



COVID-19 Aşıları için Türkçe Tweetlerle Duygu Analizi

*Sentiment Analysis with Turkish Tweets for COVID-19 Vaccines*Deniztan Ulutaş Karakol^{1*}, Çetin Cömert²¹ Karadeniz Teknik Üniversitesi, Harita Mühendisliği Bölümü, deniztanulutas@ktu.edu.tr
ORCID: <https://orcid.org/0000-0002-2131-1057>² Karadeniz Teknik Üniversitesi, Harita Mühendisliği Bölümü, ccomert@ktu.edu.tr
ORCID: <https://orcid.org/0000-0002-2019-6990>

MAKALE BİLGİLERİ

Makale Geçmişi:

Geliş 11 Eylül 2023
Revizyon 3 Aralık 2023
Kabul 7 Aralık 2023
Online 31 Aralık 2023

Anahtar Kelimeler:

Metin madenciliği, Tweet analizi,
Duygu analizi, COVID-19 aşıları,
Pandemi

ÖZ

Zamanın ve teknolojinin gereksinimleri ile "Web" anlamsal ve akıllı teknolojilerle dönüşüme uğramış ve kullanıcıların aktif olarak içerik yaratabildiği, herhangi bir konuda görüşlerini ifade edebildiği çok yönlü sanal interaktif bir platform haline gelmiştir. Bu platformlardan biri olan sosyal medya sayesinde bilgi ve fikirlerin yayılması kolaylaşmıştır. Günümüzde sosyal medya platformlarında üretilen veriler metin madenciliği yöntemleri sayesinde basit, hızlı ve etkili bir şekilde analiz edilebilmektedir. Analiz sonuçları satış tahminleri, pazarlama, çevre, sağlık, toplumun duygu ve düşüncelerini saptama gibi birçok alanda kullanılmaktadır. Bu noktadan hareketle, bu çalışmada COVID-19 global hastalığı için geliştirilen Sinovac ve Biontech aşıları için Twitter kullanıcılarının paylaştıkları tweetler üzerine zamansal duygu analizi yapılmış ve sonuçlar karşılaştırılmıştır. Böylece aşıların insanlar üzerindeki olumlu ya da olumsuz etkileri ortaya çıkarılmıştır. Sonuçta aşı olan ya da olmayan insanların aşılarına karşı duyguları ölçülmüş, aşıların olumlu ya da olumsuz etkilerinin değerlendirilmesi amaçlanmıştır. Analiz sonuçlarına göre, insanların çoğunlukla aşılarından memnun olduğu görülmektedir. Ancak nüfusun bir kısmının endişe ve korku duyması, aşıların insanlar üzerindeki etkisinin olumsuz taraflarını da yansıtmaktadır. Bu çalışma farklı aşı veya tıbbi tedavi türleri için de uyarlanıp insanlar için yol gösterici olabilir.

ARTICLE INFO

Article history:

Received 11 September 2023
Received in revised form 3 December 2023
Accepted 7 December 2023
Available online 31 December 2023Keywords: Text mining, Tweet
analysis, Sentiment analysis,
COVID-19 vaccines, Pandemic

Doi: 10.24012/dumf.1358519

* Sorumlu Yazar

ABSTRACT

With the requirements of time and technology, the "Web" has been transformed by semantic and smart technologies and has become a versatile virtual interactive platform where users can actively create content and express their opinions on any subject. Thanks to social media, which is one of these platforms, the dissemination of information and ideas has become easier. Today, data produced on social media platforms can be analyzed simply, quickly and effectively thanks to text mining methods. Analysis results are used in many areas such as sales forecasts, marketing, environment, health, and determining the feelings and thoughts of the society. From this point of view, in this study, a temporal sentiment analysis was performed on the tweets shared by Twitter users for the Sinovac and Biontech vaccines developed for the COVID-19 global disease, and the results were compared. Thus, the positive or negative effects of vaccines on humans have been revealed. As a result, the feelings of people who have or have not been vaccinated against vaccines were measured, and it was aimed to evaluate the positive or negative effects of vaccines. According to the results of the analysis, it is seen that people are mostly satisfied with the vaccines. However, the anxiety and fear of a part of the population also reflects the negative aspects of the impact of vaccines on humans. This study can be adapted for different types of vaccines or medical treatments and may guide people.

Giriş

Günümüzde internetteki sayısız gelişmeler sayesinde veri artık birçok farklı kaynaktan farklı biçimlerde üremekte ve gün geçtikçe devasa bir şekilde artmaktadır. Birçok kurumsal ve özel uygulamalar sanal platformlara taşındıkça insanlar taşınabilir bilgisayarlar, akıllı telefonlar ya da benzeri aygıtlarla zaman ve konum kısıtlaması olmadan birbiri ile iletişim kurabilmekte ve ihtiyaç duyduğu uygulamaları kullanabilmektedir. Son yıllarda sosyal medyanın toplumumuzda rolünün artması bilgi ve görüşlerin yayılmasını kolaylaştırmıştır. Instagram [1], Facebook [2] ve Twitter [3] sosyal medyada önde gelen üç büyük platformdur. Facebook ve Instagram daha çok görsel ağırlıklıyken, Twitter metin ağırlıklıdır. Twitter, insanların herhangi bir konu hakkında “tweet” olarak adlandırılan mesajlar paylaşabileceği ve duygularını yayımlayabileceği bir mikroblog sitesidir. Bu sayede Twitter, duyu yoğunluklu bir veri kümesi olarak düşünülebilir. Tipik bir tweet, hashtag'ler (etiket), referanslar (@) da dahil olmak üzere “0 – 280” karakter arasında metinden oluşur. Tweetlerle duyu analizi son yılların popüler konuları arasındadır. Şirketler kullanıcıların nabzını attıkları tweetler aracılığı ile ölçebilmekte ve markalarının genel duyarlılığını izleyebilmektedir. Tüketiciler de ürün tercihlerini yaparken duyu analizinden faydalanmaktadır.

Duyu analizi metin madenciliğinin bir türüdür. Metin madenciliği, belgelerdeki yapılandırılmamış metinsel verilerden ilginç örüntüler ve faydalı bilgiler çıkarır [4]. Duyu analizi metinsel verilerde örtülü (gizli) olan toplu görüşü ortaya çıkarmaktadır. Popüler uygulamaları belirli bir ürün, hizmet veya siyasi olaylar hakkında sosyal medyadan duyguların otomatik olarak ortaya çıkarılmasıdır. Duyu analizi, belirli bir metindeki anlamı veya duyguyu doğru bir şekilde anlamak için ortak duyu bilgisi gerektirir. Bunun için genellikle SentiTurkNet [5], SentiWordNet [6], SenticNet [7], NRC Emotion Lexicon [8], HisNet [9], SentiDraw [10] ve SentiFars [11] gibi duyu sınıflandırmasına ait kutuplu kavramları içeren sözlükler, duyu kavramlarını ve ilişkilerini içeren Emotion Ontology [12], OntoSenticNet [13] gibi ontolojiler ya da istatistiksel ve öğrenmeye dayalı yöntemler [14]-[19] kullanılır.

Literatürde tweetler üzerine metin madenciliği ve duyu analizi konusunda çeşitli çalışmalar [20]-[40] mevcuttur. Saran vd. (2019), 23 Ağustos–8 Eylül 2019 tarihleri arasındaki “Amazon yangını” ile ilgili Twitter gönderilerine duyu analizi uygulayarak Amazon ormanındaki yangın olayları için halkın ruh hali modellerini değerlendirmişlerdir. Tweetleri ve yorumları analiz etmek, metni tanımlamak ve sınıflandırmak için sözlük ve makine öğrenimi algoritmalarını kullanmışlardır. Analiz edilen İngilizce tweetlerin sonuçlarına göre 739 nötr, 503 olumlu, 258 olumsuz tweet belirlenmiştir [20]. Albayrak vd. (2017), insanların bedelli askerlik konusundaki düşüncelerini analiz etmek için konu ile ilgili belirli tarihlerde atılan Türkçe tweetler üzerinden SentiTurkNet sözlüğünü kullanarak duyu analizi gerçekleştirmişlerdir. Sonuçlara göre tweetlerin %16 oranında pozitif, %5 oranında negatif ve %79 oranında ise nötr olduğu görülmektedir [21]. Almatar vd. (2020), Kuveyt'te insanların ne zaman hangi konular hakkında düşündüklerini, düşündükleri hakkında nerede

