kelimenin TF-IDF değeri, TF'nin IDF ile çarpımıyla elde edilir. Yani, bir kelimenin belgedeki frekansı ne kadar yüksekse ve tüm belgelerdeki yaygınlığı ne kadar düşükse, o kelimenin TF-IDF değeri o kadar yüksek olur. TF-IDF yöntemi, metin madenciliği ve bilgi erişimi alanlarında yaygın olarak kullanılır.

Toplanan veriler boş değerlerin varlığı açısından kontrol edilmiştir. Çünkü özellikle kriter derecelendirmede çoğu kullanıcı tüm kriter derecelendirmeleri girmemektedir. Şekil 6'da görüldüğü gibi örnek kullanıcı kriter derecelendirmelerinin tamamını girmemiş, 6 kriter üzerinden sadece 3'ü üzerinden değerlendirme yapmıştır. Bu nedenden dolayı toplanan veride kayıp veri (missing values) oluşmaktadır. Kayıp verinin doldurulması için her kritere ait toplam derecelendirmelerin ortalamasını alarak kayıp veri doldurulmuştur. Bu şekilde veri analizi işlemlerinin gerçekleştirilmesi ile nihai veri seti hazır hale getirilmiştir.



Şekil 6. TripAdvisor Çevrimiçi kullanıcı yorumlarına ait eksik girilmiş kriter derecelendirmeleri

Çevrimiçi incelemeler genellikle kısadır ve LDA kısa metinsel verileri işleme konusunda sınırlıdır (Zhang et al. 2021). Bu nedenle Kullanılan modelin performansını iyileştirmek için verilerin ön işleme aşaması esastır. Verinin ön işlenmesi, 4 aşamadan gerçekleştirildi ve (1) durak sözcüklerinin kaldırılmasını, (2) e-postaların ve yeni satır karakterlerinin kaldırılmasını, (3) sözcüklerin simgeleştirilmesini ve (4) metnin tokenize edilmesini içerir. NLTK paketi tarafından sağlanan durak kelimeleri listesi, listeye daha fazla durak kelimeleri eklenerek genişletildi ve belirlenen durak kelimeleri metinden çıkartıldı. Son olarak BoW ve TF-IDF kullanılarak yorumlar LDA konu modelleme aşamasına hazır hale getirildi.

Konu Modelleme (LDA)

TripAdvisor'dan toplanan çevrimiçi kullanıcı yorumları verisi işlenerek anlamsız kelimeler hariç tutuldu. Müşterilerin çevrimiçi yorumlarının veya çevrimiçi incelemelerinin memnuniyet boyutlarını keşfetmek için LDA kullanıldı. LDA konu modelleme yöntemini oluşturmak için Gensim kütüphanesi kullanıldı. Verilerin analizi üç ana aşamadan oluşur; sosyal verilerin temizlenmesi ve hazırlanması; çevrimiçi kullanıcı yorumlarından LDA yardımıyla memnuniyet boyutlarının keşfedilmesi ve boyutların görselleştirilmesi, ve segmentasyon bölümünde müşterilerin k-ortalamlar kümeleme yardımıyla segmentlere ayrılması. Bu çalışmada kullanılan metod Şekil 9'da gösterilmiştir.

LDA konu modelleme yaygın bir şekilde araştırma alanlarında kullanılmakta olup (Blei et al 2003) tarafından tanıtılmıştır. Her belge içerisinde, belgenin başlığında yazılan konu ne olursa olsun o belge çoğunlukla birden fazla konudan oluşmaktadır. Politika ile ilgili yazılan bir makalede spor konusu hakkında, bilim konusu hakkında, doğa konusu hakkında veya diğer olabilecek konular hakkında da konulara ait yazılara sahip olabilir.

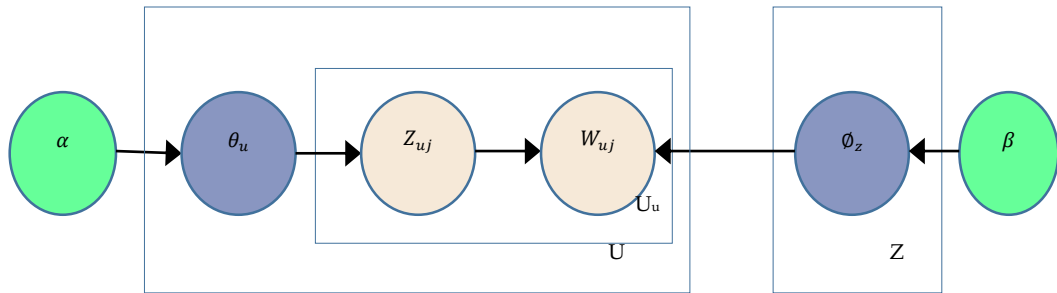
Bir belgeyi yalnızca belirli kelimelere dayanarak temsil etmek, belgelerin verimli bir şekilde karşılaştırılabilmesi için yeterince iyi değildir (Pavlinek and Podgorelec 2017). Yani, aynı bağlamdaki iki belge, benzer anlamlara sahip sözcükleri içeren farklı sözlükler kullanabilir (Pavlinek and Podgorelec 2017). LDA konu modelleme belgelerin ihtiva ettiği konu oranlarını ortaya çıkarmakta işlemlerinde büyük katkılarda bulunmaktadır. LDA yardımıyla her belge çok terimli bir konu dağılımı olarak temsil edilir ve bu konular kümelere benzeyen daha üst düzey kavramlar olarak görülebilir (Pavlinek and Podgorelec 2017). Bir koleksiyondaki her belgenin, her konunun bir kelime karışımı ile sunulduğu birkaç gizli konudan oluşturulduğu varsayımına dayanır (Pavlinek and Podgorelec 2017). LDA, kısa metinlerin konu çıkarma görevlerinde hızlı ve doğrudur (Zhang and Zhang 2022). Çevrimiçi kullanıcı yorumları genel olarak çok uzun

olmayan metinlerden oluştuğu için LDA Çevrimiçi kullanıcı yorumları konu modelleme konusunda diğer makine öğrenmesi yöntemlerine yöntemlere daha uygundur.

LDA, kelimelerin, konuların ve belgelerin üç katmanlı yapısını içeren üç katmanlı bayes olasılık modeli olarak bilinen bir belge konusu oluşturma modelidir (Li et al. 2021). Denetimsiz bir makine öğrenimi tekniği olarak, LDA konu modeli, büyük belge kümeleri veya corpora'daki gizli konuları tanımlayabilir (Li et al. 2021). Her belgeyi bir kelime frekans vektörü olarak dönüştürmek için kelime (Bag of Words) çantası yöntemini kullanır. Sonuç olarak, yapılandırılmamış metin, modellenmesi kolay olan dijital verilere dönüştürülebilir (Li et al. 2021).

Her belge konu koleksiyonundan oluşur ve her konu kelime koleksiyonu vardır ve öyle ki her bir kelime koleksiyonu spesifik bir konuyla ilişkilidir. Bir belgenin oluşmuş olduğu konular bu belgeye bazen çok miktarda bazen da az bir miktarda katkıda bulunmaktadır. Bir belge sadece bir konudan oluşmamalı bunun yanı sıra bir belge tüm konuları da içermemeli ve belge ile ilgili konular bir şekilde dağılmış olarak atanmalıdır. Yani konular matrisinde her satır sıfır olmayan değerler içermeli ve sadece bir değil birden fazla sıfır olmayan değerler içermelidir. Çünkü her belge tüm konular hakkında ilgili bir yazıya sahip değildir. Bu şekilde her belgeye ait ilgili konuların dağılım oranları konu matrisinde yer alırlar.

Bir belgenin sahip olduğu her spesifik bir konu için bir kelime koleksiyonu bulunmaktadır. Bu kelime koleksiyonu ait olduğu konu ile ilgili özel veya spesifik kelimeler içerir, bu sınırlı kelimeler o konuya özel olur çünkü bir konu tüm kelimeleri içermez. Şekil 7 LDA Üretici Prosedürü (LDA Generative Procedure) gösteriyor ve Şekil 8 LDA örnek konu ve kelime dağılım temsili göstermektedir.



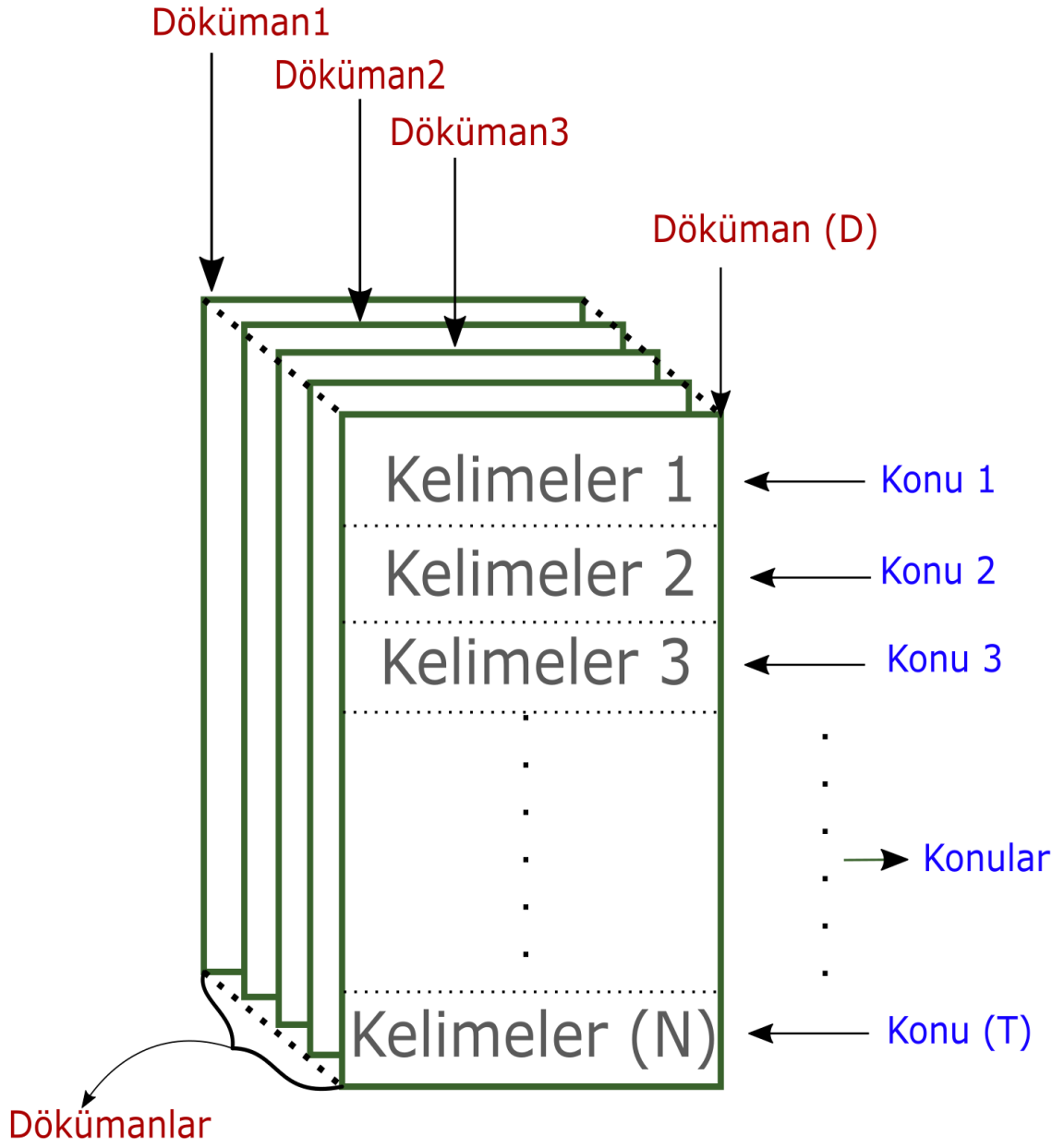
1. Her konu için $z \in Z$
 - Çok terimli bir dağılım çiz $\phi_z \sim \text{Dir}(\vec{\beta})$.
2. Her kullanıcı için $u \in U$,
 - Çok terimli bir dağılım çiz $\theta_u \sim \text{Dir}(\vec{\alpha})$.
 - Her bir kelime için $w \in D_u$,

- (a) Bir konu çiz $z \sim \text{Multinomial} \left(\vec{\theta}_u \right)$.
- (b) Bir kelime çiz $w \sim \text{Multinomial} \left(\vec{\phi}_z \right)$.

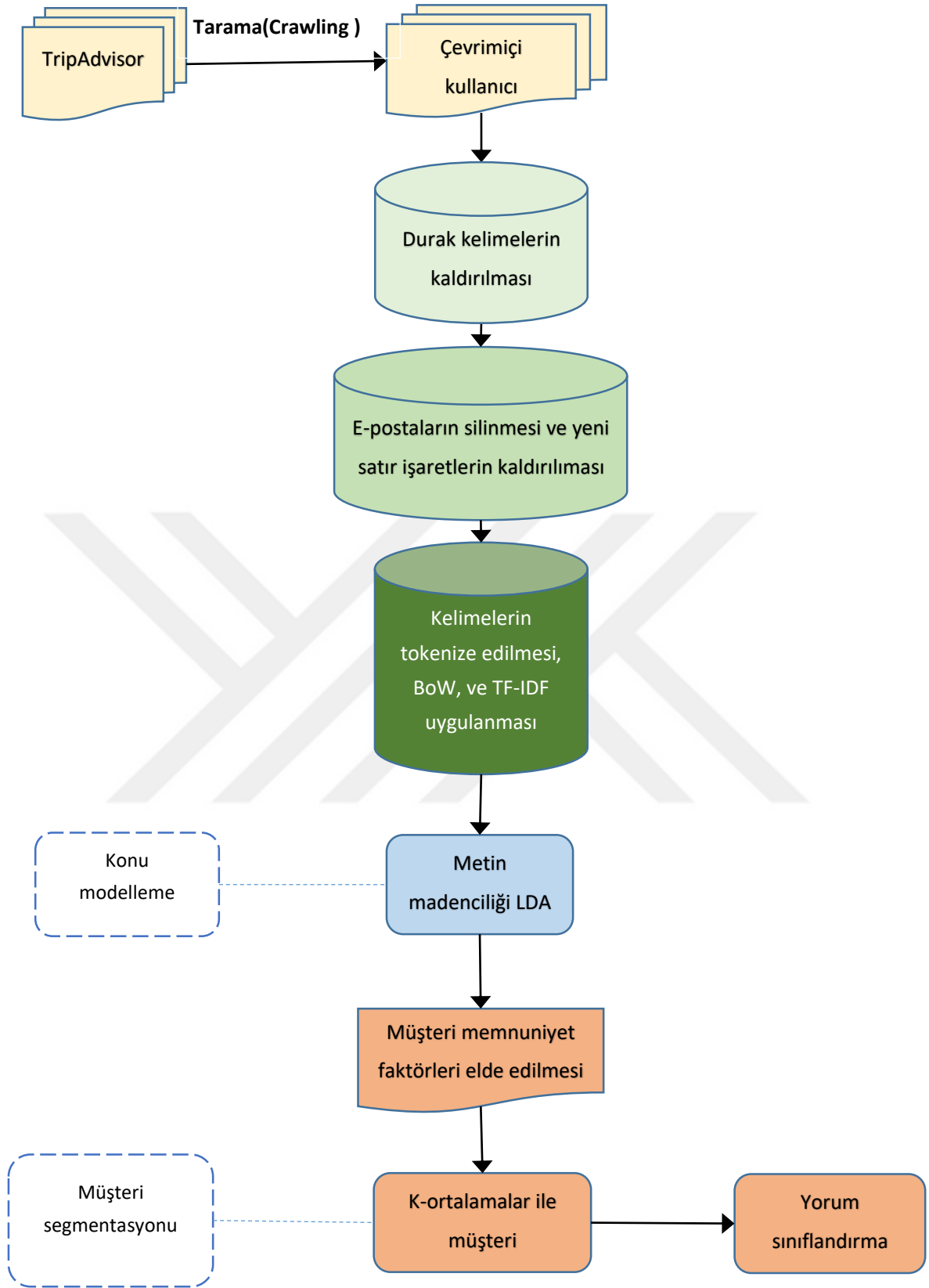
Şekil 7. LDA üretici prosedürü (LDA generative procedure)

LDA üretici prosedürü notasyonları:

1. (α) (Dirichlet prior) her belge için konuların dirichlet dağılımı, α değeri büyük olursa belge ile ilgili çok konu karışımı içereceği anlamına gelir ve eğer α değeri küçük olursa belge az sayıda konu karışımı içereceği anlamına gelir.
2. (β) (Dirichlet prior) her konu için kelimelerin dirichlet dağılımı, β değeri büyük olursa konu çoğu ilgili kelime karışımı içereceği anlamına gelir ve eğer β değeri küçük olursa konu az sayıda ilgili kelime karışımı içereceği anlamına gelir.
3. (U) Belge sayısı ya da (user) kullanıcı tarafından paylaşılan çevrimiçi yorum sayısı.
4. (U_u) Belge içindeki kelime sayısı.
5. (θ_u) u dosyası için konu dağılımını temsil eder.
6. (Z_{uj}) Her bir kelimeye atanan konuyu temsil eder. Bu şekilde her dosya bir konu dağılımı oluşturur.
7. (W) Kelimeyi temsil eder.
8. (ϕ_z) z konusu için kelime dağılımı
9. (Z) konuyu temsil eder.



Şekil 8. LDA örnek konu ve kelime dağılım temsili (Bastani et al 2019)



Şekil 9. Araştırma metodu

Python paketi; pyLDAvis, LDA konularının görselleştirmelerini sunmak için kullanıldı. Konu sayısı, örtüşmeyen veri segmentleri elde edene kadar ayarlandı, bu da araştırma bulguları bölümünde Şekil 18’de gösterildiği gibi 4 ana konuya yol açtı. Konu sayısı, örtüşmeyen veri segmentleri elde edene kadar ayarlandı.

