Journal of the Faculty of Engineering and Architecture of Gazi University 38:2 (2023) 1093-1104

Mühendislik Mimarlık Fakültesi Dergisi
Journal of The Faculty of Engineering and Architecture of Gazi University

Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Sentiment analysis of public sensitivity to COVID-19 vaccines on Twitter by majority voting classifier-based machine learning

Cihan Çılgın[1]*, Hadi Gökçen[2], Yılmaz Gökşen[3]

[1]Department of Management Information Systems, Faculty of Applied Sciences, Bolu Abant İzzet Baysal University, 14900, Bolu, Türkiye
[2]Department of Industrial Engineering, Engineering Faculty, Gazi University, 06570, Ankara, Türkiye
[3]Department of Management Information Systems, Faculty of Economics and Administrative Sciences, Dokuz Eylül University, 35160, İzmir, Türkiye

Highlights:

- Sentiment analysis on Twitter
- Measuring public sensitivity of COVID-19 vaccines
- Machine learning with majority voting classifier

Keywords:

- COVID-19
- Vaccine
- Sentiment analysis
- Majority voting classifier
- Machine learning

Article Info:

Research Article
Received: 29.11.2021
Accepted: 14.05.2022

DOI:

10.17341/gazimmfd.1030198

Correspondence:

Author: Cihan Çılgın
e-mail: cihancilgin@ibu.edu.tr
phone: +90 374 311 2111 / 5625

Graphical/Tabular Abstract

The applied model architecture includes six steps and Figure A shows that the steps of proposed model architecture; obtaining data, manually labeling for training, data preprocessing, models training and testing, developing majority voting models and labelling all tweets.

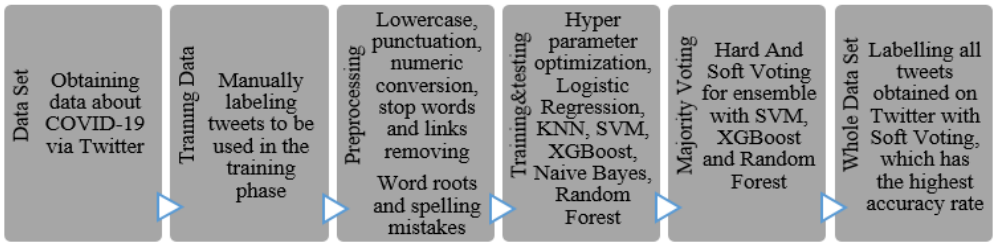


Figure A. Training and testing of the model and classification of all tweets

Purpose:
The aim of this study is to analyze public sentiment with Machine Learning of vaccine-related tweets obtained on Twitter in order to better understand the attitudes and concerns of social media users, especially regarding COVID-19 vaccines in Turkey. For this purpose, a majority voting method has been developed with machine learning methods, which are frequently used in sentiment analysis studies.

Theory and Methods:
In the study, machine learning algorithms used in six different classification tasks, which are frequently used in sentiment analysis, were compared. Then, by comparing these machine learning methods, the majority voting method, which is an ensemble learning method, was developed by using the three methods with the highest accuracy. For this purpose, both soft voting and hard voting methods were used to generate majority voting in the classification task. In addition, the data used in the study were collected between 01.04.2021 and 31.08.2021, when vaccine studies accelerated in Turkey, a total of 412,588 tweets in Turkish.

Results:
Although the SVM algorithm among the individual methods achieved a high success rate of 89.6%, it is seen that the XGBoost model is the most successful algorithm with a accuracy rate of 89.8%. Although the Random Forest approach among other machine learning approaches has achieved remarkable success, the same is not the case for other methods. For this reason, high accuracy SVM, XGBoost and Random Forest methods are used in both hard voting and soft voting majority voting approaches. Although the hard voting method achieved a higher accuracy than the individual methods with a success rate of 88.9%, the soft voting method was the most successful classification method with a relatively high accuracy rate of 90.5%. For this reason, soft voting approach was used in the labeling of daily tweets obtained in the study.

Conclusion:
As a result of the analyses carried out with the soft voting method, although there are fluctuations in the sentiment polarity of the tweets about the vaccine, it is observed that the negative sentiments and therefore the opposition to the vaccine is becoming more and more popular on social media. Particularly, when compared to previous study findings, positive sentiments in vaccine-related posts in Turkey are quite low rate. For this reason, the ongoing opposition to vaccination on social media in Turkey becomes a subject that needs to be examined more carefully. As far as we know, this study is the first in Turkey to perform sentiment analysis on COVID-19 vaccines. In addition, the findings of the study show that the proposed method is a valuable and easily applied tool to monitor the sensitivity of COVID-19 vaccines with a sentiment analysis approach via social media.

Journal of the Faculty of Engineering and Architecture of Gazi University 38:2 (2023) 1093-1104

Mühendislik Mimarlık Fakültesi Dergisi
Journal of The Faculty of Engineering and Architecture of Gazi University

Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Twitter'da COVID-19 aşılarına karşı kamu duyarlılığının çoğunluk oylama sınıflandırıcısı temelli makine öğrenmesi ile duygu analizi

Cihan Çılgın[1]*, Hadi Gökçen[2], Yılmaz Gökşen[3]

[1]Bolu Abant İzzet Baysal Üniversitesi, Gerede Uygulamalı Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, 14900, Bolu, Türkiye
[2]Gazi Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, 06570, Ankara, Türkiye
[3]Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, 35160, İzmir, Türkiye

ÖNEÇIKANLAR

- Twitter'da duygu analizi
- COVID-19 aşılarına yönelik kamu duyarlılığının ölçümü
- Çoğunluk oylama sınıflandırıcı ile makine öğrenmesi

Makale Bilgileri
Araştırma Makalesi
Geliş: 29.11.2021
Kabul: 14.05.2022

DOI:
10.17341/gazimmfd.1030198

Anahtar Kelimeler:
COVID-19,
aşı, duygu analizi,
çoğunluk oylama
sınıflandırıcısı,
makine öğrenmesi

ÖZ

Dünyada milyarlarca kullanıcısı bulunan sosyal medya platformlarının yükselişiyle birlikte bilginin yayılması her zamankinden daha kolay hale gelmiştir. COVID-19 pandemisi aşılar da dâhil olmak üzere birçok konunun tartışılmasında sosyal medya kullanımını artırmıştır. Bu çalışmanın amacı, Türkiye'de, özellikle sosyal medya kullanıcılarının COVID-19 aşılarına ilişkin tutumunu ve endişelerini daha iyi anlamak adına Twitter üzerinde elde edilen aşıyla ilgili tweetlerin makine öğrenmesi ile kamu duyarlılığını analiz etmektir. Bu amaç doğrultusunda çalışma da altı farklı sınıflandırma görevinde kullanılan makine öğrenmesi algoritması karşılaştırılarak en yüksek doğruluk oranına sahip Destek Vektör Makinesi, XGBoost ve Rastgele Orman ile bir kolektif öğrenme yöntemi olan çoğunluk oylama yöntemi geliştirilmiştir. Çoğunluk oylama yöntemlerinden birisi olan Yumuşak Oylama yöntemi hem Sert Oylama yaklaşımından hem de bireysel diğer altı makine öğrenmesi yaklaşımlarından %90,5 ile daha yüksek başarı oranına ulaşmıştır. En yüksek doğruluk oranına sahip olan Yumuşak Oylama yöntemi ile Twitter'dan elde edilen 153 güne ait 412.588 adet günlük tweet analiz edilerek sonuçlar raporlanmıştır. Çalışmanın bulguları son derece çarpıcı olup, diğer ülkeler üzerine yapılan çalışmalardan da farklılık göstermektedir. Bu çalışma bildiğimiz kadarıyla Türkiye'de COVID-19 aşılarına yönelik duygu analizi gerçekleştirilen ilk çalışmadır. Ayrıca çalışma bulguları, önerilen yöntemin sosyal medya üzerinden duygu analizi yaklaşımıyla COVID-19 aşılarına ilişkin duyarlılığı izlemek için değerli ve kolayca uygulanan bir araç olduğunu göstermektedir.

Sentiment analysis of public sensitivity to COVID-19 vaccines on Twitter by majority voting classifier-based machine learning

HIGHLIGHTS

- **S**entiment analysis on Twitter
- Measuring public sensitivity of COVID-19 vaccines
- Machine learning with majority voting classifier

Article Info
Research Article
Received: 29.11.2021
Accepted: 14.05.2022

DOI:
10.17341/gazimmfd.1030198

Keywords:
COVID-19,
vaccine,
sentiment analysis,
majority voting classifier,
machine learning

ABSTRACT

With the rise of social media platforms, which have billions of users around the World, the dissemination of information has become easier than ever. The COVID-19 pandemic has increased the use of social media to discuss many topics, including vaccines. The aim of this study is to analyze public sentiment with Machine Learning of vaccine-related tweets obtained on Twitter in order to better understand the attitudes and concerns of social media users, especially regarding COVID-19 vaccines in Turkey. For this purpose, the majority voting method, which is an ensemble learning method, was developed by comparing the machine learning algorithm used in six different classification tasks and then via Support Vector Machine, XGBoost and Random Forest having the highest accuracy, in the study. Soft Voting method, which is one of the majority voting methods, has reached a success rate of 90.5%, with a higher success rate than both the Hard Voting approach and the other six individual machine learning approaches. With the Soft Voting method, which has the highest accuracy rate, 412,588 daily tweets from 153 days obtained from Twitter were analyzed and the results were reported. The findings of the study are very striking and differ from studies on other countries. As far as we know, this study is the first in Turkey to perform sentiment analysis on COVID-19 vaccines. In addition, the findings of the study show that the proposed method is a valuable and easily applied tool to monitor the sensitivity of COVID-19 vaccines with a sentiment analysis approach via social media.