tweet attıklarını analiz etmişlerdir. Çalışma, insanların ilgi ve düşüncelerinin bir hafta boyunca nasıl değiştiğinin izlenmesine katkıda bulunmuştur. Örneğin, insanlar hafta içi mi yoksa hafta sonları mı belirli bir konuyla daha çok ilgileniyorlardı veya bu ilgi düzeylerini etkileyen hafta içi bir olay mı vardı? Çalışmada ilgi alanlarını keşfetmek için Twitter-LDA (Latent Dirichlet Allocation) modeli kullanılmıştır. Çalışma sonucunda Arapça tweetler analiz edilmiş ve Kuveyt'te din, duyu, eğitim ve kamu politikasının en popüler ilgi alanları olduğu belirlenmiştir. İnsanların din ile ilgili tweetleri çoğunlukla Cuma günü, politika ve eğitim hakkındaki tweetleri hafta içi ve duygusal ifadeler hakkındaki tweetleri çoğunlukla hafta sonları attığı görülmüştür [22]. Song ve Xia (2016), Avustralya'daki Curtin Üniversitesi, Bentley kampüsünde coğrafi referanslı Twitter duyu polaritesinin kurumsal ve zamansal dağılımını araştırmışlardır. Çalışmada Twitter duygularının farklı bölgeler ve zaman dilimleri arasında nasıl değiştiğinin araştırılması amacıyla kampüs, fen ve mühendislik binaları, sosyal bilimler binaları, kütüphane, amfi, yurt, eğlence ve park alanları olmak üzere altı bölgeye ve yarıyıl başı, yarıyıl ortası, yarıyıl sonu ve sınav sonrası incelemeler için dört zaman dilimine ayrılmıştır. Çalışmada İngilizce tweetler analiz edilmiş ve analiz sonucunda olumlu tweetlerin en yüksek yüzdesinin sosyal bilimler alanında, olumsuz paylaşımların en yüksek yüzdesinin fen bilimleri ve mühendislik ile yurt alanlarında gerçekleştiği görülmüştür. Olumsuz tweetlerin sayısının kütüphane ve fen-mühendislik alanlarında dönem sonuna doğru artarak sınav döneminde zirveye ulaşırken, dönem sonunda eğlence, spor ve yurt alanlarında düştüğü görülmüştür [23]. Häberle vd. (2019), Los Angeles'ta coğrafi etiketli İngilizce Twitter metin mesajlarına kurumsal analiz uygulayarak binaların türlerini ticari ve konut olarak sınıflandırmışlardır. Kentsel yapı türlerini sınıflandırmak için metin özelliklerini uzaktan algılama görüntüleri ile birleştirmişlerdir. Çalışmada yapılan sınıflandırmada makine öğrenme yöntemlerinden Destek Vektör makineleri, Naive Bayes ve Evrimsel Sinir Ağları kullanılmıştır. Sınıflandırma sonuçları, saf sosyal medya metninden bina sınıflandırma işleminin oldukça zor olduğunu göstermiştir [24]. Alowaidi vd. (2017), eğitilmiş makine öğrenme yöntemlerine (Naive Bayes ve Destek Vektör makineleri) dayanan Arapça semantik Twitter duyu analizi modeli geliştirmişlerdir. Tweetleri Arapça WordNet içeriğindeki kavramlarla zenginleştirerek semantik analiz yapmışlardır. Deneysel sonuçlar, kavram özelliklerinin kullanılmasının, temel BoW (Bag Of Words) gösteriminden daha iyi performans verdiğini göstermiştir [25]. Resch vd. (2018), İngilizce tweetleri analiz ederek Napa'da 2014 yılında meydana gelen depremin sebep olduğu izleri ve hasarları semantik makine öğrenme teknikleri ile birlikte kurumsal ve zamansal analiz yaparak ortaya çıkarmışlardır. Sonuçlar deprem ayak izlerinin güvenilir ve doğru bir şekilde tanımlanabileceğini ve ilgili semantik konuların önceden bilgi gerektirmeden otomatik olarak tanımlanabileceğini göstermiştir. Çalışmada ayrıca önemli kayıpların nerede oluştuğunu gösteren bir hasar haritası oluşturulmuştur. Analiz sonuçları, Federal Acil Durum Yönetim Ajansı (FEMA) tarafından sağlanan teknik kılavuz izlenerek resmi deprem ve hasar haritalarıyla karşılaştırılarak doğrulanmıştır [26]. Tavoschi vd. (2020), Twitter kullanarak bir aş ile ilgili Eylül 2016 ve Ağustos 2017

tarihleri arasında İtalya'da meydana gelen önemli olaylar hakkında kamuoyunun görüş kutupluluğunu eğitilmiş makine öğrenme teknikleri kullanarak izlemişlerdir. Çalışmada İtalyanca tweetler kullanılmıştır. Yapılan analiz sonuçlarına göre tweetlerin yüzde 60'ı tarafsız, yüzde 23'ü aşı karşıtı, yüzde 17'si ise aşı lehinde olarak sınıflandırılmıştır [27]. Nemes ve Kiss (2021), Twitter kullanıcılarının covid/coronavirus konusunda attıkları tweetleri oluşturan kelimeler arasındaki ilişkileri baz alarak duygularını tahmin eden bir model geliştirmişlerdir. Doğal dil işleme ve duygu sınıflandırması için Tekrarlayan Sinir Ağı (Recurrent Neural Network) tekniği kullanmışlardır. Çalışmada 24-25 Nisan 2020 ve 13-14 Mayıs 2020 tarihlerindeki İngilizce tweetler kullanılmıştır. Çalışma sonucunda sosyal medyada aşılara ilgili duyguların çoğunlukla olumlu olduğu ancak zamanla olumlu duygularda artış olurken olumsuz duyguların da arttığı gözlemlenmiştir [28]. Beşkirli vd. (2021), Covid-19 aşı duyurusundan önceki ve sonraki zaman diliminde Twitter'da aşı ile ilgili atılan toplam 10000 adet İngilizce tweet üzerinde duygu analizi gerçekleştirmişlerdir. Çalışmada kullanılan ilk veri seti COVID-19 aşısının 3. faz denemeleri esnasında elde edilen twitter verilerinden oluşmaktadır. İkinci veri seti COVID-19 aşısının 3. faz denemeleri esnasında seri duyurusu yapıldıktan sonra elde edilen Twitter verilerinden oluşmaktadır. Aşının seri üretim duyurusundan önce atılan tweetlerin duygusal içeriği %31 olumlu, %29 olumsuz iken %41 nötr olarak belirlenmiştir. Ancak aşı çalışmalarının seri üretim duyurusu yapıldıktan sonra olumlu düşünceler %53'e yükselirken, olumsuz düşünceler %14'e ve nötr düşünceler %33'e gerilemiştir [29]. Çakmak ve Oğuzlar (2022), Ekim-Aralık 2020 tarihleri arasındaki "5G" ve "vaccination (aşılama)" etiketlerini birlikte kullanarak Twitter'dan elde edilen İngilizce tweetlerle COVID-19 aşılari için duygu analizi gerçekleştirmişlerdir. Çalışmada tweetlere Sınıflandırma ve Regresyon Ağaçları (CART), Naive Bayes (NB), K-En Yakın Komşu (KNN) ve Rastgele Orman (RF) algoritmaları uygulanmıştır. Analiz sonuçlarına göre dünya genelinde Twitter üzerinden fikrini beyan eden kişilerin %36,4'ünün 5G ile COVID-19 aşısı hakkında pozitif algıya sahip olduğu, %35,6'sının pozitif ya da negatif görüşe sahip olmadığı ve %28'inin de negatif görüş bildirdiği görülmüştür [30]. Aygün vd. (2021), COVID-19 aşılari hakkında Twitter kullanıcılarının yaklaşımını ölçmek için Türkiye ve bazı ülkeler için duygu analizi gerçekleştirmişlerdir. Tweetler politika, sağlık, medya ve diğer olmak üzere dört farklı sınıfa ayrılarak derin öğrenme yöntemleri ile analiz edilmiştir. Çalışmada 15 Kasım-15 Mart 2020 aralığındaki Türkçe ve İngilizce tweetler kullanılmıştır. Analiz sonuçlarına göre tweetlerde aşılari hakkında %48,5 olumlu, %31,7 nötr ve %19,8 olumsuz görüş hakimdir [31]. Nemes ve Kiss (2021), Beşkirli vd. (2021), Çakmak ve Oğuzlar (2022), Aygün vd. (2021), bu çalışma ile benzer çalışmalar yapmışlardır. Ancak COVID-19 aşılari ülkemizde uygulanmaya başlamadan önceki tarihlere ait İngilizce tweetleri kullanmaları, duygu sınıflandırması için makine öğrenme veya derin öğrenme yöntemlerini kullanmaları ile bu çalışmadan farklıdır. Bu çalışmada aşılari ülkemizde uygulanmaya başladığı tarihten itibaren atılan Türkçe tweetler kullanılarak sözlük tabanlı bir duygu analizi gerçekleştirilmiştir. Lyu vd. (2021), pandemi başlangıcından itibaren ilk 10 ay içerisinde kamuoyunun

COVID-19 aşılariya yönelik tutumlarını ve aşırı kabullenmelerini veya tereddütlerini öğrenmek için tweetleri kullanarak konu modelleme ve duygu analizi gerçekleştirmişlerdir. Çalışmada 11 Mart 2020-31 Ocak 2021 tarihleri aralığındaki İngilizce tweetler kullanılmıştır. Kanada Ulusal Araştırma Konseyi Duygu Sözlüğü kullanılarak duygu analizi yapılmış ve konu modelleme için gizli Dirichlet tahsisi kullanılmıştır. Konu modelleme sonucunda 16 konu belirlenmiştir. Aşılama ile ilgili görüşler en çok tweet atılan ve en çok tartışılan konu olarak belirlenmiştir. Rusya'nın dünyanın ilk COVID-19 aşısını onayladığı 11 Ağustos 2020 tarihinde aşı konusundaki ilerlemeler dünya çapında en çok tartışılan konu haline gelmiştir. Aşı yaptırma talimatı konusu Ocak 2021'in ilk haftasından sonra en çok tartışılan konu haline gelmiştir. Haftalık ortalama duygu puanları, duyguların olumlu yönde arttığını göstermiştir. Duygu analizi sonucu güvenin en baskın duygu olduğunu, ardından beklenti, korku, üzüntü vb. duyguların geldiğini göstermiştir. Güven duygusunun, Pfizer'in aşısının %90 etkili olduğunu açıkladığı 9 Kasım 2020'de zirveye ulaştığı görülmüştür [32]. Liu ve Liu (2021), Pfizer'in ilk COVID-19 aşısını duyurusunun ardından kamuoyu duyarlılığını ölçmek ve konuma göre duyarlılığın değişimini tespit etmek için duygu analizi gerçekleştirmişlerdir. Çalışmada 1 Kasım 2020-31 Ocak 2021 tarihleri arasındaki COVID-19 aşılariyla ilgili İngilizce tweetler kullanılmıştır. Tweetlerde geçen duygunun gerçek olup olmadığını belirlemek amacıyla Valence Aware sözlüğü ve duygu çıkarımcısı kullanılmıştır. Tweetlerde geçen başlıca konuları çıkarmak için gizli Dirichlet tahsisi kullanılmıştır. Hem zamansal hem de konumsal analiz gerçekleştirerek farklı konularda gönderilen tweetlerdeki duygu farklılıkları belirlenmiştir. Sonuçta analiz edilen tweetlerin yüzde 42,8 olumlu, yüzde 26,9 nötr ve yüzde 30,3 olumsuz duygular içerdiği belirlenmiştir. Olumlu duygu içeren tweetler için deneme sonuçları, yönetim, yaşam, bilgi, etkililik ve olumsuz duygu içeren tweetler için deneme sonuçları, komplo, güven, etkililik ve yönetim olmak üzere 5'er konu belirlenmiştir [33]. Lyu vd. (2021), Liu ve Liu (2021), İngilizce tweetleri kullanmaları ve 31 Ocak 2021'den sonraki tweetleri analize dahil etmemeleri nedeniyle bu çalışmadan farklılık göstermektedirler. Marcec ve Likic (2022), Twitter'da COVID-19 aşılari ile ilgili atılan İngilizce tweetlere duygu analizi uygulamışlardır. 1 Aralık 2020-31 Mart 2021 tarihi aralığındaki 4 aylık döneme ait tweetleri kullanmışlardır. Çalışma sonuçları Pfizer ve Moderna aşılariya ilişkin duyguların 4 ay boyunca olumlu ve istikrarlı görüldüğünü buna karşılık, AstraZeneca/Oxford aşısına ilişkin olumlu duyguların Aralık ayı ile Mart ayı karşılaştırıldığında önemli bir düşüş göstererek zamanla azaldığını göstermiştir [34]. İlgili çalışma duygu sınıflandırmasında AFINN [35] sözlüğünü kullanması, İngilizce tweetler kullanması ve 31 Mart 2021 tarihinden sonraki tweetleri analiz etmemesi yönleriyle bu çalışmadan farklıdır. Villavicencio vd. (2021), Filipinler'de COVID-19 aşılariya yönelik duygular üzerine olumlu, olumsuz ve tarafsız olmak üzere analiz yapmışlardır. Naive Bayes sınıflandırma modelini kullanmışlardır. Çalışmada 1-31 Mart 2021 tarihlerindeki İngilizce ve Filipince tweetler analiz edilmiştir. Analiz sonuçlarına göre tweetlerin %83,38'i olumlu, %8,26'sı olumsuz, %8,36'sı nötr duygu içermektedir [36]. Ansari ve Khan (2021), COVID-19 aşılari ile ilgili kamuoyu