Müşteri Segmentasyon

Konu modelleme yaklaşımının ardından, kullanıcı tarafından oluşturulan içeriği birkaç parçaya ayırmak için bir kümeleme tekniği kullanıldı. Kümeleme yaklaşımları, kullanıcı tarafından oluşturulan içerik analizi ile ilgili çeşitli araştırmalarda kullanılmıştır (Nilashi, et al. 2022; Nilashi, et al. 2021). Denetimsiz öğrenme tekniği olan K-means kümeleme yaklaşımı, birkaç yineleme yoluyla en uygun küme merkezini bulan yinelemeli bir kümeleme tekniği olarak konuşlandırıldı (Zhang, Li, et al. 2022). Bu çalışmada kullanılmış k-ortalama kümeleme algoritması aşağıdaki gibidir. Burada, N gözlemleri k kümeye ayırmak için K-means algoritması veri noktaları arasındaki ortalama mesafeyi bulur ve ona göre verileri kümelere ayırır. Algoritma ile karesi alınmış hata değerinin (SSE) toplamında sunulduğu gibi, sağlanan k kümenin şu şekilde en aza indirilmesi gerekmektedir:

K-ortalamalar kümeleme algoritması:

$$SSE = \sum_{z=1}^k \sum_{x_i \in z} x_i - \mu_z$$

- SSE (Sum of Squared Errors): SSE, kümeleme işlemi sırasında ortaya çıkan hataların karelerinin toplamını ifade eder. Bu değer, kümeleme algoritmasının performansını değerlendirmek için kullanılır. SSE değeri ne kadar düşükse, kümeleme işleminin daha iyi bir sonuç verdiği anlamına gelir. Daha düşük SSE değerleri, verilerin kümeler arasında daha iyi ayrıldığını ve kümelerin daha homojen olduğunu gösterir.
- k: k, K-ortalamalar kümeleme algoritmasında kullanılan küme sayısını temsil eder. Bu değer, kullanıcı tarafından belirlenir ve genellikle deneme-yanılma yöntemi veya istatistiksel tekniklerle belirlenir. Küme sayısı, veri setinin doğasına, amaçlara ve analiz hedeflerine bağlı olarak seçilir.
- z: z, küme indeksidir. Bu indeks, her bir veri noktasının hangi kümeye ait olduğunu belirtir. Kümeleme algoritması, veri noktalarını kümeler halinde gruplandırırken, her bir veri noktasına bir küme indeksi atanır.
- x_i : x_i , veri setindeki her bir veri noktasını temsil eder. Her veri noktası, öznelikleri veya ölçümleri içeren bir gözlemi ifade eder.
- μ_z : μ_z , z kümesinin merkezini temsil eder. Bu, z kümesine ait veri noktalarının merkezini ifade eder. K-ortalamalar algoritması, küme merkezlerini hesaplamak için genellikle örneklerin ortalamasını kullanır.

Müşterilere ait TripAdvisor sitesindeki kriter derecelendirmeleri üzerinde segmentasyon yapmak oldukça önemlidir. Müşterilerin kriter derecelendirmeleri üzerinde kümelemek ve her müşteri grubunun beklentilerini ve memnuniyet boyutlarını anlamak sadece otel sahipleri ve karar verici konumundaki kişiler için değil aynı zamanda müşteriler için de önemlidir. Bu çalışmada müşterilerin kriter derecelendirmeleri bazı gruplara ayırma işleminde k-ortalamlar kullanıldı ve k grup merkezleri olarak 3 ayarlandı.

Tablo 1’de TripAdvisor sitesi üzerindeki Türkiye’deki 7 çevre dostu otellerinden toplanan müşteri kriter derecelendirmelerinin birkaç örneği gösterilmiştir.

Tablo 1. Toplanan Kriter Derecelendirmelerin Bir Örneği

Genel değerlendirme	Değer	Konum	Hizmet	Odalar	Temizlik	Uyku kalitesi
5	4	5	5	4	5	4
5	3	5	5	4	5	4
5	5	5	5	5	5	5
5	5	5	5	4	5	5
5	5	5	5	5	5	5
5	5	5	5	5	5	5
5	5	5	5	5	5	5
5	5	5	5	5	5	5
4	5	4	5	4	5	5
5	5	5	5	4	5	5
5	5	5	5	5	5	5
5	5	5	5	5	5	5
5	4	5	5	5	5	4
2	1	1	1	2	2	1
5	5	5	5	5	5	5
5	5	5	4	4	5	4
5	5	5	5	5	5	5
5	5	5	5	5	5	5

Kümelemenin sonuçları araştırma bulguları bölümünde Tablo 7’de gösterilmiştir ve detaylı olarak bahsedilmiştir. Birinci, ikinci ve üçüncü segmentlerin ağırlık merkezleri sırasıyla şu şekilde elde edilmiştir. [2.76171867, 3.69912439, 2.41256159, 3.32787149, 3.7668611, 3.54576991], [4.0044922 , 4.26684662, 4.39947376, 4.26356397, 4.66544133, 4.23735696],

[4.78443376, 4.80887392, 4.96762968, 4.79764783, 4.87002358, 4.79857728]. Elde edilen sonuçlar Tablo 7’te gösterilmiştir.

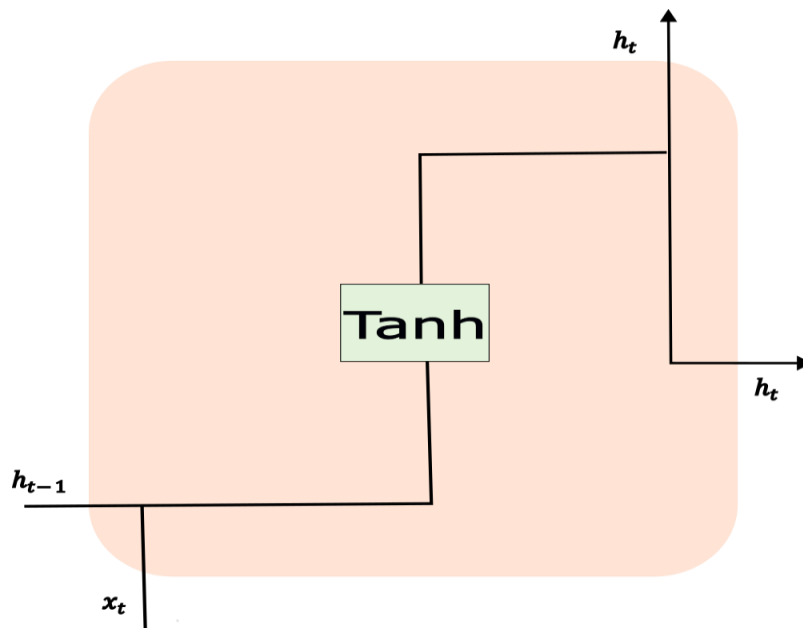
Yorum Sınıflandırma

Yorum sınıflandırma aşamasında LSTM, BiLSTM ve gated recurrent unit (GRU) modelleri uygulanmıştır.

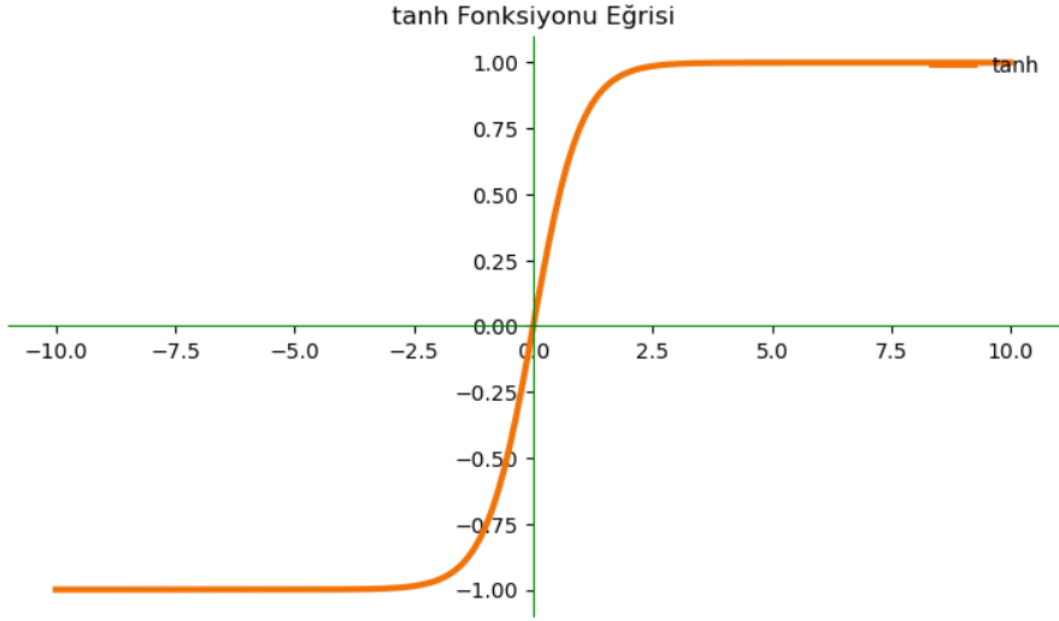
Tekrarlayan sinir ağları (RNN)

Tekrarlayan Sinir Ağı (RNN) modelleri görüntü/video alt yazısı, doğal dil, ve tahmin dahil olmak üzere sıralı veri modellemede en yaygın makine öğrenimi modellerinden biridir (Khaldi et al. 2023; Chimmula and Zhang 2020). RNN, NLP ve metin sınıflandırmada artan bir önem kazanmaktadır. En basit RNN hücresi Elman'dır, sadece bir gizli katman ihtiva eder. Şekil 10’da Elman yapısına ait bir hücre resmi gösterilmiştir.

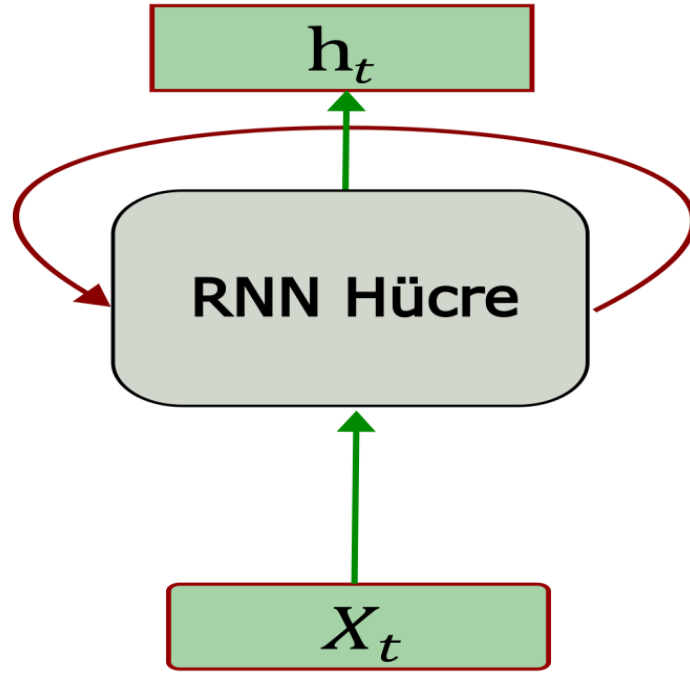
Şekil 10’de görüldüğü gibi RNN’deki gizli katmandan çıkan çıkış ayrıca sonraki değer girişi için giriş değeriyle birlikte giriş olarak kullanılır (Chen et al. 2018). Bu şekilde RNN dizi bağımlılığı tanımı içerir, örnek olarak Şekil 12’de gösterildiği gibi (h_t) çıkışı önceki çıkışlara bir bağımlılık oranı taşımaktadır. Dolayısıyla sonraki değer tahmin konusunda RNN başarılı bir tekrarlayan sinir ağlarıdır (Wu and Noels 2022). RNN çıkış verileri üzerine, çıkış verileri -1 ile 1 aralığında indirmek için Şekil 13’de görüldüğü gibi tanh fonksiyonu kullanır. RNN’in sıralı bir makine öğrenimi olması, metin tahmin ve sınıflandırma açısından oldukça kullanışlıdır çünkü RNN’de bir metindeki kelimelerin birbirlerine bağımlılık oranı hesaplamada söz konusudur.



Şekil 10. Elman hücre yapısı (Khaldi et al. 2023)



Şekil 11. Tanh fonksiyonu (Zhang et al. 2023)



Şekil 12. Tekrarlayan sinir ağının çalışma prensibi (Khaldi et al. 2023)

Uzun kısa süreli bellek (LSTM)

LSTM, RNN'nin bir uzantısı olan Uzun Kısa Süreli Bellek Ağı modelidir (Hochreiter and Schmidhuber 1997). RNN, daha uzun dizi modelleriyle uğraşırken gradyan kaybolma probleminden muzdarip olabileceğinden, LSTM, her tekrarlayan gövdede "geçit unutma", "giriş kapısı" ve "çıkış kapısı" dahil olmak üzere birkaç tür kapı biriminden oluşan bellek hücrelerini sunar (Liang and Niu 2022). Durumda saklanan geçmiş bilgileri makul bir şekilde

atmayı öğrenerek, mevcut girişi yeni bellek olarak yenileyerek, durum ve çıktı bilgileri üretebilir (Liang and Niu 2022).

LSTM birimi, Denklemler (2.1)-(2.4)'te gösterildiği gibi, RNN'e dayalı olarak giriş kapısı i , unutulmuş kapı f , bellek birimi c ve çıkış kapısını ekler, bu da uzun dizi sürecinin performansını büyük ölçüde artırır (Wu et al. 2022). LSTM biriminin işlemi Denklemler (2.1)-(2.5) ile ifade edilir.

$$i_t = \sigma(W_i X_t \times [h_{t-1}] + b_i) \quad (2.1)$$

Giriş Kapısı (Input Gate - i): Giriş kapısı, mevcut zaman adımında hangi bilginin hafızaya ekleneceğini kontrol eder. Bir sigmoid aktivasyon fonksiyonu kullanılarak önceki gizli durum (h_{t-1}) ve mevcut giriş (X_t) ile çarpılan bir ağırlık matrisi (W_i) ve bir bias terimi (b_i) üzerinden geçirilir. Bu, hangi bilginin hafıza hücresine eklenip eklenmeyeceğini kontrol eden bir geçirgenlik değeri üretir.

$$f_t = \sigma(W_f X_t \times [h_{t-1}] + b_f) \quad (2.2)$$

Unutma Kapısı (Forget Gate - f): Unutma kapısı, hafıza hücresindeki bilgilerin ne kadarının silineceğini kontrol eder. Bir sigmoid aktivasyon fonksiyonu kullanılarak önceki gizli durum (h_{t-1}) ve mevcut giriş (X_t) ile çarpılan bir ağırlık matrisi (W_f) ve bir bias terimi (b_f) üzerinden geçirilir. Bu, hafıza hücresindeki her bir ögenin ne kadarının korunacağını belirleyen bir geçirgenlik değeri üretir.

$$c_t = f_t \times c_{t-1} + i_t \times \tanh(W_c X_t \times [h_{t-1}, x_t] + b_g) \quad (2.3)$$

Hafıza Hücresi (Cell State - c_t): Hafıza hücresi, mevcut zaman adımında tutulan bilgileri temsil eder. Önceki hücre durumu (c_{t-1}), unutma kapısından gelen değer (f_t), giriş kapısından gelen değer (i_t) ve bir gizli durum ve girişin birleşimi olan bir tanh aktivasyon fonksiyonu kullanılarak hesaplanır. Bu, mevcut giriş ve önceki gizli durumun birleşiminden elde edilen bilgiyi günceller.

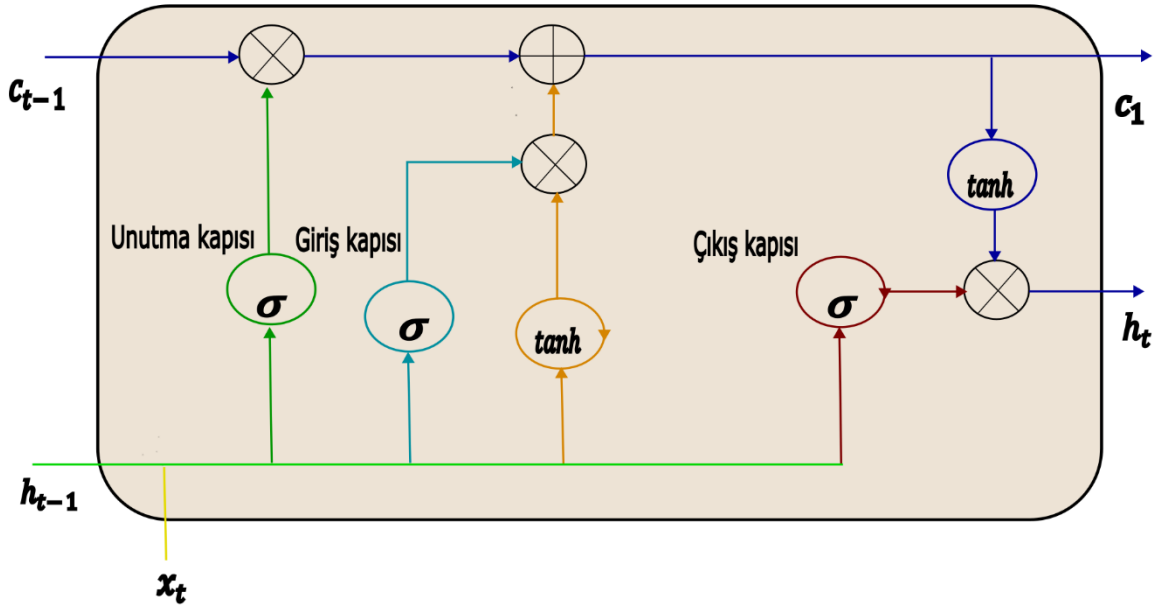
$$o_t = \sigma(W_o X_t \times [h_{t-1}, x_t] + b_o) \quad (2.4)$$

Çıkış Kapısı (Output Gate - o): Çıkış kapısı, mevcut zaman adımında hangi bilginin dışarı çıkacağını kontrol eder. Bir sigmoid aktivasyon fonksiyonu kullanılarak önceki gizli durum (h_{t-1}) ve mevcut giriş (X_t) ile çarpılan bir ağırlık matrisi (W_o) ve bir bias terimi (b_o)

üzerinden geçirilir. Bu, mevcut hücre durumunu (c_t) bir tanh aktivasyon fonksiyonundan geçirerek elde edilen bilgiyle çarparak, mevcut gizli durumu (h_t) üretir.

$$h_t = o_t \times \tanh(c_t) \quad (2.5)$$

LSTM, giriş kapısı ve unutma kapısı aracılığıyla bilginin kontrolünü sağlar ve hafıza hücresi aracılığıyla bilginin korunmasını veya güncellenmesini sağlar. Çıkış kapısı, güncellenen hafıza hücresinden uygun bilgileri çıkararak sonucu üretir. Şekil 13'te bulunan notasyonlar, t zamanında i_t, f_t, c_t, o_t giriş kapısını, unutma kapısını, bellek birimini ve çıkış kapısını sırasıyla temsil eder. h (hidden layer) gizli katmanı temsil eder. x, w, b, c, giriş, ağırlık, sapma, birim olarak sırasıyla temsil edilir. (σ) Sigmoid fonksiyonu temsil eder.