*Sorumlu Yazar/Yazarlar / Corresponding Author/Authors : *cihancilgin@ibu.edu.tr, hgokcen@gazi.edu.tr, yilmaz.goksen@deu.edu.tr / Tel: +90 374 311 2111 / 5625

1. Giriş (Introduction)

İnsanlarda olduğu gibi hayvanlarda da hastalığa neden olabilen ve birçok türü bulunan Coronavirüs'ün ilk olarak 2019 yılının sonlarında Çin'in Vuhan kentinde [1] ortaya çıktığı ve oradan da COVID-19 (SARS-CoV) olarak dünyaya yayılmaya başladığı düşünülmektedir. Teyit edilen vakalardaki hızlı artış, COVID-19'un önlenmesini ve kontrolünü son derece ciddi hale getirmiştir [2]. Mevcut durum göz önüne alındığında virüs 230 milyon insanı etkilemiş ve 4,5 milyondan daha fazla ölüme neden olmuştur. Tüm Dünyayı etkisi altına alan COVID-19 pandemisi ile birlikte bir aşının geliştirilmesi ve yaygınlaştırılmasına yönelik küresel çabalar da hızla ilerlemektedir. Mayıs 2020'de, 73. Dünya Sağlık Kurulu, SARS-CoV-22'nin bulaşmasını önlemek, kontrol altına almak ve durdurmak için küresel bir halk sağlığı hedefi olarak kapsamlı bağışıklamanın rolünü tanıyan bir karar yayınlamıştır [3]. Bu amaçla günümüzde birçok özel ve kamu destekli firma veya kuruluş tarafından çalışmalar yürütülerek geliştirilen aşılar insanlığın kullanımına sunulmuştur. Küresel olarak şu anda 125'ten fazla aşı adayı, devam eden 365 aşı denemesi ve en az bir ülke tarafından onaylanmış 18 COVID-19 aşısı mevcut durumdadır [3]. COVID-19 aşılarının yakın zamanda kullanıma sunulması ile birlikte kamuoyunda aşılara karşı temel güven düzeylerini geliştirmek ve güven kaybının tespit edilmesini sağlamak için kamuoyunun aşılara karşı duygularını sürekli olarak izlemek ve daha iyi anlamak ihtiyacı doğmuştur [4]. Aşılar geleneksel olarak halkın korkusu ve tereddüttü ile karşılaşmıştır ve COVID-19 aşıları da bir istisna değildir [5]. Doğan ve Düzel 'in [6] gerçekleştirdikleri çalışmada araştırmaya katılan bireylerin %53,6 (806)'sının salgın hastalıklara önlem olarak aşı olmayı düşündüğünü belirtirken %14,5 (217)'inin aşı olmayı düşünmediğini belirtmesi bu durumu destekler niteliktedir. Toplumda sürü bağışıklığı düzeyine ulaşmak için % 95 etkililiği olduğu düşünülen bir aşı için, gerekli sürü bağışıklığı seviyesinin % 63 ila % 76 [7] olduğu düşünüldüğünde, yeterli seviyede aşılama oranına ulaşmak için toplumun desteğinin gerekli olduğu anlaşılmaktadır. Bu nedenle, toplumun COVID-19 aşılarına yönelik görüşünü ve dolayısıyla aşı olma veya olmama isteklerini anlamak son derece önemli bir konu haline gelmektedir. Bununla birlikte, önceki çalışmaların, sosyal medyada aşı ile ilgili bilgilerin yaklaşık %30 ila %60'ının aşı karşıtlığını içerdiğini [8] ve zaman periyotları boyunca Twitter'da aşı karşıtlığının % 80 arttığını [9] göstermesi bu konuyu daha da önemli bir hale gelmektedir.

Dünyada 4,5 milyardan fazla kullanıcısı bulunan [10] sosyal medya platformlarının yükselişiyle birlikte bilginin yayılması (dolayısıyla potansiyel olarak yanlış bilgi) her zamankinden daha kolay hale gelmiştir. Sosyal medya platformları, insanların siyasetten spora, eğitimden sağlık konularına kadar fikirlerini ve duygularını geniş kitlelerle özgürce paylaşabilecekleri güçlü kamu platformları sunmaktadır. Sosyal medya kullanıcıları, hayatlarının bir yansıması olarak kullanıcı profilleri oluşturmakta ve çoğu zaman onlar hakkında kişisel bilgilerini açıkça paylaşmaktadır [11]. Twitter geçmişte de halk sağlığını tehdit eden durumların anlaşılmasında önde gelen bir araştırma aracı olarak kullanılmıştır [12, 13]. Önceki çalışmalar, Twitter'ın toplumun pandemik virüs salgınları hakkındaki görüşünü değerlendirmede oynayabileceği büyük rolü vurgulamışlardır [14]. Mevcut COVID-19 pandemisi, aşılar da dâhil olmak üzere pandemi hakkında bir dizi konuyu tartışmak için bir forum olarak sosyal medya kullanımında bir artışa neden olmuştur [15] ve şaşırtıcı olmayan bir şekilde, aşı karşıtı veya olumsuz fikirlere sahip aktivistler de görüşlerini paylaşmak için Twitter gibi platformları kullanmaya başlamışlardır. Geleneksel iletişim kanallarıyla karşılaştırıldığında, sosyal medya aşı karşıtı mesajları yaymak ve toplulukların aşı karşıtı duyarlılık etrafında şekillenmesine olanak sağlamak için benzeri görülmemiş bir fırsat sunmaktadır [16]. Geleneksel olarak, hükümetler veya özel şirketler kamuoyunun genel tutumunu anlamak için anketleri kullanır [17]; ancak bunların tipik olarak küçük örneklem büyüklükleri, kapalı sorular ve sınırlı uzay-zamansal ayrıntı düzeyi gibi sınırlamaları bulunmaktadır. Bu sınırlamaların üstesinden gelmek için, sosyal medya verilerinin, önemli uzamsal-zamansal ayrıntı düzeyiyle, kamusal duygu ve tutumlara ilişkin daha fazla, gerçek zamanlı içgörü elde etmek için kullanılabilmesi olası olmaktadır [18]. Bu çalışmanın amacı, Türkiye'de, özellikle sosyal medya kullanıcılarının COVID-19 aşılarına ilişkin tutumunu ve endişelerini daha iyi anlamak adına Twitter üzerinden elde edilen aşı ile ilgili tweetlerin Makine Öğrenimi ile kamu duyarlılığını analiz etmektir. Bu amaç doğrultusunda çalışmada literatürde metin sınıflandırma ve duygu analizinde sık kullanın altı farklı makine öğrenmesi yöntemi kullanılmıştır. Elde edilen sonuçlar neticesinde değerlendirilen altı farklı makine öğrenmesi yöntemi içerisinde en yüksek doğruluk oranına sahip Destek Vektör Makinesi, XGBoost ve Rastgele Orman ile bir kolektif öğrenme yöntemi olan çoğunluk oylama yöntemi geliştirilmiştir. Yöntemlerden elde edilen çıktılar doğrultusunda en yüksek başarı oranına sahip olan çoğunluk oylama yöntemi % 90.5 doğruluk oranı ile Yumuşak Oylama Makine Öğrenmesi yöntemi olmuştur. Daha sonrasında ise Yumuşak Oylama yöntemi ile Twitter üzerinden elde edilen 01.04.2021 ve 31.08.2021 tarihleri arasında toplamda 153 güne ait 412.588 adet günlük tweet analiz edilerek sonuçlar raporlanmıştır. Böylelikle çalışma, hükümetlerin ve sağlık kurumlarının toplumsal sürü bağışıklığı hedeflerine ulaşılmasını etkileyebilecek kamuoyu algılarını, endişelerini ve duygularını daha iyi anlamak için COVID-19 aşılarının sosyal medyada nasıl tartışıldığını anlamaya yardımcı bir araç olacaktır.

Bu çalışmanın geri kalan bölümlerinde; 2. Bölümde COVID-19 aşıları ile ilgili duygu analizi literatürü, 3. Bölümde çalışma kapsamında kullanılan metot ve yöntem, 4. Bölümde uygulama ve bulgular, son kısımda ise sonuç bölümüne yer verilecektir.

2. Literatür (Literature Review)

COVID-19 pandemi döneminde sosyal medya paylaşımları üzerine geniş bir alanda içerik analizi, duygu analizi ve metin sınıflandırma gibi doğal dil işleme çalışmaları yapılmıştır ve bu alandaki literatür COVID-19 ve aşılama çalışmalarının devam etmesiyle birlikte genişlemeye devam etmektedir. Pandeminin başlangıcı ve ortalarında COVID-19 virüsü üzerine pek çok duygu analizi çalışması mevcut olmakla birlikte [19-22] sosyal medyada COVID-19 aşılarına yönelik yapılan paylaşımların duygu analizi nispeten daha yeni bir çalışma alanı olarak karşımıza çıkmaktadır. Farklı bölgeler veya ülkeler için aşı hakkında içerik analizi [9, 23, 24], yapay zekâ ve makine öğrenmesi [18, 25, 26] veya sözlük tabanlı [5, 15, 17] duygu analizi çalışmaları COVID-19 aşısının dünya gündemine girmesi ile birlikte artmaktadır. Hussain vd. [18] Birleşik Krallık ve Birleşik Devletler özelinde Facebook ve Twitter üzerinden COVID-19 aşıları ile ilgili elde edilen 300.000 sosyal medya postu ile gerçekleştirdikleri çalışmalarında sosyal medya kullanıcılarının duygu eğilimlerini tahmin etmek için doğal dil işleme ve derin öğrenmeye dayalı teknikleri kullanmışlardır. Genel ortalama olumlu, olumsuz ve tarafsız duygular Birleşik Krallık 'ta sırasıyla %58, %22 ve %17 iken, Birleşik Devletler 'de sırasıyla %56, %24 ve %18 olarak tespit edilmiştir. Bu sonuçları destekler nitelikte Yousefinaghani vd. [15], 4,5 milyon tweet üzerinde gerçekleştirdikleri çalışmada COVID-19 aşısıyla ilgili olumlu görüşlerin Twitter'daki baskın görüş olduğunu rapor etmiş ve olumlu tweetlerdeki ana başlıkların umut, destek ve inanç olduğunu; olumsuz tweetlerin ise genellikle korku, caydırıcılık, öfke ve siyaset ile ilgili olduğu sonucuna varmışlardır. Benzer şekilde Lyu vd. 'nin [5], 1.499.421 aşı ile ilgili tweetler üzerinde gerçekleştirdikleri sözlük tabanlı çalışma haftalık ortalama duyarlılık puanlarındaki dalgalanmalara rağmen genel olarak duyarlılığın

giderek daha olumlu olduğunu göstermiştir. Duygu analizi ayrıca güvenin en baskın duygu olduğunu ve bunu beklenti, korku, üzüntü vb. 'nin izlediğini ortaya koymuştur. Filipinler özelinde Naive Bayes sınıflandırıcı temelli gerçekleştirilen bir diğer çalışmada [26] ise literatürü destekler şekilde aşıya karşı genel tutumun %83,84'ü olumlu iken %8,26'sı olumsuz, %8,36 nötr bir şekilde polarize olmuştur. Fazel vd. [27] COVID -19 aşılarına yönelik Twitter içeriği ile aşılarla ilgili önemli bilimsel haber duyuruları arasındaki bağlantıları değerlendirmek üzere Kasım 2020'den Ocak 2021'e kadar Birleşik Krallık'taki 522.893 kişiden gelen yaklaşık 2 milyon tweeti incelemişlerdir. Çalışma da negatif aşı içeriğine sahip tweetlerin oranı, özellikle aşılar ile ilgili güncel büyük haberlerle ile birlikte aynı gün %20-24'lük bir düşüş gerçekleştirmiş olsalar da olumsuz tweetlerin oranının birkaç gün içinde ortalama %40'a geri döndü gözlemlenmiştir. Benzer kapsamda Niu vd. 'nin [28] Japonya kapsamında gerçekleştirdikleri çalışmalarında büyük ölçekli aşıların başlangıcında ve öncesinde, Japonca tweetler üzerinde COVID-19 aşısına yönelik görüş ve duyarlılığı araştırmayı amaçlamışlardır. Elde edilen sonuçlar neticesinde tweetlerin yaklaşık %85 gibi büyük bir oran ile nötr duyguları yansıtırken geri kalanların büyük bir çoğunluğun negatif duygu polaritesine sahip olduğu ortaya konmuştur. Yazarlar negatif duygu polaritesinin arkasındaki nedenin büyük ölçekli aşıların başlangıcında çok sayıda olumsuz kazanın meydana gelmesi ve aşılarak karşı etkinlik ve güvenlik problemlerinden kaynaklanabileceğini vurgulamışlardır.