tepkisini ölçmek için dünya çapında atılan İngilizce tweetleri kullanarak duygu analizi gerçekleştirmişlerdir. Çalışmada 15 Mayıs-25 Haziran 2021 tarihleri arasındaki tweetler kullanılmıştır. Naïve Bayes sınıflandırıcısı kullanılarak duygu analizi gerçekleştirilmiştir. Sonuçta Türkiye için atılan tweetler olumsuz duygu ağırlıklı olarak bulunmuştur [37]. Villavicencio vd. (2021), Ansari ve Khan (2021), İngilizce ya da Filipince tweetler kullanmaları, makine öğrenme yöntemlerini kullanmaları ve 2021 yılının sadece belli aylarındaki tweetleri analiz etmeleri sebebiyle bu çalışmadan farklıdır. Çılgın vd. (2022), Türkiye’de COVID-19 aşılara ilişkin kamu duyarlılığını aşıyla ilgili tweetlere makine öğrenmesi algoritmalarını uygulayarak analiz etmişlerdir. Çalışmada 1 Nisan-31 Ağustos 2021 tarihleri arasında toplam 412.588 adet Türkçe tweet kullanılmıştır. Sonuçta incelenen tüm tweetlerin %17,70’inin pozitif, %41,97’sinin nötr ve %40,34’ünün negatif duygu polaritesinde olduğu tespit edilmiştir [38]. Aslan (2022), derin öğrenme tabanlı doğal dil işleme tekniklerini kullanarak Türk halkının aşılarda süreciyle ilgili duygularını ve tepkilerini araştırmak için duygu analizi gerçekleştirmişlerdir. Çalışmada, Twitter’da 16 Haziran-18 Eylül 2021 tarihleri arasında paylaşılan Türkçe tweetler kullanılmıştır. Analiz sonucunda tweetlerin %34,9’unun olumlu, %49,6’sının nötr, %15,5’inin olumsuz duygu dağılımına sahip olduğu görülmüştür [39]. Çılgın vd. (2022), Aslan (2022), Türkçe tweetlerle bu çalışmaya benzer duygu analizi gerçekleştirmişlerdir. İlgili çalışmalar 2021 yılının belirli aylarındaki tweetleri kullanmışlardır. Ancak bu çalışmada 2021 yılının tamamı da analize dahil edilmiştir. Mermir ve Özsezer (2023), COVID-19 aşı tartışmalarını analiz etmek için Twitter verisine duygu analizi uygulamışlardır. Çalışmada 10 Mart 2020- 18 Nisan 2022 aralığındaki 10.308 adet Türkçe tweet kullanılmıştır. Çalışma sonucunda COVID-19 aşılara ile ilgili tweet kullanıcılarının %7,5’inin olumlu, %0,59’unun olumsuz, %91,91’inin nötr duyguda olduğu belirlenmiştir. Çalışmada duygu analizi hem sözlük tabanlı hem de makine öğrenme yöntemleri kullanılarak gerçekleştirilmiştir [40]. İlgili çalışma bu çalışmaya benzer bir çalışmadır. Ancak Nisan 2022’den sonraki tweetlerin de analize dahil edilmesi ve duygu sınıflandırmasında SentiTurknet sözlüğünün kullanılmasıyla bu çalışma ilgili çalışmadan farklıdır.

Resmi olarak Mart 2020’de ülkemize giren Covid-19 sebebiyle Mart 2023 itibariyle ülkemizde toplam 102.174 [41] vefat sayısına ulaşılmıştır. Günümüzde virüsün etkinliği azalmış olsa da henüz tüm dünyada tamamen bitmemiştir. 31 Temmuz – 27 Ağustos 2023 tarihleri arasında dünya çapında yaklaşık 1,4 milyon yeni vaka ve yaklaşık 1800 yeni ölüm vakası rapor edilmiştir [42]. Virüs ile mücadelede geliştirilen aşılardan etkili çözüm olarak görülmektedir. Ülkemizde ve tüm dünyada uygulanan aşılardan olumlu ve olumsuz etkileri üzerine insanların görüşlerini alabilmek için sosyal medya platformları tercih edilmektedir. Bu amaçla çoğunlukla Twitter’da paylaşılan tweetler kullanılmaktadır. Bu çalışmada, yurt içinde insanların COVID-19 aşılara ile ilgili duygularını ölçmek için bir veri madenciliği uygulaması olan *metin madenciliği* gerçekleştirilmiştir. Bu amaçla Sinovac ve Biontech aşılara ile ilgili atılan Türkçe tweetler üzerine duygu analizi gerçekleştirilmiştir. Böylece insanların aşılara karşı duyguları ölçülmüş ve

aşılardan insanlar üzerindeki olumlu ya da olumsuz etkileri incelenmiştir.

Veri Madenciliği

Veri Madenciliği, her çeşit veriden örtülü ve anlamlı bilgilerin elde edilmesi işlemidir [43]. Elde edilen bilgiler kamu ve özel sektör ile her çeşit özel topluluk tarafından önemli kararların alınmasında hatta devlet yönetiminin izleyeceği politikaların belirlenmesinde kullanılır.

Literatürde ilişkisel, Web ve büyük veri madenciliği olmak üzere üç çeşit veri madenciliği tanımlanmıştır. İlişkisel veri madenciliği ilişkisel verilerden bilgi çıkarımını, Web madenciliği web dokümanları ve erişim kayıtlarından bilgi çıkarımını, büyük veri madenciliği ise büyük veriden bilgi çıkarımını ifade eder [43]. Büyük veri; büyük hacmi, çeşitliliği, hızı, değişkenliği ve doğruluğu ile ayırt edilir [44]. Tüm dünyada sağlık, sosyal bilimler, mühendislik, eğitim alanlarında ve sosyal medya platformlarında hızla genişlemektedir.

Metin Madenciliği

Metin madenciliği metinsel veriden örtük ve anlamlı bilgilerin çıkarılması işlemidir. Veri madenciliği, verilerin önceden yapılandırılmış bir biçimde depolandığı varsayıldığından ön işleme aşamasında veriler temizlenerek, normleştirilir ve birçok veri tabanı tablosu birleştirilir. Buna karşılık, *metin madenciliğinde* ön işleme operasyonları, doğal dil belgeleri için temsili özelliklerin tanımlanması ve çıkarılması işlemlerinden oluşur [4]. Veri madenciliği büyük ölçüde sayısal verilere bağlıysa da, *metin madenciliği* metinlerin sözcüksel ve sözdizimsel yapısına bağlıdır.

Bir *metin madenciliği* uygulaması olan *Duygu Analizi* için sırasıyla *metin ön işleme*, *ilişkilendirme* ve *katagorileme* adımları gerçekleştirilir.

Metin ön işleme

Metin ön işleme aşaması indeksleme ve kodlama işlemlerinden oluşur. Metin indeksleme, cümlelerden oluşan bir metni içerdiği kelimelere ayırma işlemidir. Bu işlem sırasıyla ‘Tokenization’, ‘Stemming’, ‘stop-words’ kaldırma ve terim ağırlıklandırma aşamalarından oluşur. ‘Tokenization’, bir metin öbeğini boşluk, noktalama işaretleri veya özel işaretlere göre kelimelere ayırma işlemidir. ‘Stemming’, ‘tokenization’ işleminden üretilen her bir kelimeyi kök haline dönüştürme sürecidir. ‘Stop-words’ kaldırma bağlaçlar, edatlar gibi dilbilgisi için gerekli ancak anlamsız sözcüklerin kaldırılmasıdır. Ağırlıklandırma, her kelimeye önem derecesine göre ağırlık hesaplama işlemidir. Terim frekansı ve TF-IDF ağırlığı, sözcüklerin ağırlıklandırılmasında kullanılan popüler şemalardır. Terim frekansı metin içinde her kelimenin geçiş sıklığını ifade eder. TF-IDF şeması ile hesaplanan kelime ağırlıkları, verilen metinde kelimelerin bulunması ile doğru orantılı, diğer metinlerde söz konusu kelimelerin bulunması ile ters orantılıdır. “ d_i ” dokümanı içinde bir “ t_i ” kelimesinin TF-IDF ağırlığı “ w_{ij} ” [45] eşitliği ile hesaplanır. “N” toplam sözcük sayısını, DF_i “ t_i ” kelimesini içeren metin sayısını, Tf_i “ t_i ” kelimesinin metin içindeki geçiş sayısını ifade eder.

$$\omega_{ij} = \log \frac{N}{DF_i} (1 + \log TF_i) \quad [45]$$

Metin kodlama işleminde, indekslemeden çıkarılan kelimelerden bazıları wrapper, temel bileşenler analizi, bağımsız bileşen analizi, tekil değer ayrışımı gibi belirli yöntemler kullanılarak seçilir ve her bir metnin kodlanmasında bunlara sayısal değerler atanır. Bu değerler kelimelerin metin içindeki önem derecelerini belirtir. Ön işleme sonucunda kelimeleri temsilen sayısal bir vektör oluşturulur.

