

Google Yorumları Üzerinden Makine Öğrenme Yöntemleri ve Amazon Comprehend ile Duygu Analizi: İç Anadoluda Bir Üniversite Örneği

Sentiment Analysis Based on Google Comments with Machine Learning Methods and Amazon Comprehend: The Case of a University in Central Anatolia

Mustafa Demirbilek ^{1*}, Sevim Özulukale Demirbilek ²

¹Gaziantep İslam Bilim ve Teknoloji Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Endüstri Mühendisliği Bölümü, Türkiye

²Yozgat Bozok Üniversitesi, Akdağmadeni Sağlık Yüksekokulu, İş Sağlığı ve Güvenliği Bölümü, Türkiye

Orcid: M. Demirbilek (0000-0002-1520-2882), S. Özulukale-Demirbilek (0000-0002-5868-5327)

Özet: Fikir madenciliği olarak da bilinen duygu analizi, metin verilerinde ifade edilen duygusal tonu, duyguyu veya öznel bilgiyi belirleme ve sınıflandırma işini içeren doğal dil işleme tekniğidir. Sosyal medya platformları aracılığı ile üretilen çok fazla sayıda yorum ve değerlendirmenin analizinin yapılabilmesi, sağlıktan üretime, eğitimden hizmete kadar pek çok alanda, yoğun rekabet şartlarında faaliyet gösteren kurumların gelecek projeksiyonu için önem arz etmektedir. Bu çalışmada, bir devlet üniversitesi hakkında Google üzerinden yapılan yorumlar için duygu analizi çalışması yapılmıştır. Web kazıma tekniği ile MS Excel'e aktarılan yorumlar, ön işleme sürecine tutulduktan sonra hem Lojistik Regresyon, Gaussian Naive Bayes ve Destek Vektör Makinesi gibi önemli denetimli makine öğrenmesi metodlarıyla hem de CatBoost gibi gelişmiş birleşik öğrenme metoduyla değerlendirilmiştir. Bununla beraber literatürde duygu analizi çalışmalarında sıkça kullanılan derin öğrenme metodu olan LSTM ve Amazon firmasının duygu analizi çalışmaları için kullanıcılara açtığı AWS Comprehend servisi, yorumların analizinde kullanılmıştır. Literatürde sıkça kullanılan performans metrikleriyle sonuçlar karşılaştırılmıştır. Kullanılan yöntemler için tüm performans metrikleri üzerinden başarı oranı % 80'in üstünde çıkmıştır. Bu da duygu analizi için seçilen tüm makine öğrenmesi tekniklerinin ve AWS Comprehend'in başarılı bir şekilde uygulanabileceğini göstermektedir. Duyarlılık hariç tüm metrikler bakımından en iyi sonucu AWS Comprehend uygulaması vermiştir.

Anahtar Kelimeler: Duygu Analizi, Makine Öğrenmesi, Derin Öğrenme, AWS Comprehend, Google Yorumları, Üniversite

Abstract: Sentiment analysis, also known as opinion mining, is a natural language processing technique that involves identifying and categorizing the emotional tone, sentiment, or subjective information expressed in text data. Analyzing a large number of comments and reviews generated through social media platforms enables analysis in various fields, from healthcare to manufacturing, education to services, particularly for organizations operating in highly competitive environments to project future outcomes. In this study, a sentiment analysis is conducted for comments made on a state university through Google. The comments are transferred to MS Excel using web scraping techniques and underwent data preprocessing. They are evaluated using both important supervised machine learning methods like Logistic Regression, Gaussian Naive Bayes, and Support Vector Machine as well as an advanced ensemble learning method as CatBoost. Additionally, LSTM, a deep learning method commonly used in sentiment analysis studies in the literature and Amazon's AWS Comprehend service are used in the analysis comments. The results are compared with frequently used performance metrics in the literature, and for all the methods, the success rate exceeds 80%. This demonstrates that all selected machine learning techniques and AWS Comprehend can be successfully applied. AWS Comprehend provided the best results in terms of all metrics except sensitivity.

Keywords: Sentiment Analysis, Machine Learning, Deep Learning, AWS Comprehend, Google Reviews, University

1. Giriş

İnternetin ortaya çıkması, bilgileri yönetme ve etkileşim kurma biçiminin değişmesine yol açmış; bu da tercih

edilen bilgilerin birden fazla kaynaktan toplanabilmesini ve doğrudan bilgisayar ekranından okunabilmesini mümkün hale gelmiştir. Bilgiyi alan kişiler yalnızca

* İletişim Yazarı / Corresponding author. Eposta/Email : mustafa.demirbilek@gibtu.edu.tr

Geliş Tarihi / Received Date: 31.10.2023 — Kabul Tarihi / Accepted Date: 03.12.2023

doi: 10.32329/uad.1383794

mevcut içeriği almakla kalmamakta, aynı zamanda bu içeriğe aktif olarak açıklama ekleyerek yeni bilgi parçaları üretmektedir. Böylece, bu kişiler okur olmanın yanı sıra yazar da olabilmektedir. Günümüzde kişiler sadece var olan bilgilere yorum yapmak, sayfalara yer işareti koymak ve puan vermekle kalmamakta, aynı zamanda fikirlerini, haberlerini ve bilgilerini de toplumla paylaşabilmekte ve internet üzerinden kişilerin kendilerini ifade edebilecekleri birçok sosyal ağlar bulunmaktadır (Tsytsarau & Palpanas, 2012).

Kişilerin yaşamlarını etkileyen ve iletişimlerinde etkili bir rol oynayan duyguları anlamak, kişilerin ihtiyaçlarını ve tercihlerini daha iyi anlamamızı sağlamaktadır. Bu nedenle, son yıllarda duygu analizi bir araştırma alanı olarak oldukça yaygınlaşmıştır (Korkmaz vd., 2023). Duygu analizi en genel tanımıyla, bir metinden yazarın düşüncelerini anlamlandırmak ve sınıflandırmaya çalışmaktır. Duygu analizi çalışması, ilk kez Nasukawa & Yi (2003)'nin "Sentiment analysis: Capturing Favorability Using Natural Language Processing" adlı çalışmasında kullanılmıştır. Dijitalleşme ile birlikte hızla gelişen sosyal ağların (Google yorumları, Facebook, Twitter, Instagram vs.) yaygınlaşmasıyla birlikte her gün çok sayıda metin üretilmekte ve birçok farklı duygu analizi tekniğiyle yapılan çalışmalar (Doaa Mohey El-Din Mohamed, 2018; Medhat vd., 2014), turizm (Zeng, 2013; Yüksel & Tan, 2018; Aksu & Karaman, 2022) ekonomi (Köksal vd., 2021), yönetim (Uyaroğlu Akdeniz & Cebeci, 2021), psikoloji (Öztürk, 2022) ve bilgisayar bilimleri (Tuna, 2022) ve film/dizi değerlendirmeleri (Gündüz, 2023) gibi farklı disiplinlerde tercih edilmektedir. Metinlerin duygusal içeriklerini belirleme amacı taşıyan bu analiz, doğal dil işleme ve makine öğrenimi gibi yöntemlerden yararlanarak yapılabilmektedir. Bu analiz, genellikle pozitif, negatif veya nötr olarak sınıflandırma yaparak duygusal tepkileri ölçmektedir. Bu analiz yöntemleriyle metinler duygusal birer ifade olarak tanımlandığında, ifadeleri anlamak ve yorumlamak mümkün olabilmektedir (Agarwal vd., 2011; Adalı, 2012).

