Eurasian Academy of Sciences
Eurasian Econometrics, Statistics & Emprical Economics Journal
2020 Volume:17 S: 88- 98
Published Online September 2020 (http://econstat.eurasianacademy.org)
http://dx.doi.org/10.17740/eas.stat.2020-V17-06

SENTIMENT ANALYSIS OF TOURISM-RELATED TWEETS DURING COVID-19 OUTBREAK THROUGH MACHINE LEARNING TECHNIQUES

Engin KARAMAN* Çiğdem ARICIGİL ÇİLAN**

*Kocaeli Üniversitesi, Bilgi İşlem Daire Başkanlığı, enginkaraman1013@gmail.com,ORCID: 0000-0002-2336-6289

**Prof. Dr., İstanbul Üniversitesi, İşletme Fakültesi, ccilan75@gmail.com, ORCID: 0000-0001-5115-9773

Received Date:15.08.2020, Revised Date:20.09.2020, Accepted Date:24.09.2020

Copyright © 2020 Engin KARAMAN & Çiğdem ARICIGİL ÇİLAN. This is an open access article distributed under the Eurasian Academy of Sciences License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Covid-19 virus which is effective all the world and is a global pandemic also affected tourism choices in Turkey. In this study, sentiment analysis study was conducted over the tourism hahstagli (#turizm) Turkish tweets posted between April and August 2020. The data was obtained from the Twitter API application. 9678 messages collected in this process were structured over the necessary pre-processing and transformation processes and made ready for analysis as 4202 messages, and the messages were labeled in three categories (neutral, positive and negative) according to the emotion expressions they contain.Classification performances were compared using Machine Learning algorithms (Logistic Regression Analysis, Decision Tree, Multinominal Naive Bayes Analysis, Cluster Analysis (k-Nearest Neighbor), Support Vector Machines and Random Forests), which are frequently used in sentiment analysis studies. As a result, Logistic Regression model was found to be the most successful model.

Keywords: Tourism, Python, Twitter API, Sentiment Analysis, Machine Learning

JEL Classification: C13, C51, C81

COVİD-19 DÖNEMİNDE TURİZM İLE İLGİLİ TWİTLERİN MAKİNE ÖĞRENMESİ YÖNTEMLERİYLE DUYGU ANALİZİ

ÖZET

Tüm dünyada etkili olan ve küresel bir pandemi özelliği taşıyan COVİD-19 virüsü, Türkiye'de ki turizm tercihleri üzerinde de etkisini göstermiştir. Bu çalışmada, Nisan- Ağustos 2020 tarihleri arasında atılan turizm hahstagli (#turizm) Türkçe twitler üzerinden duygu analizi çalışması yapılmıştır. Veriler Twitter API uygulamasından elde edilmiştir. Bu süreçte toplanan 9678 adet ileti gerekli ön işleme ve dönüştürme süreçleri üzerinden yapılandırılarak 4202 adet olarak analize hazır hale getirilmiş ve iletiler anlamsal açıdan içerdikleri duygu ifadesine göre üç kategoride etiketlenmiştir. Duygu analizi çalışmalarında Makine Öğrenmesi'nin en çok kullanılan yöntemlerinden (algoritmalarından); Lojistik Regresyon Analizi, Karar Ağacı, Multinominal Naive Bayes Analizi, Kümeleme Analizi(k-En Yakın Komşu), Destek Vektör Makineleri ve Rassal Ormanlar kullanılarak sınıflandırma performansları karşılaştırılmış ve en başarılı model olarak 0.66 doğruluk skoruyla Lojistik Regresyon modeli olmuştur. Oluşturulan model geliştirilmeye açık olmakla birlikte tahminlime çalışmalarında kullanıma uygundur.

Anahtar Kelimeler: Turizm, Python, Twitter API, Duygu Analizi, Makine Öğrenmesi

JEL Sınıflandırması: C13, C51, C81

1. GİRİŞ

Bilgiye erişimin çok daha kolay olduğu günümüz dünyasında, milyonlarca insan güncel olaylara ilişkin duygularını, görüşlerini ve eleştirilerini sosyal medya ağlarını kullanarak anında ifade etmektedirler. Tüketici görüşlerinden, politik eleştiriye, beğeni ifadelerinden eğitsel faaliyetlere ve bir çok alanda bu tür online topluluklar başkalarını da etkileyebilen interaktif bir forum özelliği taşımaktadır. Ayrıca sosyal medya, işletmeler, hükümetler ve kişiler için tüketici ya da takipçi görüşlerinin değerlendirmesi bakımından da önemli bir geri bildirim fırsatı sunmaktadır. Sosyal ağların bireylerin kararlarını ve satın alma davranışlarını değiştirdiği tartışılmaz bir gerçektir. İnternet kullanıcıların %87 si tüketici görüşleri aracılığıyla satın alma kararlarını belirlediklerini ifade etmiştir (Jose vd., 2010). Bir örgüt, politikacı ya da işletme, tüketicilerinin ya da takipçilerinin düşünce ve duygularını ne kadar hızlı izleyebilirse onların reaksiyonlarını organize etmede ve politika geliştirmede rakiplerine oranla çok daha fazla avantaj elde eder(Sarlan vd., 2014, s. 212). Twitter, Facebook vb… sosyal ağlardan elde edilen büyük veriler kullanılarak yapılan araştırmalar sayesinde, müşteri görüşleri, toplumun memnuniyet hissi, halkın siyasetçiler ya da iktidar hakkındaki görüşleri analiz edilmektedir (Zhang vd., 2011, s. 56).