Metin ilişkilendirme

Metin ilişkilendirme, kelime kümelerinden kelimelerin ilişkilendirme kurallarının üretilmesi işlemidir. Kelime ilişkilendirme, kelimeler arasındaki nedensel ilişkilerin (Eğer-sonra kuralları) çıkarılması işlemidir. İndeksleme işlemiyle içerdiği kelimelerle ilişkilendirilen her metin için kelime-doküman matrisi oluşturulur. Matriste her satır bir metne, her sütun bir kelimeye ve matrisi temsil eden her veri ilgili metinde o kelimenin ağırlığına karşılık gelir. Kelime kümeleri, matristen çıkarılır. Kelimelerin ilişkilendirme kuralları Apriori algoritması tarafından üretilir. Algoritma için kelime kümeleri, matriste ağırlıkları sıfır olmayan kelimeler alınarak satır satır oluşturulur. Algoritmaya girdi olarak matristeki sütunlara karşılık gelen kelime listesi ve kümeleri verilir, çıktı olarak ilişkilendirme kuralları oluşturulur.

Metin kategorilendirme

Metin kategorilendirme, her metne önceden tanımlanmış kategorilerden birinin veya birkaçının atanması süreci olarak tanımlanır. Bunun için makine öğrenme algoritmaları kullanılır. Veri öğelerini sınıflandırma modeli olarak bir liste veya bir kategori ağacının önceden tanımlanması gerekir. Metinler her kategoriye örnek olarak ayrılır. Tüm örnek metinler, bir kelime listesine endekslenir ve bunlardan bazıları seçilir. Kategorilere ayrılan örnek metinler sayısal vektörlere kodlanır. Örnek metinlerden kodlanmış eğitim örnekleri kullanılarak, sınıflandırma algoritmasına bağlı olarak denklemler, sembolik kurallar veya optimize edilmiş parametreler oluşturulur; buna öğrenme süreci denir. Öğrenme sürecinden sonra, örnek metinlerden ayrılmış metinler sınıflandırılır. Spam e-posta filtreleme, duygu analizi, bilgi filtreleme ve konu yönlendirme metin sınıflandırmasının türleri arasındadır.

Duygu analizi, metin sınıflandırmasının özel bir örneğidir. Bir görüşü “pozitif”, “nötr” veya “negatif” olarak sınıflandırma sürecidir. Bu süreçte bir görüşe metinsel bir girdi verisi verilir ve çıktı olarak bu üç tutumdan biri oluşturulur. Üç kategoriden biriyle etiketlenen metinler toplanır ve sayısal vektörler halinde kodlanır. Makine öğrenimi algoritması, örnek etiketli metinlerden kodlanan sayısal vektörleri öğrenir. Test için verilen bir metin sayısal bir vektöre kodlanır ve üç kategoriden birine sınıflandırılır.

Duygu analizinde diğer bir yöntem duygu sınıflandırması içeren polarite sözlüklerinin kullanılmasıdır. Bu

çalışmada sözlük tabanlı duygu analizi yapılmış ve bu amaçla SentiTurkNet kullanılmıştır. SentiTurkNet, duygu sınıflandırması ve fikir madenciliği uygulamaları için tasarlanmış Türkçe için oluşturulmuş bir polarite sözlüğüdür. WordNet'in [46] tüm eş anlamlı sınıflarının “olumluluk”, “olumsuzluk” ve “tarafsızlık” kavramlarına göre otomatik olarak açıklanması ile oluşturulmuştur. İngilizce WordNet ve SentiWordNet ile eşleştirilmiştir. Türkçe WordNet'teki tüm eş anlamlı sınıflara (synset) polarite sınıfından (pozitif, objektif/nötr veya negatif) bir kavramın atanması ile oluşturulmuştur. Dilden bağımsızdır ve farklı dillerde de uygulanabilir.

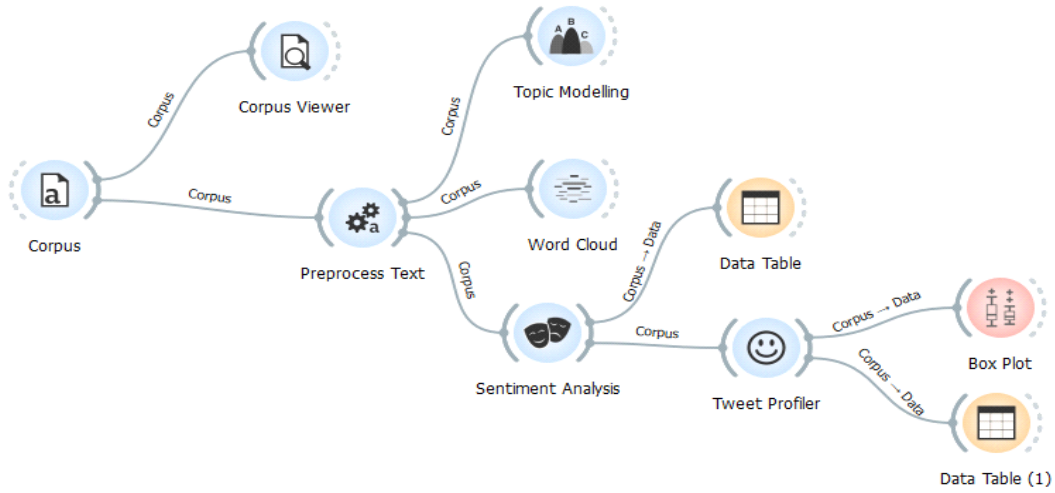
Orange

Orange [47], Slovenya'da Ljubljana Üniversitesi tarafından geliştirilen, Python dilinde yazılmış, açık kaynak kodlu bir makine öğrenimi, veri madenciliği ve veri görselleştirme yazılımıdır. Akademik çalışmalarda, endüstri alanında ve eğitim-öğretim'de kullanılmaktadır. En önemli endüstriyel ortak, Orange'ı ilaç geliştirmede kullanılan ve ilgili çeşitli bölümlerinin geliştirilmesine sponsorluk yapan bir ilaç devi olan Astra-Zeneca'dır [48]. Orange, içerdiği Twitter aracı ile Twitter API'sine erişim sağlayarak tweet analizine olanak sağlamaktadır. Ancak Orange ile en son iki hafta içinde atılmış tweetler çekilebilmektedir. Daha eski tarihlere ait tweetler çekilememektedir. API planları 30 Mart 2023'te değiştirildikten sonra Twitter artık ücretsiz erişim ve ücretsiz Tweet alımını desteklememektedir. Bu çalışmada tweetler Python kodu kullanılarak snsrape ile Twitter Api'sinden Twitter veri erişim ve ücret politikası değişmeden önce çekilmiştir. Elde edilen tweetlerle gerçekleştirilen duygu analizi uygulamasında Orange Metin Madenciliği eklentisi kullanılmıştır.

Duygu analizi gerçekleştirimi

Sosyal medya kullanımındaki büyüme oranıyla birlikte, kullanıcılar tarafından üretilen veri miktarının artması insanların güncel her konu hakkında görüşlerini ifade edebildiği Twitter'ı, duygu analizi ve fikir madenciliğini sağlayan bir araç haline getirmiştir. Bu değişim, önemli konularda kitlelerin nabzını kolayca ölçmek ve önemli kararlar alabilmek için kurum, kuruluş ve kişileri sosyal medya kullanımına yönlendirmiştir. Bu noktadan hareketle, bu çalışmada COVID-19 için geliştirilen Sinovac ve Biontech aşılı için Twitter kullanıcılarının paylaştıkları tweetler üzerine duygu analizi yapılmış ve sonuçlar karşılaştırılmıştır. Sonuçta insanların aşılardaki duygularını öğrenmek ve aşılardan insanlar üzerindeki etkilerini ortaya çıkarmak amaçlanmıştır.