Günümüzde kişiler, kurumlar hakkındaki düşüncelerini sosyal ağ aracılığıyla sıklıkla değerlendirmektedirler (Tuna vd., 2021; Uyaroğlu Akdeniz & Cebeci, 2021). Öğrencilerin eğitim kurumu olan üniversiteleri tercih ederken de sosyal ağdaki yorumların önemli bir etkiye sahip olduğu söylenebilir. Öğrencilerin gerçek zamanlı ve çevrimiçi olarak paylaştıkları yorumlar geniş ölçekte toplanabilen önemli bir bilgi kaynağıdır. Böylece, öğrenciler sosyal ağdaki yorumlarla üniversiteler hakkında bilgi toplayabilir, tecrübelerini paylaşan diğer öğrencilerle iletişime geçebilir ve üniversite yaşamıyla ilgili bilgilere ulaşabilirler. Bu yorumların duygusal ifadeler içermesi, öğrencilerin üniversitelere karşı memnuniyet düzeylerini ve beklentilerini anlama fırsatı sunabilmektedir. Ayrıca, sosyal ağ aracılığıyla üniversitelerin resmi hesaplarında tanıtım videoları ve etkinlik duyuruları gibi içerikler de yer almaktadır. Bu içerikler, potansiyel öğrencilere üniversitenin kampüs yaşamı, sosyal etkinlikler ve akademik programlar gibi birçok konuda fikir edinme im-

kânı sunabilmektedir. Öğrenciler, sosyal ağ aracılığıyla üniversite hakkında daha kapsamlı bir bilgiye ulaşarak tercihlerini yapma sürecinde daha bilinçli kararlar verebilirler. Makine öğrenme yöntemleriyle duygu analizi, bu yorumlardan elde edilen verileri otomatik olarak analiz ederek bilgilerin ortaya çıkmasında önemli bir araç haline gelmektedir.

Bu çalışmanın amacı, İç Anadolu Bölgesi'nde yer alan bir üniversitenin Google yorumları üzerinden makine öğrenme yöntemleriyle duygu analizi yapılarak, üniversitenin öğrenci memnuniyetini değerlendirmektir. Bu çalışmada, Google yorumları öncelikle pozitif ve negatif olarak sınıflandırılarak değerlendirilmiştir. Duygu analizi sıklıkla Twitter platformu üzerinde paylaşılan bilgiler üzerinden yapılırsa da (Giachanou & Crestani, 2016), gerek Twitter üzerinden çalışan kötü niyetli yazılımların (bot) sahte ve manipülatif yorumlar üretmesi gerekse de platformun el değiştirmesinden sonra veri almaya (kazıma) sınır ve yasaklar getirmesi sebebiyle diğer sosyal medya platformları üzerinden çalışmalar artmaktadır.

Google, inceleme platformu endüstrisinde önemli bir oyuncu haline gelmiştir; ABD'deki veriler, Google'ın "Yerel Rehber" inin Facebook, Yelp ve TripAdvisor'ı geride bırakarak en hızlı büyüyen inceleme platformu olduğunu göstermektedir (Bright Local, 2018). Google incelemeleri, Google'ın en yaygın kullanılan ve güvenilen arama motoru olması nedeniyle büyük bir öneme sahiptir (Ross, 2022). Müşterilerin % 60'tan fazlası iş yapmadan önce başkalarının deneyimlerini anlamak amacıyla bir şirketle etkileşime geçmeden önce Google incelemelerini kontrol ettiğini göstermektedir (Online Reviews Statistics and Trends, 2022). Büyüyen önemine rağmen, Google yorumları üzerine yapılan duygu analizi çalışmaları, özellikle havaalanları (Lee & Yu, 2018), restoranlar (Mathayomchan & Taecharungroj, 2020) ve büyük şehirler (Taecharungroj, 2019) üzerine yapılan çalışmalarla sınırlı kalmıştır.

Bu çalışmanın literatüre en önemli katkısı metodolojiktir. Lojistik Regresyon, Gaussian Naive Bayes, Destek Vektör Makinesi, CatBoost, LSTM gibi çok farklı deneysel makine ve derin öğrenme metodları yorumların duygu analizi sırasında kullanılmıştır. Bununla da yetinilmeyip son yıllarda Amazon firmasının duygu analizleri ve diğer makine öğrenme teknikleri için hizmete sunduğu AWS Comprehend servisinden yararlanılarak, her bir yorumun duygu durumu geliştirilen bir yazılımla otomatik olarak değerlendirilmiş; bu çalışmada kullanılan performans metriklerine göre diğer yöntemlerle karşılaştırılmıştır. Yapılan literatür araştırmalarına göre Google yorumları üzerinden bu kadar kapsamlı bir duygu analizi çalışmasına rastlanmamıştır.

İkinci bölümde, verilen toplanması, ön işleme ve duygu analizi çalışması sırasında kullanılan yöntem ve performans metrikleriyle ilgili detaylı bilgiler verilmiştir. Sonraki bölümde çıkan sonuçlar irdelenmiş ve yöntemlerin karşılaştırması yapılmıştır. Son bölümde ise sonuçlarla ilgili tartışma yapılmış; bu çalışmanın sınırlılıkları de-

ğlendirilmiş ve ileride yapılacak çalışmalarla ilgili araştırmacılara fikir verilmiştir.

2. Materyal ve Yöntem

Yapılacak duygu analizi çalışması; web kazıma ile verilerin toplanması, ön işleme, makine öğrenme modellerinin uygulanması ve sonuçların analizi adımlarından oluşmaktadır (Şekil 1).

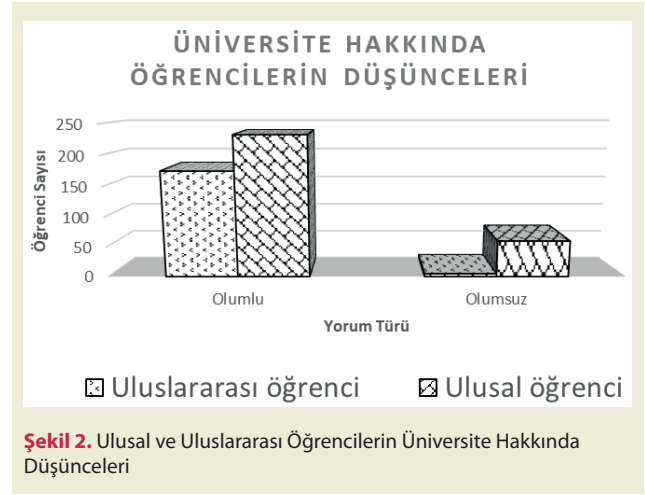
2.1. Verilerin Toplanması ve Ön İşleme

Daha önce belirtildiği gibi bu çalışma ülkemizin İç Anadolu Bölgesinde yer alan bir üniversitemiz için yapılan Google yorumları üzerinden bir duygu analizi çalışmasıdır. Google üzerinden yapılan yorumlar, daha sonraki aşamalarda girdi olarak kullanılması için "Instant Data Scraper" isimli Google uzantısı olarak çalışan bir program sayesinde MS Excel çalışma sayfasına aktarılmıştır.