Bu analizler içerisinde en yaygın kullanılanlardan biri olan Duygu Analizi, insanların yazdığı metinlerdeki konu hakkındaki duygularının olumsuz mu, olumlu mu ya da nötr mü olduğunu belirlemeye yönelik geliştirilen bir yöntemdir(Akin & Gürsoy, 2018, s. 246). Bu çalışmada Twitter kaynaklı metinlerden oluşan verilerin Duygu Analizi yapılmıştır. Kullanıcı sayısı ve günlük ortalama ileti sayısı (saniyelik ortalama 8000 adet (Murthy, 2018)) bakımından en çok kullanılan sosyal medya platformlarından biri olan Twitter, araştırmacılar için de önemli bir veri kaynağıdır. Twitter'dan elde edilen iletiler (twit) metnin uyandırdığı anlam içeriğine göre etiketlenerek analize tabii tutulur. Bunun için kullanılan modellemeler genellikle Makine Öğrenmesi teknikleri ile gerçekleştirilir. Model vasıtasıyla yeni twitlerin İçerik Analizi yapılarak kişilerin hissettikleri duygunun belirlenmesi sağlanır. Bununla birlikte Twitter'ın kendine özel bir dili olması verinin analizi için zorluklar oluşturmaktadır. Bu sorunla baş edebilmek için veri, ön işleme evrelerinden geçirilerek Veri Madenciliği veya Doğal Dil İşleme yöntemleriyle analiz için hazır hale getirilir(Ayan vd., 2019, s. 496).

Bu çalışmada, tüm dünyada etkili olan ve küresel bir pandemi özelliği taşıyan COVID-19 virüsünün Türkiye'deki tüketicilerin turizm tercihleri üzerindeki algıları incelenmiştir. Birçok ülkede uygulanan hükümet önlemleri ve seyahat kısıtlamaları, küresel hasılanın yaklaşık %8'ini, istihdamın %10'unu oluşturan yıllık ortalama %4-5 oranında büyüyen (World Tourism Organization (UNWTO), 2019) turizm sektörünü doğrudan etkilemiş, neredeyse durma noktasına getirmiştir. Turizm sektörü, geçmişte yaşanan bütün krizlerde ayakta kalmayı başarsa da, COVID-19 ile birlikte gerçekleşen olaylar sadece Türkiye için değil küresel turizm endüstrisi için ciddi sonuçlar doğurmuştur. Bütün bu yaşanan olaylar, turizm talebini tahmin edilemez hale getirirken, risk ve kriz yönetiminin önemi turizmin bütün paydaşları için artmıştır (Kirant Yozcu & Cetin, 2019). Verileri elde etmek için turizm hastaglı (#turizm) twitler toplanmıştır. Sonuç olarak Nisan- Ağustos 2020 tarihleri arasında atılan 9678 adet ileti Twitter

API uygulamasından indirilmiş ve gerekli ön işleme adımları sonucunda 4202 adet ileti örneklemimizi oluşturmuştur.

2. KAVRAMSAL ÇERÇEVE

Turizm, 2019 yılında 1,5 milyardan fazla insanın katıldığı büyük bir küresel aktivite olmasına rağmen, dünyanın herhangi bir bölgesinde oluşan krizlere karşı oldukça hassas ve kırılgandır. Kriz kavramı, arzu edilmeyen ve acil müdahale edilmesi gereken, olumlu ve olumsuz sonuçlar doğuran durumlar ve olaylar olarak tanımlanabilir(Faulkner, 2001) . Covid-19 önceki krizlerden farklı olarak küresel seyahat kısıtlamaları nedeni ile turizmi aniden ve küresel boyutta neredeyse durdurmuştur. Bunun yanında Covid-19'un uzun vadeli ve benzerlerinden çok daha yıkıcı bir ekonomik krizi tetiklediği de aşikardır. Sosyal medya, ürün ve hizmetlerin tercih edilmesi esnasında tüketicilerin görüş, eleştiri, öneri veya uyarıları hakkında bireylere bilgi akışı sağladığı için bir destek ağı olarak düşünülmektedir (Zeng & Gerritsen, 2014). Sosyal medya paylaşımları, piyasalarda yeni eğilimlerin ortaya çıkması, şirketlerin saygınlığı, tüketici davranışları, hizmet ve ürün memnuniyeti gibi konuların değerlendirilmesinde değerli, kullanılabilir bilgiler sağlamaktadır.

2.1. Duygu Analizi (Sentiment Analysis)

Doğal dil işleme alanında sıklıkla kullanılan bir teknik olan duygu analizi bir metnin duygu içerip içermediğinin belirlenmesi ve eğer duygu içeriyorsa bu duygu durumunun saptanması sürecine dayanır(Liu, 2012). Bir başka ifadeyle, duygu analizi öznel ifadeleri özellikle internette tüketici tarafından üretilen içeriklerde yer alan görüş ve düşünceleri içeren özel bir metin madenciliği türüdür(Xiang vd., 2015). Duygu Analizi'nde sözlük tabanlı ve Makine Öğrenmesi tabanlı olmak üzere iki yaklaşım mevcuttur. Sözlük tabanlı yaklaşımlar uygun bir sözcüğün önsel yapısını gerektirirken, Makine Öğrenmesi tabanlı yaklaşımlar bireylerin kodlamasından hareketle metni otomatik olarak sınıflandırır (Paltoglou & Thelwall, 2017). Genellikle metnin pozitif-negatif-nötr şeklinde sınıflandırılması Makine Öğrenmesi algoritmasına (Naive Bayes, Destek Vektör Makineleri gibi) dayanarak yapılır. Elde edilen sınıflamadan hareketle bir dağılım oluşturulur. Böylece topluluğun genel düşüncesine ulaşılır.