Çalışmada gerçekleştirilen duygu analizi iş akışı Şekil 1'de gösterilmektedir. İlk olarak Twitter'dan Sinovac ve Biontech ile ilgili tweetler elde edilmiştir. Tweetlere ön işlemler (Preprocess Text) uygulandıktan sonra konu modellemesi (Topic Modelling) yapılmıştır. Konu modelleme ile tweetlerde geçen konular ve konularda geçen popüler anahtar kelimeler incelenmiştir. Ardından WordCloud (Kelime Bulutu) aracı ile tweetlerde en sık geçen kelimeler ağırlıklarıyla bulunmuştur. Son aşamada



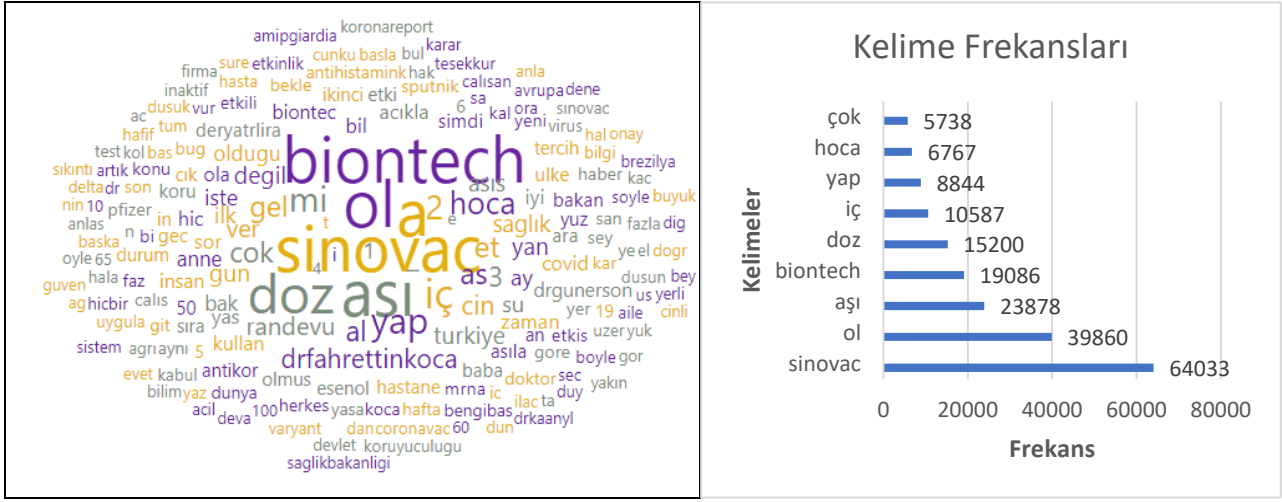
Şekil 1. Twitter duygu analizi iş akışı.

tweetler üzerine duygu analizi (Sentiment Analysis) uygulanmıştır. Duygu analizi sonuçları Orange Tweet Profiler ile hesaplanarak Box Plot'da görselleştirilmiştir. Sonuç olarak yaklaşık iki yılda COVID-19 aşıları hakkında atılan tweetlerin sahiplerinin aşı hakkındaki duygu profilleri belirlenmiştir.

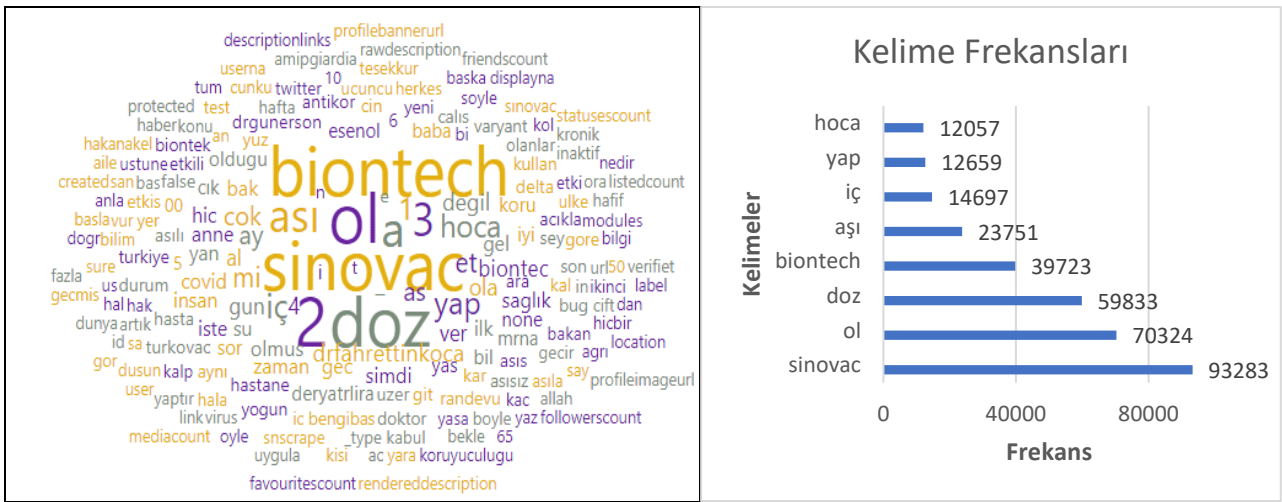
Materyal ve yöntem

Çalışmada kullanılan tweetler Python kodu kullanılarak sncscrape ile Twitter Api'sinden çekilmiştir. İlk olarak "Sinovac" anahtar kelimesini içeren tweetler aşının ülkemizde uygulanmaya başladığı tarihten itibaren 13 Ocak 2021-16 Şubat 2023 zaman aralığında elde edilmiştir. Daha sonra "Biontech" anahtar kelimesini içeren tweetler Biontech aşısının ülkemizde uygulanmaya başladığı tarihten itibaren 28 Temmuz 2021-23 Şubat 2023 zaman aralığında elde edilmiştir. Elde edilen tweetlerden retweetler ve yinelenen tweetler ayıklandıktan sonra Sinovac için toplam 326085 tweet ve Biontech için toplam 213474 tweet analiz edilmiştir. İkinci aşamada Tweetlere metin ön işleme analizi uygulanmıştır. Ön işlemler, kelimelerdeki harfleri küçük harfe dönüştürme, aksanları ortadan kaldırma, HTML ve URL etiketlerini kaldırma, hashtag'lerin (#) kaldırılması, metni kelimelere ayırma (Tokenization) ve filtreleme işlemlerini içerir. Filtreleme işlemleri noktalama işaretlerinin, tekrar eden kelimelerin ('ve', 'veya', içinde) ve bağlaçların kaldırılması aşamalarını içerir. Bu amaçla Apache Lucene Türkçe "Stop Words" (Durak Kelimeleri) listesi genişletilerek kullanılmıştır. Ön işlemlerden sonra metin içeriğindeki aşı ile ilgili olmayan kelimeler, site linkleri, yanlış yazılmış kelime ekleri temizlenmiştir. Üçüncü aşamada ön işlemlerden geçmiş olan tweetler Word Cloud aracı ile analiz edilmiş ve tweetlerde en çok kullanılan kelimeler belirlenmiştir (Şekil 2-9). Word Cloud, tweetlerde geçen tüm kelimeleri, kelimelerin boyutlarını, sözcüklerin ağırlığını (sıklığını) gösterir. Sözcükler, ağırlıklarına göre listelendiği için hangi kelimelerin kaç kez kullanıldığı ve kelimelerin popülerlik

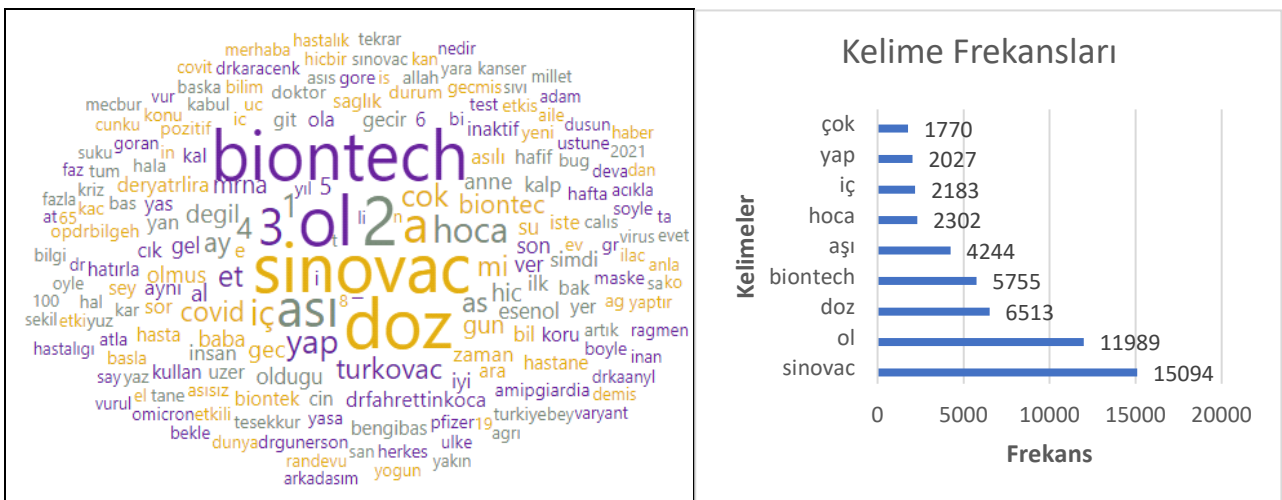
durumu görülür. Örneğin, Şekil 2'de Sinovac ile ilgili tweetlerde en çok kullanılan üç kelime sırasıyla Sinovac (64033), ol (39860) ve aşı (23878)'dir. Şekil 6'da Biontech ile ilgili tweetlerde en çok kullanılan üç kelime ise sırasıyla Biontech (16967), ol (12914) ve doz (8798)'dur. Tweetlerde bulunma sıklıklarına göre görselleştirilen kelimeler araştırılan konuların içerikleri hakkında kullanıcıya bilgi verir. Dördüncü işlem adımında Konu Modelleme ile her metindeki kelime kümelerine ve sıklıklarına göre konular bulunmuştur. Bir tweet farklı oranlarda birden çok konu içerebildiği için aynı zamanda belge başına konu ağırlığı da hesaplanmıştır. Bu amaçla 'Latent Dirichlect Allocation' (LDA) tekniği kullanılmıştır. Bu teknik ile bir konuda ilgili olan veya olmayan sözcükler ve aynı anda olumlu veya olumsuz olabilen konu ağırlıkları hesaplanır. Bu şekilde tweetlerde geçen 10 konu ve bu konularda geçen popüler anahtar kelimeler belirlenmiştir. Sıradaki adımda Duygu Analizi gerçekleştirilerek metinlerdeki duygular tahmin edilmiştir. Orange Duygu Analizi aracı "Liu & Hu" [49] ve Vader [50] duygu modüllerini ve tekniklerini kullanır. "Liu & Hu" tekniği sözlük tabanlı çalışır ve nihai puan hesaplanırken pozitif kelimelerin toplamı ile negatif kelimelerin toplamı arasındaki fark alınarak belgenin uzunluğuna göre normalleştirilip 100 ile çarpılır. Sonuç nihai puan, belgedeki duygu farkının yüzdesini yansıtır. Vader tekniği ise; sözlük ve kural tabanlı duygu analizini destekler. İki teknik arasındaki fark; Vader tekniğinin sadece İngilizce metinlerin üzerinde çalışmasıdır. Çalışmada Türkçe duygu sınıflarının oluşturulması için Orange'a SentiTurknet sözlüğü özel sözlük olarak import edilerek kullanılmıştır. Duygu analizi sonucunda pozitif ve negatif olmak üzere iki adet öznelik hesaplanmıştır. Elde edilen sonuçlar -1'e ne kadar yakın ise duygu o kadar olumsuz, +1'e ne kadar yakın ise duygu o kadar olumludur. Bu şekilde tweetler, bir sorgu terimiyle ilgili olarak olumlu veya olumsuz şekilde sınıflandırılmaktadır. Sınıflandırma sonucunda oranlama yapılarak hangi duygunun yüksek olduğu belirlenmektedir.