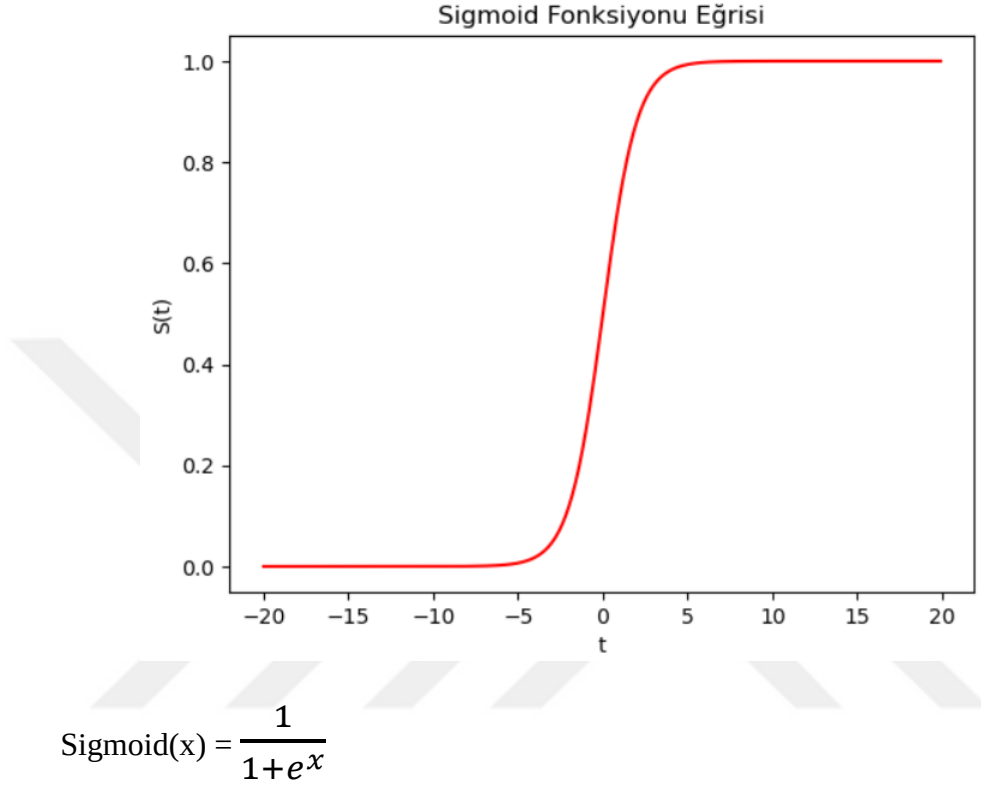


Şekil 13. LSTM birim yapısı (Wu et al. 2022)

Unutma kapısında (Forget gate) h_{t-1} ve x_1 bilgileri birleştirildikten sonra sigmoid fonksiyonundan geçirilirler. Sigmoid fonksiyonu sonucu 0 ile 1 aralığındadır. Sigmoid fonksiyonu Şekil 14'de verilmiştir. Bu şekilde 0 yakın sonuç çıkarsa unutma anlamı taşır yani o verilerin kaydedilmesi önemli değildir.

Giriş kapısında birleştirilmiş h_{t-1} ve x_1 verileri sigmoid fonksiyonundan geçirilirler ve sigmoid fonksiyonu bilgilerin hafızaya kaydetmek için önemli olup olmadığına karar verir. Tanh fonksiyonu verilerin önem boyutunu 1 ve -1 arasında hesaplar, sigmoid ve tanh'tan çıkan sonuçlara çarpma işlemi uygulandıktan sonra c_t 'ye eklenirler. Bu şekilde yeni hücre durumu (cell state) elde edilir.

Çıkış kapısı, sonraki gizli katman (hidden state h) değeri ne olacağına dair karar verir. Çıkış kapısında birleştirilmiş h_{t-1} ve x_t giriş verileri sigmoid fonksiyonundan geçirilirler. Sonra güncel hücre durumu (cell state c), tanh fonksiyonuna giriş olarak verilip önceki hesaplanan sigmoid çıkışıyla çarpılır. Şekil 13'te LSTM yapısı gösterilmiştir ve şekilde görüldüğü gibi h_t değeri elde edilir.



Şekil 14. Sigmoid fonksiyonu (Liu 2020)

Çift yönlü uzun kısa süreli bellek (BiLSTM)

LSTM, verilerin yalnızca ileri bilgi akışını öğrenebilirken, BiLSTM modelinin çıktı katmanını aynı anda hem ileri (forward) durumdan hem de geri (backward) durumdan bilgi alabileceğinden dolayı daha iyi bir tahmin doğruluğu verebilir (Bhattacharjee, Kumar, and Senthilkumar 2022). BiLSTM iki gizli katmanını ileri ve geri yönlerde olmak üzere tek bir çıkışa bağlar (Bhattacharjee et al 2022). BiLSTM ağ yapısı Şekil 15'de verilmiştir. İleri gizli katman ($H_{f(t)}$) ve geriye doğru gizli katman ($H_{b(t)}$) aşağıdaki gibi hesaplanır (Mousa and Schuller 2017; Wang et al. 2023):

$$H_{f(t)} = \tanh(W_{fh}x_t + W_{fhh}H_{f(t-1)} + b_f) \quad (2.6)$$

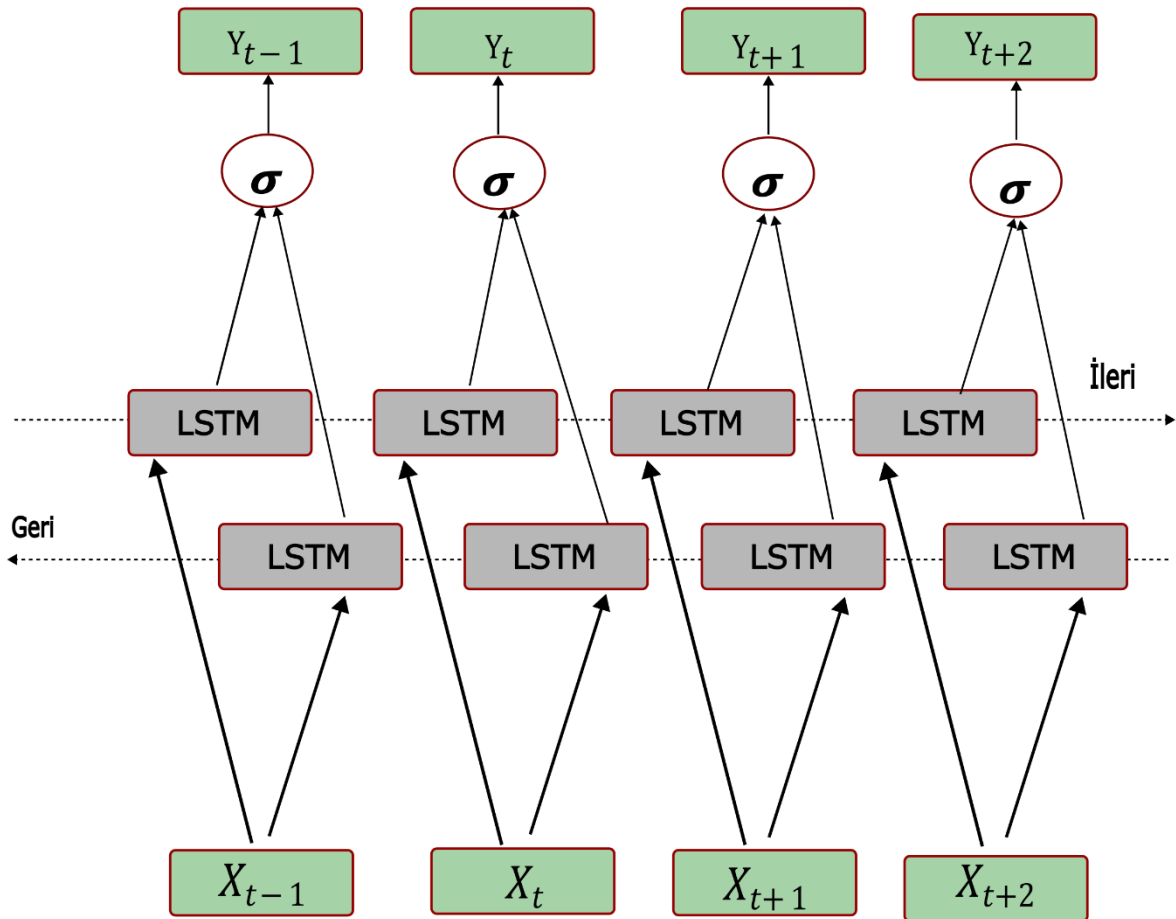
Bu denklemde, W_{fh} ve W_{fhh} giriş ve önceki durum vektörlerinin ağırlıklarını, b_f ise bias'ı temsil eder. σ , sigmoid aktivasyon fonksiyonunu gösterir. Forget gate f_t , hücrenin önceki durumundan hangi bilgileri unutması gerektiğini belirler.

$$H_{b(t)} = \tanh(W_{bh}x_t + W_{bhh}H_{f(t+1)} + b_b) \quad (2.7)$$

$H_{b(t)}$ denklemi, geri yönlü LSTM'deki hücre durumunu $H_{b(t)}$ hesaplar. Bu durum, geri yönde işlenen giriş verisini temsil eder.

$$Y_{(t)} = \sigma(W_{fhy}H_{f(t)} + W_{bhy}H_{b(t)} + b_y) \quad (2.8)$$

$Y_{(t)}$ denklemi, ileri ve geri yönlü LSTM ağlarının çıktılarını kullanarak son çıktıyı $Y_{(t)}$ hesaplar. Bu denklem, ileri ve geri yönlü LSTM çıktılarını birleştirerek bir sigmoid aktivasyon fonksiyonu (σ) üzerinden birleştirir.



Şekil 15. BiLSTM ağ yapısı(Wang et al. 2023)

Kapılı tekrarlayan hücre (gated recurrent unit GRU)

LSTM'nin bir çeşidi olarak kapılı tekrarlayan birim (GRU) performansı LSTM'ye benzer, ancak hesaplama karmaşıklığı çok daha düşüktür, hücre durumunu kaldırır ve bilgi iletmek için gizli durumu kullanır (Jung et al 2018). LSTM'nin popüler ve daha basit bir çeşidi olan GRU, LSTM ile aynı kapı kontrol mekanizmasını kullanır (Zhao et al. 2017). GRU Gradyan kaybolma sorununun çözmekle birlikte, LSTM'deki giriş kapısı ve unutma kapısını güncelleme kapısına birleştirir. GRU iki kapıdan oluşur sıfırlama kapısı (reset gate r_t) ve güncelleme kapısı (update gate z_t). Şekil 16, GRU'nun yapısını gösterir ve güncelleme denklemleri şu şekildedir:

Güncelleme kapısı z_t , önceki zaman adımının (h_{t-1}) gizli durumuna ve mevcut zaman adımındaki (x_t) girdiye dayalı olarak hesaplanır.

$$z_t = \sigma(W_z x_t + U_z h_{t-1}) \quad (2.9)$$

GRU'nun sıfırlama kapısı, r_t , ayrıca önceki zaman adımının (h_{t-1}) gizli durumuna ve mevcut zaman adımındaki (x_t) girişe göre güncellenir.

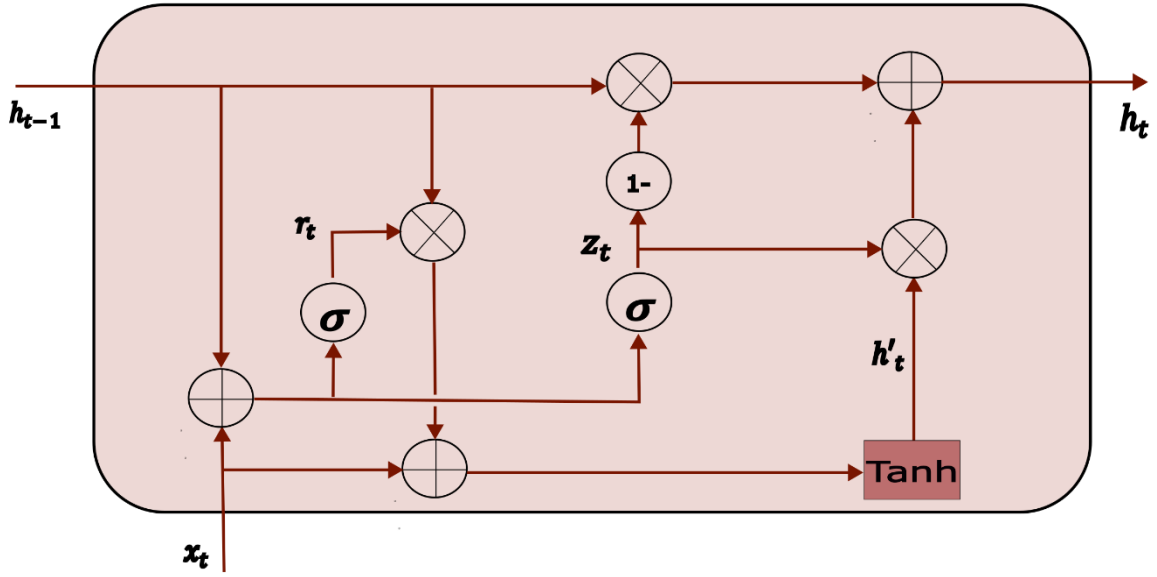
$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (2.10)$$

r_t 'ye bağlı olarak, bir güncelleme aday vektörü aşağıdaki gibi hesaplanabilir:

$$h'_t = \tanh(W x_t + r_t \odot U h_{t-1}) \quad (2.11)$$

Son olarak, bir GRU hücresi için, t zaman adımındaki çıkış h_t şu şekilde hesaplanır:

$$h_t = z_t \odot U h_{t-1} + (1 - z_t) \odot h'_t \quad (2.12)$$



Şekil 16. GRU birim yapısı (Zhang et al. 2022)

t zamanı için, x_t ve h_t giriş vektörleridir. $(W_z, U_z), (W_r, U_r), (W_{h'}, U_{h'})$ ağırlık matrisleri güncelleme kapısını, sıfırlama kapısını ve aday gizli durumunu (h') için sırasıyla ağırlıkları temsil ederler. $\odot$ hadamard ürünü (hadamard product), σ sigmoid fonksiyonu, ve $\tanh$ hiperbolik teğet (hyperbolic tangent) fonksiyonunu temsil eder.

Sıfırlama kapısı r_t , h'_t 'deki bilgi miktarının önceki zaman adımı h_{t-1} 'in gizli durumuna bağlı olarak ne kadar olduğunu kontrol edebilir. $r_t = 1$ ise, h'_t hem geçerli zaman adımı x_t 'deki girişe hem de önceki zaman adımı h_{t-1} 'in gizli durumuna bağlı olacaktır. Aksine, r_t sıfıra eşit ise; h'_t tamamen, önceki zaman adımı h_{t-1} 'in gizli durumundan bağımsız olarak x_t şimdiki zaman girişine bağlıdır. Bu şekilde kısa vadeli bağımlılıkları hedefleyen birimler, çoğunlukla aktif bir sıfırlama kapısına sahip olma eğilimindedir.

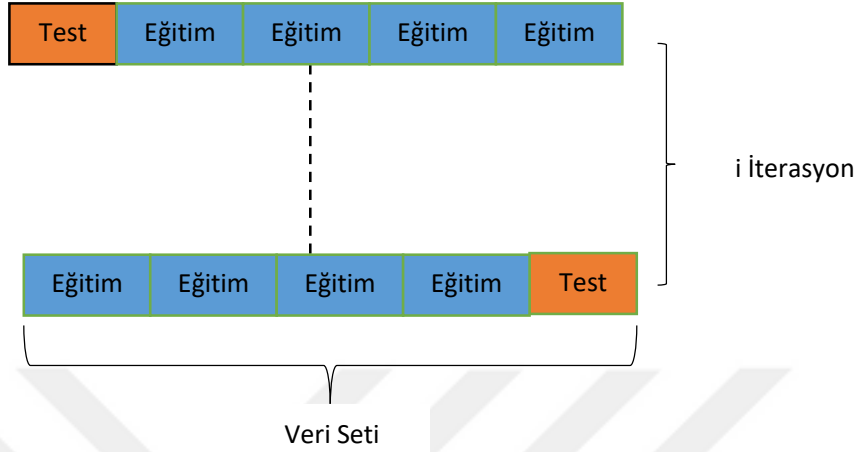
Güncelleme kapısı z_t , mevcut gizli durum h_t 'deki bilginin önceki zaman adımı h_{t-1} 'in gizli durumundan ne kadar etkileneceğini kontrol eder. z_t her zaman 1'e eşit olursa, GRU ağı mevcut x_t girişi ne olursa olsun her adımda aynı bilgiyi çıkış olarak verir. Bu nedenle, daha uzun vadeli bağımlılıkları hedefleyen birimler için sıklıkla aktif bir güncelleme kapısına sahiptirler.

Böylece, güncelleme kapısı ve sıfırlama kapısı, GRU biriminin uzun kısa vadeli bilgileri işlemesine yardımcı olabilir ve orijinal RNN gradyanının kaybolmasının sorununu azaltabilir.

Çapraz doğrulama

Çapraz doğrulama, bir modelin performansını objektif bir şekilde değerlendirmek ve genelleme yeteneğini tahmin etmek için kullanılan bir yöntemdir. Veri kümesi, eğitim ve test

verisi olarak ayrılır. Model, eğitim verisinde öğrenilir ve test verisinde performansı ölçülür. Bu işlem, veri kümesini farklı alt kümelerde bölerek tekrarlanır (Cheng et al. 2022). Her bir bölmede model eğitilir ve test edilir. Sonuç olarak, birden fazla performans ölçümü elde edilir. Çapraz doğrulama, modelin aşırı öğrenmeyi tespit etme yeteneğini artırır ve verilerin daha etkili bir şekilde kullanılmasını sağlar. Çapraz doğrulama süreci Şekil 17’de gösterilmiştir.