Duygu Analizi bireylerin ve pazar gruplarının algılarını ve karakteristiklerini öğrenmek, paylaşılan yorumların içerik ve hareket noktasını belirlemek amacıyla yaygın kullanılmaktadır (Chandra Pandey vd., 2017). Turizm endüstrisinde de son dönemde sıklıkla uygulanmaya başlanan Duygu Analizi ile ilgili öne çıkan çalışmalardan bazıları şöyledir: Yu, vd. (2019), Japonya turizmi için Duygu Analizi'ni çok boyutlu olarak, online yorumlara dayanarak incelemişlerdir(Yu vd., 2019). Neidhardt vd. (2017), turizm web sitelerindeki yorumlardan hareketle turizm alanındaki faaliyetlere bakış açılarını, kullanıcıların gelecek seyahatleri ile ilgili duygularını analiz etmişlerdir (Neidhardt vd., 2017). Ayrıca, Duygu Analizi turizm endüstrisinde özellikle gezginlerin sosyal medya yorumlarını değerlendirmek için sıklıkla kullanılmaktadır. Örneğin, turizm destinasyon araştırmalarında ve markaların itibarını izlemek için (Marchiori & Cantoni, 2011; Sigala vd., 2012), konaklama yerleri hakkındaki yorumlardan hareketle müşteri yönetimi için uygun stratejiler oluşturmak konusunda (Pantelidis, 2010) uygulamalar kullanılmıştır.

Sosyal medya, ürün ve hizmetlerin tercih edilmesinde tüketicilerin görüş, eleştiri, öneri veya uyarıları hakkında bireylere bilgi akışı sağladığı için bir destek ağı olarak düşünülmektedir (Xiang ve Gretzel, 2010; Zeng ve Gerritsen, 2014). Sosyal medya paylaşımları, son dönemde turizm endüstrisine yönelik çalışmalarda piyasalarda yeni eğilimlerin ortaya çıkması, şirketlerin saygınlığı, tüketici davranışları, hizmet ve ürün memnuniyeti gibi konuların değerlendirilmesinde değerli bilgiler olarak kabul edilmektedir (McGarrity, 2016). Twitter'dan elde edilen verilerin Duygu Analizi teknikleri kullanılarak kişilerin duygularının teşhisi amacıyla birçok çalışma yapılmaktadır.

İlk uygulamalarının yurtdışında başlamasıyla birlikte Türkiye'de de Twitter üzerinden Duygu Analizi tekniğinin kullanımı ile ilgili örnekler mevcuttur. Özellikle 2012 yılı ve sonrasında Duygu Analizi'nin kullanım sıklığı artmıştır: Kaya ve arkadaşları, Türkçe kaynaklardan edindikleri politika haberleri metinlerinden Duygu Analizi çalışması gerçekleştirmiştir(Kaya vd., 2012). Aynı yıl; Şimşek ve Özdemir, Amazon'un Mechanical Turk (MTurk, 2017) servisinden yararlanarak üzüntü ve mutluluk ifadelerinde kullanılmak üzere 113 sözcüklük bir Türkçe sözlük yaratmış ve bu sözcüklerin Twitter iletilerinde geçme frekanslarına göre mesajları mutlu ve mutsuz şeklinde sınıflandırmışlardır(Şimşek & Özdemir, 2012). Çetin ve Amasyalı, Türkçe Twitter mesajları üzerinden eğiticili 6 yöntemin ve geleneksel 2 yöntemin performanslarını 5 algoritma kullanarak 2 veri kümesi üzerinde karşılaştırmışlardır (Çetin & Amasyalı, 2013) . Meral ve Diri, Twitter'dan toplanan verileri kullanarak Destek Vektör Makinesi (DVM), Rastgele Orman ve Naïve Bayes Makine Öğrenmesi yöntemlerini kullanılarak, akıllı bir sistem oluşturulmuşlardır. Bu üç sınıflandırma yönteminin doğruluk oranları birbirlerine çok yakın olmakla birlikte, en iyi sınıflayıcının DVM yöntemi olduğunu tespit etmişlerdir. Ayan vd. çalışmalarında Twitter üzerindeki tweetlerin İslamofobik olup olmadığını Duygu Analizi ile saptamışlardır.