Genel olarak yapılan çalışmalarda, aşılara karşı tutumun sosyal medya üzerinde olumlu bir kamuoyu oluşturduğu görülmekte olup Kwok vd. [25] İngilizce twetler üzerinde gerçekleştirdikleri çalışmalarında, tüm tweetlerin yaklaşık üçte ikisinin COVID-19 aşısı hakkında olumlu bir kamuoyu görüşü ifade ettiği; 8 temel duygudan güven ve beklentinin tweetlerde gözlenen iki belirgin olumlu duygu olduğu ve korkunun ise en olumsuz duygu olduğu ortaya konmuştur. Diğer çalışmalardan farklı olarak farklı aşı firma ve türlerine karşı tutumları ölçmek amacı ile Marcec ve Likic [17] 701.891 tweet ile ve günlük duygu analizi gerçekleştirerek Pfizer ve Moderna aşılarına ilişkin görüşün, 4 ay boyunca olumlu ve istikrarlı olduğunu, aylar arasında önemli bir duygu farkı olmadığını tespit etmişlerdir. Buna karşılık, AstraZeneca/Oxford aşısına ilişkin duyarlılığın zaman içinde azaldığı ve Aralık ile Mart ayı karşılaştırıldığında önemli bir düşüş olduğu görülmüştür. Benzer bir çalışma ise Nezhad ve Deihimi [29] tarafından İran'ın yerli ve ithal COVID-19 aşılarına yönelik görüşlerini karşılaştırmak amacıyla Farsça tweetler üzerinden gerçekleştirilmiştir. Çalışma sonuçlarına göre yabancı aşılara yönelik tweetlerin % 43'ü olumlu duygu polaritesine sahipken yerli aşılarak yönelik bu oran % 40 olarak tespit edilmiştir. Ayrıca yabancı aşılara yönelik paylaşılan tweetlerin sadece % 10'luk bir kısmı nötr tweetler içerirken yerli aşılara yönelik tweetlerin % 20'lik bir kısmının nötr duygu polaritesine sahip olduğu sonucuna varılmıştır. Çalışmada, yerli aşılara yönelik olumlu tepkilerin daha az olma sebebi, yerli aşıya yönelik zorunlu aşılamanın olumsuz düşüncelere yol açması olarak açıklanmıştır. Bu çerçevede Türkiye'de de uygulanmaya başlayan COVID-19 aşılarından sonra sosyal medya üzerinden olumlu ve olumsuz birçok içerik paylaşılmaktadır. Bilhassa son aylarda aşının etkinliğinin sorgulanması ve aşı karşıtlığı kavramlarının daha çok tartışılıyor olması kamuoyu algısının, endişelerinin ve duygularının daha iyi anlaşılmasını gerekli hale getirmektedir. Bu nedenle çalışmamız, gözden geçirdiğimiz literatür kapsamında Türkiye özelinde Covid-19 aşılarına verilen tepkiler üzerine duygu analizi gerçekleştiren ilk çalışmadır.

3. Materyal ve Metot (Materials and Method)

Günümüzde sosyal medya platformların kullanımının ve çevrimiçi alışverişin hızla yaygınlaşması, kullanıcıların yorumlanabilir ve analize açık birçok düşüncelerini elektronik mecralarda paylaşmalarına sebep olmaktadır. Günümüzde büyük miktarlardaki verinin üretildiği elektronik ortamda ortaya çıkan metin verilerinin hızla artması nedeniyle bu verilerin otomatik olarak analiz edilmesi ihtiyacı da aynı şekilde artmaktadır. Bu sebeple, duygu analizi veya fikir madenciliği kavramı, Doğal Dil İşlemenin bir alt alanı olarak karşımıza çıkmaktadır. Duygu analizi veya fikir madenciliği, bir metnin (yorum, görüş, haber vb.) tutumunun ne kadar olumlu, tarafsız veya olumsuz olduğunu değerlendirmek için çeşitli Makine Öğrenimi algoritmaları veya Derin Öğrenme ağları kullanan bir sınıflandırma işlemi olarak ifade edilebilmektedir. Temelde bir sınıflandırma süreci olan Duygu Analizi yaklaşımı elektronik formdaki metin belgelerinin yaygınlaşması ve artması ile birlikte metin belgelerinin içeriğini analiz etmek için otomatik yöntemlerin kullanılmasının önemini artırmaktadır; çünkü yeni metin belgelerini belirlemek ve bunları iyi tanımlanmış kategorilere ayırmak için alan uzmanlarını kullanan yöntem zaman alıcı ve pahalı olduğu gibi sınırları da bulunmaktadır [30]. Sonuç olarak, metin belgelerinin içeriklerine göre tanımlanması ve sınıflandırılması zorunlu hale gelmektedir. Sınıflandırıcının temel amacı, ayrı metin belgeleri kategorileri için nispeten sabit kalan bir dizi özellik elde etmek ve çok sayıda metin belgesini birden çok ilgili metin belgesi içeren belirli kategorilere (veya klasörlere) sınıflandırmaktır [31]. Ayrıca duygu analizi çeşitli sosyal medya içeriklerinin metinsel tabanlı olarak duygu durumlarına göre sınıflandırılmasını sağlamaktadır [32]. Bu amaçla bu çalışma kapsamında Doğal Dil İşleme alanının sınıflandırma görevinin bir alt dalı olan duygu analizi literatüründe sıklıkla başvurulan makine öğrenmesi yöntemleri tercih edilmiştir.

3.1. Veri (Data)

Çalışmada kullanın metin verilerinin tümü Twitter üzerinden elde edilirken günlük tam doz aşısını tamamlayan kişi sayı [33] Our World In Data organizasyonu tarafından sağlanan verilerden elde edilmiştir. Bu amaç doğrultusunda *'aşı'* ve *'covid-19'* başlık etiketleri ile Türkiye'de aşı çalışmalarının hız kazandığı 01.04.2021 ve 31.08.2021 tarihleri arasında günlük olarak toplamda 412.588 adet Türkçe tweet çekilmiştir. Verilerin toplanmasından sonraki ilk adım, veri ön işleme aşamasıdır. Öncelikli olarak çalışmanın amacı doğrultusunda metin veri haricinde tüm alanlar elde edilen veri setinden kaldırılmıştır. Daha sonra, tüm büyük harfler küçük harflere dönüştürülmüş ve sayı karakterleri ise ya metin olarak dönüştürülmüş ya da kaldırılmıştır. Ayrıca Türkçe'de yer alan durak kelimeler ve tüm URL ve E-posta ile ilgili kelimeler, gürültülü kelimeler, fazla satırlar, fazla boşluklar ve noktalama işaretleri de tüm tweetlerden silinmiştir. Tüm bunlara ek olarak tweetlerin içerisinde anlam ifade etmeyen link veya benzeri öbeklerin kaldırılması amacı ile 25 karakterden daha uzun olan tüm kelimeler de kaldırılmıştır. Türkçe dilinin sondan eklemeli bir dil olması doğal dil işleme alanında birçok zorluğa yol açmakla birlikte özellikle çalışmamızda olduğu gibi sosyal medya mecralarında kullanılan dilin birçok yazım yanlışına sahip olması geliştirilecek olan model performanslarını olumsuz yönde etkileyebilmektedir. Bu nedenle hem kelime köklerinin tespit edilmesinde hem de yazım yanlışlarının düzeltilerek model performanslarının artırılması amacı ile Akın ve Akın [34] tarafından Türkçe diller için geliştirilen bir açık kaynak doğal dil işleme kütüphanesi olan Zemberek kullanılmıştır.

Veri ön işleme süreci tamamlandıktan sonra makine öğrenmesi algoritmalarında kullanmak üzere eğitim veri setinin hazırlanması gerekmektedir. Bu nedenle her ayın ilk gününe ait kayıtlar her ayın farklı bir gündemi olabilmesi nedeniyle eğitim veri seti olarak kullanılmıştır. Böylelikle her ay Twitter üzerindeki Covid-19 aşılarına yönelik gündemin birbirinden farklı olması nedeniyle oluşabilecek model performansındaki olumsuz sonuçların önüne geçilmesi amaçlanmıştır. Elde edilen veri seti incelendiğinde Nisan ayında *"Aşının yetersizliği", "Randevu Alma Sorunu", "Hangi tür aşının seçilmesi*"; Mayıs ayında *"Aşının devlet tarafından sağlanma*

durumu"; Haziran *ayında "Aşı öncelik sıralaması"*; Temmuz ayında *"Aşı karşıtlığı"* ve *"Aşının yan etkileri"*; Ağustos ayında ise *"Aşı etkinliği"* gibi konular COVID-19 aşıları kapsamında gündemi belirlemişlerdir. Daha sonrasında hey ayın ilk gününe ait tweetler manuel olarak değerlendirmeye tabi tutularak her bir tweete tek tek olumlu, olumsuz ve nötr görüşlere göre etiket atamaları yapılmıştır. Toplamda 10.118 tweet sınıflandırılmıştır; fakat tweetlerin çok yüksek yüzdesinin nötr tweet içermesi nedeni ile performans ölçümlerinde yanlılığı önlemek amacıyla maksimum sayıda pozitif veya negatif tweet içerecek şekilde eğitim veri seti 1.047 nötr, 1.041 olumlu ve 1.047 adet olumsuz tweet olacak şekilde düzenlenmiştir. Ayrıca Tablo 1'den de anlaşılacağı üzere COVID-19 aşısı üzerine genel bir duygu bildirmeyen tweetler nötr olarak etiketlenirken, aşıya karşı olumsuz, şüpheci ve karşıt duygular olumsuz; aşının toplum sağlığı için ne kadar önemli olduğuna ve gerekli olduğuna yönelik tweetler ise olumlu olarak etiketlenmiştir.

Çalışma kapsamında kelimelerin vektörize edilmesi amacı ile doğal dil işleme (NLP) alanında çok popüler olan TF-IDF (Term Frequency-Inverse Document Frequency) yaklaşımı benimsenmiştir. TF-IDF yöntemi, belirli bir belgedeki kelimelerin göreceli sıklığını, kelimenin tüm belge bütünü üzerinde ters orantılı olarak belirlemektedir. Daha basit bir ifade ile TF-IDF, belirli bir belgedeki kelimelerin göreceli sıklığını, o kelimenin tüm belge bütünü üzerindeki ters oranına kıyasla belirleyerek çalışmaktadır. Sezgisel olarak, bu hesaplama belirli bir kelimenin belirli bir belgede ne kadar alakalı olduğunu belirlemede yardımcı olmaktadır [30]. Bu amaç doğrultusunda, yöntem iki öğe kullanır: *TF,* j belgesindeki i teriminin sıklığı ve *IDF,* - i teriminin ters belge sıklığıdır. [35]. TF-IDF'yi, Eş. 1'de de sunulduğu üzere formüle etmek mümkün olacaktır:

$$tfidf(w) = tf * \log\frac{N}{df(w)} \quad (1)$$

burada, $tf(w)$, terim sıklığıdır (bir belgede geçen ilgili kelimenin kelime sayısı), $df(w)$, belge sıklığıdır (ilgili kelimeyi içeren belge sayısı), *N*, tüm belge veya kayıt sayısıdır, $tfidf(w)$ ise vektördeki özelliğin göreli ağırlığıdır.