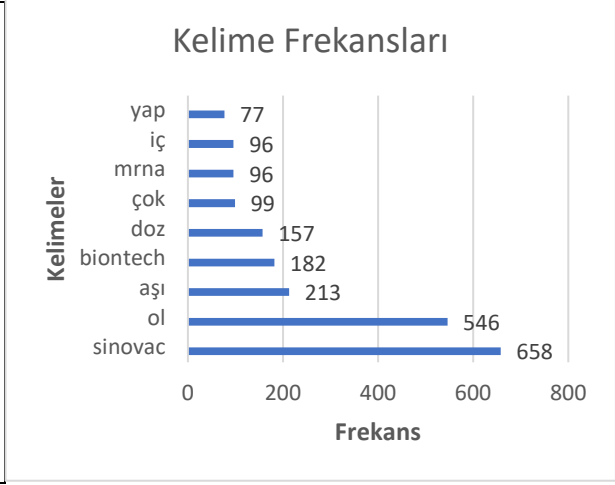
Şekil 2. Sinovac ile ilgili tweetlerin kelime bulutu ve frekansları (13.01.2021- 30.06.2021).



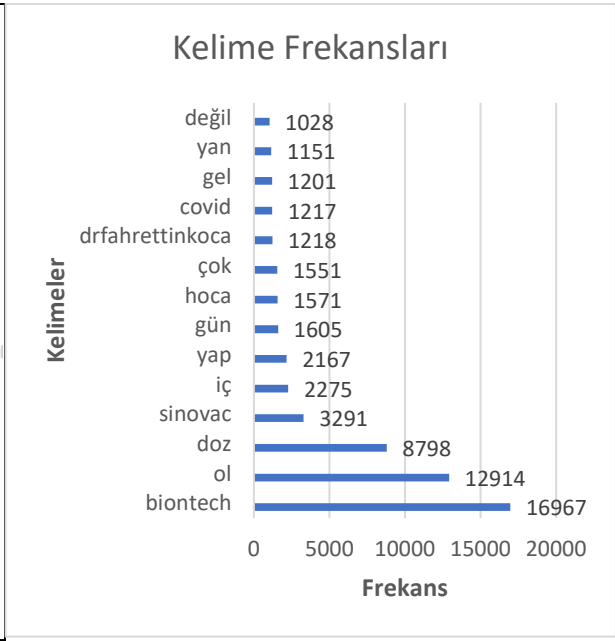
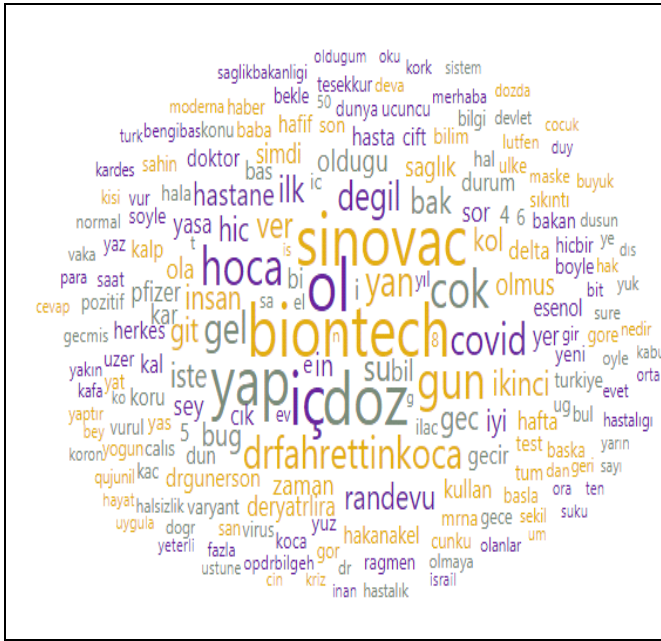
Şekil 3. Sinovac ile ilgili tweetlerin kelime bulutu ve frekansları (01.07.2021-31.12.2021).



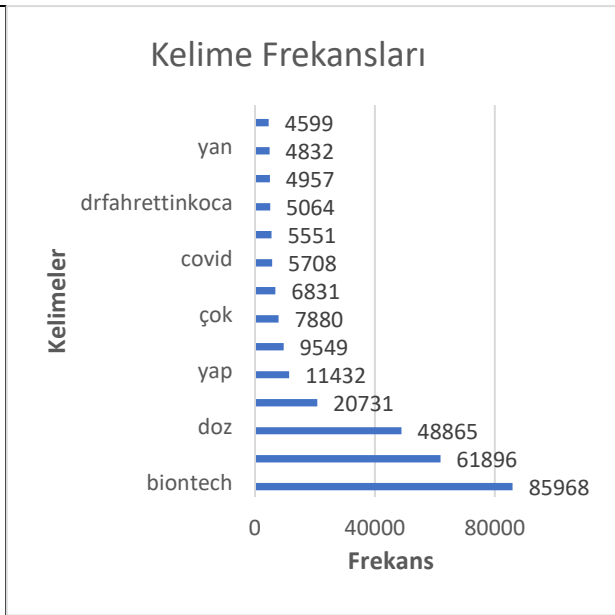
Şekil 4. Sinovac ile ilgili tweetlerin kelime bulutu ve frekansları (01.01.2022- 31.12.2022).



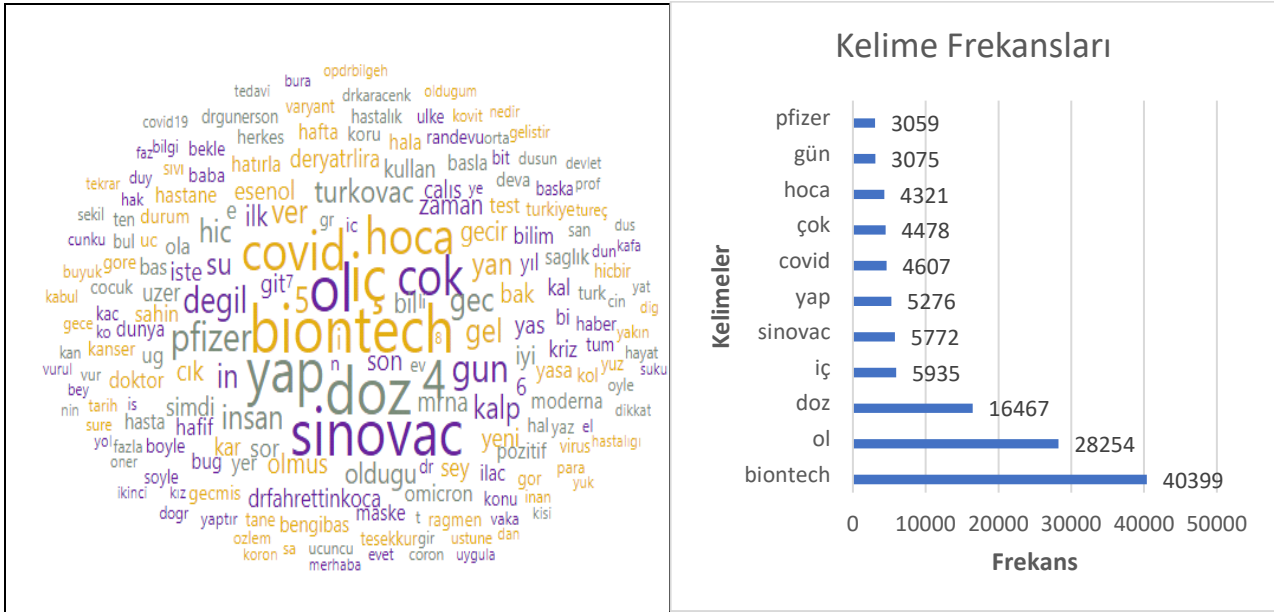
Şekil 5. Sinovac ile ilgili tweetlerin kelime bulutu ve frekansları (01.01.2023-16.02.2023).



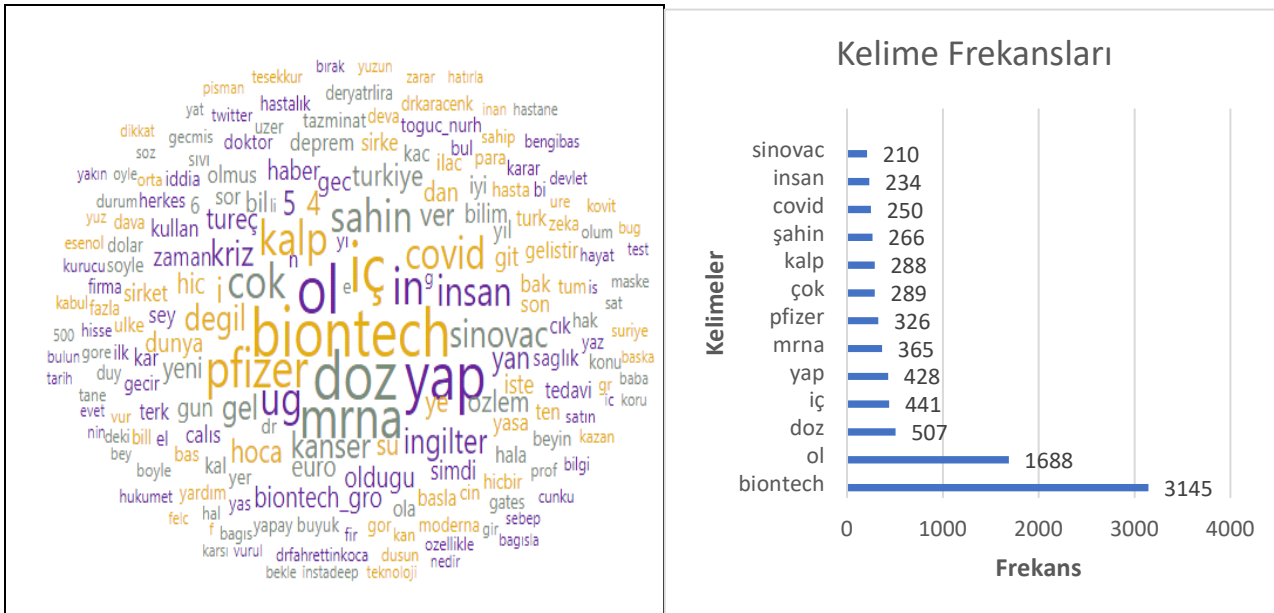
Şekil 6. Biontech ile ilgili tweetlerin kelime bulutu ve frekansları (28.07.2021-12.08.2021).