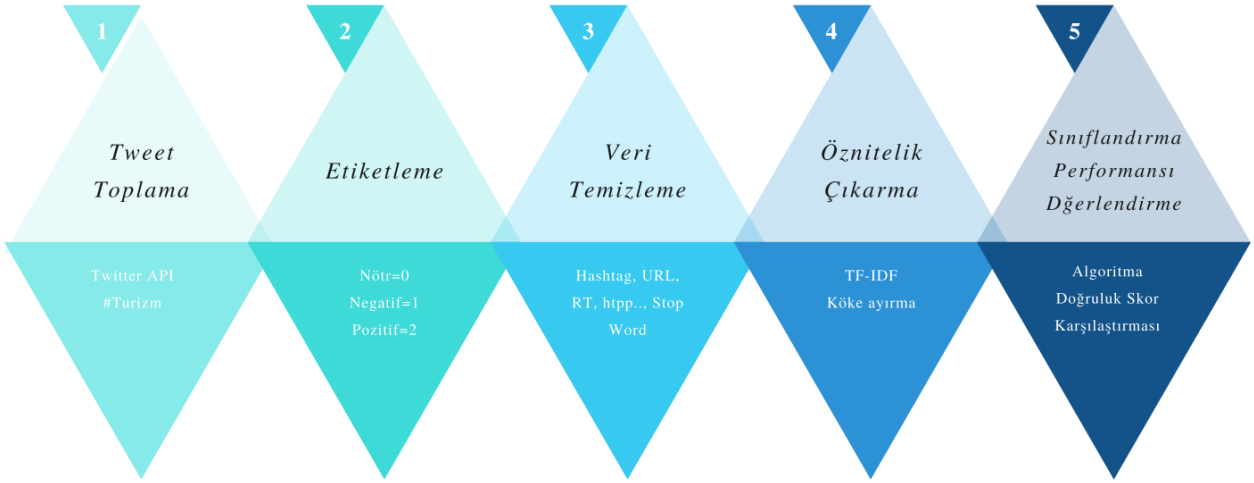
Çalışmada Naive Bayes Sınıflandırıcı ve Lineer Ridge regresyonu ile eğitilen modeller üzerinden Recall, precision, F1 ölçütlerinde hesaplamalarda bulunmuşlardır. Sonuç olarak pozitif Tweetler için Ridge Regresyon modelinden Naive Bayes sınıflandırıcıya göre daha iyi sonuçlar alınmıştır (Ayan vd., 2019). Akın ve Şimşek çalışmalarında Yarı Denetimli Öğrenme ile pozitif ve negatif duygu sözlükleri genişletip, sözlük temelli performansı yüksek bir Duygu Analizi modeli sunmuşlardır (Akin & Gürsoy, 2018). Kırcı ve Gülbak, Instagram verilerini kullanarak Duygu Analizi yöntemini çalışmışlardır (Kirci & Gülbak, 2020). Çoban vd., Türkçe Twitlerden oluşan veri setini kullanarak Multinominal Naive Bayes, Destek Vektör Makinesi (DVM) ve KNN algoritmalarını kullanarak Duygu Analizi yöntemini kullanmışlardır. Sonuç olarak veri üzerinde karakter seviye N-Gram modelinin, BoW modelden tüm sınıflandırıcılar için daha iyi sonuç verdiğini tespit etmişlerdir(Çoban vd., 2015). Nizam ve Akın, Bazı gıda firmalarının çeşitli ürünleri hakkında atılan Twitter veri seti üzerinden Duygu Analizi çalışması yapmış ve Random Forest , Naive Bayes , Decision Tree , Sequential Minimal Optimization ve 1-Nearest Neighbors (IB1) sınıflandırma algoritmalarının başarılarını incelemişlerdir. Sonuç olarak, dengeli veri setinin, dengesiz veri setine göre daha iyi performans gösterdiğini belirlemişlerdir(Nizam & Akın, t.y.). İlhan ve Sağaltıcı, Twitter verileri ile duygu analizi gerçekleştirmişlerdir. Bunun için Naive Bayes ve Destek Vektör Makinesi (DVM) gibi Makine

Öğrenmesi yöntemlerini kullanarak tweetleri pozitif ve negatif sınıflara ayırmak için akıllı bir model oluşturmuşlardır (İlhan & Sağaltici, 2020).

Bununla birlikte son zamanlarda özellikle Doğal Dil İşlemi alanındaki hızlı ilerlemeler, birçok tekniğin ve uygulamanın Duygu Analizi'nde optimal çözümler oluşturmasına olanak sağlamıştır. Son olarak OpenAI ekibi tarafından geliştirilen GPT3 uygulaması, Derin Öğrenme yalşaşımını kullanarak kendi kendine içerik üretebilen bir özbağlanımlı dil modeli sunulmuştur. Bu alanda yapılacak çalışmaların büyük gelişmeler sağlayacağı aşikardır.

3. ARAŞTIRMANIN METODOLOJİSİ

Çalışmada kullanılan veri seti Twittter üzerinden elde edilmiştir. Veriler Makine Öğrenmesi'ne uygun olarak belirli bir süreçten geçirilerek, analize hazır hale getirilmiştir. Ardından birçok sınıflandırma algoritması uygulanarak performans değerlendirmeleri yapılmıştır. Yapılan işlemleri özetleyen şema Şekil 1'de sunulmuştur.



Şekil 1. Duygu Analizi Süreci

3.1. Verilerin Toplanması

Twitter sosyal medya verilerinin toplanmasında en sık kullanılan uygulamadır. Kullanıma sunduğu Twitter Streaming API arayüzü sayesinde sınırlı da olsa araştırmacılara istenilen konu ya da hashtag üzerinden JSON formatında metinler sunmaktadır(Ramteke vd., 2016). Kullanıcı ihtiyacına göre veri istenilen formata (CSV, XLSX vb…) dönüştürülebilmektedir. Çalışmamızda COVİD 19 pandemisinin toplumsal hayatı etkilediği Nisan- Eylül 2020 tarihleri arasında atılan #Turizm hashtag ile atılan twitler toplanmıştır. Ham veri seti 9678 adet iletiden oluşmuştur. Veriler Excel dosyalarında toplanmıştır.