3.2. Yöntem (Method)

Metin verileri üzerinde duygu analizi, makine öğrenmesi alanında metin verileri ile bir sınıflandırma görevi olarak tanımlanmaktadır. Bu amaç ile bu çalışmada da sınıflandırma görevinde sıklıkla kullanılan ve literatürde mevcut çalışmalarda başarılı sonuçlar vermiş olan altı farklı makine öğrenmesi yaklaşımı kullanılmıştır. Ayrıca bu çalışmanın en temel amaçlarından birisi olarak sınıflandırma performansını artırmak amacıyla bu altı yöntem ile bir topluluk öğrenme yaklaşımı olan çoğunluk oylama mimarisi geliştirilmiştir.

3.2.1. Naive bayes (Naive bayes)

Naive Bayes, çok uzun zamandır kullanılan istatistik temelli popüler makine öğrenme yöntemlerinden birisi olmuştur. Özellikle basit kullanım yapısı çeşitli görevler için yöntemi çekici kılmakla birlikte birçok problem için de makul performanslar sunmaktadır. Alanda pek çok araştırmacı tarafından [36, 37, 38] metin sınıflandırma ve duygu analizi için de tercih edilen Naive Bayes yöntemi, Bayes olasılık teoremine dayanan bir sınıflandırıcıdır. Sınıfın bağlamı, yani bir bağımsızlık varsayımı göz önüne alındığında, örneklerin tüm niteliklerinin (yani özelliklerin) birbirinden bağımsız olduğunu varsaymaktadır. Naive Bayes bağımsızlık varsayımına rağmen birçok alanda şaşırtıcı derecede iyi performans ortaya koymaktadır [38]. Metin sınıflandırma bağlamında, d_j belgesinin bir c sınıfına ait olma olasılığı Bayes teoremi ile Eş. 2'de sunulduğu üzere hesaplanabilmektedir.

$$p(c|d_j) = \frac{p(d_j|c)p(c)}{p(dj)} \quad (2)$$

3.2.2. Destek vektör makinesi (Support vector machine)

Destek Vektör Makinesi, doğrusal veya doğrusal olmayan sınıflandırma, regresyon ve hatta aykırı değer tespiti yapabilen çok güçlü ve çok yönlü bir makine öğrenimi modelidir. Destek Vektör Makineleri özellikle karmaşık ancak küçük veya orta ölçekli veri kümelerinin sınıflandırılması için çok uygundur [39]. Destek vektör makinesi kullanılan uzayın boyutuna göre doğrusal, düzlemsel ve hiper-düzlem şeklindeki ayırma mekanizmaları kullanarak veriyi hem iki hem de daha çok sınıfa ayırmada kullanılabilen bir sınıflandırma yöntemidir [40]. Vapnik [41] tarafından önerilen Destek Vektör Makineleri, istatistiksel öğrenme teorisine ait olan yapısal risk minimizasyonu teorisinden türetilmiştir. Destek Vektör Makinesinin temel fikri, girdi vektörlerini yüksek boyutlu bir öznitelik uzayına eşlemek ve bu uzayda optimal ayırma hiperdüzlemini oluşturmaktır. Destek Vektör Makinesi, ayırıcı hiperdüzlem ve veriler arasındaki marjı maksimize ederek genelleme hatasının üst sınırını en aza indirmeyi amaçlamaktadır [42]. Birçok duygu analizi ve metin sınıflandırma görevinde [42, 43, 44] başarı ile kullanılan Destek Vektör Makinesi, orijinal eğitim verilerini daha yüksek bir boyuta dönüştürmek için doğrusal olmayan bir eşleme kullanmaktadır. Bu yeni boyut içinde, doğrusal optimal ayırıcı hiperdüzlemi (yani, bir sınıfın demetlerini diğerinden ayıran bir "karar sınırı") aramakta ve yeterince yüksek bir boyuta uygun bir doğrusal olmayan eşleme ile, iki sınıftan gelen veriler her zaman bir hiperdüzlem ile ayrılabilmektedir [45].

3.2.3. Lojistik regresyon (Logistic regression)

Lojistik Regresyon, girdi alarak ve girdi değerini ağırlık değeriyle çarparak çalışan birçok makine öğrenme yönteminden biridir. Farklı olası sınıfları ayırt etmek için girdiden hangi özelliklerin en yararlı olduğunu öğrenen bir sınıflandırıcıdır [46]. Bu model bir değişken vektörünü kavrar ve her bir girdi değişkeni için katsayıları veya ağırlıkları değerlendirir ve ardından belirtilen tweetin sınıfını bir kelime vektörü olarak tahmin etmektedir [47]. Lojistik Regresyon, çok boyutlu uzayda eğitim veri noktalarının her birinden minimum mesafeye sahip bir grafik gösterimi olan bir fonksiyonu hesaplamak için çeşitli lojistik fonksiyonları kullanmaktadır [48].

3.2.4. K-en yakın komşu (K-nearest neighbor)

K-En Yakın Komşu, birçok durumda basit ancak etkili olan, parametrik olmayan denetimli bir sınıflandırma algoritmasıdır. K-En

Tablo 1. Etiketlenmiş tweetler (Labeled tweets)

Text	Class
Aşı olanlara maske takma zorunluluğu kalktı da bizim mi haberimiz yok turizm için açıldık evet ama niye bu kadar gevşemelerine müsaade ediyorsunuz	0
Aşı ruhsatı olmayan ve yan etkileri belli olmayan sıvıya milleti zorlamayın gün olur hesabını veremezsiniz	1
Aşı elimizdeki tek silah okulların açılmasını sağlayacak şey de ve ben okula gitmek istiyorum	2

Not: 0: Nötr, 1: Olumsuz, 2: Olumlu

Yakın Komşu sınıflandırıcı, verimli sonuçlarla etkili performansı ve basitliği nedeniyle örüntü tanıma için en popüler sınıflandırıcı olarak kabul edilmektedir [49]. Örüntü tanıma, makine öğrenimi, veri madenciliği, nesne tanıma gibi birçok alanda da kullanılan metin sınıflandırma ve duygu analizi araştırmalarında da sıklıkla kullanılmaktadır [35, 49, 50]. En yakın komşu sınıflandırıcılar, analoji yoluyla, yani belirli bir test seti ile ona benzer eğitim setlerini karşılaştırarak öğrenmeye dayanmaktadır. Eğitim seti n özniteliği ile tanımlanır ve her set, n boyutlu uzayda bir noktayı temsil etmektedir. Böylelikle, tüm eğitim gözlemleri n-boyutlu bir model uzayında saklanmaktadır [45]. Bilinmeyen bir yeni bir gözlem seti verildiğinde, bir k-en yakın komşu sınıflandırıcısı, bilinmeyen sete en yakın olan k eğitim gözlemleri için desen uzayını genellikle Eş. 3'te de görülmekte olan Öklid mesafesi gibi bir uzaklık metriği ile aramaktadır. Ardından arama sonucu ile elde edilen k adet komşu gözlem noktasına göre yeni veri setinin sınıfını belirlemektedir.

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i-} x_{2i})^2} \quad (3)$$

3.2.5. XGBoost (XGBoost)

Chen ve Guestrin [51] tarafından geliştirilen XGBoost, ağaç büyütme için yüksek etkili, ölçeklenebilir bir makine öğrenme algoritmasıdır. XGBoost, ağaç büyütme için en etkili makine öğrenme yöntemlerinden biridir ve veri bilimcileri tarafından birçok araştırma alanında son teknoloji ürünü sonuçlara ulaşmak için yaygın olarak kullanılmaktadır. Geleneksel bir ağaç öğrenmenin aksine, XGBoost yalnızca öncüllerinden bilgi almakla kalmaz, aynı zamanda önceki ağacın hatalarını azaltmak ve sonunda daha doğru sonuçlar elde etmek için karşılık gelen yapraklardaki puanları toplar. XGBoost, öğrenmeyi hızlandıran ve yüksek doğruluğu tahmin edebilen paralel ve dağıtılmış hesaplama gibi birçok başka özelliğe sahiptir [52] Bu algoritma birden fazla karar ağacından oluşur ve her bir ağacı oluşturmak için gradyan iniş yöntemi kullanılır. Tüm tek karar ağaçlarına dayalı olarak optimizasyon, hedef olarak kayıp işlevini en aza indirerek gerçekleştirilir [53] ve kayıp işlevini en aza indirmek için aşağıda verilen Eş. 4 kullanmaktadır:

$$L(f_t) = \sum_{i=1}^{n} l(y_i, \hat{y_i}^{t-1}) + \Omega(f_t) + C \quad (4)$$

burada $\Omega(f_t)$, tahmin edilen $\hat{y_i}^{t-1}$ değeri ile gerçek değer y_i arasındaki farkı açıklamak için kullanılan kayıp fonksiyonudur.

3.2.6. Rastgele orman (Random forest)

Rastgele Orman, eğitim verilerinin rastgele seçiminden yapılan budanmamış sınıflandırma veya regresyon ağaçlarından oluşan bir gruptur [54]. Rastgele Orman, genellikle torbalama yöntemiyle eğitilmiş, tipik olarak eğitim setinin boyutuna ayarlanmış maksimum numunelerle oluşturulan bir Karar Ağaçları topluluğudur [55]. Rastgele Orman, bu tür algoritmaların performansının biyoinformatik ve hesaplamalı biyoloji gibi bazı alanlarda sınıflandırma görevi için olağanüstü olması nedeniyle son yıllarda popülerlik kazanmaktadır [54]. Benzer şekilde yöntem duygu analizi alanında [56, 57, 58] da sıklıkla kullanılmakta olup başarılı uygulama sonuçları elde edilmiştir. Rastgele orman, her ağacın rastgele bir vektörün değerlerine bağlı olduğu bir ağaç tahmincileri topluluğudur. Bu vektör, ormandaki sayısız ağaç için aynı dağılımdan bağımsız olarak örneklenmektedir [54]. Bir girdi vektöründen yeni bir nesneyi sınıflandırmak için, girdi vektörü ormandaki ağaçların her birine sunulmaktadır. Her ağaç bir sınıflandırma sonucu verir ve bu durum ağacın o sınıf için "oy verdiğini" ifade etmektedir. Orman en çok oyu alan sınıflandırmayı seçerek süreci tamamlamaktadır [55].