Şekil 17. Çapraz doğrulama süreci

Sınıflandırma değerlendirme yöntemleri

Denetimli makine öğrenmesi yöntemlerinin sonuçlarını değerlendirmek için birden fazla yöntem bulunmaktadır. Denetimli makine öğrenmesi modellerinin performansları en yaygın şekilde kesinlik (precision), duyarlılık (recall) ve doğruluk (accuracy) metrikleri kullanılarak değerlendirilmektedir. Bu yöntemlerin değerlerini elde etmek için 4 faktör elde edilmesi gerekiyor TP (True Positive), FP (False Positive), TN (True Negative) ve FN (False Negative). TP, FP, TN ve FN, pozitif sınıfın pozitif olarak sınıflandırıldığını, negatif sınıfın pozitif olarak sınıflandırıldığını, negatif sınıfın negatif olarak sınıflandırıldığını ve pozitif sınıfın negatif olarak sınıflandırıldığını sırasıyla belirtir (Urtnasan, Park, and Lee 2018). Bu faktörler Tablo 2’de gösterilmiştir.

Tablo 2. Karışıklık Matrisi

Kategori		Kategori Gerçek Değer	
		Evet	Hayır
Denetimli ML Modeli Kararı	Evet	TP	FP
	Hayır	FN	TN

Kesinlik, bir modelin diğer sınıflara ait olan gerçek sınıfları sınıflandırmama yeteneğini gösterir. Duyarlılık, bir modelin gerçek sınıflara ait olan gerçek sınıfları sınıflandırma

yeteneğini gösterir. Kesinlik (precision), duyarlılık (recall) ve doğruluk (accuracy) metrikleri Denklem 2.13, 2.14 ve 2.15'te sırasıyla gösterilmiştir.

$$Kesinlik = \frac{TP}{TP + FP} \quad (2.13)$$

$$Duyarlılık = \frac{TP}{TP + FN} \quad (2.14)$$

$$Doğruluk = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.15)$$



ARAŞTIRMA BULGULARI

Veri seti üzerindeki veri temizleme ve makine öğrenimi uygulaması, LDA yardımıyla müşterilerin dikkatini çeken ana konuları ve ana özellikleri ortaya çıkarıldı. Müşterilerin memnuniyet faktörleri elde edildikten sonra k-ortalamlar kullanarak müşteriler 3 gruba ayrıldı. Son olarak denetimli makine öğrenmesi yöntemi ile yorum sınıflandırma işlemi uygulandı.

Veri Analizi ve Konu Modelleme

Konular arasında yeterli boşluk ile kesişmeyen kümeler gösteren uygun konu sayısı olduğunu bulana kadar LDA'ya verilen konu sayısı değiştirildi (Zhang et al. 2021). Bu çalışmada elde edilen ve çevrimiçi kullanıcı yorumlarından oluşan veri seti için 4 konunun uygun olduğu görüldü. 4 ana konunun sonuçları Şekil 18, ve Şekil 19'da gösterilmiştir. Tablo 3'de elde edilen her konuya ait en fazla ağırlık taşıyan ilk 20 kelime ve o kelimelerin olasılıkları gösterildi.

Bu da Şekil 18'de gösterildiği gibi 4 ana konuya yol açtı. Ayrıca, konuları görselleştirmek için Şekil 19'da gösterildiği gibi her bir konuya bir kelime bulutu oluşturuldu.

Şekil 18'deki daireler konuları temsil eder ve her konunun boyutu, konu önemini gösterir. Anlamlı konu modelleri, görselleştirmede büyük, örtüşmeyen kümeler sergilemeli, bu da konuların birbirinden farklı olduğunu ve kolayca ayrılabilceğini (yani, kümelerin "iyi ayrılmış") olduğunu belirtir. Bu tür sonuçlar, küçük, örtüşen veya dağınık kümelere sahip konulardan daha yorumlanabilir (Sievert and Shirley 2014; Nicolas, Kim, and Chi 2021). Sonuç olarak birbirinden ayrılmış 4 konu elde edilmiştir. Şekil 18'de görüldüğü gibi sol tarafta 10 konu daire şeklinde görselleştirildi; daha büyük daireler, bir konunun daha yaygın olduğunu gösterir. Şeklin sağ tarafında, elde edilen belirli bir konuyla ilgili en önemli 30 ifade görüntülendi. Şekil 18, çevrimiçi inceleme veri setindeki 4 ana konuyu ve belirli bir konuyla ilgili en alakalı kelime dağılımını sunmaktadır. Bu çalışmada 4 konu oluşturulmuş ve her konu için 30 anahtar kelime elde edilmiştir. Elde edilen her bir konu için veri bulutu gösterimi birinci, ikinci, üçüncü ve dördüncü konu için sırasıyla Şekil 19'da sunulmuştur ve o her bir konuya ait sıklık açısından daha yüksek olan 15 kelime gösterilmiştir.

Tablo 3'de elde edilen her bir konu için en fazla ağırlığa sahip ilk 20 kelime gösterilmiştir. Tablonun ilk sütununda konu, ikinci sütununda kelimeler ve üçüncü sütununda ise kelimelerin olasılığı gösterilmiştir.

Müşterilerin dikkatini çeken ana faktörleri ve ana özellikleri Şekil 19'da gösterilen kelime bulutundan anlayabiliriz. Konu 1, havuzlar, restoranlar, teras manzaraları gibi mülklere odaklandı. Konu 2, içecekler, çay, kahve gibi içecek hizmetlerine odaklandı. Konu 3 kelime bulutu, vakit, saat, gece, gündüz gibi farklı kalite özelliklerine odaklandı ve son olarak, konu 4 personel, kahvaltı, konaklama gibi genel hizmetleri dikkate aldı.

Şekil 18. Elde edilen başlıca konular

Şekil 19. Konu1’de oluşturulan kelime bulutu

Tablo 3. Konulardaki Kelimelerin Ağırlık Olasılığı

Konu	Kelimeler	Olasılık değeri
Konu 1	1. room	1. 0.043084126
	2. four_season	2. 0.03221914
	3. pool	3. 0.023576122
	4. bosphorus	4. 0.023127267
	5. restaurant	5. 0.015259054
	6. area	6. 0.015104373
	7. terrace	7. 0.012945139
	8. spa	8. 0.012623256
	9. view	9. 0.012565276
	10. property	10. 0.010079036
	11. night	11. 0.00924857
	12. small	12. 0.008850245
	13. water	13. 0.008702054
	14. well	14. 0.008508602
	15. good	15. 0.007881225
	16. little	16. 0.0076405243
	17. walk	17. 0.0075065056
	18. price	18. 0.007224763
	19. bed	19. 0.0070347763
	20. large	20. 0.0067437827
Konu 2	1. drink	1. 0.037317242
	2. boat	2. 0.027895337
	3. stunning	3. 0.023480326
	4. fresh	4. 0.019243483
	5. coffee	5. 0.019110547
	6. tea	6. 0.015837856
	7. lunch	7. 0.015449807
	8. fruit	8. 0.014099809
	9. evening	9. 0.011889618
	10. river	10. 0.01125279
	11. marble	11. 0.010399811
	12. ground	12. 0.010331888
	13. turkish	13. 0.009963535
	14. hot	14. 0.009835182
	15. dish	15. 0.009276011
	16. ride	16. 0.009266182
	17. turkish_delight	17. 0.009125408
	18. menu	18. 0.008956711
	19. top_notch	19. 0.0089031365
	20. cold	20. 0.00877111

Tablo 3. (Devamı)

Konu 3	1. take	1. 0.015070304
	2. day	2. 0.0134060215
	3. go	3. 0.013352919
	4. leave	4. 0.010119528
	5. think	5. 0.00877672
	6. find	6. 0.008386071
	7. check	7. 0.008304651
	8. get	8. 0.00800398
	9. night	9. 0.0074142083
	10. time	10. 0.0072205695
	11. hour	11. 0.007115378
	12. know	12. 0.0069272993
	13. early	13. 0.0066962796
	14. way	14. 0.006652014
	15. book	15. 0.0064572585
	16. see	16. 0.006444127
	17. first	17. 0.0063382396
	18. ask	18. 0.006041153
	19. detail	19. 0.0058514113
	20. guest	20. 0.0058147386
Konu 4	1. stay	1. 0.03927564
	2. staff	2. 0.030648166
	3. service	3. 0.026199857
	4. great	4. 0.022764143
	5. location	5. 0.020434555
	6. good	6. 0.020203317
	7. room	7. 0.018950969
	8. breakfast	8. 0.017546473
	9. make	9. 0.014815984
	10. well	10. 0.012604807
	11. view	11. 0.012135054
	12. beautiful	12. 0.011391731
	13. excellent	13. 0.011329619
	14. amazing	14. 0.010515659
	15. city	15. 0.010406575
	16. place	16. 0.010385024
	17. helpful	17. 0.009217482
	18. friendly	18. 0.009115987
	19. recommend	19. 0.009111981
	20. time	20. 0.0090325195

Müşteri Segmentasyonu

K-means kümeleme yardımıyla elde edilen Segment1, Segment2 ve Segment3'e ait istatistiksel bilgiler Tablo 4, Tablo 5 ve Tablo 6'de sırasıyla gösterilmiştir.

Tablo 4. Az memnun grubu istatistiksel bilgileri

Özellik	Değer	Konum	Hizmet	Odalar	Temizlik	Uyku Kalitesi
Sayı	428	428	428	428	428	428
Ortalama	2.748096	3.690937	2.401026	3.332545	3.757971	3.547473
Standard Sapma	1.177635	1.123444	1.091948	1.189388	1.135368	1.288573
Min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
Max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Tablo 5. Orta memnun grubu istatistiksel bilgileri

Özellik	Değer	Konum	Hizmet	Odalar	Temizlik	Uyku Kalitesi
Sayı	1682	1682	1682	1682	1682	1682
Ortalama	4.004264	4.267242	4.396503	4.259593	4.665032	4.234868
Standard Sapma	0.692060	0.774266	0.694926	0.613923	0.466450	0.717602
Min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
Max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Tablo 6. Çok memnun grubu istatistiksel bilgileri

Özellik	Değer	Konum	Hizmet	Odalar	Temizlik	Uyku Kalitesi
Sayı	7197	7197	7197	7197	7197	7197
Ortalama	4.784434	4.808874	4.96763	4.797648	4.870024	4.798577
Standard Sapma	0.316615	0.313264	0.17990	0.302326	0.223211	0.294943
Min	4.000000	3.000000	3.000000	3.000000	1.000000	3.000000
Max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

Standart sapma, istatistikte bir veri setindeki değişkenliğin veya yayılmanın ölçüsüdür. Yüksek bir standart sapma, veri noktalarının daha geniş bir aralıkta dağıldığını ve daha fazla çeşitlilik veya yayılma gösterdiğini ifade eder.

Düşük standart sapma, istatistikte bir veri setindeki değişkenliğin veya yayılmanın düşük olduğunu ifade eder. Düşük bir standart sapma, veri noktalarının ortalama değere daha yakın ve daha sıkı bir şekilde bir araya geldiğini gösterir.

Tablo 6’da görüldüğü gibi çok memnun grubundaki standart sapma değeri diğer gruplara göre düşük görünüyor çünkü bu grupta müşteri memnuniyet yüksek ve bu şekilde derecelendirmelerin çoğu yüksek ve veri noktalarının ortalama değere daha yakın ve daha sıkı bir şekilde bir araya geldiği görülüyor.

Tablo 7. Müşteri Gruplarının Merkezleri (Centroids)

Özellik	Az memnun	Orta memnun	Çok memnun
Değer (Value)	2.761718	4.004492	4.784433
Konum (Location)	3.699124	4.266846	4.808873
Hizmet (Service)	2.412561	4.399473	4.967629
Odalar (Rooms)	3.327871	4.263563	4.797647
Temizlik (Cleanliness)	3.766861	4.665441	4.870023
Uyku Kalitesi (Sleep Quality)	3.545769	4.237356	4.798577

Merkezler (centroids), derecelendirmeleri aracılığıyla yeni müşterilerin memnuniyetini tahmin ederken önemlidir (Nilashi, et al. 2022). Merkezler (centroids) ayrıca müşteriler için hangi faktörlerin daha önemli olduğunu ortaya koyuyor (Nilashi, et al. 2022). Bu çalışmada Tablo 7’de görüldüğü gibi K-means kümeleme kullanılarak 3 müşteri grubu (az memnun, orta memnun ve çok memnun) elde edildi. Az memnun grubundaki hizmet ve değer (value) kriter derecelendirmelerinin diğer kriterlere göre oldukça düşük olduğu ortaya çıkarılmıştır. Orta memnun grubunda temizlik kriteri diğer kriterlere göre kısmen yüksektir ve genel olarak orta bir derecelendirme gösterilmiştir. Çok memnun grubunda müşteriler Hizmet kriterinden 4.967629 olarak en yüksek kriter derecelendirmesi gösterildi ve nispeten bu grupta tüm kriterler için derecelendirmeler çok yüksektir. Genel olarak, Tablo 7’te verilen her grup için kayıt sayısı sırasıyla 428, 1682 ve 7197’tür.

Yorum Sınıflandırma

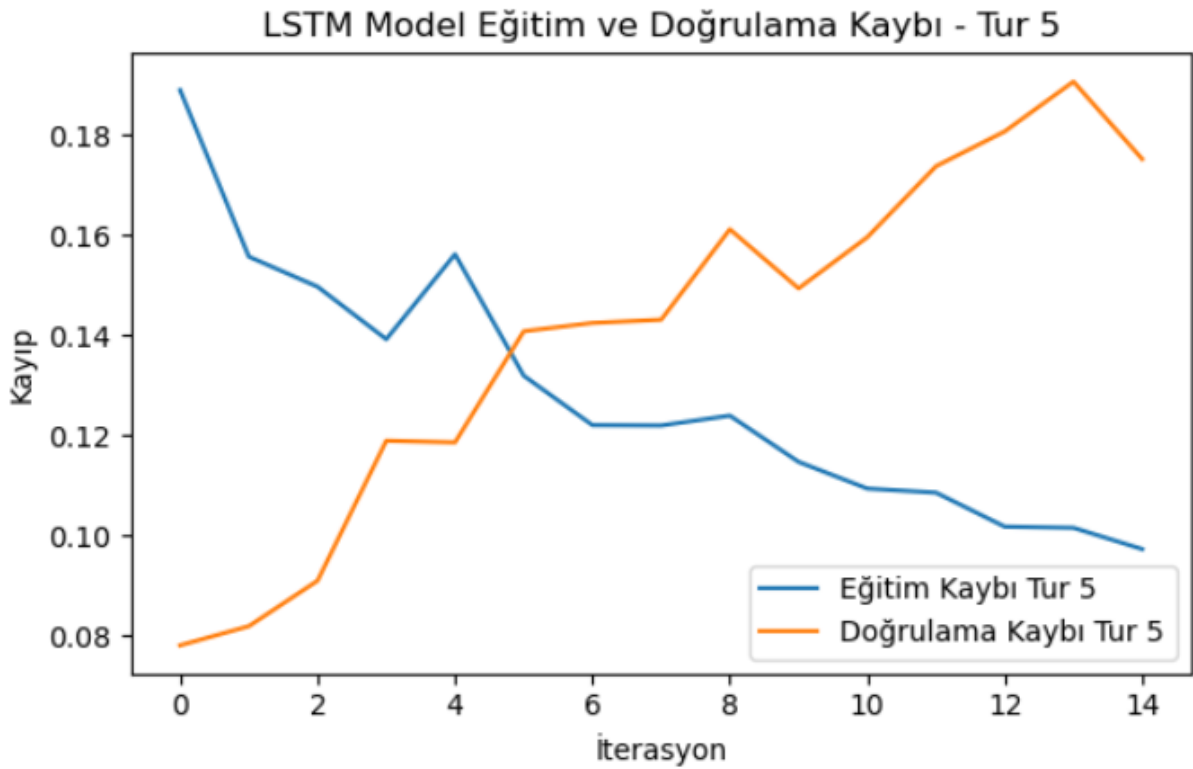
Yorum sınıflandırma işleminde LSTM, BiLSTM, GRU modelleri kullanılmıştır. Bu çalışmada tüm modellerde optimize edici adam ve sigmoid aktivasyon fonksiyonu kullanılmıştır. Eğitim sürecinde çapraz doğrulama uygulanmıştır. Çapraz doğrulama için tur sayısı 5 olarak ayarlanmıştır. Bu çalışmada (Polat B 2021) LSTM ve GRU kullanılarak yorum sınıflandırma işlemi için iterasyon (epoch) sayısı 15 olarak uygun görülmüştür. Benzer bir şekilde bu çalışmada modellerin eğitiminde iterasyon (epoch) sayısı 15 olarak ayarlanmıştır. Her bir modelin en yüksek elde edilen çapraz doğrulama turu sonuçları detaylı bir şekilde anlatılmıştır. En iyi sonuç veren model sonuç ve öneriler bölümünde ele alınmıştır.

LSTM model ile sınıflandırma

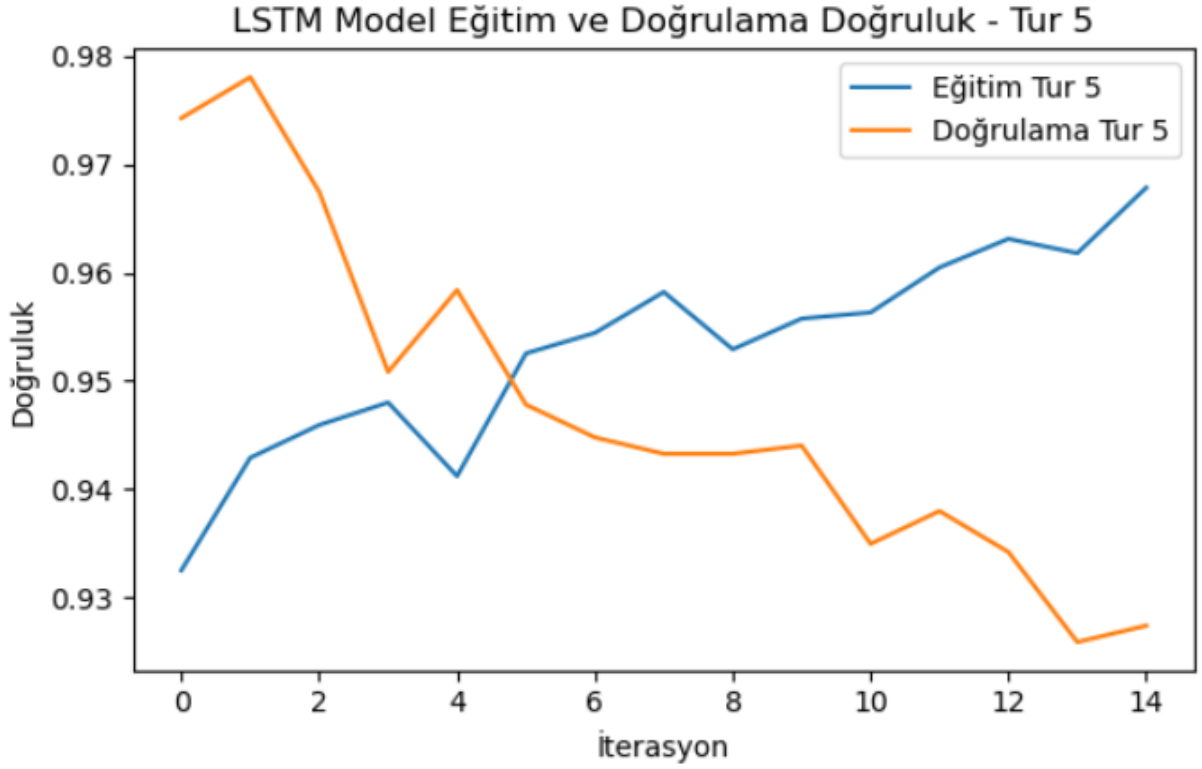
Bu bölümde LSTM modeli eğitimi ve elde sonuçları detaylı olarak anlatılmıştır. Bu modelin eğitimi sürecinde çapraz doğrulama uygulanmıştır. Çapraz doğrulama için tur sayısı 5 olarak ayarlanmıştır. Eğitim ve doğrulamada model 15 iterasyon olarak eğitilmiştir. Eğitim sonucunda elde edilen en yüksek çapraz doğrulama sonucu eğitim doğrulama kaybı bölümünde, ve karışıklık matrisi bölümünde anlatılmıştır.