3.2. Etiketleme (Labelling)

Sürecin bu aşamasında, ham veri seti içerisindeki gereksiz, tekrar eden ve amaç dışı veriler elenmiştir. Bu elemenin sonucunda 4202 adet tekrarsız ve özgün ileti etiketlenmeye hazır hale getirilmiştir. Bu sürecin ardından her ileti, içerdiği anlamlara göre; nötr ise 0, negatif ise 1 ve pozitif 2 olarak işaretlenmiştir.

Tablo 1: Covid19 Döneminde Turizm İle İlgili Görüşlerin Dağılımı

Duygu Durumu	Frekans	Yüzde %
Pozitif	1472	35
Nötr	1158	28
Negatif	1572	37

3.3. Veri Temizleme

Twitlerin içerisinde; @, # , http, simge vb. sayısız anlam içermeyen karakterler bulunabilmektedir. Veri setinin doğru sınıflandırılabilmesi için bu karakterler veri setinden temizlenmiştir. Bununla birlikte Türkçe'de sık kullanılan sıfat, bağlaç vb. (çok, bazen, ve, ama) ifadeler (stop words) metinde anlamlı değilse metinden temizlenmiştir. Böylece kelime matrisinin boyutu gereksiz sütunlardan ayıklanmıştır.

3.4. Öznitelik Çıkarma

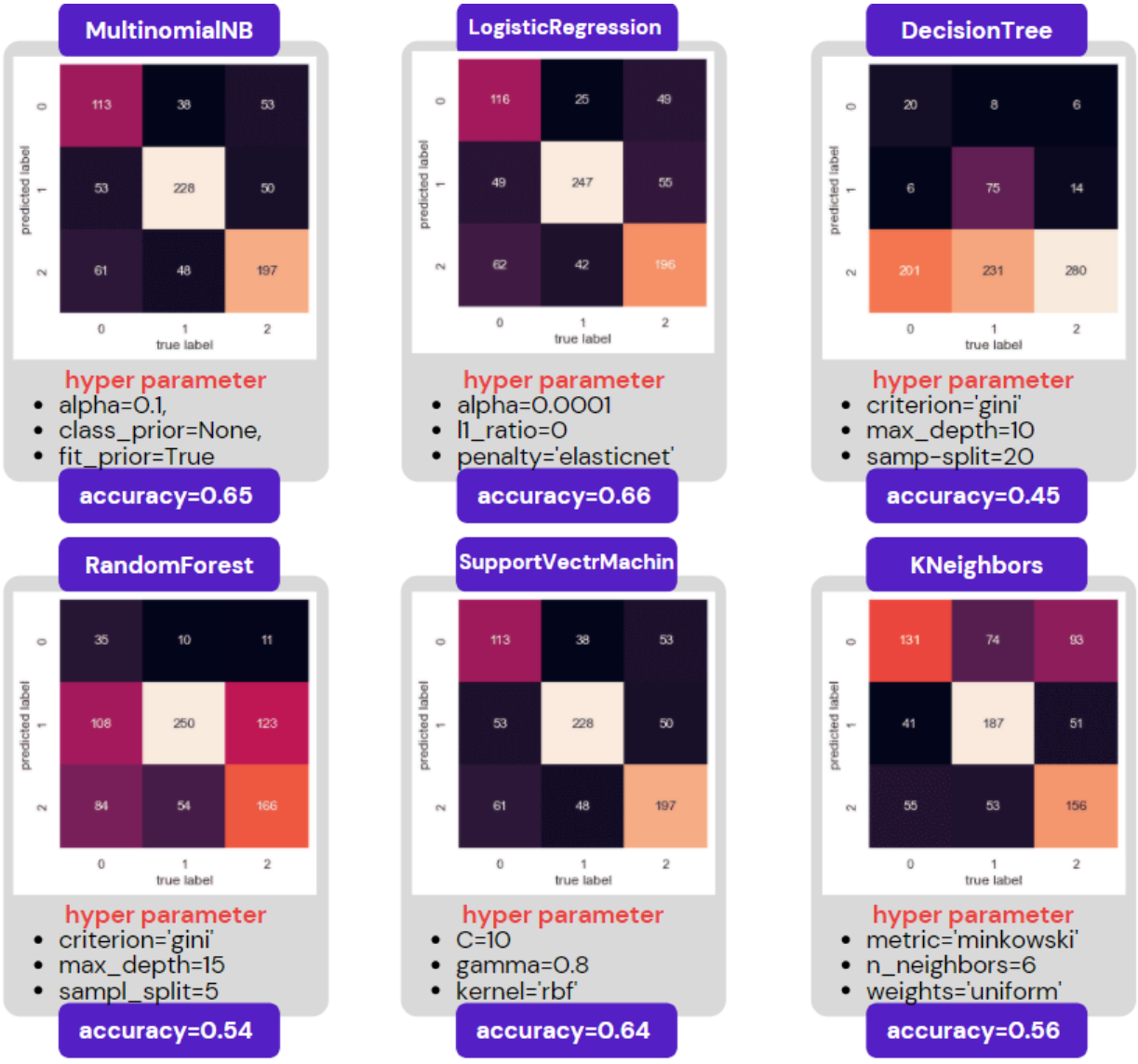
Öznitelik çıkarma, algoritmaların ayırt edici gücünü daha belirgin şekilde ortaya çıkarmayı amaçlamaktadır (Sariman & Mutaf, 2020, s. 142). Bu sürece geçmeden önce öznitelik uzay boyutunu indirgemek amacıyla kelime kök analizi () yapılmıştır(Çoban vd., 2015). Bunun için çekim ve yapım eklerinden oluşan kelimeler köküne indirgemek için üretilen bir Türkçe Doğal Dil İşleme kütüphanesi olan Zemberek kullanılmıştır. Kütüphanenin Python reposu araştırmacılar tarafından sıklıkla tercih edilmektedir. Bu çalışmada metin sınıflandırmada en çok kullanılan yöntemlerden biri olan TF-IDF (Terim Frekansı-Ters Doküman Frekansı, Terim Ağırlıklandırma Yöntemi kullanılmıştır. TF-TDF, verideki bir kelimenin frekansını, o kelimeyi içeren toplam doküman sayısı ve tüm dokümanların sayısına dayalı olarak hesaplayan ağırlıklandırma yöntemidir(Özyurt & Akçayol, 2018, s. 672). Böylece verideki önemsiz değerler elenerek önemli öznitelikler tespit edilir ve sınıflandırma işlemi sırasında performans artışı sağlanır(Alpkoçak vd., 2019, s. 722). Son olarak veri seti Vektör Uzay Modeline dönüştürülerek bilgisayar diline uygun hale getirilmiştir.