Instant Data Scraper web sitelerinden veri kazımak ve bu verileri excel veya csv uzantılı dosyalara aktarmak için kullanılmaktadır. İçerdiği yapay zekâ sayesinde veri kazıma yapılacak web sitelerindeki gereksiz verilerden kurtularak, kullanıcıya en ilgili verileri aktarmasına olanak sağlar (Instant Data Scraper, 2023).

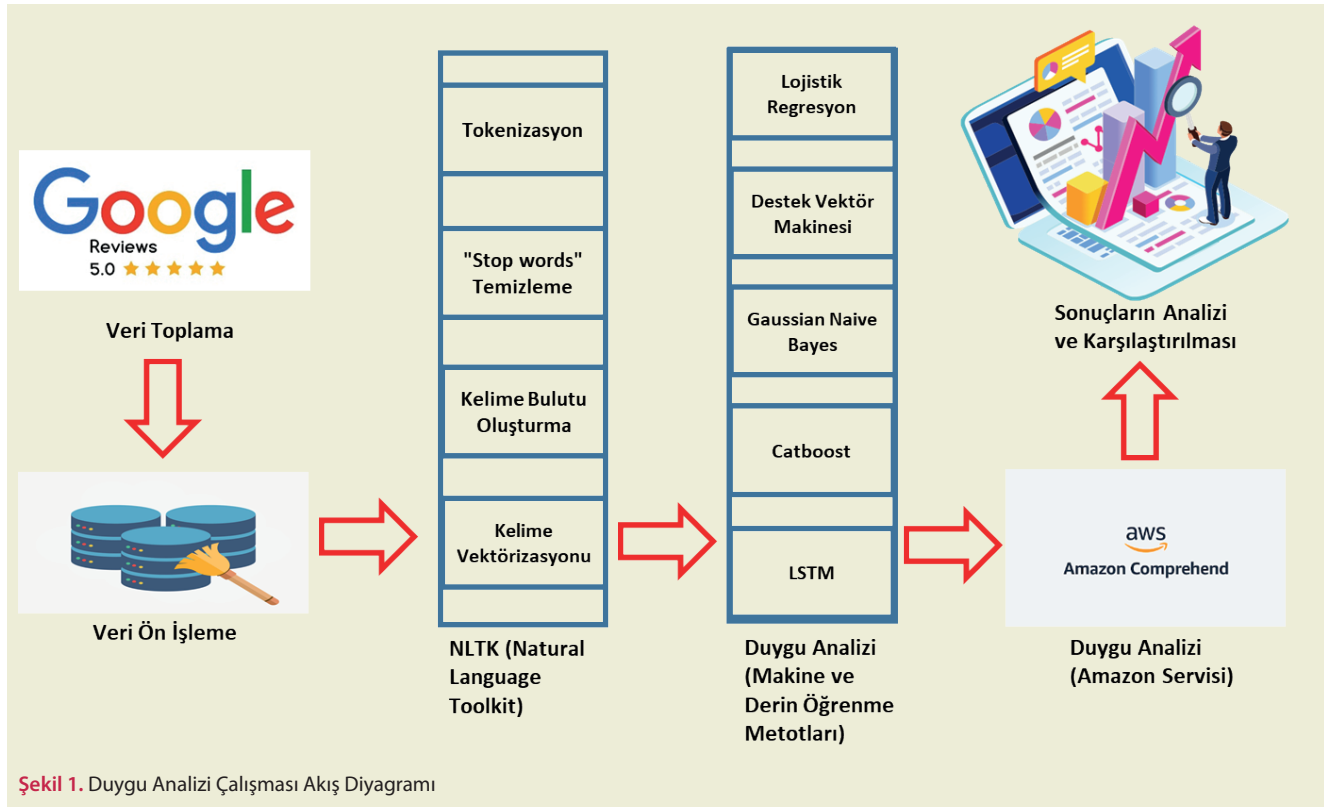
Üniversite ile ilgili yapılan toplamda 643 yorum dosyaya aktarılmış, bunlardan boş olan, anlamsız kelimeler içeren ve sadece emojilerden oluşanlar ayıklanmıştır. Ayıklama işleminden sonra kalan yorum sayısı 293 olmuştur. Kalan yorumlardan İngilizce olanları Türkçeye çevrilmiştir. Daha sonra yorumlar olumlu ve olumsuz olmak üzere manuel olarak sınıflandırılmıştır. Şekil 2'de ulusal ve uluslararası öğrencilerin üniversite hakkında düşün-

celeri yorum türüne göre dağılımı gösterilmiştir.



Şekil 2. Ulusal ve Uluslararası Öğrencilerin Üniversite Hakkında Düşünceleri

Öncelikle metindeki tüm harfler küçük harflere dönüştürülmüştür. Etkisiz kelimeler olarak adlandırılan "stopwords" Türkçe için sıklıkla kullanılan açık kaynak kodlu bir siteden çekilerek "ne", "nasıl", "ve", "ama", "fakat", vb. kelimelerin elenmesi sağlanmıştır. Kelimelerin işlenmesi bir Python kütüphanesi olan NLTK (Natural Language Toolkit) ile yapılmıştır. NLTK, Python programlama dilinde doğal dil işleme (NLP-Natural Language Processing) projeleri için kullanılan popüler bir kütüphanedir. NLTK, dil analizi, metin sınıflandırma, dil modellemesi, çeviri, kelime kök çıkarma, etiketleme, ayrıştırma gibi birçok NLP görevini gerçekleştirmek için kullanılabilir. NLTK, metin verilerini kolayca işlemek için bir dizi metod, sınıf ve kaynak içerir. Ayrıca, metinler üzerinde çeşitli ön işleme işlemlerini (örneğin, tokenizasyon, stop word

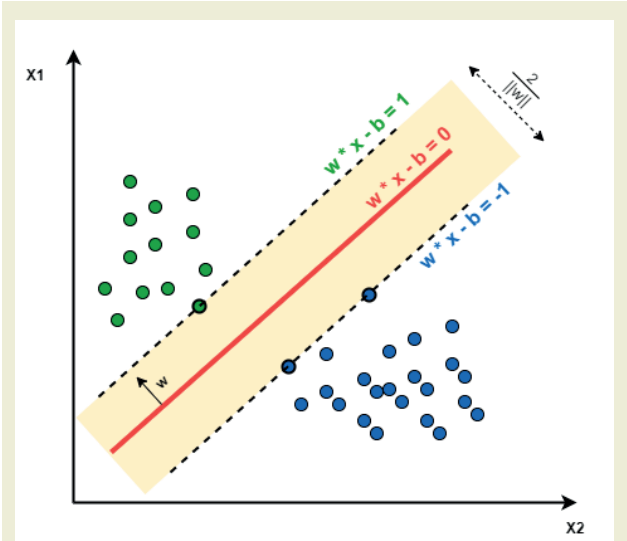


Şekil 1. Duygu Analizi Çalışması Akış Diyagramı

sınırı, veri noktalarının en iyi şekilde sınıflara ayrılmasını sağlar. Verilerin sadece iki boyutlu uzayda temsil edildiği durumlarda doğrusal, üç boyut düşünüldüğünde düzlemsel ve daha yüksek boyutlarda hiper-düzlemsel ayrımlar yapabilmektedir (Kaynar vd., 2016).