Eğitim ve doğrulama kaybı

LSTM modelinin eğitime başlamasıyla Şekil 20’de gösterildiği gibi eğitim hata oranı yüksek başladı fakat eğitim adımlarının ilerlemesi ile birlikte hata oranı azalmıştır. LSTM modeli eğitim ve doğrulama doğruluk eğrisi Şekil 21’de gösterilmiştir. Bu model 15 adım olarak eğitilmiştir. Eğitim sonucunda kayıp: 17% ve doğruluk: 92% olarak elde edilmiştir. Şekil 20’de görüldüğü gibi eğitim kaybı eğrisi sıfıra doğru azalıyordur.



Şekil 20. LSTM modeli eğitim ve doğrulama kayıp eğrisi



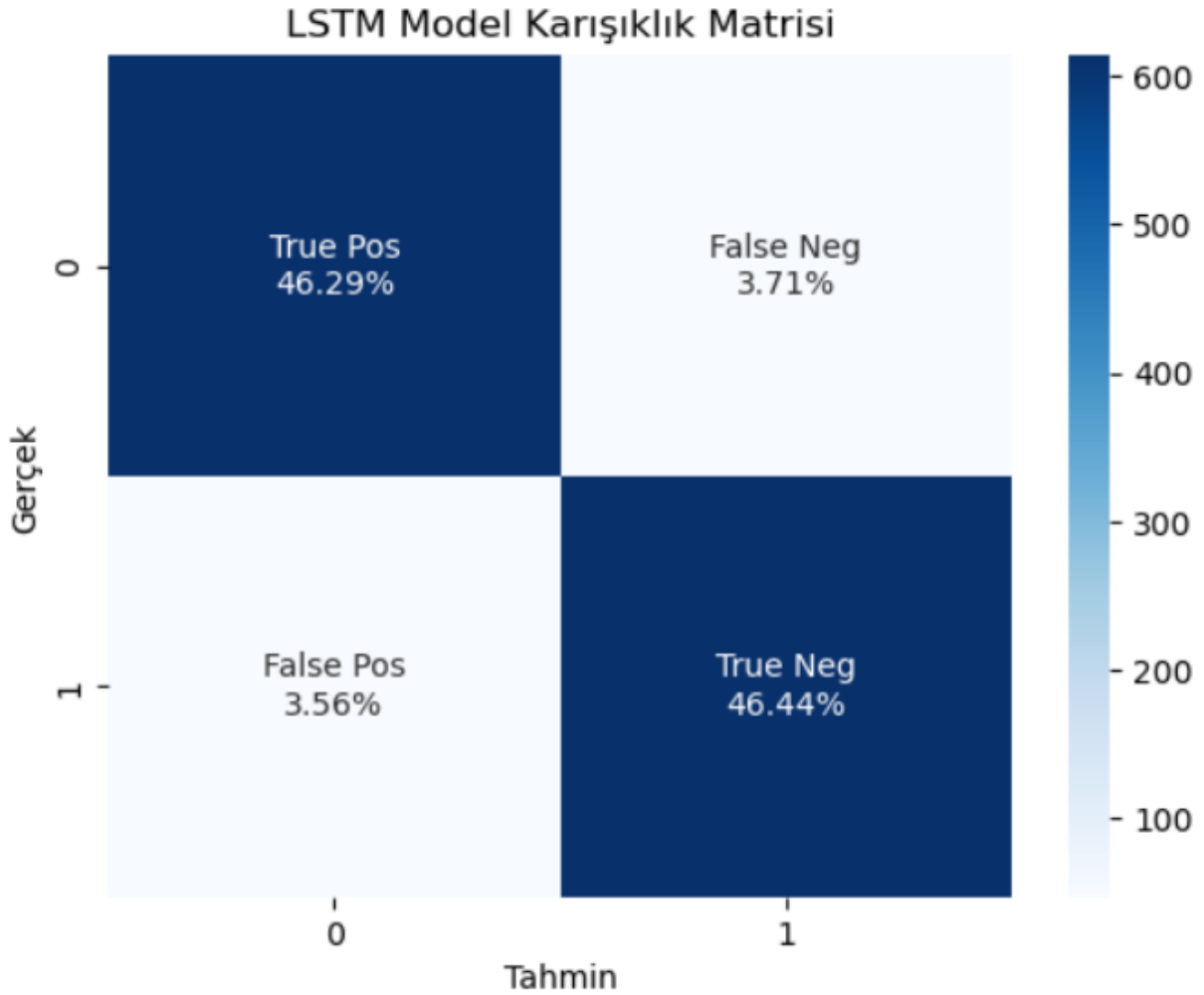
Şekil 21. LSTM modeli eğitim ve doğrulama doğruluk eğrisi

Karışıklık matrisi

Şekil 22’de verilen LSTM modeli karışıklık matrisi incelendiğinde 46.29% true-pozitif, 3.71% false-negatif, 3.56% false-pozitif, ve 46.44% true-negatif olarak elde edilmiştir. Bu çalışmada true pozitif ve true negatif yüksek olarak görülmektedir. Bu da modelin başarılı olduğunu göstermektedir. Bu modeli kullanarak elde edilen precision, recall ve F-1 score Tablo 8’de gösterilmiştir.

Tablo 8. LSTM Modeli Sonuçları

Kategori	Kesinlik	Duyarlılık	F-1 score
Olumsuz	0.93	0.93	0.93
Olumlu	0.93	0.93	0.93



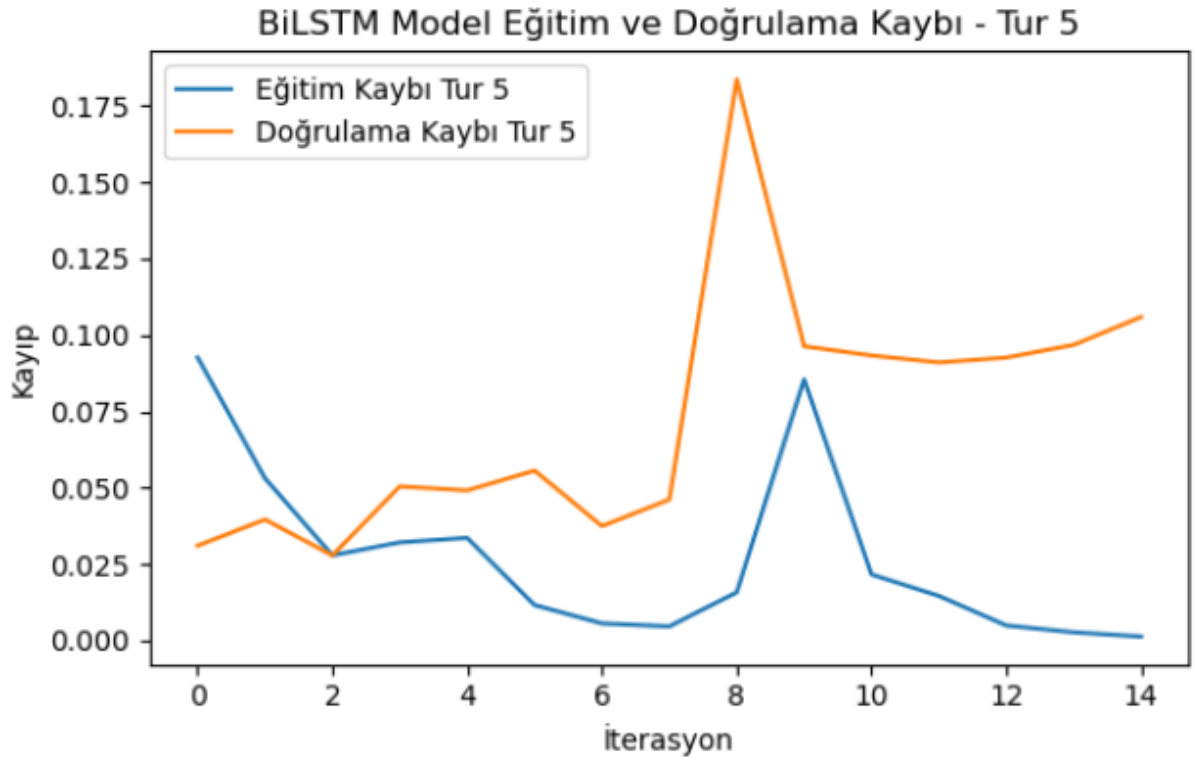
Şekil 22. LSTM modeli karışıklık matrisi

Çift yönlü LSTM model ile sınıflandırma

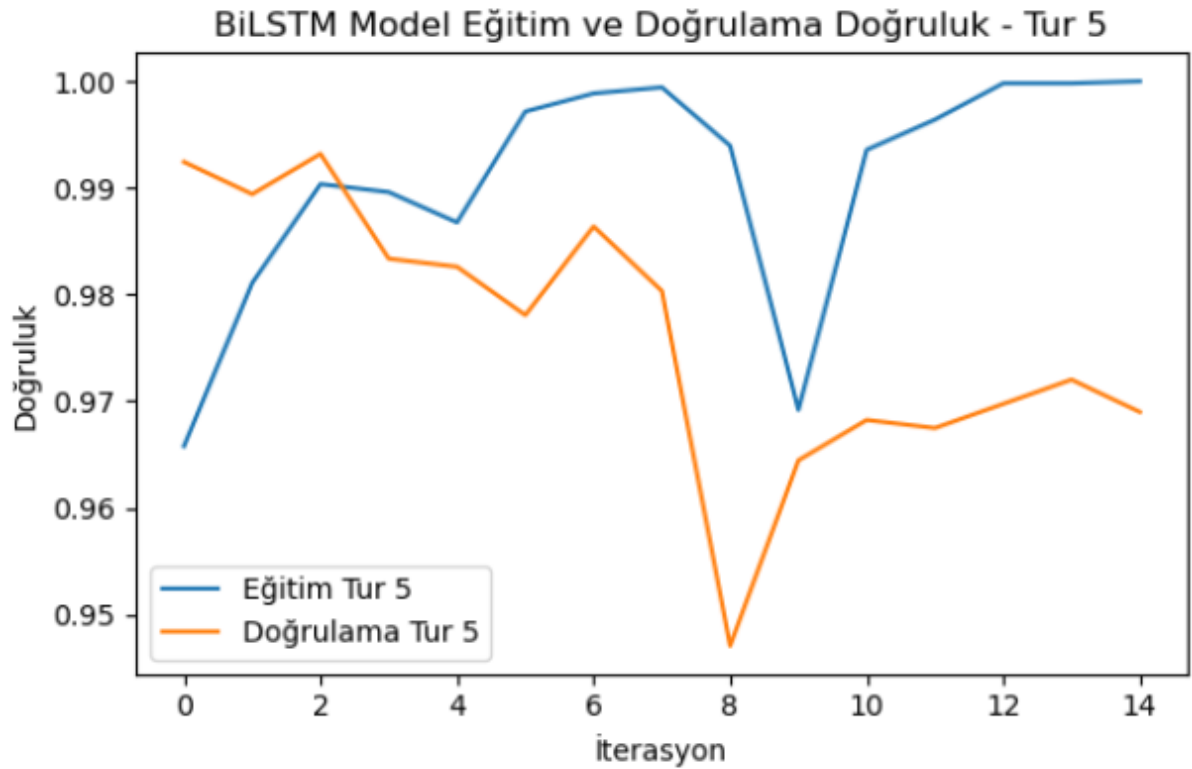
Bu bölümde çift yönlü LSTM modeli elde edilen veri seti üzerindeki eğitimi ve elde sonuçlar detaylı olarak anlatılmıştır. Bu modelin eğitimi sürecinde çapraz doğrulama uygulanmıştır. Çapraz doğrulama için tur sayısı 5 olarak ayarlanmıştır. Eğitim ve doğrulamada model 15 epok olarak eğitilmiştir. Eğitim sonucunda elde edilen en yüksek çapraz doğrulama sonucu eğitim doğrulama kaybı bölümünde, ve karışıklık matrisi bölümünde anlatılmıştır.

Eğitim ve doğrulama kaybı

BiLSTM modeli eğitime başlamasıyla Şekil 23'te gösterildiği gibi eğitim hata oranı yüksek başladı fakat eğitim adımlarının ilerlemesi ile birlikte hata oranı azalmıştır. BiLSTM modeli eğitim ve doğrulama doğruluk eğrisi Şekil 24'te gösterilmiştir. Bu model 15 adım olarak eğitilmiştir. Eğitim sonucunda kayıp:10% ve doğruluk: 96% olarak elde edilmiştir. Şekil 23'de görüldüğü gibi eğitim kaybı eğrisi sıfıra doğru azalmaktadır.



Şekil 23. Çift yönlü LSTM modeli eğitim ve doğrulama eğrisi



Şekil 24. Çift yönlü LSTM modeli eğitim ve doğrulama eğrisi

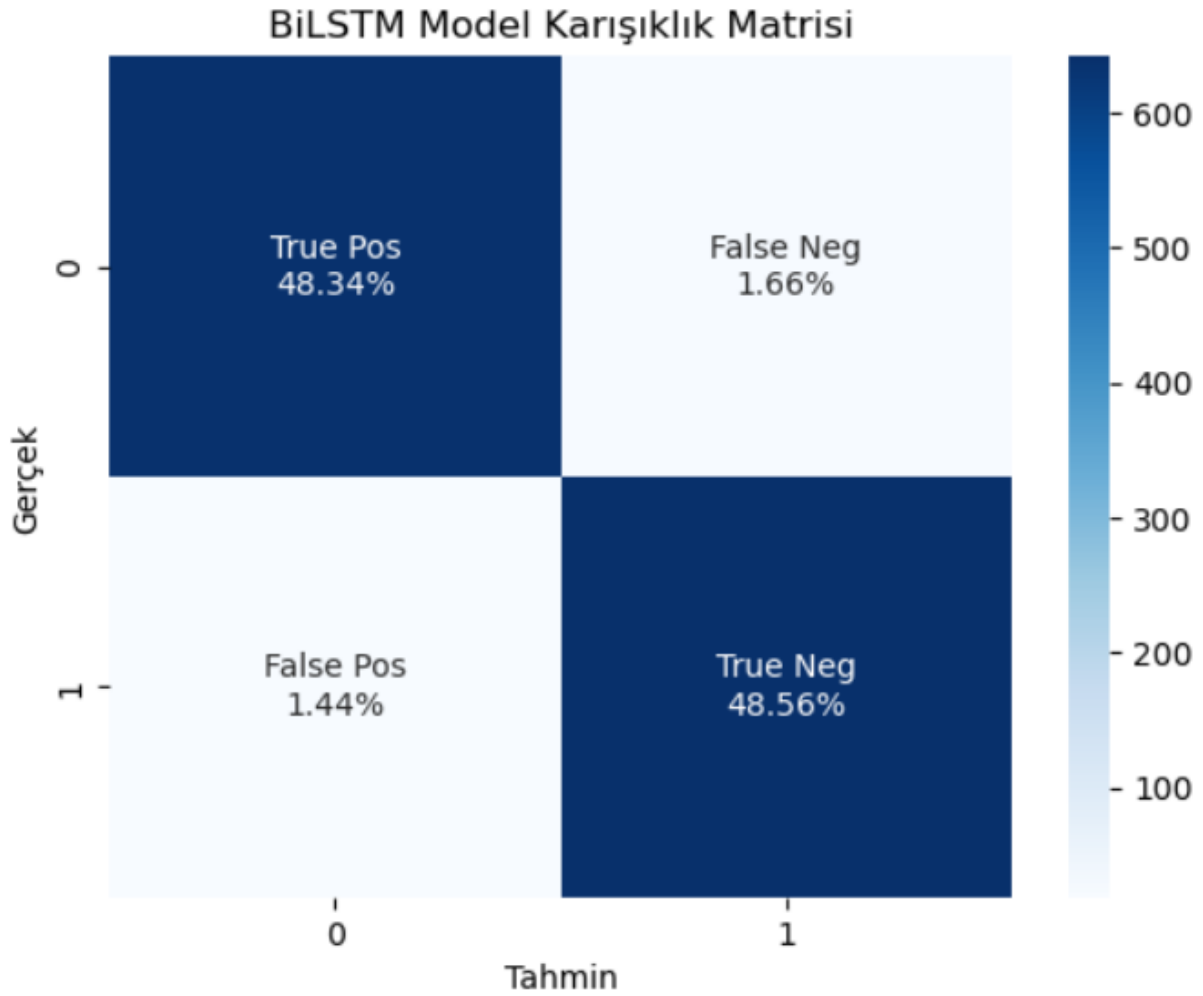
Karışıklık matrisi

Şekil 25'te verilen BiLSTM modeli karışıklık matrisi incelendiğinde 48.34% true pozitif, 1.66% false negatif, 1.44% false pozitif, ve 48.56% true negatif olarak elde edilmiştir.