3.5. Sınıflandırma Performansının Değerlendirilmesi

Sınıflandırma işlemi, sonradan edinilen ve kategorilendirilmemiş metinlerin, modelin eğitim verisi setindeki kategorilerinden en uygun olana atanma işlemidir(Çoban vd., 2015). Böylece başarılı bir şekilde eğitilmiş ve performansı yüksek olan model aracılığıyla, ilk defa karşılaşılan örnekler metin içeriğine bağlı olarak etiketlenir. Doğal Dil İşleme ve özel olarak Duygu Analizi alanında çeşitli algoritmalar günümüzde kulanılmaktadır(Ramteke vd., 2016). Verilerin, üçlü kategorik özellik göstermesi nedeniyle çoklu sınıflayıcı algoritmalar kullanılmıştır. Çalışmada bu algoritmalar arasında en sık kullanılan;Lojistik Regresyon Modeli, Karar Ağacı, Multinominal Naive Bayes, Kümeleme Analizi(k-En Yakın Komşu), Destek Vektör Makineleri ve Rassal Ormanlar, Python programlama dilindeki pandas, nltk ve scikit learn kütüphaneleri kullanılarak uygulanmıştır. Bütün bu işlemlerin gerçekleştirilmesinde. % 80 eğitim ve %20 test olarak ayrılan veri seti belirtilen algoritmalarca sınıflandırılmış ve performans skorları doğruluk değerlerine göre belirlenmiştir. Ayrıca her sınıflandırıcı için en uygun hiper parametrenin belirlenmesinde GridSearchCV yöntemi kullanılmıştır.

3.6. Bulgular

Ham verinin işlenerek analize uygun hale getirilmesinden sonra, eğitim-test olarak ayrılan veri seti Vektör Uzay Modeli'nde sayısal hale getirilerek çeşitli sınıflama algoritmalarıyla etkinlikleri karşılaştırılmıştır. Buna göre sınıflandırma için kullanılan Makine Öğrenmesi algoritmalarının karşılaştırma matrisi ve doğruluk katsayıları Şekil 2'de sunulmuştur.



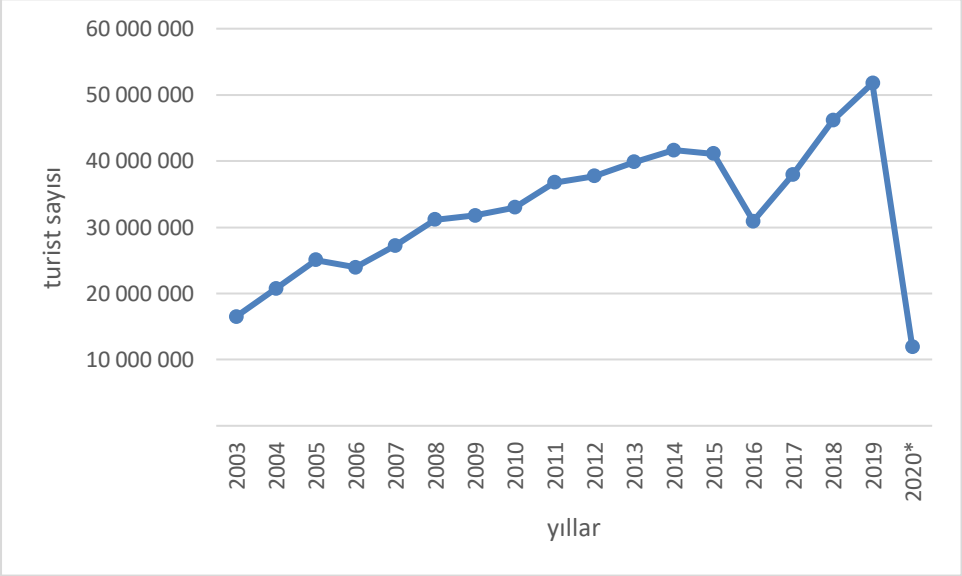
Şekil 2. Sınıflandırma Algoritmalarının Performanslarının Karşılaştırılması