En genel formda, DVM iki boyutlu uzayda veri kümesini iki ayrı sınıfa bölecek maksimum ayırım doğrusunu bulmaya çalışır (Şekil 4'teki kırmızı hat). Bu doğruya bağlı kalarak sınıflar arasında bir karar sınırı (hiper-düzlem) oluşturmayı hedefler. Bu sınır, veri noktalarını sınıflara ayıran en iyi ayrımı sağlamalıdır. İki sınıf arasındaki karar sınırı genellikle en geniş marjı olan bir sınıflandırma hiper-düzlemi olarak seçilir ($w \cdot x - b = +/-1$). Bu hiper-düzleme en yakın olan noktalar "Destek Vektörü" olarak adlandırılır (Géron, 2022).

Destek Vektör Makinesi ve İki Boyutlu Doğrusal Hiper-Düzlem Seçimi Şekil 4'te ve bu çalışma kapsamında DVM için kullanılan parametreler Tablo 1'de gösterilmiştir.



Şekil 4. Destek Vektör Makinesi ve İki Boyutlu Doğrusal Hiper-Düzlem Seçimi (Multiclass Classification Using Support Vector Machines, 2023).

Tablo 1. DVM için Kullanılan Parametreler

C	1.0	Probability	False
Kernel	Radial basis function	Tol	10^{-3}
Degree	3	Cache_size	200
Gamma	Scale	Class_weight	None
Coef0	0.0	verbose	False
Shrinking	True	max_iter	-1

CatBoost (Categorical Boosting)

CatBoost, 2017 yılında Yandex mühendisleri tarafından geliştirilen, kategorik özelliklere sahip büyük ve karmaşık veri setlerinde yüksek performans sağlamak için sıralı artırma, rastgele permutasyonlar ve gradyan tabanlı

optimizasyon gibi teknikleri bir araya getiren açık kaynaklı bir kütüphanedir. Bu kütüphane, sınıflandırma, regresyon ve sıralama gibi çeşitli görevler için kullanılabilir (Dorogush vd., 2018).

CatBoost, Gradient Boost tekniğinin doğruluğunu ve verimliliğini artırmak için çeşitli teknikler kullanır. Bu teknikler arasında özellik mühendisliği (feature engineering), karar ağacı optimizasyonu ve sıralı artırma adlı yenilikçi bir algoritma bulunur. Algoritmanın her adımında, CatBoost mevcut tahminlere göre kayıp fonksiyonunun negatif gradyanını hesaplar. Bu gradyanı kullanarak tahminleri güncelleyerek mevcut tahminlere ölçeklenmiş bir gradyan ekler. Bu ölçek faktörü, kayıp fonksiyonunu en aza indiren bir çizgi arama algoritmasıyla seçilir. Karar ağaçlarını oluştururken CatBoost, ağaçları kayıp fonksiyonunun negatif gradyanına uyumlu hale getirmek için gradyan tabanlı bir optimizasyon tekniği kullanır. Bu yaklaşım, ağaçların kayıp fonksiyonu üzerinde en büyük etkisi olan özellik alanlarında odaklanmasını sağlar ve daha doğru tahminler elde edilir. Diğer yandan, öğrenme amaç fonksiyonunu optimize etmek için özellikleri belirli bir sıra ile permutasyon yapan sıralı artırma adlı yeni bir algoritma sunar. Bu yaklaşım özellikle büyük özellik sayısına sahip veri setleri için daha hızlı yakınsama ve daha iyi model doğruluğu sağlayabilir (CatBoost, 2023; What Is CatBoost, 2023).

CatBoost tekniğini diğer makine öğrenme tekniklerine göre öne çıkaran noktalar aşağıdaki gibi sıralanabilir (CatBoost, 2023; What Is CatBoost, 2023):

- Eğitim ve Test aşamalarından önce "One Hot Encoding" ve "Label Encoder" uygulamaları gerektirmeden ham veri üzerinde çalışmaya olanak sağlar,
- Eksik verilerle modelleme yapmaya izin verir,
- Çapraz doğrulama özelliği tanımlı olarak gelir,
- L1 ve L2 düzenleme tekniklerini kullanarak modelde oluşacak aşırı uyum (overfitting) durumunu ortadan kaldırır,
- Daha hızlı eğitim olanağı ve daha doğru bir model sağlar.

LSTM (Long Short Term Memory) Ağı Metodu

LSTM, zaman serileri, konuşma ve metin gibi sıralı verileri işlemek için özelleştirilmiş bir Yinelemeli Sinir Ağı (RNN-Recurrent Neural Network) türüdür. LSTM ağları, sıralı verilerde uzun süreli desenleri yakalama konusunda başarılıdır ve bu nedenle dil çevirisi, konuşma tanıma ve zaman serileri tahmini gibi görevler için son derece uygundur (Srinivas, 2021).

Tek bir gizli durumun zaman içinde geçtiği geleneksel RNN'nin aksine, LSTM uzun vadeli ilişkileri anlama konusunda zorlanabilen sorunu bir bellek hücresi ekleyerek çözer. Bu hücre, bilgiyi uzun süre boyunca saklamak için

kullanılan bir depo görevi görür. Bellek hücrelerini yönlendiren üç kapı bulunur: giriş kapısı, unutma kapısı ve çıkış kapısı. Bu kapılar, bellek hücrelerine hangi bilginin dahil edileceğini, sonuç olarak çıkarılacağını ve çıkartılacağını belirler. Tablo 2, bu çalışmada için kullanılan LSTM Ağı Yöntemi için modellemede kullanılan parametreleri göstermektedir.

Parametreler	Değerler
Parti Boyutu (Batch Size)	100
Tekrar (Epoch)	250
Gizli Katman Aktivasyon Fonksiyonu	ReLU
Çıktı Katmanı Aktivasyon Fonksiyonu	Sigmoid
Optimizör /Öğrenme Oranı	Adam/0.001
Kayıp Fonksiyonu	Binary cross-entropy
Her kelimeye Karşılık Gelen Vektör Uzunluğu	50

Bu çalışmada, LSTM modeli TensorFlow v2.13 Keras uygulaması üzerinden çalıştırılmıştır.

2.3. Amazon Comprehend

Amazon Comprehend, AWS'den (Amazon Web Services) sunulan bir hizmettir ve NLP için kullanılmaktadır. Bu hizmet, yapılandırılmamış metin verilerinden değerli iç görüler ve bilgiler çıkarmak için tasarlanmıştır. Amazon Comprehend ile geliştiriciler ve veri bilimcileri, belgelerin, sosyal medya gönderilerinin, müşteri geri bildirimlerinin, e-postaların ve diğer metin tabanlı kaynakların içeriğini kolayca analiz edip anlayabilirler. Ayrıca sentiment modellerinin daha önce eğitilmiş olması, her bir yorumun bireysel olarak değerlendirilmesine olanak sağlamaktadır. Amazon Comprehend hizmetini kullanılan Python diline uyarlayabilmek ve tüm yorumlar için duygu analizi çalışmasını gerçekleştirmek için "Boto3" kütüphanesi kullanılmıştır. Boto3, AWS için Python Yazılım Geliştirme Kiti olarak geliştirilmiştir ve Python geliştiricilerine Amazon S3 ve Amazon EC2 gibi hizmetleri kullanan yazılım geliştirme imkânı sunar. Ancak servisin Türkçe desteği ile ilgili bir bilgi bulunmadığından, Türkçe yorumlar Boto3 aracılığıyla İngilizce'ye çevriler değerlendirilmiştir. Yapılan yorumlar servis tarafından 0-1 olasılık dağılımları üzerinden pozitif, negatif ve nötr skorları olarak döndürülmektedir (AWS, 2023; Qaisi & Aljarah, 2016). Şekil 5'te AWS Comprehend servisi

ile bir yorumun duygu analizi sürecini gösterilmektedir.