Bu çalışmada true pozitif ve true negatif yüksek olarak görülmektedir. Bu da modelin başarılı olduğunu göstermektedir. Bu modeli kullanarak elde edilen precision, recall ve F-1 score Tablo 9’da gösterilmiştir.

Tablo 9. BiLSTM modeli sonuçları

Kategori	Kesinlik	Duyarlılık	F-1 score
Olumsuz	0.97	0.97	0.97
Olumlu	0.97	0.97	0.97



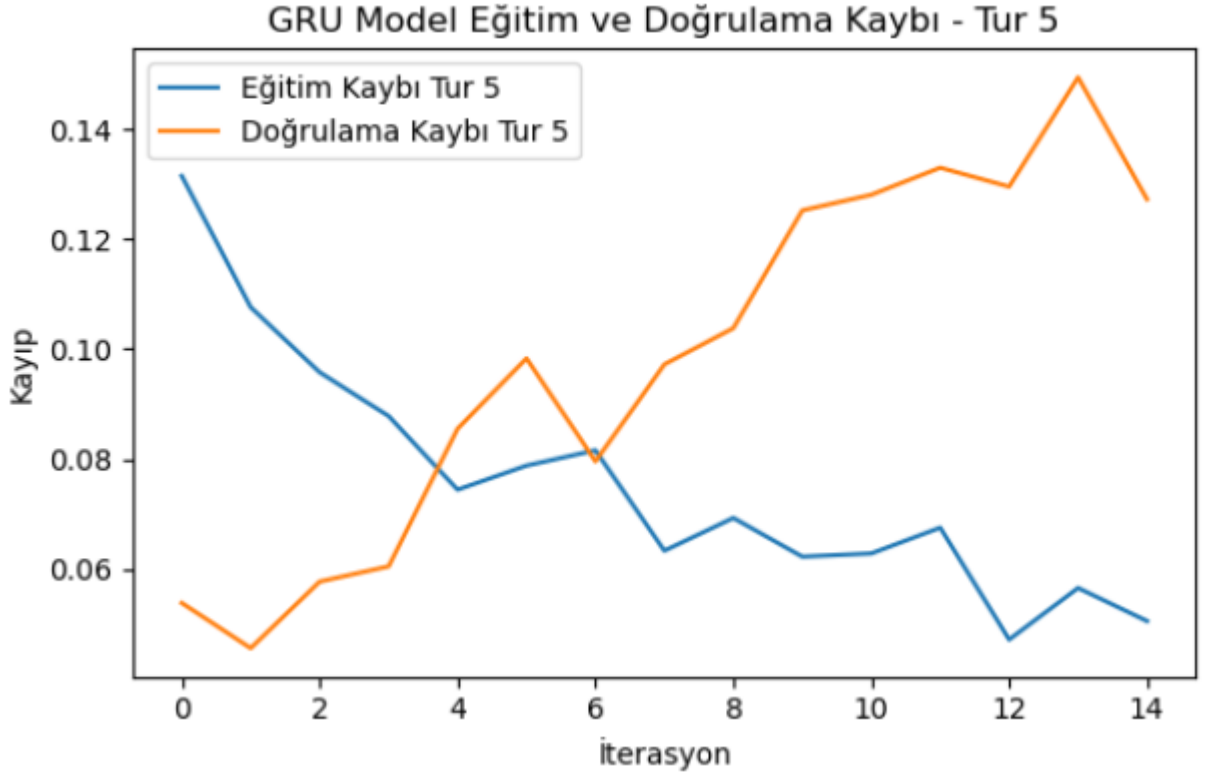
Şekil 25. Çift yönlü LSTM modeli karışıklık matrisi

GRU model ile sınıflandırma

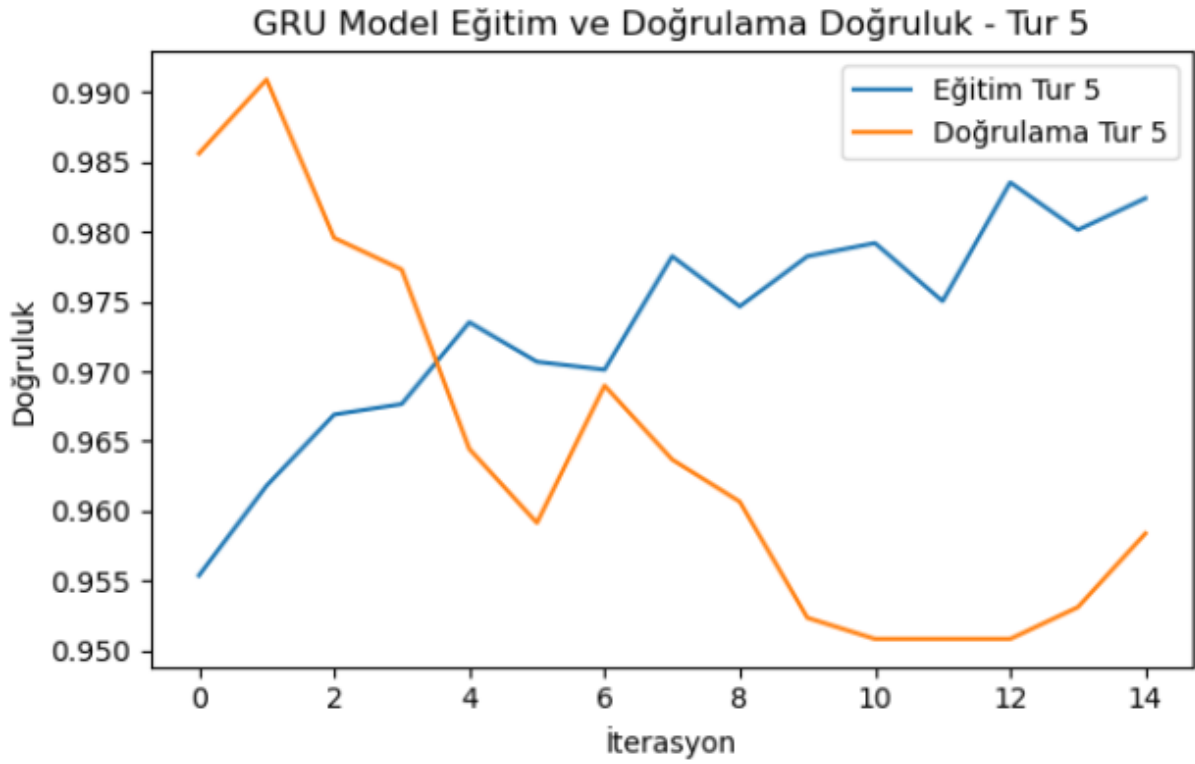
Bu bölümde GRU modeli eğitimi ve elde sonuçlar detaylı olarak anlatılmıştır. Bu modelin eğitimi sürecinde çapraz doğrulama uygulanmıştır. Çapraz doğrulama için tur sayısı 5 olarak ayarlanmıştır. Eğitim ve doğrulamada model 15 iterasyon (epoch) olarak eğitilmiştir. Eğitim sonucunda elde edilen en yüksek çapraz doğrulama sonucu eğitim doğrulama kaybı bölümünde, ve karışıklık matrisi bölümünde anlatılmıştır.

Eğitim ve doğrulama kaybı

GRU modeli eğitime başlamasıyla Şekil 26’da gösterildiği gibi eğitim hata oranı yüksek başlamış fakat eğitim adımlarının ilerlemesi ile birlikte hata oranı azalmıştır. GRU modeli eğitim ve doğrulama doğruluk eğrisi Şekil 27’de gösterilmiştir. Bu model 15 adım olarak eğitilmiştir. Eğitim sonucunda kayıp: 12% ve doğruluk: 95% olarak elde edilmiştir. Şekil 26’da görüldüğü gibi eğitim kaybı eğrisi sıfıra doğru azalmaktadır.



Şekil 26. GRU modeli eğitim ve doğrulama eğrisi



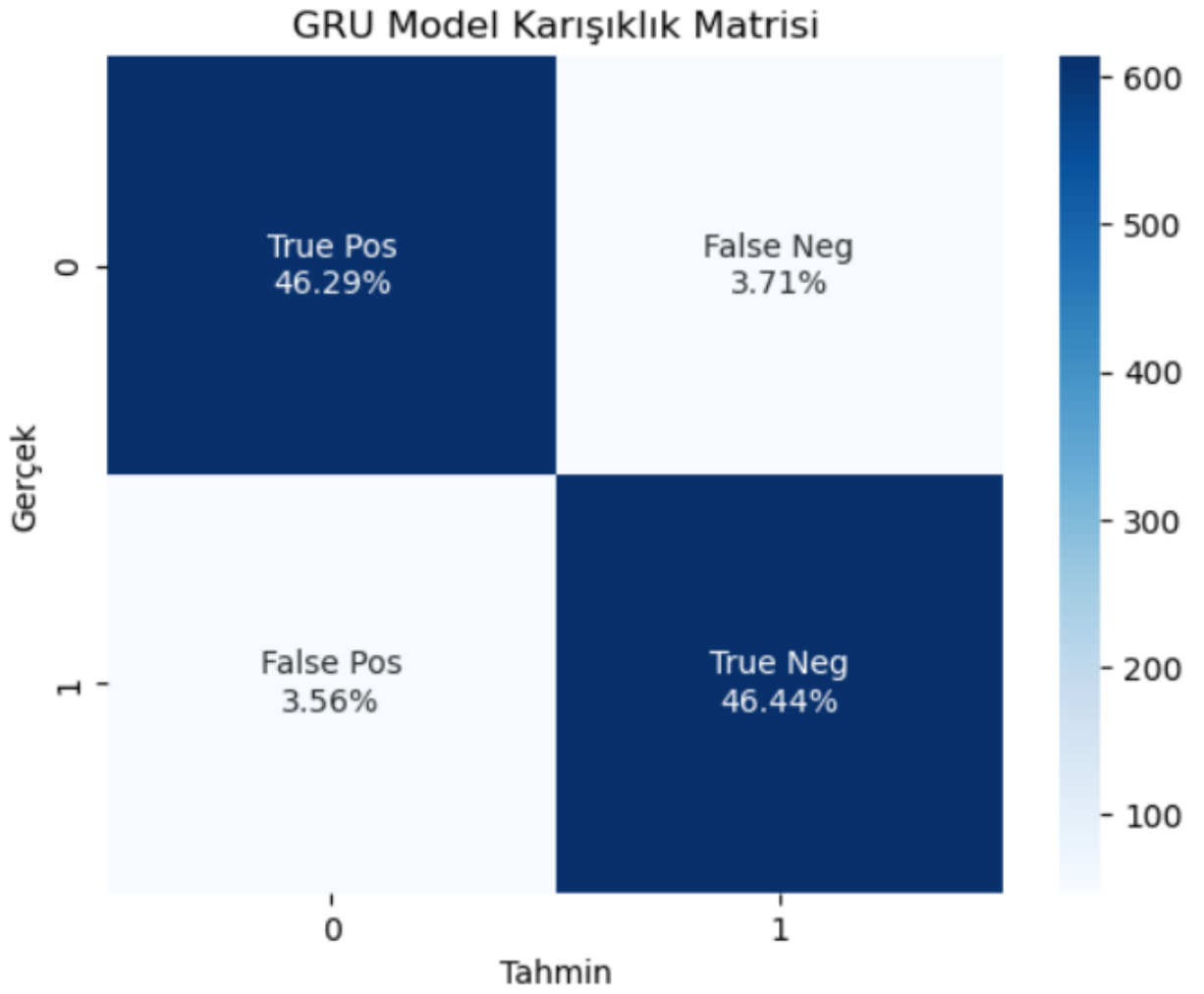
Şekil 27. GRU modeli eğitim ve doğrulama eğrisi

Karışıklık matrisi

Şekil 28’de verilen GRU modeli karışıklık matrisi incelendiğinde 46.29% true pozitif, 3.71% false negatif, 3.56% false pozitif, ve 46.44% true negatif olarak elde edilmiştir. Bu çalışmada true pozitif ve true negatif yüksek olarak görülmektedir. Bu da modelin başarılı olduğunu göstermektedir. Bu modeli kullanarak elde edilen precision, recall ve F-1 score Tablo 10’de gösterilmiştir.

Tablo 10. GRU Modeli Sonuçları

Kategori	Kesinlik	Duyarlılık	F-1 score
Olumsuz	0.97	0.95	0.96
Olumlu	0.95	0.97	0.96



Şekil 28. GRU modeli karışıklık matrisi

Sınıflandırma model performans değerlendirme

Yorumları sınıflandırma aşamasında birden fazla model eğitilmiştir. LSTM, BiLSTM ve GRU modelleri eğitildikten sonra performans değerlendirmesi için doğruluk ve diğer bazı performans değerlendirme yöntemleri kullanılmıştır. BiLSTM modeli ile en yüksek doğruluk oranı 96% olarak elde edilmiştir. Bu sonuç, LSTM ve GRU ile elde edilen sonuçlardan daha yüksektir. Bu sonuç diğer modellere göre daha yüksek olduğu için bu çalışmada BiLSTM modeli atandı. Elde edilen modelin sonucunu görmek için manuel olarak modelin görmediği negatif ve pozitif iki yorum modele giriş olarak verilmiştir. Bu iki yorum TripAdvisor sitesinden Türkiye’deki çevre dostu bir otele aittir. Bu yorumlar Tablo 11’de metin ve BiLSTM ile elde edilen sınıflandırma sonucu ile gösterilmiştir. BiLSTM modeli ile pozitif yorumu 1 olarak sınıflandırma değeri elde edildi. Bu değer doğru bir tahmin ve model bu yorum için doğru bir sonuç vermiştir. Negatif yorum için ayrıca 0.00019 sınıflandırma değeri elde edildi. Bu değer ayrıca negatif bir tahmin olduğu görülmüştür.

Tablo 11. Negatif ve Pozitif Test Yorumların Sınıflandırılması Örnek Sonuç

Yorum	Kategori
This hotel as everything you could wish for and it's all inclusive even the mini bar that's stocked daily and the safe the food is first class we wouldn't go to any other hotel in turkey we love it stayed three years on trot.	Pozitif
Absolutely the worst hotel for customer service, customer care, and any sort of people's skills. On arrival we did not receive a warm welcome; no introduction to the hotel and what facilities are available, especially as I have a baby. We were not given any information about what there was on the resort or about the restaurants, about room service, about children's entertainment or about the ironing rule. Which is that they don't have one in the room and you have to bring down two items before 10 am. I found out about this at a later time, when asked where the Iron was. When we checked in it was around 4pm, We were so hungry. The restaurant was closed but we was not told about the so called snack bar or where it even was. Instead we asked a cleaner who told us about the snack bar and how it was closing in 10 mins. He escorted all 3 of us there, we were all running as fast as we could behind this cleaner who told us to hurry even more, as it was closing. If it wasn't for him we would have gone without food for 10 hours. I asked about the mircowave to heat of my daughter's food and nobody knew what a mirowave meant. Until I eventually found someone who told me to go to the main restaurant. Which we eventually found and saw that they had a dedicated kids restaurant there, where they had a mirco wave, sterilizer and fresh food that they blend for your baby. How amazing and why was we not told about This??? I called for milk from my room I was told someone was bringing it within 5 mins, 45mins later no show, I called again they said 5 mins, 15 mins later no show. So I went downstairs, there I was told about the iron service and I asked to talk to a manager. Who basically found me to be more amusing then actually show any empathy. He said they should have told me at reception about the restaurants. Although I told him I had a one year old he still did not tell me about this kids club and what they had to offer my daughter or us as a family. He said if I need anything to ask him. However, everytime I saw him after that he never made eye contact with me not even when I smiled. Not one person who worked at the hotel ever said hello to us apart from the cleaners. We were made to feel invisible the entire time. There was a kids club in the resort we was never told about this or if it was free. I am used to travelling around 3/4 times a year; especially to 5 star hotels and deluxe 4 stars and never have I felt more out of place or unwanted in my entire life. As a new mother I felt even more disappointed and let down. Just saying hello to someone makes u feel better but here they only showed customer care to their Turkish or russian German customers and only catererd for them. The food was boring and the a le carte was so horrible. To make it worse we did not have a rep to talk to. We was stuck in a hotel that did not care for us and did not want to give us any information about what was available in the hotel and outside to do with my baby. I would never recommed this hotel to English customers. They do not care. The room was not properly cleaned on arrival with dust everywhere, sand and hair on the floor. I called for the cleaner who came and could not	Negatif

Tablo 11. (Devamı)

understand any english. I called my sister who thankfully knows how to speak Turkish and she explained about my asthma and my daughter's health condition and asked that the room is properly cleaned everyday. However, it was still poor. We even got given a bed runner with unsavoury stains, blood and other boderly fluids! I spoke to a few people who were visiting the hotel and none was happy with their service all saying the standards had droped and they don't care about you or your stay. When leaving the hotel we was not thanked or asked how we found our stay. They didn't even take off our wrist bands or ask to take our room keys. Just said you are checked out and off he went. He could not wait to get rid of us. Horrible staff and that is what makes or breaks a holiday experience for me. The hotel complex is massive. We would often get lost. We found a games room with bowling ally however, we stumbled across this on our last day. There should be hardcopy maps given to guests on arrival and explained in detail what is available to you and your family, as part of an introduction . Again I can not stress enough how disgusting the customer service is in this place, they do not value your custom in the slightest...! Stay away from this hotel. Because of this hotel I will never return to Turkey. Please do not respond to me with your generic response saying how disappointed you are to read this and how you try your upmost etc. Because you do not try what so ever!