Şekil 2'de görüldüğü gibi, en iyi performansı gösteren sınıflayıcı 0.66 doğruluk oranı ile Stochastic Gradient Descent opsiyonlu Lojistik Regresyon yöntemidir. Yani modelimiz tahminlemede %66'lık doğrulukla çalışmaktadır. Bu oran mükemmel bir verim olmasa da tatmin edici bir sayıdır. Skoru arttırmak için çeşitli varyasyonlar geliştirilebilir. Ayrıca istenilen amaca uygun olarak diğer sınıflayıcılar da kullanılabilir. Örneğin amaç sadece olumlu

görüşlerin baz alındığı bir model elde etmek olduğunda Karar Ağacı algoritması, olumsuz görüşlerin hedef alındığı bir model olduğunda ise Rassal Ormanlar algoritması kullanılabilir.

4. SONUÇ

Türkiye ekonomisi için önemli bir kaynak olan turizm sektörü Covid19 pandemisi ve kısıtlamalar nedeniyle 2020 yaz sezonunda olumsuz bir seyir izlemiştir. Şekil 3'de görüldüğü gibi 2020 Eylül ayına kadar gelen turist sayısı keskin bir düşüş göstermiştir (yigm.ktb.gov.tr) .



Şekil 3. 2003 ve 2020 Yılları Arasında Türkiye'deki Turist Sayısı

Bu tercihler doğal olarak insanların sosyal medya iletilerinde de kendini göstermiştir. Dünya çapında kısıtlamaların başladığı Nisan-Eylül ayları arasındaki Twitter mesajlarında da bunun yansıması görülmektedir. Doğal Dil İşleme araçlarındaki gelişmeler analistlere hızlı araştırma süreçleri ve uygulama alanları sunmaktadır. Bu alanların en önemli yöntemlerinden biri olan Twitter'da Duygu Analizi literatürde kendine sıklıkla yer bulmaktadır. Bu çalışmada Makine Öğrenmesi algoritmaları kullanılarak duygu analizi uygulanmıştır. Gerekli ön işleme ve yapılandırma adımlarından sonra veri 6 farklı algoritma üzerinden işlenmiş ve performansları karşılaştırılmıştır.

Sonuç olarak Lojistik Regresyon modeli bu algoritmalar içerisinde performansı en yüksek olan modeldir. Modelin test verilerini başarıyla tespit etme oranı % 66'dır. Bu oranı artırmak ve daha başarılı bir model oluşturmak mümkündür. Öncelikle örnek sayısı arttırılarak daha başarılı sonuçlar elde edilebilir. Ayrıca Derin Öğrenme algoritması kullanılarak da model geliştirilebilir. Türkçe dili üzerinde geliştirilen metin analizi uygulamaları İngilizce ile karşılaştırıldığında henüz ueyerince gelişmemiştir. Bunun üzerine son yıllarda başarılı çalışmalar (Zemberek kütüphanesi, SentiTurkNet vb…) yapılmakla birlikte üzerinde yoğunlaşmak temel bir ihtiyaç olarak karşımıza çıkmaktadır. Çalışma, Türkiye turizmi alanında ilk olma özelliği açısından özgünlük taşımaktadır.

REFERENCES

- Akin, B., & Gürsoy, U. (2018). Adaptif Öğrenme Sözlüğü Temelli Duygu Analiz Algoritması Önerisi. *Bilişim Teknolojileri Dergisi*. https://doi.org/10.17671/gazibtd.342419
- Alpkoçak, A., Tocoglu, M. A., Çelikten, A., & Aygün, İ. (2019). Türkçe Metinlerde Duygu Analizi için Farklı Makine Öğrenmesi Yöntemlerinin Karşılaştırılması. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, *21*(63), 719-725. https://doi.org/10.21205/deufmd.2019216303
- Ayan, B., Kuyumcu, B., & Ceylan, B. (2019). Twitter Üzerindeki İslamofobik Twitlerin Duygu Analizi ile Tespiti. *Journal of the Faculty of Engineering and Architecture of Gazi University*, *7*(2), 495-502. https://doi.org/10.29109/gujsc.561806
- Chandra Pandey, A., Singh Rajpoot, D., & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. *Information Processing & Management*, *53*(4), 764-779. https://doi.org/10.1016/j.ipm.2017.02.004
- Çetin, M., & Amasyalı, M. F. (2013). Supervised and traditional term weighting methods for sentiment analysis. *2013 21st Signal Processing and Communications Applications Conference (SIU)*, 1-4. https://doi.org/10.1109/SIU.2013.6531173
- Çoban, Ö., Özyer, B., & Özyer, G. T. (2015). Sentiment analysis for Turkish Twitter feeds. *2015 23nd Signal Processing and Communications Applications Conference (SIU)*, 2388-2391. https://doi.org/10.1109/SIU.2015.7130362
- Faulkner, B. (2001). Towards a framework for tourism disaster management. *Tourism Management*, *22*(2), 135-147. https://doi.org/10.1016/S0261-5177(00)00048-0
- İlhan, N., & Sağaltici, D. (2020). Twitter'da Duygu Analizi. *Harran Üniversitesi Mühendislik Dergisi*, *5*(2), 146-156. https://doi.org/10.46578/humder.772929
- Jose, A. K., Bhatia, N., & Krishna, S. (2010). Twitter sentiment analysis. *National Institute of Technology Calicut*.
- Kaya, M., Fidan, G., & Toroslu, I. H. (2012). Sentiment Analysis of Turkish Political News. *2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, *1*, 174-180. https://doi.org/10.1109/WI-IAT.2012.115
- Kirant Yozcu, O., & Cetin, G. (2019). A Strategic Approach to Managing Risk and Crisis at Tourist Destinations. Içinde N. Kozak & M. Kozak (Ed.), *Tourist Destination*