Her bir yorum için skorlar alındıktan sonra bunlar negatif "0" ve pozitif "1" olarak değerlendirilecektir. Daha sonra bu sonuçlar gerçek değerlerle karşılaştırılıp ilgili metrikler hesaplanacaktır.

2.4. Model Başarım Ölçütleri

Makine öğrenme modellerinin performansını karşılaştırmak için bazı ölçütler kullanılmaktadır. Bunları açıklamadan önce bazı temel kavramları açıklamak faydalı olacaktır.

TP (Doğru Pozitif): Eldeki veriler doğruyken (pozitif), bunlar için model tarafından yapılan sınıflandırma tahminlerinin doğru olduğu durumların sayısını verir.

TN (Doğru Negatif): Eldeki veriler yanlışken (negatif), bunlar için model tarafından yapılan sınıflandırma tahminlerinin yanlış olduğu durumların sayısını gösterir.

FP (Yanlış Pozitif): Eldeki veriler yanlışken, bunlar için model tarafından yapılan sınıflandırma tahminlerinin doğru olduğu durumların sayısını gösterir.

FN (Yanlış Negatif): Eldeki veriler doğruyken, bunlar için model tarafından yapılan sınıflandırma tahminlerinin yanlış olduğu durumların sayısını verir.

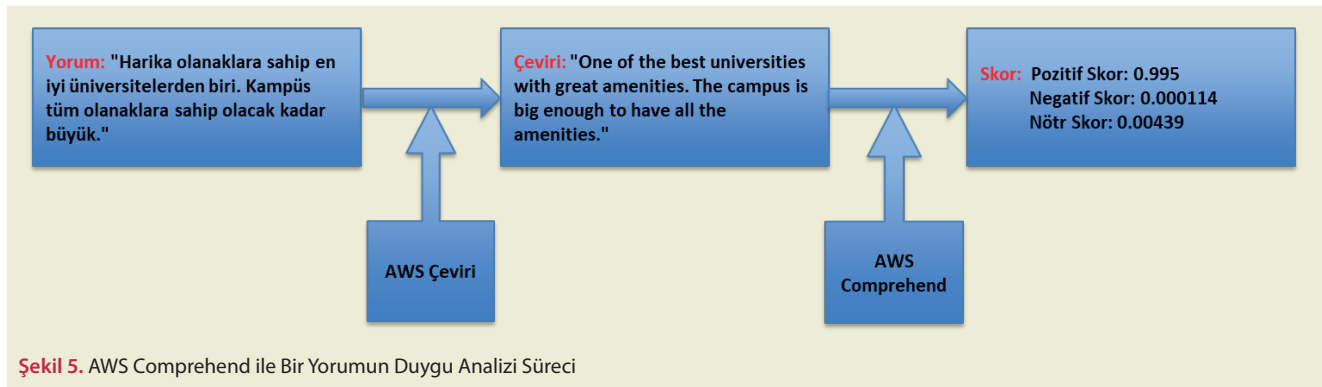
Sadece doğru ve yanlış olarak ikili bir sınıflandırmadan bahsettiğimiz durumlarda yukarıdaki metrikler kullanılarak karışıklık matrisi (confusion matrix) Şekil 6'daki gibi elde edilir.

Bu metrikleri kullanarak hesaplayacağımız model performans ölçütleri ise Doğruluk (Accuracy), Duyarlılık (Recall), Kesinlik (Precision) ve F Ölçüsü olacaktır (Géron, 2022).

Doğruluk (Acc): Doğruluk değeri modelin doğru olarak tahmin ettiği TP ve TN sayılarının tüm örnekleme oranı olarak hesaplanmaktadır. Çoğunlukla modelin başarısı olarak nitelendirilmektedir. Denklem 1'de gösterilmiştir.

$$Acc = \frac{TP + TN}{Popülasyon} \quad (1)$$

Duyarlılık (R): Modelin pozitif verileri sınıflandırmadaki başarısını göstermektedir. TP sayısının toplam pozitif



Şekil 5. AWS Comprehend ile Bir Yorumun Duygu Analizi Süreci

Populasyon=P+N		Tahmini Durum	
		Pozitif	Negatif
Gerçek Durum	Pozitif (P)	TP	FN
	Negatif (N)	FP	TN

Şekil 6. Karışıklık Matrisi

sayısına (P) bölünmesiyle hesaplanır. Denklem 2’de gösterilmiştir.

$$R = \frac{TP}{TP + FN} \quad (2)$$

Kesinlik (P): Modelin pozitif olarak tahmin ettiği verilerin (TP + FP) gerçekten pozitif olma olasılığını gösteren ölçektir. Denklem 3’te gösterilmiştir.

$$P = \frac{TP}{TP + FP} \quad (3)$$

F Ölçütü (F Score): Kesinlik ve Duyarlılık değerlerini birleştirerek, ikisi arasındaki dengeyi temsil eden tek bir değer sunar. Denklem 4’te gösterilmiştir.

$$F\ Score = \frac{2 * R * P}{R + P} \quad (4)$$

AUC ve ROC: ROC (İşlem Karakteristik) eğrisi ve AUC (Eğri Altında Kalan Alan) sınıflandırma problemlerini değerlendirmede önemli metriklerdir. ROC eğrisi, bir sınıflandırma modelinin duyarlılık ile özgüllük arasındaki ilişkiyi gösteren bir grafikdir. Duyarlılık, gerçek pozitif oranıdır ve doğru şekilde tahmin edilen pozitif örneklerin toplam pozitif örnek sayısına oranını temsil eder. Denklem 5’te gösterilen özgüllük ise hatalı pozitif oranıdır ve doğru şekilde tahmin edilen negatif örneklerin toplam negatif örnek sayısına oranını temsil eder. AUC, ROC eğrisinin altında kalan alanı ifade eder ve sınıflandırma modelinin genel performansını ölçer. AUC değeri 0 ile 1 arasında bir değer alır, 1’e ne kadar yakınsa modelin performansı o kadar iyidir. Bir modelin AUC değeri 0.5’e yakınsa, rastgele tahmin etmekle aynı performansa sahiptir. AUC değeri 1’e yaklaştıkça, modelin sınıflandırma yeteneği artar.