Şekil 7. Biontech ile ilgili tweetlerin kelime bulutu ve frekansları (13.08.2021-31.12.2021).



Şekil 8. Biontech ile ilgili tweetlerin kelime bulutu ve frekansları (1.1.2022-31.12.2022).



Şekil 9. Biontech ile ilgili tweetlerin kelime bulutu ve frekansları (1.1.2023-23.02.2023).

İş akışının son aşamasında Tweet Profiler ile duygu analizi sonuçları görselleştirilmiştir. Tweet Profiler “Ekman’s”, “Plutchik’s” ve “Mood States Profili” (POMS) olmak üzere üç tür duygu sınıflandırmasını destekler.

Bulgular

Çalışmada COVID-19 aşılı ile ilgili zamana bağlı duygu değişimini ölçmek için zamansal duygu analizi gerçekleştirilmiştir. Duygu analizi sonuçları incelendiğinde, Şekil 10’da görüldüğü üzere 2021 yılının ilk yarısında Sinovac için araştırılan 67289 tweet analizi sonucunda tweetlerden 33805’inin sevinç, 26509’unun

şaşkınlık, 4590’nın korku, 294’ünün nefret, 1993’ünün üzüntü, 98’inin kızgınlık içerdiği görülmüştür. Şekil 11’e bakıldığında 2021 yılının ikinci yarısında Sinovac aşısı için ele alınan 96869 tweet analizi sonuçlarına göre tweetlerin 54339’unun sevinç, 33512’sinin şaşkınlık, 5960’nın korku, 347’sinin nefret, 92’sinin kızgınlık ve 2619’unun üzüntü içerdiği tespit edilmiştir. Şekil 12’de görüldüğü üzere 2022 yılında Sinovac aşısı için ele alınan 16157 tweet analizi sonuçlarına göre tweetlerin 9118’inin sevinç, 5542’sinin şaşkınlık, 461’inin üzüntü, 967’sinin korku, 58’inin nefret, 11’inin kızgınlık içerdiği görülmüştür. Şekil 13 incelendiğinde 2023 yılının ilk iki ayında Sinovac aşısı için ele alınan 770 tweet analizi sonuçlarına göre tweetlerin 360’nın sevinç, 326’sının şaşkınlık, 27’sinin üzüntü, 50’sinin korku, 5’inin nefret ve

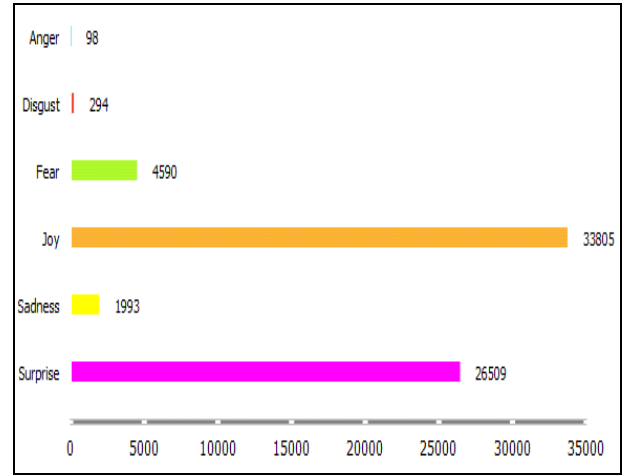
2'sinin kızgınlık içerdiği görülmüştür. 2021 yılında Sinovac için atılan tweetlerin duygu analizi sonuçları incelendiğinde en yüksek duygunun %53,69 ile sevinç olduğu görülmüştür. İkinci sırada %36,56 ile şaşkınlık duygusu, üçüncü sırada ise %6,43 ile korku duygusu yer almaktadır. 2022 yılında Sinovac için atılan tweetlerin analiz sonuçları incelendiğinde %56,43 sevinç, %34,3 şaşkınlık, %5,98 korku içerdiği görülmektedir. 2022 yılında aşı hakkındaki sevinç duygusunun önceki yıla göre arttığı görülmektedir. Ancak şaşkınlık ve korku duygusunun 2021 yılına göre azaldığı görülmektedir. Şekil 14'te görüldüğü üzere Biontech aşısı ülkemizde uygulanmaya başlandığı tarihten itibaren 2021 yılında ilk iki ay (Temmuz-Ağustos) için atılan 18172 tweet analiz edildiğinde 9559'unun sevinç, 6991'inin şaşkınlık, 485'inin üzüntü, 1019'unun korku, 71'inin nefret, 47'sinin kızgınlık içerdiği görülmüştür. Şekil 15'e bakıldığında Biontech aşısı için 2021 yılının son dört ayında atılan 138556 tweet incelendiğinde 52522'sinin sevinç, 38834'ünün şaşkınlık, 40354'ünün üzüntü, 6299'unun korku, 377'sinin nefret, 170'inin kızgınlık içerdiği görülmüştür. 2021 yılının son dört ayında Biontech aşısıyla ilgili "üzüntü" ifadesinin önceki iki aya kıyasla çok yüksek çıkmasının sebebi aşı olma oranlarının artması, zaman ilerledikçe aşılardan yan etkilerinin ortaya çıkma ihtimalleri ve bu dönemde diğer dönemlere kıyasla çok fazla tweet atılması olduğu düşünülmektedir. Şekil 16'da görüldüğü üzere Biontech aşısı için 2022 yılında atılan toplam 52362 tweet analiz edildiğinde 26550'sinin sevinç, 20450'sinin şaşkınlık, 1657'sinin üzüntü, 3414'ünün korku, 188'inin nefret, 102'sinin kızgınlık içerdiği görülmüştür. Şekil 17'de görüldüğü üzere Biontech aşısı için 2023 yılının ilk ayında atılan 4384 tweet analiz edildiğinde 2035'inin sevinç, 1812'sinin şaşkınlık, 140'ının üzüntü, 350'sinin korku, 30'unun nefret, 17'sinin kızgınlık içerdiği görülmüştür.

2021 yılında Biontech hakkında atılan tweetlerin duygu analizi sonuçları incelendiğinde en yüksek duygunun %39,61 ile sevinç olduğu görülmüştür. İkinci sırada %29,24 ile şaşkınlık duygusu, üçüncü sırada ise %4,67 ile korku duygusu yer almaktadır. 2022 yılında Biontech hakkında atılan tweetlerin duygu analizi sonuçları incelendiğinde en yüksek duygunun %50,70 ile sevinç olduğu görülmektedir. İkinci sırada %39,05 ile şaşkınlık, üçüncü sırada ise %6,52 ile korku duygusu yer almaktadır. 2022 yılında Biontech aşısı hakkındaki sevinç, şaşkınlık ve korku duygularının önceki yıla kıyasla arttığı görülmektedir. Şekil 18, 19 ve 20'de Sinovac ve Biontech aşılarının 2021-2022 yıllarındaki duygu değişiminin karşılaştırılması gösterilmektedir.

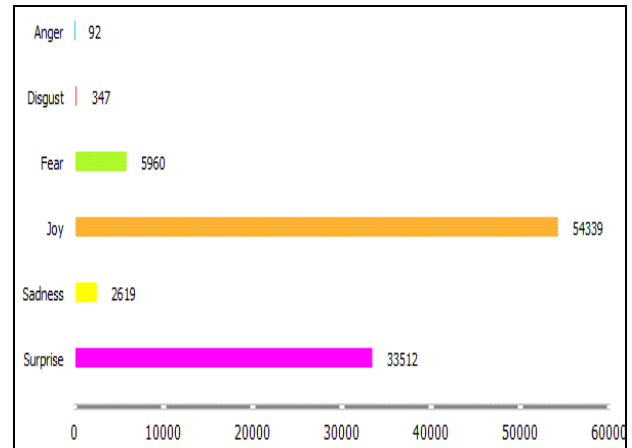
Genel olarak her iki aşı hakkında atılan tweetlerin analizi sonucunda COVID-19 aşılarının insanlar üzerinde olumlu etkisi olduğu, aşılardan memnun kalındığı görülmektedir. Sinovac ve Biontech aşıları karşılaştırıldığında Sinovac aşısından daha çok memnun kalındığı görülmektedir. Ancak analiz sonuçlarının bir kısmı COVID-19 aşılarının insanlar üzerinde olumsuz etkilerinin de olduğunu göstermektedir.

Çalışmada elde edilen tweetlerin büyük bir kısmının konum bilgilerinin eksik olduğu görülmüştür. Bu sebeple konuma göre duygu analizi gerçekleştirilememiştir.