3.2.7. Çoğunluk oylama yöntemi (Majority voting method)

Bireysel sınıflandırma modelleri, son zamanlardaki gelişmelerle birlikte genellikle daha iyi performans gösteren topluluk öğrenme sistemleri tarafından sınırları zorlanmaktadır. Bu tür sistemlerde, öncelikli olarak optimal sınıflandırıcı seti seçilimi yapılır ve daha sonra belirli bir füzyon yöntemiyle bu sınıflandırıcılar birleştirilir [59]. Oylama, birçok sınıflandırıcının sonuçlarını birleştiren bir toplama tekniğidir [60]. Bazı durumlarda her biri kendi hataları olan birkaç iyi tahmincimiz mevcuttur. Genel amaç tahmincilerin önyargılarını veya farklılıklarını azaltmak olmayıp, hepsinin aynı hataları yapmamaları umuduyla tahminlerini birleştirmektir [61]. Birçok bireysel sınıflandırıcıdan daha da iyi bir sınıflandırıcı oluşturmanın en basit yolu, her sınıflandırıcının tahminlerini toplamak ve en çok oyu alan sınıfı tahmin etmektir. Bu çoğunluk oy sınıflandırıcısı, sert oylama sınıflandırıcısı olarak adlandırılmaktadır [62]. Genellikle bu tür oylama sınıflandırıcısı topluluktaki en iyi bireysel sınıflandırıcıdan yüksek bir doğruluk elde edebilmektedir. Çalışma kapsamında kullanılan bir başka çoğunluk oylama sınıflandırıcısı ise yumuşak oylama sınıflandırıcısıdır. Yumuşak oylama sınıflandırıcısı, girdi verilerini farklı bireysel sınıflandırıcılar tarafından oluşturulan tüm tahminlerin olasılıklarına göre sınıflandırmaktadır. Yumuşak oylama topluluğu, sınıflandırma görevinde sınıf etiketleri için bireysel sınıflandırma modelleri ile tahmin edilen olasılıkların toplanmasını ve en büyük toplam olasılığa sahip sınıf etiketini tahmin etmeyi amaçlamaktadır.

4. Uygulama ve Bulgular (Application and Findings)

Çalışmada daha önce bahsedildiği üzere sosyal medya kullanıcılarının COVID-19 aşılarına ilişkin tutumunu ve endişelerini daha iyi anlamak adına Twitter üzerinde elde edilen aşı ile ilgili tweetlerin altı farklı makine öğrenimi yaklaşımı ve çoğunluk oylama sınıflandırıcıları ile kamu duyarlılığını analiz etmek amaçlanmıştır. Duygu analizi ve metin sınıflandırma literatüründe sıklıkla kullanılan ve başarılı sonuçlar veren bu yöntemlerin öncelikli olarak başarı karşılaştırmaları yapılmış ve en yüksek başarı oranına sahip üç bireysel yöntem seçilerek hem yumuşak oylama hem de sert oylama yaklaşımları ile birleştirilerek değerlendirilmiştir. Hem topluluk öğrenmesi hem de bireysel yöntemler içerinde en yüksek başarı oranına sahip yöntem ile toplum eğilimini keşfetmek adına günlük tweetler analiz edilmiştir. Bu amaç doğrultusunda manuel olarak sınıflandırılmış 3.135 adet eğitim verisinden oluşan veri setinin %75'i eğitim ve %25'i test veri seti olacak şekilde kullanılmıştır.

Kullanılan yöntemlerin bazıları herhangi bir parametreye ihtiyaç duymazken Destek Vektör Makinesi, Rasgele Orman, XGBoost ve K-En Yakın Komşu yöntemleri daha yüksek bir başarı oranı için belirli hiperparametrelere gereksinim duymaktadır. Bu nedenle makine öğrenmesi alanında hiperparametre optimizasyonunda kullanılan '*Izgara Arama* (*Grid Search)'* yaklaşımı ile olası parametre kombinasyonları denenerek en iyi hiperparametre kümesi her bir yöntem için belirlenmiştir. Bu belirleme sürecinde 10 kat çapraz doğrulama ile metrik olarak doğruluk oranları dikkate alınmıştır. Çalışmada kullanılan veri seti özelinde ızgara arama sonucu elde edilen ve algoritmaların çalışmasında kullanılan hiperparametreler aşağıda sunulan Tablo 2' de özetlenmektedir.

En yüksek başarı oranını sağlayan hiperparametrelerin belirlenmesi süreci sonrasında eğitim veri seti her bir yöntem ile koşturulmuştur. Ayrıca en yüksek başarıya sahip üç model olarak Destek Vektör Makineleri, XGBoost ve Rastgele Orman, sınıflandırma modelleri çoğunluk oyu esas alınarak yapıldığı bir topluluk yönteminde

birleştirilmiştir. Bu amaçla hem sert oylama hem de yumuşak oylama yöntemi kullanılmıştır. Çalışma kapsamında kullanılan sınıflandırma algoritmalarının test veri setinden elde edilen doğruluk oranları ve başarı metrikleri aşağıda yer alan Tablo 3'te detaylı olarak gösterilmektedir.

Tablo 2. Parametrik yöntemlerde kullanılan hiperparametreler (Hyperparameters used in parametric methods)

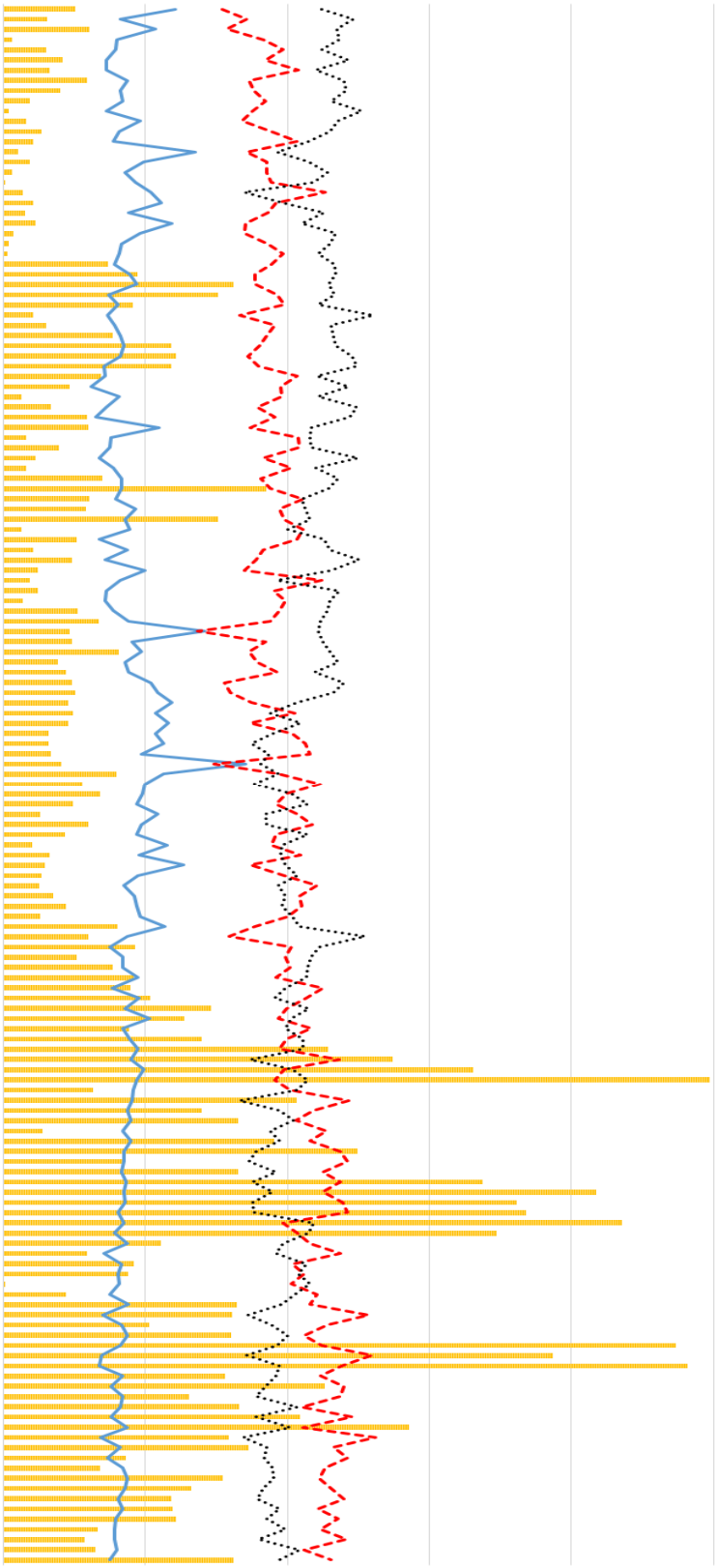
Yöntem	En uygun Hiperparametreler
K-En Yakın Komşu	k= 9
Destek Vektör Makinesi	c= 5 , gamma= 0.5 , Kernel=Linear
Rastgele Orman	N_estimater= 200, Max_depth=None
XGBoost	N_estimater= 1000, Max_depth=6

Tablo 3'te yer alan sınıf değerlerinden 0 değeri nötr, 1 değeri negatif ve 2 değeri pozitif tweetleri temsil etmektedir. Tablo 3'ten de anlaşılacağı üzere bireysel yöntemlerden Destek Vektör Makinası algoritması %89,6 ile yüksek bir başarı oranı elde etmiş olsa da %89,8'lik başarı oranı ile XGBoost modelinin en başarılı algoritma olduğu görülmektedir. Diğer makine öğrenimi yaklaşımları arasında Rastgele Orman yaklaşımı da kayda değer bir oran ile başarı elde etmiş olmasına rağmen diğer yöntemler için aynı durum çok söz konusu değildir. Bu nedenle hem sert oylama hem de yumuşak oylama çoğunluk oylama yaklaşımlarında yüksek doğruluk oranına sahip Destek Vektör Makinesi, XGBoost ve Rastgele Orman yöntemleri kullanılmıştır. Tablo 3'ten de anlaşılacağı gibi çoğunluk oylama yaklaşımları, beklendiği üzere bireysel makine öğrenmesi yaklaşımlarından daha yüksek bir başarı oranı sunmaktadır. Ayrıca her ne kadar sert oylama yöntemi %88,9 başarı oranı ile bireysel yöntemlerden daha yüksek bir doğruluk oranına ulaşsa da yumuşak oylama yöntemi %90,5 gibi göreceli olarak yüksek bir doğruluk oranı ile en başarı sınıflandırma metodu olmuştur. Bu nedenle çalışmada elde edilen günlük tweetlerin yani sınıflandırılmamış tweet'lerin etiketlenmesi aşamasında bir çoğunluk oylama yöntemi olan yumuşak oylama yaklaşımı kullanılmıştır. En yüksek doğruluk oranına sahip modelin belirlenmesinin ardından toplumun sosyal medya üzerinden aşıya yönelik duyarlılığını tahmin etmek için günlük olarak elde edilen tweetler tek tek yumuşak oylama yaklaşımı ile değerlendirilmiştir. Elde edilen sonuçlar ile Şekil 1'de görülmekte olduğu gibi günlük yapılan tam doz aşısı tamamlayan kişi sayısının yanı sıra nötr, olumlu ve olumsuz tweetlerin oransal olarak günlük ilerlemesini gösteren bir grafik oluşturulmuştur. Şekil 1 üzerinde yukarıdan aşağıya doğru 1 Nisan 2021'den başlayarak 31 Ağustos 2021'e kadar günlük tam doz aşılı kişi sayısı sarı sütunlarla, negatif tweetlerin oranı kırmızı kesikli çizgiyle, nötr tweetlerin oranı gri noktalı çizgiyle ve pozitif tweetlerin oranı mavi çizgiyle gösterilmektedir. Grafikteki zamansal eğilimler, aşılarla ilgili politikaların, kararların ve önemli olayların insanların duygularını nasıl değiştirdiğini gözlemlemeye yardımcı olan bir araç olarak kullanılabilmektedir.