$$\text{Özgüllük} = \frac{FP}{FP + TN} \quad (5)$$

3. Deneysel Sonuçlar

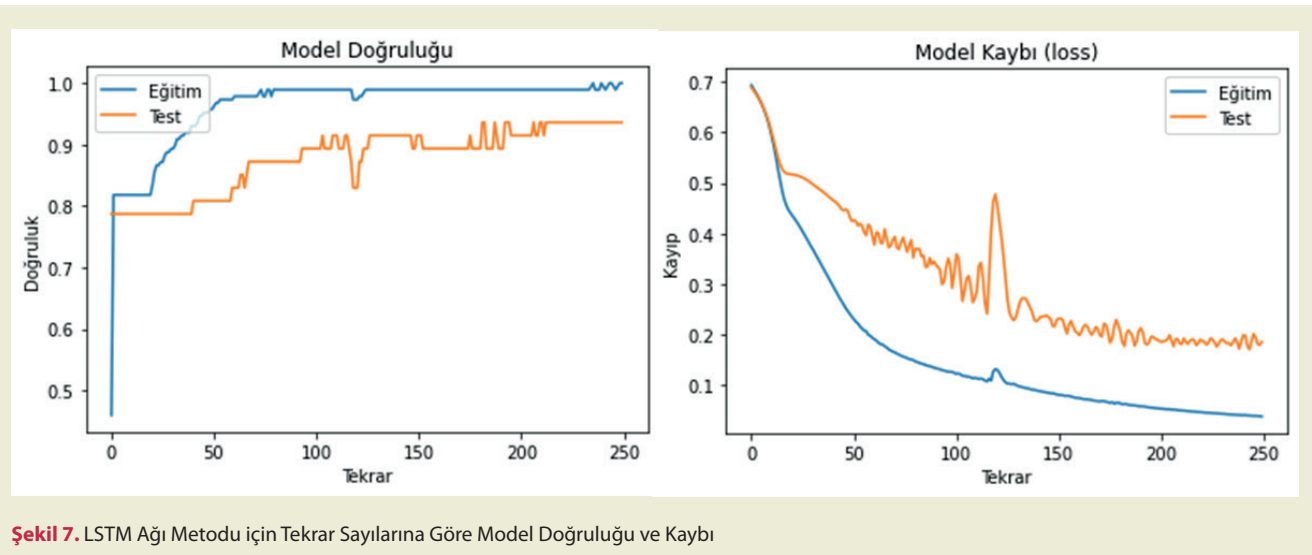
Modelin eğitimi ve testi sırasında farklı veri kümelerinin kullanılabilmesi için çapraz doğrulama (cross validation) metodu kullanılmıştır. Çapraz doğrulama, makine öğrenmesi modellerinin performansını değerlendirmek için kullanılan bir tekniktir. Bu yöntem, mevcut veri setini birden fazla alt kümeye böler. Her bir bölme sırasında, bir alt küme doğrulama seti olarak kullanılırken, geri kalan alt kümeler eğitim seti olarak kullanılır. Model, her bir bölme için eğitim seti üzerinde eğitilir ve doğru-

lama seti üzerinde test edilir. Bu işlem bölme sayısı kadar tekrarlanır ve her bir doğrulama sonucu kaydedilir. Sonunda, elde edilen sonuçlar için istatistiksel metotlar ile (ortalama veya diğer metriklerle) modelin genel performansı değerlendirilir. Çapraz doğrulama, bir modelin genelleştirme yeteneğini değerlendirmek için kullanılır. Modelin, yeni ve görünmeyen verilere nasıl tepki verebileceğini tahmin etmek için kullanılan veriye dayalı bir ölçü sağlar. Ayrıca, aşırı uyum problemlerini tespit etmek ve modelin hiper-parametrelerini ayarlamak için de faydalıdır (Cross Validation in Machine Learning, 2023). Her yöntemin ortalama performans sonuçlarını alabilmek adına rastgele seçilen 10 eğitim/test seti için çapraz doğrulama yapılmıştır. Her veri setinin %80’i eğitim için %20’si test için kullanılmıştır. Hem çapraz doğrulama hem de makine öğrenmesi tekniklerinin kodlanması için Python programlama dili için geliştirilmiş ve birçok makine öğrenme tekniğini barındıran “scikit-learn” kütüphanesindeki ilgili modüller kullanılmıştır. Her performans skoru için ortalama sonuçlar ve %95’lik güven aralığı için iki standart sapma değerleri (+/-) Tablo 3’te verilmiştir. LSTM ve AWS için çapraz doğrulama yapılmadığından her metrik için tek bir skor gösterilmektedir.

Tablo 3. Tanımlanan Metriklere Göre Makine Öğrenme Tekniklerinin Performansları

	Acc	R	P	F-Ölçütü
GNB	0,89 ±0,07	0,97 ±0,08	0,91 ±0,07	0,94 ±0,04
LR	0,82 ±0,07	1 ±0	0,82 ±0,07	0,90 ±0,04
DVM	0,82 ±0,14	0,93 ±0,13	0,87 ±0,08	0,90 ±0,09
CatBoost	0,81 ±0,03	1 ±0	0,81 ±0,03	0,90 ±0,02
LSTM	0,94	0,83	0,91	0,87
AWS	0,94	0,96	0,96	0,96

Tablo 3’te görüleceği gibi yöntemler farklı metriklere göre farklı sonuçlar vermiştir. LSTM ve AWS metodu doğruluk (Acc) metriğine göre diğer yöntemlerden daha iyi sonuç vermişken; AWS metodu duyarlılık (R) metriği hariç tüm metotlardan daha iyi değerler vermiştir. Tabii diğer yöntemler için standart sapma miktarlarına da dikkat etmek gerekir. Güven aralıkları oluşturulduğunda çakışmalar olabilecektir. Standart sapmanın yüksek çıkmasının sebebi çapraz doğrulama sırasında sadece 10 farklı veri seti için performans metriklerinin hesaplanmasıdır. Veri seti sayısını artırdıkça sapmalar da doğal olarak düşecektir. Diğer yöntemlerin performanslarına baktığımızda birbirlerine yakın oldukları görülmektedir. Bunun yanında DVM için hesaplanan sonuçlarda sapmaların yüksek olduğunu gözlemlerken, sonuçlar içinde en küçük sapma değerlerine sahip olan yöntemin CatBoost olduğu tespit edilmiştir. Bunun sebebi CatBoost yönteminin daha önce belirtilen düzenleme tekniklerini kullanması olabilir. Sınıflandırma problemlerinde sadece Acc üzerinden değerlendirme yapmak doğru sonuçlar vermemektedir. Özellikle P ve R metriklerinin yüksek çıkması nadir vakaların doğru tespit edilmesi bakımından önem arz etmektedir. Bu çalışma için önerilen mo-



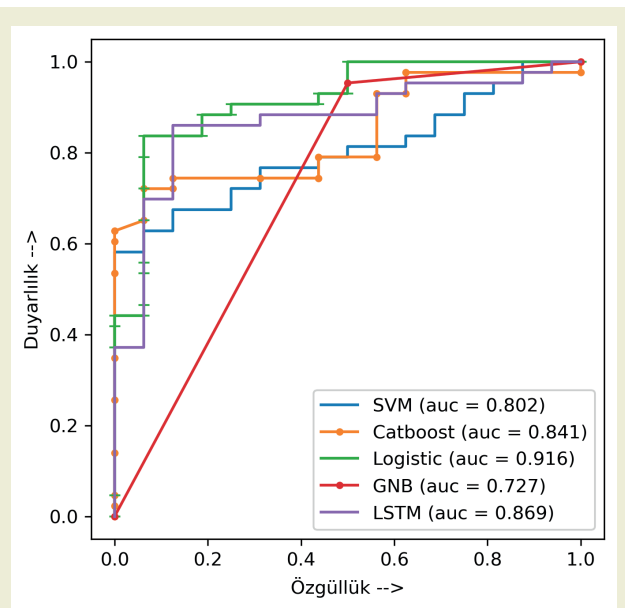
Şekil 7. LSTM Ağı Metodu için Tekrar Sayılarına Göre Model Doğruluğu ve Kaybı

dellerin hem R hem de P için de yüksek performans sonuçları verdiği gözlenmiştir.