TARTIŞMA

Veri toplama ve makine öğrenimi müşterilerin ilgisini etkileyen ana konuları ve ana özellikleri ortaya çıkarmaktadır. LDA'ya verilen konu sayısını değiştirerek, konuların birbirleriyle kesişmeyen ve yeterli boşluk bırakan uygun sayının 4 olduğu tespit edildi. 4 ana konunun sonuçları Şekil 18 ve Şekil 19'da gösterilmiştir. Şekil 19'dan müşterilerin ilgisini çeken ana kriterleri ve ana özellikleri çıkarabiliriz. 1. Konu, havuzlar, restoranlar, teras manzaraları gibi özelliklere odaklanırken, 2. Konu, içecek hizmetlerine (içecekler, çay, kahve vb.) odaklanmaktadır. 3. Konu kelime bulutu, zaman, saat, gece, gün vb. gibi diğer kalite özelliklerine odaklanmaktadır. Ve son olarak, 4. Konu, personel, kahvaltı, konaklama vb. gibi genel hizmetlere odaklanmaktadır. Elde edilen veri setinden müşteri memnuniyeti boyutlarını ortaya çıkarma işleminin ardından, müşteriler k-means kümeleme tekniği kullanılarak 3 segmente ayrıldı. Elde edilen segmentler, her bir değerlendirme kriteri için müşteri davranışlarını ortaya çıkarmaktadır. Tablo 7'deki küme merkezlerinden çıkarılan gruplar, müşteri memnuniyetinin her grup için sırasıyla düşük, orta ve yüksek olduğunu gösteriyor. Az memnun müşteri grubunda müşteri memnuniyeti temizlik ve konum kriterlerinde yüksek, hizmet kriterinde düşük olmuştur. Genel olarak az memnun müşteri grubunda müşteri memnuniyeti, diğer iki gruba kıyasla nispeten düşüktür. Orta memnun grubundaki müşteriler, diğer kriterlere kıyasla temizlik kriterinde daha fazla memnun olurken, değer kriterinde daha az memnun olmuşlardır ve orta memnun grubundaki müşteri memnuniyeti diğer iki gruba kıyasla nispeten orta düzeydedir. Son olarak, çok memnun müşteri grubundaki müşteri memnuniyeti çok yüksektir, özellikle hizmet kriterinde müşteriler 4.96 küme merkeziyle çok memnun olmuşlardır. Elde edilen 4 konuda sunulan özellikler ve her bir grup için çıkarılan küme merkezleri, otel yöneticilerine ve karar vericilere çevre dostu otellerle ilgili müşteri endişelerini anlama konusunda yardımcı olabilir. Otel yöneticileri ve karar vericiler, bu konu ve özellikler sayesinde çevre dostu otellerinde hangi alanları geliştirmeleri gerektiğini anlayabilirler. Bu 4 konu ve bu konularla ilgili özellikler, otel yöneticilerinin ve karar vericilerin müşterilerin çevre dostu otellerle ilgili endişelerini anlamalarına yardımcı olabilir. Otel yöneticileri ve karar vericiler bu konu ve özellikler sayesinde çevre dostu otellerinde hangi bölümleri geliştirmeleri gerektiğini de bilebilir. Kelime bulutlarında görüldüğü gibi, çevrimiçi kullanıcı yorumlarından dört ana konu çıkarıldı, Konu 1'i tesis merkezli, Konu 2'yi içecek merkezli, Konu 3'ü zamanlama merkezli ve son olarak Konu 4'ü hizmet merkezli olarak işaret edilebilir.

Bu çalışmada elde edilen bulguların benzer bağlamlardaki literatür sonuçlarıyla uyumu doğrulanabilir. Otellerdeki odalar, havuzlar, restoranlar gibi tesisler otel seçimi ve genel turizm

deneyiminin değerlendirilmesi için oldukça önem taşımaktadır (Zhang et al. 2021). İçecekler de turizm ve konaklama işletmelerinde araştırmacıların ilgisini çekmiştir (Park et al 2016; Türker and Süzer 2022). Otelden çıkış işlemleri, yemek ve oda servisi gibi hizmetler açısından zamanlama turist memnuniyeti için önemlidir ve literatürde incelenmiştir (De Palma et al 2018). Son olarak, otel endüstrisinde hizmet kalitesinin önemli rolü, ampirik sonuçlar aracılığıyla önceki literatürde onaylanmıştır (Fan et al 2022; Harif et al 2022; Perramon et al 2022).

Tablo 2'de gösterildiği gibi müşteri segmentasyonu, merkezlerin kriter derecelendirmeleri aracılığıyla yeni müşterilerin memnuniyetini tahmin etmedeki önemi nedeniyle çeşitli araştırmalarda kullanılmıştır (Nilashi, et al. 2022; Nilashi, et al. 2021).

Bu çalışmada müşteri yorumlarını olumlu ya da olumsuz olarak sınıflandırmak için LSTM, BiLSTM ve GRU modelleri eğitildi. Modellerin veri seti üzerinde elde edilen eğitim sonuçları Tablo 12'de gösterildi.

Tablo 12. LSTM, BiLSTM, ve GRU model sonuçları

Model	Kayıp	Doğruluk
LSTM	17%	92%
BiLSTM	10%	96%
GRU	12%	95%

Tablo 12'de görüldüğü gibi BiLSTM diğer eğitilen LSTM ve GRU modellerine göre en yüksek doğruluk değeri 96% ve en az kayıp değeri 10% verdi. Bu modelleri eğitmek için 15 iterasyon ve çapraz doğrulama için 5 tur uygulandı.

Müşteri memnuniyetini ML yöntemleri kullanılarak daha önce literatürde araştırılmış ve ML metodolojisinde LDA, K-means, ve bulanık mantık bazlı yaklaşımlar kullanılmıştır (Nilashi, et al. 2022). Bu tez çalışması sadece Türkiye'deki çevre dostu oteller üzerinde odaklanmıştır. Ayrıca LDA konu modellemede gensim kütüphanesi kullanılmıştır. Gensim kütüphanesi sayısı ile konu modelleme işleminin ne kadar başarılı olduğunu gösterilebilir. Ayrıca elde edilen her konu için bir kelime bulutu gösterildi. Çevre dostu otellerin müşterileri 3 kümeye ayrılarak her bir küme için ve her bir kriter derecelendirme için müşteri memnuniyeti ortaya çıkarıldı. Son olarak bu çalışmada RNN tabanlı LSTM, BiLSTM, ve GRU modelleri müşteri yorum sınıflandırılması için uygulandı. Bu çalışma (Onan 2021) İngilizce ürün yorumlarından oluşan Twitter veri seti kullanılarak CNN, RNN, LSTM, ve GRU ML yöntemleri eğitildi ve 93.85% doğruluk değeri ile CNN-LSTM diğer modeller göre daha başarı göstermiştir. Fakat bu tez çalışmasında Türkiye'deki çevre dostu otellerden ve İngilizce yorumlarından oluşan veri seti eğitiminde 15 iterasyon ve 5 tur çapraz doğrulama

kullanılmasıyla BiLSTM diğer LSTM, ve GRU modellerine göre en yüksek doğruluk değeri 96% ve en az kayıp değeri 10% göstererek daha yüksek performans sergiledi.

Bu çalışmanın toplanan veriler ve dağıtılan yöntem açısından birkaç sınırlılığı vardır. Veriler, bir çevrimiçi sosyal platformdan toplandı; TripAdvisor; turistlerin algılarını daha geniş bir şekilde araştırmak için diğer portallardan yararlanılabilir. Ayrıca kullanılan yöntem, LDA kullanarak gezginlerin memnuniyetinin boyutlarını keşfetmeye, müşterileri benzer derecelendirme davranışına sahip gruplara ayırmaya ve toplanan müşteri yorumlarının sınıflandırılması için LSTM, BiLSTM ve GRU ML modellerinin performansı karşılaştırılmasına odaklandı. Fakat, ankete dayalı bir yaklaşım içeren diğer araştırma modelleri ve bulanık mantık yaklaşımı kullanan müşteri tahmini, keşfedilen memnuniyet boyutlarının önem düzeylerinin sıralamasına ilişkin daha geniş bir algı sunabilir.



SONUÇ ve ÖNERİLER

Web platformlarındaki çevrimiçi yorumların büyük veri olarak kullanılması ve ML yöntemleri ile analiz edilmesi oldukça önemlidir. Müşteri beklentilerinin ortaya çıkarılması, turizm sektörü ve özellikle yeşil otellerdeki müşteri memnuniyetinin etkileyen faktörlerin analiz edilmesi açısından çok önemlidir. Yeşil turizm kapsamında müşteri memnuniyetini etkileyen faktörlerin elde edilmesi; karar-vericiler için, turizm gelirlerinin iyileştirilmesi ve turistler için önem taşımaktadır. Çalışmanın amacına ulaşmak için veri toplama ve temizleme işlemleri yapıldıktan sonra müşterilerin memnuniyetini etkileyen faktörlerin ortaya çıkarılması için LDA konu modelleme tekniği uygulandı. Müşterilerin TripAdvisor sitesindeki verdiği derecelendirmelere göre onları farklı gruplara ayırmak amacı ile K-means kümeleme tekniği uygulandı. Çevrim içi kullanıcı yorumlarının sınıflandırma işleminde LSTM, BiLSTM ve GRU ML yöntemleri uygulandı.

Bu çalışmada, TripAdvisor sitesinde kullanıcılar tarafından oluşturulan önemli bir büyük sosyal veri türü olarak kabul edilen çevrimiçi yorum verileri toplandı. Verileri toplamak için özel bir crawler python programlama dili ve selenium kütüphanesi kullanıldı. Veri, URL'lerini kullanarak otel sitelerindeki çevrimiçi yorumları tarayan tarama tekniği kullanılarak toplanmıştır.

Toplanan veriler ön işleme tabi tutuldu, durak sözcükleri silindi ve uzatıldı, uzatılmış durak sözcükleri anlamsız sözcükler ve tekrarlanan sözcükler içerir, e-postalar ve yeni satır simgeleri kaldırıldı, sözcükler tokenize edildi ve metin temizlendi. Turistlerin ilgisini çeken en önemli özellikler, LDA konu modelleme tekniği kullanılarak keşfedildi. Gezin memnuniyeti turizm sektöründe ve özellikle çevre dostu otellerde temel olarak önemlidir. Bu çalışmada, gezginlerin memnuniyet boyutlarını keşfetmek için çevre dostu otel sitelerinde çevrimiçi kullanıcı yorumları kullanıldı ve NLP ve makine öğrenme teknikleri uygulandı. Araştırma bulguları, tartışma bölümünde ele aldığımız 4 ana memnuniyet boyutunu göstermektedir.

Elde edilen kullanıcı memnuniyet boyutları yeşil oteller için dikkate alınması oldukça önemlidir. Oteller, gezginlerin çevrimiçi yorumlarından elde edilen bu boyutlara dayalı olarak otellerindeki ana özellikleri geliştirebilirler. Araştırma, turistlerin deneyimleri için önemli faktörleri ortaya çıkararak turizm endüstrisindeki karar vericilere içgörüler sunmuştur.

Bu aşamada müşteriler derecelendirmiş oldukları kriter derecelendirmelerine göre k-ortalamlar kullanılarak 3 gruba ayrıldı. Az memnun, orta memnun ve çok memnun olarak müşteriler 3 gruba ayrıldı. Değer, konum, hizmet, odalar, temizlik, ve uyku kalitesi

derecelendirme faktörleri ile elde edilen gruplara ve küme merkezlerine göre işletme sahiplerine hangi bölümleri iyileştirebilecekleri ile ilgili bir fikir vermiştir.

Son adımda yorum sınıflandırma işlemi uygulanmıştır. LSTM, BiLSTM ve GRU modelleri olarak veri seti üzerine eğitilmiş ve karşılaştırılmıştır. 96% doğruluk değeri ile en iyi performans gösteren BiLSTM modelidir.

Bu çalışma, toplanan veriler ve uygulanan yöntem açısından birkaç sınırlamaya sahiptir. Veriler tek bir çevrimiçi sosyal platformdan toplanmıştır, turistler arasındaki popülerliğinden kaynaklı olarak TripAdvisor hedef alınmıştır, turistlerin algılarını daha geniş bir şekilde araştırmak için diğer portallar kullanılabilir.



KAYNAKLAR

- Afrizal, Ahimsa Denhas, Nur Aini Rakhmawati, and Aris Tjahyanto. 2019. 'New filtering scheme based on term weighting to improve object based opinion mining on tourism product reviews', *Procedia Computer Science*, 161: 805-12.
- Ahadh, Abdhul, Govind Vallabhasseri Binish, and Rajagopalan Srinivasan. 2021. 'Text mining of accident reports using semi-supervised keyword extraction and topic modeling', *Process Safety and Environmental Protection*, 155: 455-65.
- Albahri, Ahmed Shihab, Rula A Hamid, ZT Al-qays, AA Zaidan, BB Zaidan, A OS Albahri, AH AlAmoodi, Jamal Mawlood Khlaf, EM Almahdi, and Eman Thabet. 2020. 'Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): a systematic review', *Journal of medical systems*, 44: 1-11.
- Alghamdi, Rubayyi, and Khalid Alfalqi. 2015. 'A survey of topic modeling in text mining', *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 6.
- Bahmanyar, Reza, Shiyong Cui, and Mihai Datcu. 2015. 'A comparative study of bag-of-words and bag-of-topics models of EO image patches', *IEEE Geoscience and Remote Sensing Letters*, 12: 1357-61.
- Bastani, Kaveh, Hamed Namavari, and Jeffrey Shaffer. 2019. 'Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints', *Expert Systems with Applications*, 127: 256-71.
- Bhattacharjee, Biplab, Rajiv Kumar, and Arunachalam Senthilkumar. 2022. 'Unidirectional and bidirectional LSTM models for edge weight predictions in dynamic cross-market equity networks', *International Review of Financial Analysis*, 84: 102384.
- Bian, Yiwen, Rongsheng Ye, Jing Zhang, and Xin Yan. 2022. 'Customer preference identification from hotel online reviews: A neural network based fine-grained sentiment analysis', *Computers & Industrial Engineering*: 108648.
- Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. 'Latent dirichlet allocation', *Journal of machine Learning research*, 3: 993-1022.
- Borkar, Gautam M, Leena H Patil, Dilip Dalgade, and Ankush Hutke. 2019. 'A novel clustering approach and adaptive SVM classifier for intrusion detection in WSN: A data mining concept', *Sustainable Computing: Informatics and Systems*, 23: 120-35.
- Bozkurt, Ferhat. 2012. 'Mekansal - zamansal veri madenciliği yörengelerin durma ve hareket algoritmaları', Yüksek lisans tezi, Atatürk University.
- Bozkurt, Ferhat, Önder Çoban, Faruk Baturalp Günay, and Şeyma Yücel Altay. 2019. 'High performance twitter sentiment analysis using CUDA based distance kernel on GPUs', *Tehnički vjesnik*, 26: 1218-27.
- Chen, Haihua, Lei Wu, Jiangping Chen, Wei Lu, and Junhua Ding. 2022. 'A comparative study of automated legal text classification using random forests and deep learning', *Information Processing & Management*, 59: 102798.
- Chen, Qimei, Miao Hu, Yi He, Ingrid Lin, and Anna S Mattila. 2022. 'Understanding guests' evaluation of green hotels: The interplay between willingness to sacrifice for the environment and intent vs. quality-based market signals', *International Journal of Hospitality Management*, 104: 103229.

- Chen, Weiling, Chai Kiat Yeo, Chiew Tong Lau, and Bu Sung Lee. 2018. 'Leveraging social media news to predict stock index movement using RNN-boost', *Data & Knowledge Engineering*, 118: 14-24.
- Chen, Zhenyu, Jianping Li, Liwei Wei, Weixuan Xu, and Yong Shi. 2011. 'Multiple-kernel SVM based multiple-task oriented data mining system for gene expression data analysis', *Expert Systems with Applications*, 38: 12151-59.
- Cheng, Jianhong, Hulin Kuang, Qichang Zhao, Yahui Wang, Lei Xu, Jin Liu, and Jianxin Wang. 2022. 'DWT-CV: Dense weight transfer-based cross validation strategy for model selection in biomedical data analysis', *Future Generation Computer Systems*, 135: 20-29.
- Chimmula, Vinay Kumar Reddy, and Lei Zhang. 2020. 'Time series forecasting of COVID-19 transmission in Canada using LSTM networks', *Chaos, Solitons & Fractals*, 135: 109864.
- De Palma, Andre, Carlos Ordás Criado, and Laingo M Randrianarisoa. 2018. 'When Hotelling meets Vickrey. Service timing and spatial asymmetry in the airline industry', *Journal of Urban Economics*, 105: 88-106.
- Ekinci, Ekin, and Sevinç İlhan Omurca. 2017. 'Ürün Özelliklerinin Konu Modelleme Yöntemi ile Çıkartılması', *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 9: 51-58.
- Emerson, Joseph, Bo Kara, and Jarka Glassey. 2020. 'Multivariate data analysis in cell gene therapy manufacturing', *Biotechnology Advances*, 45: 107637.
- Fan, Hua, Wei Gao, and Bing Han. 2022. 'How does (im) balanced acceptance of robots between customers and frontline employees affect hotels' service quality?', *Computers in Human Behavior*, 133: 107287.
- Filimonau, Viachaslau, Jorge Matute, Mirosław Mika, Magdalena Kubal-Czerwińska, Kinga Krzesiwo, and Aneta Pawłowska-Legwand. 2022. 'Predictors of patronage intentions towards 'green'hôtels in an emerging tourism market', *International Journal of Hospitality Management*, 103: 103221.
- Frye, Maik, Johannes Mohren, and Robert H Schmitt. 2021. 'Benchmarking of Data Preprocessing Methods for Machine Learning-Applications in Production', *Procedia CIRP*, 104: 50-55.
- Gao, Baojun, Minyue Zhu, Shan Liu, and Mei Jiang. 2022. 'Different voices between Airbnb and hotel customers: An integrated analysis of online reviews using structural topic model', *Journal of Hospitality and Tourism Management*, 51: 119-31.
- García, Hilario López, and Iván Machón González. 2004. 'Self-organizing map and clustering for wastewater treatment monitoring', *Engineering applications of artificial intelligence*, 17: 215-25.
- Ghiasi, Alireza, Ching-Tai Ng, and Abdul Hamid Sheikh. 2022. "Damage detection of in-service steel railway bridges using a fine k-nearest neighbor machine learning classifier." In *Structures*, 1920-35. Elsevier.
- Godnov, Uroš, and Tjaša Redek. 2016. 'Application of text mining in tourism: case of Croatia', *Annals of Tourism Research*, 58: 162-66.
- Guo, Hongfeng, Shengxiang Xia, Qiguang An, Xin Zhang, Weihua Sun, and Xinyao Zhao. 2020. 'Empirical study of financial crises based on topological data analysis', *Physica A: Statistical Mechanics and its Applications*, 558: 124956.