Şekil 1 üzerinde de rahatlıkla görülebildiği üzere aşı ile ilgili atılan tweetlerin duygu polaritesinde günden güne dalgalanmalar mevcut olsa da negatif duygunun ve dolayısıyla aşı karşıtlığının gün geçtikçe sosyal medya üzerinde daha fazla gündem oluşturduğu gözlemlenmektedir. Analizler sonucunda ulaşılan dikkat çekici bir diğer bulgu ise tam doz aşılı kişi sayısının ve dolayısıyla aşılama oranlarının daha çok artığı Temmuz ve Ağustos aylarında aşı karşıtı düşüncelerin yoğun olduğu negatif tweetlerin oranında da artış gözlemlenmesidir. Bu durum tam doz aşılı kişi sayısı arttıkça aşı karşıtı bireylerin üzerinde daha fazla baskı hissetmiş olabileceklerini ve dolayısıyla sosyal medyada aşı karşıtı görüşlerini daha fazla gündeme getirmek istediklerini düşündürmektedir. Yine tam doz aşı sayısı arttığında pozitif tweetlerin oranının azaldığı ve istikrarlı bir görünüm kazandığı da söylenebilir. Diğer yandan, nötr ve negatif tweetlerin oranı görece daha dalgalıdır. Ayrıca aşı ile ilgi nötr paylaşımların oranının zaman içerisinde azaldığı belirgin bir şekilde görülmektedir. Bu durum aşı uygulamasının belirli bir döneminden sonra aşı ile ilgili paylaşımların aşı karşıtlığı ve aşıyı onaylama üzerine kutuplaşmaya başladığına bir kanıt olarak sunulabilir. Her ne kadar pozitif tweetler Haziran ayı itibariyle bir artış göstermiş olsalar da bu o dönemde gerçekleştirilen yüksek aşılama oranları ile açıklanabilir, genel bir artış veya azalış trendi yansıtmamaktadır. Nisan ayı itibari ile değerlendirilmeye alınan 78.078 adet tweetin

Tablo 3. Sınıflandırma yöntemlerinin performans metrikleri (Performance metrics of classification methods)

Yöntemler	Sınıf	Precision Skoru	F1 Skoru	Recall Skoru	Doğruluk Oranı
Lojistik Regresyon	0	0,881	0,842	0,902	0,874
	1	0,856	0,916	0,851	
	2	0,868	0,878	0,876	
K-En Yakın Komşu	0	0,663	0,747	0,625	0,670
	1	0,600	0,645	0,778	
	2	0,630	0,692	0,693	
Destek Vektör Makinesi	0	0,900	0,875	0,908	0,894
	1	0,888	0,916	0,879	
	2	0,893	0,894	0,893	
Naive Bayes	0	0,886	0,890	0,775	0,847
	1	0,793	0,869	0,887	
	2	0,837	0,879	0,827	
XGBoost	0	0,888	0,877	0,918	0,898
	1	0,881	0,908	0,907	
	2	0,890	0,892	0,913	
Rastgele Orman	0	0,866	0,868	0,927	0,885
	1	0,888	0,892	0,875	
	2	0,877	0,880	0,900	
Sert Oylama	0	0,880	0,888	0,935	0,899
	1	0,902	0,916	0,879	
	2	0,890	0,902	0,906	
Yumuşak Oylama	0	0,907	0,888	0,923	**0,905**
	1	0,888	0,916	0,915	
	2	0,897	0,902	0,919	



Şekil 1. Analiz sonuçlarına göre günlük tweetlerin duygu dağılımı
(Sentiment distribution of daily tweets according to analysis results)

%18,05'i pozitif, % 36,81'i negatif, %45,13'ü nötr duygu polaritesine sahiptir. Mayıs ayı itibari ile değerlendirilmeye alınan 83.229 adet tweetin %15,89'u pozitif, % 38,30'u negatif, %45,81'i nötr duygu polaritesine sahiptir. Haziran ayı itibari ile değerlendirilmeye alınan

84.085 adet tweetin %21,09'u pozitif, % 38,19'u negatif, % 40,72'si nötr duygu polaritesine sahiptir. Temmuz ayı itibari ile değerlendirilmeye alınan 84.086 adet tweetin %17,49'u pozitif, % 42,59'u negatif, %39,92'si nötr duygu polaritesine sahiptir. Ağustos ayı itibari ile değerlendirilmeye alınan 83.110 adet tweetin %15,96'sı pozitif, %45,81'i negatif, % 38,28'i nötr duygu polaritesine sahiptir. Çalışmada incelenen 5 aylık dönem boyunca atılan tüm tweetlerin %17,70'ı pozitif duygu polaritesine sahipken, nötr ve negatif duygu polaritesinde bu durum sırsıyla %41,97 ve %40,34 olarak tespit edilmiştir.

N-gram n karakterden daha uzun bir metnin n karakterli bir bölümüdür. Genel olarak, metni bir dizi örtüşen n-gram'a böler [63]. N-gram modeli veya algoritması konuşma ve dil işlemede hem ön işleme aşamasında hem de metin içeriklerinin daha iyi anlaşılmasında alanda tercih edilen önemli araçlardan birisidir. Bu çalışma kapsamında da yumuşak oylama yöntemiyle sınıflandırmaya tabi tutulan tüm tweetlerin, hem pozitif hem de negatif olarak sınıflandıranları için kelime bazında 2 gram, 3 gram ve 4 gram kullanılmıştır. Böylelikle sınıflandırılan tweetlerin içerikleri hakkında bir içgörü elde etmek mümkün olacaktır. Bu amaç doğrultusunda elde edilen en yüksek on sıklık derecesine sahip n-gram sonuçları Tablo 4'te verilmektedir. Çalışmada geliştirilen sınıflandırma modeli kapsamında pozitif olarak sınıflandırılan tweetler içerisinde *"Aşı, zorunluluk"*, *"Zorunluluk, getir"*, *"Okul, aç"* 2-gram sonuçları arasında oldukça dikkat çekicidir. Ayrıca 3-gram ve 4-gram sonuçları içerisinde ise sırasıyla *"İçin, aşı, ol"*, *"Aşı, zorunluluk, getir"* ve *"Antikor, düzey, yüksek, ol"*, *"Sağlık, için, aşı, ol"*, *"Antikor, düzey, yüksek, ol"* kelime grupları pozitif duyguları yansıtan tweetlerin genel yapısı hakkında bir izlenim elde edilmesinde faydalı bilgiler sunmaktadır. Ayrıca pozitif tweetler dâhilinde n-gram sonuçları incelendiğinde aşı olunmasına yönelik bilgilerin oldukça fazla sıklığa sahip olduğu ayrıca Sağlık Bakanı tarafından sağlanan bilgilerin veya şahsının pozitif yaklaşımı desteklemede kullanıldığı rahatlıkla gözlemlenebilmektedir. Belki de bu çalışmanın en çarpıcı kısmının oluşturan negatif olarak sınıflandırılmış tweetler içerinde de *"Aşı, olma"*, *"Aşı, karşıt"*, *"Aşı, dayatma"*, *"Aşı, öl"*, *"Kalp, kriz"* gibi 2-gram sonuçları Covid-19 aşılarına karşı olumsuz, şüpheci ve karşıt duyguların ne düzeyde olduğunu ortaya koymaktadır. Benzer şekilde *"Aşı, ol, öl"*, *"Zorla, aşı, yap"*, *"Çocuk, aşı, yapma"* ve *"Aşı, ol, virüs, bulaş"*, *"Kalp, kriz, geçir, öl"*, *"Aşı, değil, deney, sıvı"* gibi 3-gram ve 4-gram sonuçları aşı karşıtı olarak nitelendirebileceğimiz negatif tweetlerin hangi savlar ile bu durumu desteklediklerini daha belirgin olarak ortaya koymaktadır. Tablo 4 kapsamında sunulmasa da Covid-19 aşılarına yönelik yapılan nötr paylaşımlar ise Zhang vd. 'nin [24] gerçekleştirdikleri konu modellemesi kapsamında elde edilen *aşı erişimi, aşı etkinliği ve kullanıma sunulması, aşı geliştirme ve insanların görüşleri ve aşılama durumu* gibi başlıklar ile benzerlik ortaya koymaktadır.

Çalışma kapsamında elde edilen sonuçlar farklı ülkeler üzerine gerçekleştirilen duygu analizi çalışmaları ile benzerlik göstermemektedir. Hussain vd.'nin [18] Birleşik Krallık ve Birleşik Devletler kapsamında, benzer örneklem büyüklüğü ile gerçekleştirdikleri çalışmada tweetlerin %58'si pozitif duyguyu yansıtırken Türkiye özelinde bu oranın %19,60'da kalmış olması oldukça çarpıcı olduğu gibi negatif duygu yansıtan tweetlerin zaman içerisinde artış göstermesi de ayrıca değerlendirilmesi gereken başka bir konudur. Benzer durum, diğer çalışmalarda da olduğu gibi Villavicencio vd. [26] Filipinler üzerinde gerçekleştirdikleri çalışmalarında aşıya karşı %81'lik, ayrıca Nezhad ve Deihimi [29] İran kapsamında gerçekleştirdikleri çalışmalarında ise yerli aşılar için %40'lık ve yabancı aşılar içinse %43'lük pozitif tutum bulguları ile görülmektedir. Ayrıca Doğan ve Düzel 'in [6] Türkiye'de gerçekleştirdikleri çalışmalarında araştırmaya katılan bireylerin %53,62'sı aşı olmayı düşündüğünü belirtirken %14,52'inin aşı olmayı düşünmediğini belirtmiştir. Fakat bu durum bizim çalışmamız kapsamında elde edilen %17,70 pozitif ve %40,34 negatif duygu polaritesi neticesinde taban tabana bir zıtlık ortaya koymaktadır. Bu farklılık, anketlerin doğası gereği oluşabilecek küçük örneklem büyüklükleri, kapalı sorular ve sınırlı uzay-zamansal sınırlamalarında kaynaklanabileceği gibi Türkiye genelinde sosyal medya paylaşımlarında ve bilhassa Twitter paylaşımlarında bireylerin daha çok olumsuz görüş ve fikirleri paylaşmaya yönelik eğilimleri ile de açıklanabilir.