LSTM özelinde tekrar sayılarına göre hem eğitim hem de test verileri için model doğruluğu ve kaybını gözlenmesi önem arz etmektedir. Şekil 7’de 250 tekrara göre hem eğitim hem de test verilerindeki doğruluk ve kayıplar gösterilmiştir. Görüleceği gibi hem eğitim verileri hem de test verileri için doğruluk tekrar sayısına bağlı olarak yükselmekte; yaklaşık 200. tekrardan sonra sabit değerlere yakınsamaktadır. Kayıp, aynı zamanda amaç veya maliyet fonksiyonu olarak da bilinir ve tahmin değerleri ile gerçek değerler arasındaki benzerliğin bir ölçüsüdür. Eğitim sırasında LSTM modelinin amacı bu kayıp değerini en aza indirmektir. Eğitim süreci boyunca, içsel parametrelerini (ağırlıklar ve sapmalar) bir optimizasyon algoritması kullanarak (gradient descent) kayıp değerini en aza indirmeyi amaçlar. Model, parametreleri en iyi yapılandırmayı bulmak için parametreleri tekrarlayarak günceller. Şekil 7, tekrar sayısına göre model kaybını göstermektedir. Beklendiği gibi, kayıp değeri

dönem sayısına göre kademeli olarak azalmaktadır. Yapılan denemelerde 250 tekrardan sonra test verileri için kayıp oranı arttığından, tekrar sayısının 250 olarak belirlenmesinin uygun olduğu düşünülmüştür.

Şekil 8’de ROC ve AUC değerleri gösterilmiştir. GNB yöntemi, tüm metrikleri bakımından AWS hariç diğer yöntemlere göre ortalamanın üstü sonuçlar vermesine rağmen AUC değerine göre değerlendirildiğinde diğer yöntemlerden da kötü sonuçlar vermektedir. ROC’un diğer metriklerle göre farkı pozitif ve negatif sonuçları beraber düşünerek değerlendirmeye olanak sağlamasıdır. Bir tarafta Duyarlılık (R) değerleri Şekil 8’deki gibi dikey eksen oluştururken, yanlış tahmin edilen negatif sonuçların tüm negatif sonuçlara oranını gösteren özgüllük değeri yatay eksenini oluşturmaktadır. Sonuç olarak GNB yönteminin negatif sınıflandırmada başarısız olduğu sonucuna ulaşabiliriz. AUC değeri olarak en iyi sonucu veren yöntemin LR olduğu, LSTM yönteminin ise LR’yi takip ettiği görülebilir.



Şekil 8. ROC ve AUC Değerleri

4. Sonuç ve Tartışma

Fikir ve duygular karar verme sürecini etkileyen önemli faktörlerdir. Güncel bir çalışma alanı olan duygu analizi ile birçok kurum ve araştırmacı, internetteki çeşitli platformlar (Facebook, Google, Twitter) aracılığıyla toplanan çok büyük mertebedeki işlenmemiş verinin analizi gerçekleştirilerek toplulukların yönelimini anlamaya çalışmaktadırlar.

Bu çalışma, İç Anadolu’daki bir şehirde yer alan bir devlet üniversitesi için yapılan Google yorumları duygu analizi ile irdelenmiştir. Veri kazıma yöntemi ile internetten çekilen yorumlar önce ön işlemeden geçirilip sınıflandırılmış; sonrasında Gaussian Naive Bayes (GNB), Lojistik Regresyon (LB), Destek Vektör Makinesi (DVM) ve CatBoost makine öğrenme teknikleri kullanılarak duygu analizi çalışması gerçekleştirilmiştir. Bunun yanında duygu analizleri için literatürde sıklıkla kullanılan derin öğrenme metodu LSTM ve bulut üzerinden duygu analizi çalışmalarını yapmayı olanaklı kılan AWS Comprehend servisi, bu ça-

İşmada duygu analizi için kullanılmıştır. Bu yöntemler literatürde en çok kullanılan performans metrikleri ile karşılaştırılmıştır.

Tüm yöntemlere genel olarak baktığımızda, tüm performans metrikleri için başarı oranı %80'in üstündedir. Bu da duygu analizi için seçilen tüm makine öğrenmesi tekniklerinin ve AWS Comprehend'in başarılı bir şekilde uygulanabileceğini göstermektedir. Duyarlılık hariç tüm metrikler bakımından en iyi sonucu AWS Comprehend uygulaması vermiştir. Çok uzun zamandır Amazon firmasının müşteri yorum ve değerlendirmelerine göre duygu analizi modelleri üzerinde çalıştığı düşünülürse sonuçlar şaşırtıcı değildir. Ayrıca bu çalışmada incelenen yorum sayısının makine öğrenmesi açısından biraz düşük kaldığı düşünülürse, ön eğitim gerekmeyen AWS Comprehend gibi platformların kullanılması daha uygun olabilir. Ancak bu servis tamamen ücretsiz değildir. Üye olduktan sonra aylık ancak belirli bir metin sayısına kadar kullanıcılar bu servisten ücretsiz yararlanabilmektedirler. Diğer

tarafından LSTM Ağı yöntemi de AWS hariç diğer yöntemlere göre ortalamadan üstünde sonuçlar vermiştir. LSTM ve CatBoost yöntemlerinin dezavantajı ise eğitimleri diğer yöntemlere göre daha fazla zaman almıştır. Son olarak Lojistik Regresyon metodu, Gaussian Naive Bayes'e göre doğruluk, kesinlik ve F-ölçütü gibi performans metriklerine göre daha düşük performans gösterse de AUC değerine göre Gaussian Naive Bayes en kötü, Lojistik Regresyon en iyi skoru veren yöntem olmuştur.

Bu çalışmanın en önemli sınırlılığı, duygu analizi için değerlendirilen yorum sayılarının makine öğrenimi için görece az olmasıdır. Bu konu da daha fazla üniversite üzerine yapılan yorumlar değerlendirmeye alınarak daha stabil ve performansı yüksek modeller oluşturulabilir. Ayrıca bu ortak veri tabanı ve üzerinden geliştirilen modeller, AWS Comprehend servisine benzer şekilde Türkiye'deki tüm üniversiteler için yapılan bireysel yorumların hızlı bir biçimde duygu analiziyle değerlendirilmesi için kullanılabilir.