- Han, Heesup, Jin-Soo Lee, Ho Le Thu Trang, and Wansoo Kim. 2018. 'Water conservation and waste reduction management for increasing guest loyalty and green hotel practices', *International Journal of Hospitality Management*, 75: 58-66.
- Han, J., J. Pei, and M. Kamber. 2011. *Data Mining: Concepts and Techniques* (Elsevier Science).
- Harif, Mohd Amy Azhar Mohd, Muhammad Nawaz, and Waseem Ul Hameed. 2022. 'The role of open innovation, hotel service quality and marketing strategy in hotel business performance', *Heliyon*: e10441.
- Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. 'Long short-term memory', *Neural computation*, 9: 1735-80.
- Hu, Nan, Ting Zhang, Baojun Gao, and Indranil Bose. 2019. 'What do hotel customers complain about? Text analysis using structural topic model', *Tourism Management*, 72: 417-26.
- Huang, Jing, Yulang Guo, Cheng Wang, and Lei Yan. 2019. 'You touched it and I'm relieved! The effect of online review's tactile cues on consumer's purchase intention', *Journal of Contemporary Marketing Science*.
- Ida, Yoshiaki, Eisuke Fujita, and Takashi Hirose. 2022. 'Classification of volcano-seismic events using waveforms in the method of k-means clustering and dynamic time warping', *Journal of Volcanology and Geothermal Research*, 429: 107616.
- Imberman, Susan P. 2001. "Effective use of the kdd process and data mining for computer performance professionals." In *Int. CMG Conference*, 611-20. Citeseer.
- Jelodar, Hamed, Yongli Wang, Rita Orji, and Shucheng Huang. 2020. 'Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach', *IEEE Journal of Biomedical and Health Informatics*, 24: 2733-42.
- Jung, Minju, Haanvid Lee, and Jun Tani. 2018. 'Adaptive detrending to accelerate convolutional gated recurrent unit training for contextual video recognition', *Neural Networks*, 105: 356-70.
- Kamilaris, Andreas, Andreas Kartakoullis, and Francesc X Prenafeta-Boldú. 2017. 'A review on the practice of big data analysis in agriculture', *Computers and Electronics in Agriculture*, 143: 23-37.
- Khalidi, Rohaifa, Abdellatif El Afia, Raddouane Chiheb, and Siham Tabik. 2023. 'What is the best RNN-cell structure to forecast each time series behavior?', *Expert Systems with Applications*, 215: 119140.
- Krishnamoorthy, M, and R Karthikeyan. 2022. 'Pattern mining algorithms for data streams using itemset', *Measurement: Sensors*: 100421.
- Lepakshi V. 2022. *Computational Approaches for Novel Therapeutic and Diagnostic Designing to Mitigate SARS-CoV-2 Infection* (Academic Press,: United Kingdom: London).
- Li, Jingjing, Jianling Jiao, Yuwen Xu, and Chuxi Chen. 2021. 'Impact of the latent topics of policy documents on the promotion of new energy vehicles: empirical evidence from Chinese cities', *Sustainable Production and Consumption*, 28: 637-47.
- Liang, Minhui, and Tiansen Niu. 2022. 'Research on Text Classification Techniques Based on Improved TF-IDF Algorithm and LSTM Inputs', *Procedia Computer Science*, 208: 460-70.

- Lin, Guancen, Aijing Lin, and Danlei Gu. 2022. 'Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient', *Information Sciences*, 608: 517-31.
- Liu, Hui. 2020. *Robot systems for rail transit applications* (Elsevier).
- Luke, James. 2015. 'Data mining of automatically promotion Tweet for products and services using Naïve Bayes Algorithm to increase Twitter engagement followers AtPT. Bobobobo', *Procedia Computer Science*, 59: 254-61.
- Ma, Guanhua, Junhua Ma, Hao Li, Yiming Wang, Zhaohua Wang, and Bin Zhang. 2022. 'Customer behavior in purchasing energy-saving products: Big data analytics from online reviews of e-commerce', *Energy Policy*, 165: 112960.
- Mansoor, Muvazima, Kirthika Gurumurthy, and VR Prasad. 2020. 'Global sentiment analysis of COVID-19 tweets over time', *arXiv preprint arXiv:2010.14234*.
- Martin-Rodilla, Patricia, Jose Ignacio Panach, Cesar Gonzalez-Perez, and Oscar Pastor. 2018. 'Assessing data analysis performance in research contexts: An experiment on accuracy, efficiency, productivity and researchers' satisfaction', *Data & Knowledge Engineering*, 116: 177-204.
- Mengutaycı Ü. 2021. 'Yapay Sinir Ağları ile Otel Müşteri Yorumlarının Metinsel Yolla Sınıflandırılarak Duygu Analizlerinin Tespiti', Yüksek Lisans Tezi, Lisansüstü Eğitim Enstitüsü, Kütahya Dumlupınar Üniversitesi.
- Mohamed, Ehsan, and Tahir Celik. 2022. 'Early detection of failures from vehicle equipment data using K-means clustering design', *Computers and Electrical Engineering*, 103: 108351.
- Mousa, Amr, and Björn Schuller. 2017. 'Contextual bidirectional long short-term memory recurrent neural network language models: A generative approach to sentiment analysis'.
- Mythily, R, Aisha Banu, and Shriram Raghunathan. 2015. 'Clustering models for data stream mining', *Procedia Computer Science*, 46: 619-26.
- Narasimhulu, K, and KT Meena Abarna. 2022. 'High performance social data computing with development of intelligent topic models for healthcare', *Microprocessors and Microsystems*, 95: 104690.
- Nemes, László, and Attila Kiss. 2021. 'Social media sentiment analysis based on COVID-19', *Journal of Information and Telecommunication*, 5: 1-15.
- Nicolas, Clément, Jinwoo Kim, and Seokho Chi. 2021. 'Natural language processing-based characterization of top-down communication in smart cities for enhancing citizen alignment', *Sustainable Cities and Society*, 66: 102674.
- Nilashi, Mehrbakhsh, Rabab Ali Abumalloh, Abdullah Alghamdi, Behrouz Minaei-Bidgoli, Abdulaziz A Alsulami, Mohammed Thanoon, Shahla Asadi, and Sarminah Samad. 2021. 'What is the impact of service quality on customers' satisfaction during COVID-19 outbreak? New findings from online reviews analysis', *Telematics and Informatics*, 64: 101693.
- Nilashi, Mehrbakhsh, Rabab Ali Abumalloh, Ahmed Almulihi, Mesfer Alrizq, Abdullah Alghamdi, Muhammed Yousoof Ismail, Abul Bashar, Waleed Abdu Zogaan, and Shahla Asadi. 2021. 'Big social data analysis for impact of food quality on travelers' satisfaction in eco-friendly hotels', *ICT Express*.
- Nilashi, Mehrbakhsh, Rabab Ali Abumalloh, Behrouz Minaei-Bidgoli, Waleed Abdu Zogaan, Ashwaq Alhargan, Saidatulakmal Mohd, Sharifah Nurlaili Farhana Syed Azhar, Shahla

- Asadi, and Sarminah Samad. 2022. 'Revealing travellers' satisfaction during COVID-19 outbreak: moderating role of service quality', *Journal of Retailing and Consumer Services*, 64: 102783.
- Nilashi, Mehrbakhsh, Rabab Ali Abumalloh, Masoumeh Zibarzani, Sarminah Samad, Waleed Abdu Zogaan, Muhammed Yousoof Ismail, Saidatulakmal Mohd, and Noor Adelyna Mohammed Akib. 2022. 'What Factors Influence Students Satisfaction in Massive Open Online Courses? Findings from User-Generated Content Using Educational Data Mining', *Education and Information Technologies*: 1-35.
- Nilashi, Mehrbakhsh, Behrouz Minaei-Bidgoli, Mesfer Alrizq, Abdullah Alghamdi, Abdulaziz A Alsulami, Sarminah Samad, and Saidatulakmal Mohd. 2021. 'An analytical approach for big social data analysis for customer decision-making in eco-friendly hotels', *Expert Systems with Applications*, 186: 115722.
- Onan, Aytuğ. 2021. 'Sentiment analysis on product reviews based on weighted word embeddings and deep neural networks', *Concurrency and Computation: Practice and Experience*, 33: e5909.
- Park, Sohyun, E Lundeen, and HM Blanck. 2016. 'Knowledge of Health Conditions Related to Drinking Sugar-Sweetened Beverage and Sugar-Sweetened Beverage Intake Among US Adults', *Journal of Nutrition Education and Behavior*, 48: S98.
- Pavlinek, Miha, and Vili Podgorelec. 2017. 'Text classification method based on self-training and LDA topic models', *Expert Systems with Applications*, 80: 83-93.
- Perramon, Jordi, Marc Oliveras-Villanueva, and Josep Llach. 2022. 'Impact of service quality and environmental practices on hotel companies: An empirical approach', *International Journal of Hospitality Management*, 107: 103307.
- Pliakos, Konstantinos, and Celine Vens. 2018. 'Mining features for biomedical data using clustering tree ensembles', *Journal of biomedical informatics*, 85: 40-48.
- Polat B. 2021. 'TÜRKÇE ÜRÜN YORUMLARI VERİSİ İLE DUYGU ANALİZİ', Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Ankara Üniversitesi.
- Prihayati, Yuni, and Thomas Oni Veriasa. 2021. "Developing green tourism to create the sustainable landscape: evidence from Community-based Coffee Tourism (CbCT) in Puncak, Bogor, Indonesia." In *IOP Conference Series: Earth and Environmental Science*, 012027. IOP Publishing.
- Rita, Paulo, Sérgio Moro, and Gabriel Cavalcanti. 2022. 'The impact of COVID-19 on tourism: Analysis of online reviews in the airlines sector', *Journal of Air Transport Management*, 104: 102277.
- Sato, Yuki, Kazuhiro Izui, Takayuki Yamada, and Shinji Nishiwaki. 2019. 'Data mining based on clustering and association rule analysis for knowledge discovery in multiobjective topology optimization', *Expert Systems with Applications*, 119: 247-61.
- SEC. 2019. 'TripAdvisor Inc 2019 Form 10-K Annual Report'. https://www.sec.gov/ix?doc=/Archives/edgar/data/1526520/000156459020005269/trip-10k_20191231.htm.
- Shaheen, Musarrat, Farrah Zeba, Namrata Chatterjee, and Raveesh Krishnankutty. 2019. 'Engaging customers through credible and useful reviews: the role of online trust', *Young Consumers*.
- Shalvi, Doron, and Nicholas DeClariss. 1998. "An unsupervised neural network approach to medical data mining techniques." In *1998 IEEE International Joint Conference on*

- Neural Networks Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98CH36227)*, 171-76. IEEE.
- Shu X, Ye Y. 2022. 'Knowledge Discovery: Methods from data mining and machine learning', *Social Science Research*.
- Sievert, Carson, and Kenneth Shirley. 2014. "LDAvis: A method for visualizing and interpreting topics." In *Proceedings of the workshop on interactive language learning, visualization, and interfaces*, 63-70.
- Sim, Youngseok, Seul Ki Lee, and Ian Sutherland. 2021. 'The impact of latent topic valence of online reviews on purchase intention for the accommodation industry', *Tourism Management Perspectives*, 40: 100903.
- Srivastava, Arpita, and Vivek Kumar. 2021. 'Hotel attributes and overall customer satisfaction: What did COVID-19 change?', *Tourism Management Perspectives*, 40: 100867.
- Streimikiene, Dalia, Biruta Svagzdiene, Edmundas Jasinskis, and Arturas Simanavicius. 2021. 'Sustainable tourism development and competitiveness: The systematic literature review', *Sustainable development*, 29: 259-71.
- Sun, Aixun, Ee-Peng Lim, and Ying Liu. 2009. 'On strategies for imbalanced text classification using SVM: A comparative study', *Decision Support Systems*, 48: 191-201.
- Tian, Ying, and Yao Zhang. 2022. 'Pricing of crowdfunding products with strategic consumers and online reviews', *Electronic Commerce Research and Applications*, 54: 101169.
- Tiwari, Sanjib, Michael Lane, and Khorshed Alam. 2019. 'Do social networking sites build and maintain social capital online in rural communities?', *Journal of rural studies*, 66: 1-10.
- Tolba, Amr, and Zafer Al-Makhadmeh. 2021. 'Predictive data analysis approach for securing medical data in smart grid healthcare systems', *Future Generation Computer Systems*, 117: 87-96.
- Tuna, Hakan, and Melek Başdal. 2021. 'Curriculum evaluation of tourism undergraduate programs in Turkey: A CIPP model-based framework', *Journal of Hospitality, Leisure, Sport & Tourism Education*, 29: 100324.
- Türker, Nuray, and Özkan Süzer. 2022. 'Tourists' food and beverage consumption trends in the context of culinary movements: The case of Safranbolu', *International Journal of Gastronomy and Food Science*, 27: 100463.
- UNEP. 2013. 'World's Largest Travel Site Awards Qualifying Accommodations Across the U.S. with Bronze, Silver, Gold or Platinum Status', Accessed October. <https://www.unep.org/es/node/6002>.
- Urtnasan, Erdenebayar, Jong-Uk Park, and Kyoung-Joung Lee. 2018. 'Multiclass classification of obstructive sleep apnea/hypopnea based on a convolutional neural network from a single-lead electrocardiogram', *Physiological measurement*, 39: 065003.
- van Altena, Allard J, Perry D Moerland, Aeilko H Zwinderman, and Silvia D Olabarriaga. 2016. 'Understanding big data themes from scientific biomedical literature through topic modeling', *Journal of Big Data*, 3: 1-21.
- Wang, Heshan, Yiping Zhang, Jing Liang, and Lili Liu. 2023. 'DAFA-BiLSTM: Deep Autoregression Feature Augmented Bidirectional LSTM network for time series prediction', *Neural Networks*, 157: 240-56.
- Wang, Qiang, Wen Zhang, Jian Li, Feng Mai, and Zhenzhong Ma. 2022. 'Effect of online review sentiment on product sales: The moderating role of review credibility perception', *Computers in Human Behavior*, 133: 107272.

- Wu, Chunrong, Qinglan Peng, Jia Lee, Kenji Leibnitz, and Yunni Xia. 2021. 'Effective hierarchical clustering based on structural similarities in nearest neighbor graphs', *Knowledge-Based Systems*, 228: 107295.
- Wu, Hao, Zhuangzhuang He, Weitao Zhang, Yunsheng Hu, Yunzhi Wu, and Yi Yue. 2021. "Multi-class text classification model based on weighted word vector and bilstm-attention optimization." In *Intelligent Computing Theories and Application: 17th International Conference, ICIC 2021, Shenzhen, China, August 12–15, 2021, Proceedings, Part I* 17, 393-400. Springer.
- Wu, Hao, Zhichao Zhang, Xiaoyong Li, Kai Shang, Yongming Han, Zhiqiang Geng, and Tingrui Pan. 2022. 'A novel pedal musculoskeletal response based on differential spatio-temporal LSTM for human activity recognition', *Knowledge-Based Systems*: 110187.
- Wu, Ling, and Ludovic Noels. 2022. 'Recurrent Neural Networks (RNNs) with dimensionality reduction and break down in computational mechanics; application to multi-scale localization step', *Computer Methods in Applied Mechanics and Engineering*, 390: 114476.
- Wu, Meifen, Ruyin Long, Feiyu Chen, Hong Chen, Yun Bai, Kun Cheng, and Han Huang. 2023. 'Spatio-temporal difference analysis in climate change topics and sentiment orientation: Based on LDA and BiLSTM model', *Resources, Conservation and Recycling*, 188: 106697.
- Xie, Chaorun, Xiaolong Tian, Xiaochun Feng, Xiaoni Zhang, and Junhu Ruana. 2022. 'Preference Characteristics on Consumers' Online Consumption of Fresh Agricultural Products under the Outbreak of COVID-19: An Analysis of Online Review Data Based on LDA Model', *Procedia Computer Science*, 207: 4486-95.
- Xu, Guixian, Yueting Meng, Xiaoyu Qiu, Ziheng Yu, and Xu Wu. 2019. 'Sentiment analysis of comment texts based on BiLSTM', *Ieee Access*, 7: 51522-32.
- Yeşiltaş, Murat, Mert Gürlek, and Gökhan Kenar. 2022. 'Organizational green culture and green employee behavior: Differences between green and non-green hotels', *Journal of Cleaner Production*, 343: 131051.
- Yu, Meng, Mingming Cheng, Lin Yang, and Zhicheng Yu. 2022. 'Hotel guest satisfaction during COVID-19 outbreak: The moderating role of crisis response strategy', *Tourism Management*, 93: 104618.
- Zhang, Chuanfang, Kaixiang Peng, Jie Dong, and Long Miao. 2022. 'A comprehensive operating performance assessment framework based on distributed Siamese gated recurrent unit for hot strip mill process', *Applied Soft Computing*: 109889.
- Zhang, En, Huimin Li, Yuchen Huang, Shuangxi Hong, Le Zhao, and Congmin Ji. 2022. 'Practical multi-party private collaborative k-means clustering', *Neurocomputing*, 467: 256-65.
- Zhang, Ning, Rong Liu, Xiao-Yang Zhang, and Zhi-Liang Pang. 2021. 'The impact of consumer perceived value on repeat purchase intention based on online reviews: by the method of text mining', *Data Science and Management*, 3: 22-32.
- Zhang, Qiangqiang, Shiyuan Wang, Dongyuan Lin, and Shanmou Chen. 2023. 'Robust Affine Projection Tanh Algorithm and Its Performance Analysis', *Signal Processing*, 202: 108749.
- Zhang, Yilin, and Lingling Zhang. 2022. 'Movie Recommendation Algorithm Based on Sentiment Analysis and LDA', *Procedia Computer Science*, 199: 871-78.

- Zhao, Rui, Dongzhe Wang, Ruqiang Yan, Kezhi Mao, Fei Shen, and Jinjiang Wang. 2017. 'Machine health monitoring using local feature-based gated recurrent unit networks', *IEEE Transactions on Industrial Electronics*, 65: 1539-48.
- Zhong, Chenghao, Hao Wang, and Qingchun Yang. 2022. 'Hydrochemical interpretation of groundwater in Yinchuan basin using self-organizing maps and hierarchical clustering', *Chemosphere*: 136787.
- Zhou, Zhi-Hua. 2003. "Three perspectives of data mining." In.: Elsevier.
- Zibarzani, Masoumeh, Rabab Ali Abumalloh, Mehrbakhsh Nilashi, Sarminah Samad, OA Alghamdi, Fatima Khan Nayer, Muhammed Yousoof Ismail, Saidatulakmal Mohd, and Noor Adelyna Mohammed Akib. 2022. 'Customer satisfaction with Restaurants Service Quality during COVID-19 outbreak: A two-stage methodology', *Technology in Society*, 70: 101977.



ÖZGEÇMİŞ

Kişisel Bilgiler	
Adı Soyadı	: Mahmud Alrahhal
Doğum tarihi	:
Doğum Yeri	:
Uyruğu	:
Adres	:
Tel	:
E-mail	:
Eğitim	
Lisans	:
Yüksek lisans	:
Tezden Üretilmiş Yayınlar	