Mevcut çalışmalar içerisinde, bu çalışma ile her ne kadar birebir benzer sonuçlar mevcut olmamakla birlikte Fazel vd.'nin [27] bulgu sonuçları ile bu çalışmada elde edilen sonuçlar yaklaşık %40 olan negatif duygu polaritesi bakımından benzerlik göstermektedir. Fakat

Tablo 4. İlk 10 N-gram sonuçları (Top 10 N-gram results)

2-gram	3-gram	4-gram
Pozitif		
Aşı, ol	Doz, aşı, ol	Aşı, ol, aşı, ol
Doz, aşı	İçin, aşı, ol	Yüksek, ol, aşı, ol
Aşı, yap	Bilim, kurul, üye	Antikor, düzey, yüksek, ol
Sağlık, bakan	Sağlık, bakan, koca	Düzey, yüksek, ol, aşı
Aşı, zorunluluk	Git, aşı, ol	Bakan, koca, aşı, ol
Git, aşı	Aşı, zorunluluk, getir	Sağlık, için, aşı, ol
Bakan, koca	Aşı, randevu, al	Sağlık, bakan, koca, aşı
Okul, aç	Aşı, ol, koru	Dünya, yaygın, uygula, göre
Zorunluluk, getir	İki, doz, aşı	Toplum, ölçek, aşıla, antikor
Aşı, koruma	Acil, kullan, onay	Sağlık, çalış, önce, grup
Negatif		
Aşı, karşıt	Aşı, olma, aşı	Aşı, olma, aşı, olma
Aşı, olma	Aşı, karşıt, değil	Aşı, karşıt, değil, aşı
Olma, aşı	Aşı, ol, öl	Aşı, ol, maske, tak
Değil, aşı	Aşı, ol, virüs	Aşı, ol, virüs, bulaş
Aşı, dayatma	Zorla, aşı, yap	Aşı, olma, aşı, karşıt
Aşı, değil	Aşı, karşıt, yap	Kalp, kriz, geçir, öl
Zorla, aşı	Yoğun, bakım, yat	Aşı, konu, zorla, yan
Aşı, öl	Çocuk, aşı, yapma	Aşı, değil, deney, sıvı
Ol, öl	Aşı, karşıt, ol	Geçen, sene, aşı, yok
Kalp, krizi	Aşı, olma, reddet	Aşı, ol, kalp, kriz

her iki çalışmada negatif duygu polaritesi bakımından oldukça benzerlik göstermesine rağmen olumlu duygu polaritesi açısından benzerlik göstermemekle birlikte oldukça farklı sonuçlar ortaya koymaktadır. Her ne kadar Niu vd.'nin [28] çalışmalarında negatif ve olumlu tweet dağılımına yönelik net bir oran verilmemiş olsa da özellikle negatif duygu dağılımının olumlu tweetlerden çok daha yüksek bir orana sahip olması açısından bu çalışma sonuçları ile benzer sonuçlar ortaya koyduğu görülmektedir. Ansari ve Khan'ın [64] 74 farklı ülke için TextBlob kütüphanesi vasıtasıyla her ülke için sınırlı sayıda örneklem büyüklüğü ve sadece İngilizce tweetler ile gerçekleştirdikleri çalışmada Türkiye için elde edilen bulgular bizim çalışma bulgularımızı destekler nitelikte olup twitter üzerinde paylaşılan içeriklerde COVID-19 aşılarına yönelik negatif duygu polaritesinin yoğun olduğunu göstermektedir. Bir diğer üzerinde durulması gereken elzem husus ise geliştirilen modellerin, performanslarının literatürde COVID-19 aşılara yönelik yapılan çalışmalar içerisinde hem sözlük tabanı hem de makine öğrenmesi tabanlı çalışmalar [17, 26, 65] çerçevesinde oldukça tatminkâr ve daha yüksek veya yarışabilir başarı oranlarına ulaşmış olmasıdır. Ayrıca COVID-19 aşılarına karşı duygu analizi gerçekleştiren çalışmaların [5, 15, 18, 23, 25], az sayıdaki istisna hariç, birçoğu manuel bir tweet değerleme süreci gerçekleştirmemiş olup sözlük tabanlı veya hazır doğal dil işleme yapıları kullanmaları nedeniyle de elde edilen sınıflandırma sonuçlarının performanslarına dair herhangi bir bulgu sonucu paylaşmamaktadır. Bu nedenle çalışmamız sonucunda elde edilen bulgu sonuçları diğer çalışmalardan farklı olarak emek yoğun bir süreç sonucunda manuel olarak değerlendirilen bir veri seti ile sınanmıştır.

Çalışmamızın sonuçları doğrultusunda Türkiye'de paylaşılan COVID-19 aşısı ile ilgi sosyal medya içeriklerinin yüksek oranda negatif duyguları yansıttığı görülmektedir. Bu nedenle bilhassa literatürde yer alan diğer çalışmaların bulguları göz önüne alındığında, bu çalışma Türkiye'de sosyal medya platformları üzerinde gerçekleştirilecek analizlerin, aşı kabulünü güçlendirmek ve aşı tereddüt ve karşıtlığını azaltmak ve halk sağlığının teşviki için değerli bir araç olabileceğini göstermektedir.

5. Sonuçlar (Conclusions)

COVID-19 bir pandemi olarak kabul edilmesi ile birlikte, desteklenmesi gereken aşı tanıtımı için COVID-19 aşısı konusundaki tartışmalara ilişkin kamuoyu ve duyarlılığın araştırılması ve özetlenmesi gerekmektedir. Bu çerçevede çalışmanın amacı, Türkiye'de, özellikle sosyal medya kullanıcılarının COVID-19 aşılarına ilişkin tutumunu ve endişelerini daha iyi anlamak adına Twitter üzerinden elde edilen aşı ile ilgili tweetlerin makine öğrenimi ile kamu duyarlılığını analiz etmektir. Bu tür bir içgörü, COVID-19 aşılarının alımını artırmayı ve aşı tereddütleriyle mücadele etmeyi amaçlayan sağlık hizmetleri müdahalelerinin planlanması ve uygulanmasına olanak sağlamada değerli olacak [17] ve aynı zamanda bu tür müdahalelerin potansiyel etkisinin tahmin edilmesine de hizmet edebilecektir. Bu doğrultuda bu çalışmada metin sınıflandırma ve duygu analizi literatüründe sıklıkla başvurulan ve yöntem kısmında da detayları açıklanan altı farklı makine öğrenmesi algoritması karşılaştırılmıştır. Elde edilen sonuçlar neticesinde değerlendirilen altı farklı makine öğrenmesi yöntemi içerisinde en yüksek doğruluk oranına sahip Destek Vektör Makinesi, XGBoost ve Rastgele Orman ile bir kolektif öğrenme yöntemi olan çoğunluk oylama yöntemi geliştirilmiştir. Bu kapsamda çoğunluk oylama yöntemlerinden yumuşak oylama ve sert oylama yöntemleri oluşturularak karşılaştırılmıştır. Gerçekleştirilen analizler sonucunda Twitter'dan elde edilen verilerin duygu analizi için her ne kadar sert oylama yöntemi %88,9 doğruluk oranı ile bireysel yöntemlerden daha yüksek bir doğruluk oranına sahip olsa da yumuşak oylama yöntemi %90,5 gibi yüksek bir doğruluk oranı ile en başarılı yöntem olmuştur. Yumuşak oylama yöntemi ile gerçekleştirilen analizler sonucunda aşı ile ilgili atılan tweetlerin duygu polaritesinde günden güne dalgalanmalar mevcut olsa da negatif duygunun ve dolayısıyla aşı karşıtlığının gün geçtikçe sosyal medya üzerinde daha fazla gündem oluşturduğu gözlemlenmektedir. Bilhassa önceki çalışma bulguları ile karşılaştırıldığında aşı ile ilgili paylaşımlarda pozitif duyguların oldukça az bir orana sahip olması, Türkiye'de sosyal medya üzerinden süren aşı karşıtlığını daha dikkatli incelenmesi gereken bir konu haline getirmektedir. Ayrıca bulgularımız neticesinde özellikle COVID-19 aşılarına karşıtlık gösteren tweetlerin kelime analizlerinde COVID-19 aşılarına karşı yüksek bir güvensizlik ve bilimsel desteği olmayan dezenformasyon olduğu görülmektedir. Bu çalışma bildiğimiz kadarıyla Türkiye'de COVID-19 aşılarına yönelik duygu analizi gerçekleştiren ilk çalışma olmakla birlikte elde edilen sonuçlar neticesinde, sosyal medya üzerinden duygu analizi yaklaşımı ile COVID-19 aşılarına ilişkin duyarlılığı izlemenin değerli ve kolayca uygulanan bir araç olduğunu da göstermektedir. Ayrıca çalışma Türkiye'de sosyal medya platformları üzerinde gerçekleştirilecek analizlerin, aşı kabulünü güçlendirmek ve aşı tereddüt ve karşıtlığını azaltmak ve halk sağlığının teşviki için değerli bir araç olabileceğini de göstermektedir. Bu nedenlerle özellikle toplumsal bağışıklığın kazanımı noktasında toplumun yüksek oranda aşılanma ihtiyacı düşünüldüğünde kamuoyu algılarını, endişelerini ve duygularını daha iyi anlamak politika yapıcılar ve sağlık süreci yöneticileri açısından önem arz etmektedir.

Kaynaklar (References)

1. Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., Zhao, Y., Li, Y., Wang, X. and Peng, Z., Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China, Jama, 323 (11), 1061-1069, 2020.
2. Zheng, Y. Y., Ma, Y. T., Zhang, J. Y. and Xie, X., COVID-19 and the cardiovascular system, Nature Reviews Cardiology, 17 (5), 259-260, 2020.
3. Machingaidze, S., & Wiysonge, C. S., Understanding COVID-19 vaccine hesitancy, Nature Medicine, 27 (8), 1338-1339, 2021.
4. Horder, J., Toll of vaccine hesitancy, Nature human behaviour, 4 (4), 335-335, 2020.
5. Lyu, J. C., Le Han, E., & Luli, G. K., COVID-19 vaccine–related discussion on Twitter: topic modeling and sentiment analysis, Journal of medical Internet research, 23 (6), e24435, 2021.
6. Doğan, M. M., & Düzel, B., Fear-anxiety levels specific to Covid-19, Electronic Turkish Studies, 15 (4), 739-752, 2020.
7. Kadkhoda, K., Herd Immunity to COVID-19: Alluring and Elusive, American Journal of Clinical Pathology, 155 (4), 471–472, 2021.
8. Hussain, A., Ali, S., Ahmed, M., & Hussain, S., The anti-vaccination movement: a regression in modern medicine, Cureus, 10 (7), 2018.
9. Bonnevie, E., Gallegos-Jeffrey, A., Goldbarg, J., Byrd, B., & Smyser, J., Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic, Journal of communication in healthcare, 14 (1), 12-19, 2021.
10. Dean, B., Social network usage & growth statistics: How many people use social media in 2021, Published August, 12, 2020.
11. Öztürk, N., & Ayvaz, S., Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis, Telematics and Informatics, 35 (1), 136-147, 2018.
12. Fung, I. C. H., Fu, K. W., Ying, Y., Schaible, B., Hao, Y., Chan, C. H. and Tse, Z. T. H., Chinese social media reaction to the MERS-CoV and avian influenza A (H7N9) outbreaks, Infectious diseases of poverty, 2 (1), 31, 2013.
13. Chew, C. and Eysenbach, G., Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak, PloS one, 5 (11), e14118, 2010.
14. Noor, S., Guo, Y., Shah, S. H. H., Fournier-Viger, P., & Nawaz, M. S., Analysis of public reactions to the novel Coronavirus (COVID-19) outbreak on Twitter, Kybernetes, 2020.
15. Yousefinaghani, S., Dara, R., Mubareka, S., Papadopoulos, A., & Sharif, S., An Analysis of COVID-19 Vaccine Sentiments and Opinions on Twitter, International Journal of Infectious Diseases, 108, 256-262, 2021.