Kaynakça

- Adalı, E. (2012). Doğal Dil İşleme. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi, 5(2).
- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment Analysis of Twitter Data. LSM '11 Proceedings of the Workshop on Languages in Social Media, 30-38.
- Aksu, M. Ç., & Karaman, E. (2022). Turistik Mekanlara Yönelik Sosyal Medya Paylaşımlarının Yapay Zekâ Yöntemleriyle Değerlendirilmesi: Artvin İli Örneği. Journal of Tourism and Gastronomy Studies, 10(1), 505-524. <http://doi.org/10.21325/jotags.2022.1001>.
- Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python: O'Reilly Media, Inc. ISBN: 9780596516499
- AWS, (2023), <https://aws.amazon.com/tr/comprehend/>. Erişim Tarihi: 20/05/2023.
- BrightLocal, (2018). Comparison of Local Review Sites: Which Platform Is Growing the Fastest? <https://www.brightlocal.com/research/comparison-of-local-review-sites/>. Erişim Tarihi: 20/05/2023.
- Büyükeke, A., Sökmen, A., & Gencer, C. (2020). Metin Madenciliği ve Duygu Analizi Yöntemleri ile Sosyal Medya Verilerinden Rekabetçi Avantaj Elde Etme: Turizm Sektöründe Bir Araştırma. Journal of Tourism and Gastronomy Studies, 8(1), 322-335. <http://doi.org/10.21325/jotags.2020.550>
- CatBoost. (2023). <https://catboost.ai/>. Erişim Tarihi: 20/05/2023.
- Cross Validation in Machine Learning, (2023). <https://www.geeksforgeeks.org/cross-validation-machine-learning/>. Erişim Tarihi: 20/05/2023.
- Doaa Mohey El-Din Mohamed, H. (2018). A survey on sentiment analysis challenges. Journal of King Saud University - Engineering Sciences, 30(4), 330-338. <https://doi.org/10.1016/j.jksues.2016.04.002>
- Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: Gradient Boosting with Categorical Features Support. arXiv:1810.11363v1, 1-7.
- Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd Edition ed.): O'Reilly Media, Inc. ISBN: 9781098125974.
- Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. ACM Computing Surveys (CSUR), 49(2), 1-41. <https://doi.org/10.1145/2938640>.
- Gündüz, H. (2023). Derin Transformatörlerden Çift Yönlü Kodlayıcı Temsilleri ve Destek Vektör Makineleri ile Türkçe Film Yorumları Üzerine Duygu Analizi. KSÜ Mühendislik Bilimleri Dergisi, 26(2), 542-549. <https://doi.org/10.17780/ksujes.1241043>.
- Instant Data Scraper, (2023). <https://chrome.google.com/webstore/detail/instant-data-scraper/foaokhiedipichpaobibbnahnkdoiiah>. Erişim Tarihi: 20/05/2023.
- Jahromi, A. H., & Taheri, M. (2017). A Non-Parametric Mixture of Gaussian Naive Bayes Classifiers Based on Local Independent Features. Paper presented at the 2017 Artificial Intelligence and Signal Processing Conference (AISP), Shiraz, Iran. <https://doi.org/10.1109/AISP.2017.8324083>.
- Kaynar, O., Görmez, Y., Yıldız, M., & Albayrak, A. (2016). Makine Öğrenmesi Yöntemleri ile Duygu Analizi. Paper presented at the In International Artificial Intelligence and Data Processing Symposium (IDAP'16), Malatya/TURKEY.
- Korkmaz, A., Aktürk, C., & Talan, T. (2023). Analyzing the User's Sentiments of ChatGPT Using Twitter Data. Iraqi Journal For Computer Science and Mathematics, 4(2), 202-214. <http://dx.doi.org/10.52866/ijcsm.2023.02.02.018>.
- Köksal, B., Erdem, G., Türkeli, C., & Kamlıoğlu Öztürk, Z. (2021). Twitter'da Duygu Analizi Yöntemi Kullanılarak Bitcoin Değer Tahminlemesi. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 9, 280-297. doi:10.29130/dubited.792909
- Lee, K., & Yu, C., (2018). Assessment of airport service quality: a complementary approach to measure perceived service quality based on Google reviews. J. Air Transp. Manag. 71, 28-44. <https://doi.org/10.1016/j.jairtraman.2018.05.004>.
- Mathayomchan, B., & Taecharungroj, V. (2020). "How was your meal?" Examining customer experience using Google maps reviews. International Journal of Hospitality Management, 90, 102641. <https://doi.org/10.1016/j.ijhm.2020.102641>.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment Analysis Algorithms and Applications: A Survey. Ain Shams En-

- gineering Journal, 5, 1093-1113. <http://dx.doi.org/10.1016/j.asej.2014.04.011>
- Multiclass Classification Using Support Vector Machines, (2023). <https://www.baeldung.com/cs/svm-multiclass-classification>. Erişim Tarihi: 23/06/2023.
- Nasukawa, T., & Yi, J. (2003). Sentiment Analysis: Capturing Favorability Using Natural Language Processing. Paper presented at the Proceedings of the 2nd International Conference on Knowledge Capture (K-CAP), Florida. <https://doi.org/10.1145/945645.945658>.
- NLTK, (2023). <https://www.nltk.org/>. Erişim Tarihi: 20/05/2023.
- Online Reviews Statistics and Trends: A 2022 Report by ReviewTrackers, <https://www.reviewtrackers.com/reports/online-reviews-survey/>, Erişim tarihi: 06.07.2023
- Öztürk, S. (2022). Bipolar Bozukluk Manik Atak Tanılı Hastaların Atak Şiddetinin Video Tabanlı Duygu Analizi ile Değerlendirilmesi. (Doktora Tezi), Trakya Üniversitesi Edirne.
- Ross, P. (2022), What Every Small Business Needs to Know About Google Reviews, <https://www.forbes.com/sites/allbusiness/2022/07/27/what-every-small-business-needs-to-know-about-google-reviews/?sh=41d1b9995e18>, (Erişim tarihi: 06.07.2023)
- Qaisi, L. M., & Aljarah, I. (2016, July). A twitter sentiment analysis for cloud providers: A case study of Azure vs. AWS. In 2016 7th International Conference on Computer Science and Information Technology (CSIT) (pp. 1-6). IEEE. <https://doi.org/10.1109/CSIT.2016.7549473>.
- Rish, I. (2001). An Empirical Study of The Naive Bayes Classifier. Paper presented at the In IJCAI 2001 workshop on empirical methods in artificial intelligence.
- Srinivas, A. C. M. V., Satyanarayana, C., Divakar, C., & Sirisha, K. P. (2021, February). Sentiment analysis using neural network and LSTM. In IOP conference series: materials science and engineering (Vol. 1074, No. 1, p. 012007). IOP Publishing. <http://doi.org/10.1088/1757-899X/1074/1/012007>.
- Taecharunroj, V. (2019). User-generated place brand identity: harnessing the power of content on social media platforms. *J. Place Manag. Dev.* 12 (1), 39–70. ISSN: 1753-8335.
- Tsytsarou, M., & Palpanas, T. (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3), 478-514. doi:10.1007/s10618-011-0238-6
- Tuna, M. F. (2022). Mobil Uygulama Müşteri Geri Bildirimindeki Duyguların Makine Öğrenmesi Yöntemleriyle Sınıflandırılması. *Journal of Business and Communication Studies*, 1(1), 83-103. <http://dx.doi.org/10.29228/jobacs.63080>
- Tuna, M. F., Kaynar, O. ve Akdoğan, M. Ş. (2021). Otellere İlişkin Çevrimiçi Geribildirimlerin Makine Öğrenmesi Yöntemleriyle Duygu Analizi. *İşletme Araştırmaları Dergisi*, 13(3), 2232-2241. <https://doi.org/10.20491/isarder.2021.1258>
- Uyaroğlu Akdeniz, F. N., & Cebeci, H. İ. (2021). Belediye Hizmetlerin Değerlendirilmesinde Duygu Analizi Yaklaşımı: Sakarya İli Örneği. *Zeki Sistemler Teori ve Uygulamaları Dergisi*, 4(2), 127-135. doi:10.38016/jista.932762
- What Is CatBoost. (2023). <https://builtin.com/machine-learning/catboost>. Erişim Tarihi: 20/05/2023.
- Yüksel, A. S., & Tan, F. G. (2018). Metin Madenciliği Teknikleri ile Sosyal Ağlarda Bilgi Keşfi. *Mühendislik Bilimleri ve Tasarım Dergisi*, 6(2), 324-333. doi:http://dx.doi.org/10.21923/jesd.384791
- Zeng, B. (2013). Social Media in Tourism. *J Tourism Hospit* 2(1), 1-2. <http://dx.doi.org/10.4172/2167-0269.1000e125>