Management: Instruments, Products, and Case Studies (ss. 273-287). Springer International Publishing. https://doi.org/10.1007/978-3-030-16981-7_16

- Kirci, P., & Gülbak, E. (2020). Instagram Verileri ile Duygu Analizi. *Avrupa Bilim ve Teknoloji Dergisi*, 360-364. https://doi.org/10.31590/ejosat.780129
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, *5*(1), 1-167. https://doi.org/10.2200/S00416ED1V01Y201204HLT016
- Marchiori, E., & Cantoni, L. (2011). The Online Reputation Construct: Does it Matter for the Tourism Domain? A Literature Review on Destinations' Online Reputation. *Information Technology & Tourism*, *13*(3), 139-159. https://doi.org/10.3727/109830512X13283928066715
- Murthy, D. (2018). *Twitter: Social communication in the twitter age* (Second edition). Polity.
- Neidhardt, J., Rümmele, N., & Werthner, H. (2017). Predicting happiness: User interactions and sentiment analysis in an online travel forum. *Information Technology & Tourism*, *17*(1), 101-119. https://doi.org/10.1007/s40558-017-0079-2
- Nizam, H., & Akın, S. S. (t.y.). *Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması*. 6.
- Özyurt, B., & Akçayol, M. A. (2018). *Fikir madenciliği ve duygu analizi, yaklaşımlar, yöntemler üzerine bir araştırma*. http://acikerisim.selcuk.edu.tr:8080/xmlui//handle/123456789/14106
- Paltoglou, G., & Thelwall, M. (2017). Sensing Social Media: A Range of Approaches for Sentiment Analysis. Içinde J. A. Holyst (Ed.), *Cyberemotions: Collective Emotions in Cyberspace* (ss. 97-117). Springer International Publishing. https://doi.org/10.1007/978-3-319-43639-5_6
- Pantelidis, I. S. (2010). Electronic Meal Experience: A Content Analysis of Online Restaurant Comments. *Cornell Hospitality Quarterly*, *51*(4), 483-491. https://doi.org/10.1177/1938965510378574
- Ramteke, J., Shah, S., Godhia, D., & Shaikh, A. (2016). Election result prediction using Twitter sentiment analysis. *2016 International Conference on Inventive Computation Technologies (ICICT)*, *1*, 1-5. https://doi.org/10.1109/INVENTIVE.2016.7823280

- Sariman, G., & Mutaf, E. (2020). Covıd-19 Sürecinde Twıtter Mesajlarının Duygu Analizi Sentıment Analysıs Of Twıtter Messages In Covıd-19 Process. *Euroasia Journal of Mathematics Engineering Natural and Medical Sciences*, *7*, 137-148.
- Sarlan, A., Nadam, C., & Basri, S. (2014). Twitter sentiment analysis. *Proceedings of the 6th International Conference on Information Technology and Multimedia*, 212-216. https://doi.org/10.1109/ICIMU.2014.7066632
- Sigala, M., Christou, E., & Gretzel, U. (2012). *Social Media in Travel, Tourism and Hospitality: Theory, Practice and Cases*. Ashgate Publishing, Ltd.
- Şimşek, M. U., & Özdemir, S. (2012). Analysis of the relation between Turkish twitter messages and stock market index. *2012 6th International Conference on Application of Information and Communication Technologies (AICT)*, 1-4. https://doi.org/10.1109/ICAICT.2012.6398520
- World Tourism Organization (UNWTO) (Ed.). (2019). *International Tourism Highlights, 2019 Edition*. World Tourism Organization (UNWTO). https://doi.org/10.18111/9789284421152
- Xiang, Z., Schwartz, Z., Gerdes, J. H., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, *44*, 120-130. https://doi.org/10.1016/j.ijhm.2014.10.013
- Yu, C., Zhu, X., Feng, B., Cai, L., & An, L. (2019). Sentiment Analysis of Japanese Tourism Online Reviews. *Journal of Data and Information Science*, *4*(1), 89-113. https://doi.org/10.2478/jdis-2019-0005
- Zeng, B., & Gerritsen, R. (2014). What do we know about social media in tourism? A review. *Tourism Management Perspectives*, *10*, 27-36. https://doi.org/10.1016/j.tmp.2014.01.001
- Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear". *Procedia - Social and Behavioral Sciences*, *26*, 55-62. https://doi.org/10.1016/j.sbspro.2011.10.562