16. Muric, G., Wu, Y., & Ferrara, E., COVID-19 Vaccine Hesitancy on Social Media: Building a Public Twitter Dataset of Anti-vaccine Content, Vaccine Misinformation and Conspiracies, arXiv preprint arXiv:2105.05134, 2021.

17. Marcec, R., & Likic, R., Using Twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. Postgraduate Medical Journal, Published Online First: 09 August 2021, 2021.

18. Hussain, A., Tahir, A., Hussain, Z., Sheikh, Z., Gogate, M., Dashtipour, K., ... & Sheikh, A., Artificial intelligence–enabled analysis of public attitudes on facebook and twitter toward covid-19 vaccines in the united kingdom and the united states: Observational study, Journal of medical Internet research, 23 (4), e26627, 2021.

19. Dubey, A. D., Twitter Sentiment Analysis during COVID-19 Outbreak, Available at SSRN 3572023, 2020.

20. Bhat, M., Qadri, M., Noor-ul-Asrar Beg, M. K., Ahanger, N., & Agarwal, B., Sentiment analysis of social media response on the Covid19 outbreak, Brain, Behavior, and Immunity, 87, 136, 2020.

21. Manguri, K. H., Ramadhan, R. N., & Amin, P. R. M., Twitter sentiment analysis on worldwide COVID-19 outbreaks, Kurdistan Journal of Applied Research, 54-65, 2020.

22. Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., & Choi, G. S., A performance comparison of supervised machine learning models for Covid-19 tweets sentiment analysis, Plos one, 16 (2), e0245909, 2021.

23. Thelwall, M., Kousha, K., & Thelwall, S., Covid-19 vaccine hesitancy on English-language Twitter, Profesional de la información (EPI), 30 (2), 1-13, 2021.

24. Zhang, J., Wang, Y., Shi, M., & Wang, X., Factors Driving the Popularity and Virality of COVID-19 Vaccine Discourse on Twitter: Text Mining and Data Visualization Study, JMIR Public Health and Surveillance, 7 (12), 1-13, 2021.

25. Kwok, S. W. H., Vadde, S. K., & Wang, G., Tweet topics and sentiments relating to COVID-19 vaccination among Australian Twitter users: Machine learning analysis, Journal of medical Internet research, 23 (5), e26953, 2021.

26. Villavicencio, C., Macrohon, J. J., Inbaraj, X. A., Jeng, J. H., & Hsieh, J. G., Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes, Information, 12 (5), 204, 1-16, 2021.

27. Fazel, S., Zhang, L., Javid, B., Brikell, I., & Chang, Z., Harnessing Twitter data to survey public attention and attitudes towards COVID-19 vaccines in the UK. Scientific reports, 11 (1), 1-5, 2021.

28. Niu, Q., Liu, J., Nagai-Tanima, M., Aoyama, T., Masaya, K., Shinohara, Y., & Matsumura, N., Public Opinion and Sentiment Before and at the Beginning of COVID-19 Vaccinations in Japan: Twitter Analysis, medRxiv, 2021.

29. Nezhad, Z. B., & Deihimi, M. A., Twitter sentiment analysis from Iran about COVID 19 vaccine, Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 16 (1), 1-5, 2022.

30. De Vel, O., Mining e-mail authorship, In Proc. Workshop on Text Mining, ACM International Conference on Knowledge Discovery and Data Mining (KDD'2000), Boston Massachusetts-USA, August, 2000.

31. Yun-tao, Z., Ling, G., & Yong-cheng, W., An improved TF-IDF approach for text classification, Journal of Zhejiang University-Science A, 6 (1), 49-55, 2005.

32. Güran A., Ateş E., Pearson correlation and Granger causality analysis of Twitter sentiments and the daily changes in Bist30 index returns, Journal of the Faculty of Engineering And Architecture of Gazi University, 36 (3), 1687-1702, 2021.

33. Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., ... & Roser, M., Coronavirus pandemic (COVID-19), Our World in Data, 2020.

34. Akın, M. D., & Akın, A. A., An Open Source Natural Language Processing Library for Turkic Languages: Zemberek, Electrical Engineering, 431, 38-44, 2007.

35. Trstenjak, B., Mikac, S., & Donko, D., KNN with TF-IDF based framework for text categorization, Procedia Engineering, 69, 1356-1364, 2014.

36. McCallum, A., & Nigam, K, A comparison of event models for naive bayes text classification, In AAAI-98 workshop on learning for text categorization, 752 (1), 41-48, July, 1998.

37. Frank, E., & Bouckaert, R. R., Naive bayes for text classification with unbalanced classes, In European Conference on Principles of Data Mining and Knowledge Discovery, Springer, Berlin-Germany, 503-510, September, 2006.

38. Kim, S. B., Han, K. S., Rim, H. C., & Myaeng, S. H., Some effective techniques for naive bayes text classification, IEEE transactions on knowledge and data engineering, 18 (11), 1457-1466, 2006.

39. Géron, A., Hands-on machine learning with scikit-learn and tensorflow: Concepts. Tools, and Techniques to build intelligent systems, 2017.

40. Dönmez İ., Aslan Z., Document Sentiment classification using hybrid wavelet methodologies, Journal of the Faculty of Engineering And Architecture of Gazi University, 36 (2), 701-714, 2021.

41. Vapnik, V., The nature of statistical learning theory, Springer science & business media, 2013.

42. Lin, Y., & Wang, J., Research on text classification based on SVM-KNN, In 2014 IEEE 5th International Conference on Software Engineering and Service Science, IEEE, Beijing- China, 842-844, June, 2014

43. Huq, M. R., Ali, A., & Rahman, A., Sentiment analysis on Twitter data using KNN and SVM, International Journal of Advanced Computer Science and Applications, 8 (6), 19-25, 2017.

44. Colas, F., & Brazdil, P., Comparison of SVM and some older classification algorithms in text classification tasks, In IFIP International Conference on Artificial Intelligence in Theory and Practice, Springer, 169-178, Boston-USA, August, 2006.

45. Han, J., Pei, J., & Kamber, M, Data mining: concepts and techniques, Elsevier, 2011.

46. Indra, S. T., Wikarsa, L., & Turang, R., Using logistic regression method to classify tweets into the selected topics, In 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS), IEEE, 385-390, Malang- Indonesia, October, 2016.

47. Prabhat, A., & Khullar, V., Sentiment classification on big data using Naïve Bayes and logistic regression, In 2017 International Conference on Computer Communication and Informatics (ICCCI), IEEE ,1-5, Coimbatore- India, January, 2017

48. Salazar, D. A., Vélez, J. I., & Salazar, J. C., Comparison between SVM and logistic regression: Which one is better to discriminate?, Revista Colombiana de Estadística, 35 (SPE2), 223-237, 2012.

49. Hota, S., & Pathak, S., KNN classifier based approach for multi-class sentiment analysis of twitter data, International Journal of Engineering & Technology, 7 (3), 1372-1375, 2018.

50. Bilal, M., Israr, H., Shahid, M., & Khan, A., Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques, Journal of King Saud University-Computer and Information Sciences, 28 (3), 330-344, 2016.

51. Chen, T. ve Guestrin, C., XGBoost: "A Scalable Tree Boosting System", Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery And Data Mining, 785-794, San Francisco California-USA, August, 2016

52. Zhao, Y., Chetty, G., & Tran, D, Deep Learning with XGBoost for Real Estate Appraisal, In 2019 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 1396-1401, Xiamen- China, December, 2019

53. Liang, Y., Wu, J., Wang, W., Cao, Y., Zhong, B., Chen, Z., & Li, Z., Product marketing prediction based on XGboost and LightGBM algorithm, In Proceedings of the 2nd International Conference on Artificial Intelligence and Pattern Recognition, 150-153, Beijing-China, August, 2019

54. Breiman, L., Random forests, Machine learning, 45 (1), 5-32, 2001.

55. Ho, T. K., Random decision forests, In Proceedings of 3rd International Conference on Document Analysis And Recognition, IEEE, 278-282, Montreal, Canada, August, 1995

56. Fauzi, M. A., Random Forest Approach for Sentiment Analysis in Indonesian, Indonesian Journal of Electrical Engineering and Computer Science, 12 (1), 46-50, 2018

57. Gupte, A., Joshi, S., Gadgul, P., Kadam, A., & Gupte, A., Comparative study of classification algorithms used in sentiment analysis, International Journal of Computer Science and Information Technologies, 5 (5), 6261-6264, 2014.

58. Da Silva, N.F., Hruschka, E. R., & Hruschka Jr, E.R., Tweet sentiment analysis with classifier ensembles, Decision Support Systems, 66, 170-179, 2014

59. Ruta, D., & Gabrys, B., Classifier selection for majority voting, Information fusion, 6 (1), 63-81, 2005

60. Gandhi, I., & Pandey, M., Hybrid ensemble of classifiers using voting, In 2015 international conference on green computing and Internet of Things (ICGCIoT), IEEE, 399-404, Greater Noida-India, October, 2015.

61. Amr, T., Hands-On Machine Learning with scikit-learn and Scientific Python Toolkits: A practical guide to implementing supervised and unsupervised machine learning algorithms in Python, Packt Publishing, Limited, 2020.
62. Géron, A., Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems, O'Reilly Media, 2019.
63. Cavnar, W. B., & Trenkle, J. M., N-gram-based text categorization, In Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval, Las Vegas-USA, April, 1994
64. Ansari, M. T. J., & Khan, N. A., Worldwide COVID-19 Vaccines Sentiment Analysis Through Twitter Content. Electronic Journal of General Medicine, 18 (6), 1-10, 2021.
65. Nwafor, E., Vaughan, R., & Kolimago, C., Covid Vaccine Sentiment Analysis by Geographic Region, In 2021 IEEE International Conference on Big Data, IEEE, 4401-4404, Jeju Island-Korea, December, 2021.