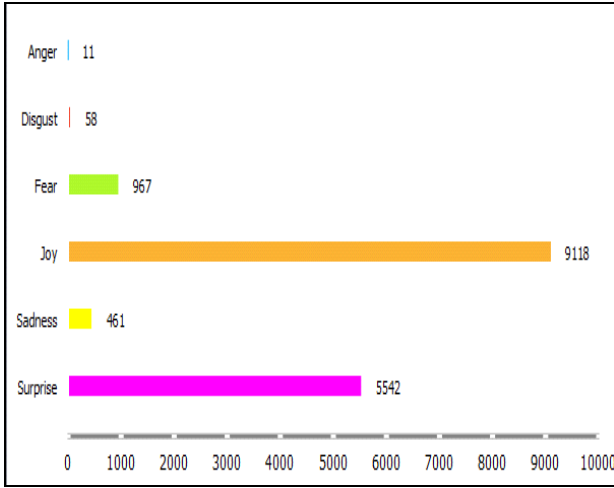
Bu çalışma ile benzer duygu analizi yapan Çılgın vd. (2022) ve Aslan (2022) analiz sonuçlarında bu çalışmadan farklı olarak aşılarda nötr duygunun daha yüksek olduğunu belirlemişlerdir. Bu farkın sebebi olarak ilgili çalışmaların 2021 yılının sadece belli aylarına ait tweetleri değerlendirmeleri ve bu çalışmadan farklı olarak makine öğrenme yöntemlerini kullanmaları olduğu düşünülmektedir. Mermer ve Özsezer (2023) de bu çalışmaya benzer duygu analizi yapmış olmalarına rağmen aşılarda nötr duygunun daha yüksek olduğunu belirlemişlerdir. Bunun sebebinin ilgili çalışmanın Nisan 2022'den sonraki tweetleri analize dahil etmemesi, makine öğrenme yöntemlerini kullanması ve bu çalışmadakinden farklı bir duygu sözlüğü kullanmasından kaynaklı olabileceği düşünülmektedir.



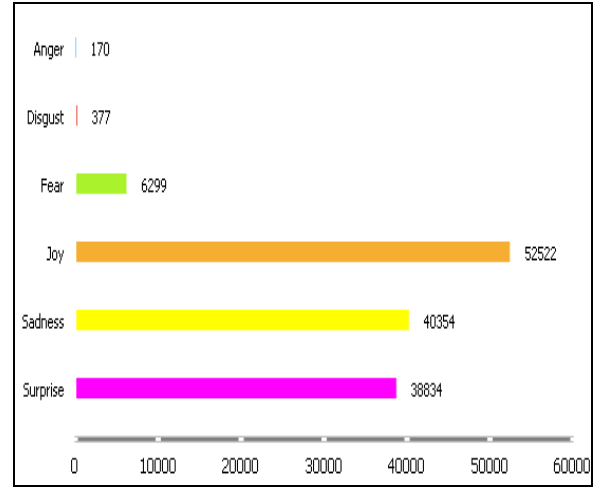
Şekil 10. Sinovac için Duygu Analizi sonuçları (13.01.2021- 30.06.2021).



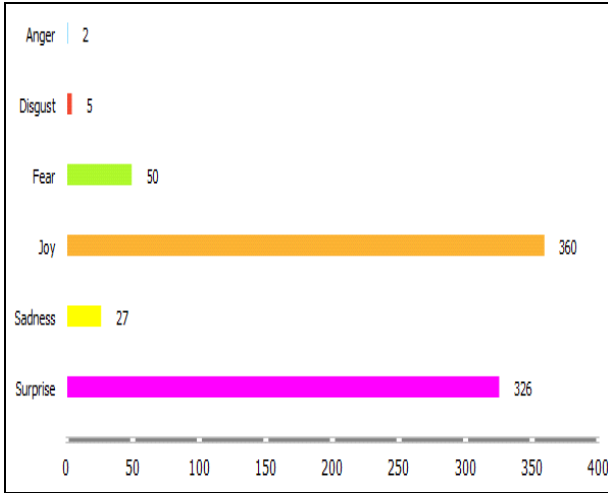
Şekil 11. Sinovac için Duygu Analizi sonuçları (01.07.2021-31.12.2021).



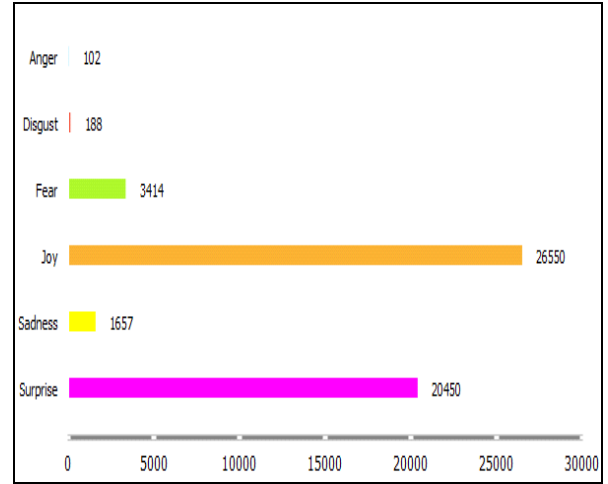
Şekil 12. Sinovac için Duygu Analizi sonuçları (01.01.2022- 31.12.2022)



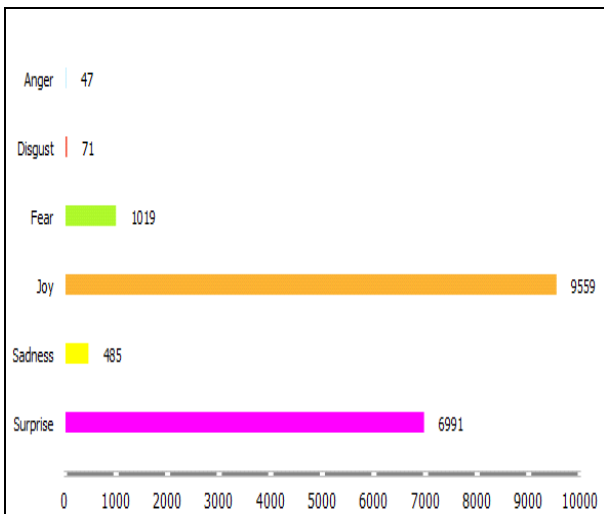
Şekil 15. Biontech için Duygu Analizi sonuçları (13.08.2021-31.12.2021).



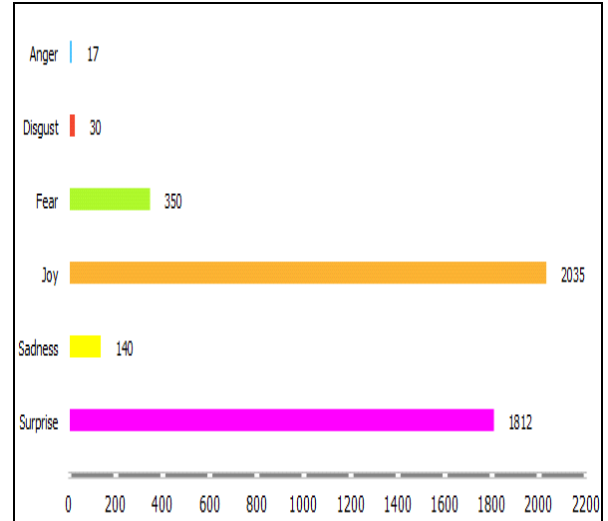
Şekil 13. Sinovac için Duygu Analizi sonuçları (01.01.2023-16.02.2023)



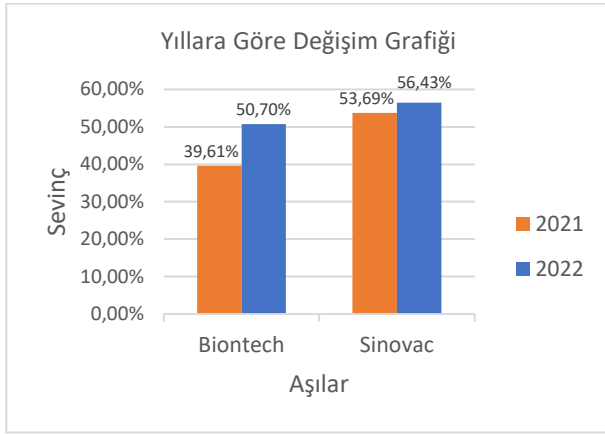
Şekil 16. Biontech için Duygu Analizi sonuçları (1.1.2022-31.12.2022).



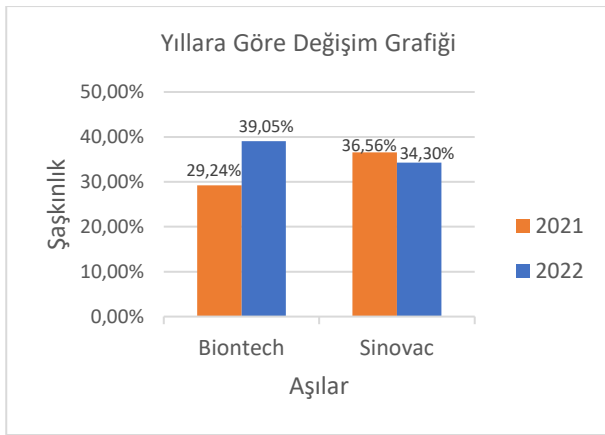
Şekil 14. Biontech için Duygu Analizi sonuçları (28.07.2021-12.08.2021).



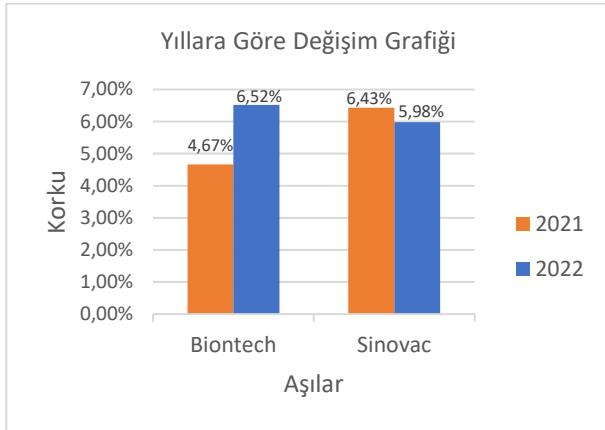
Şekil 17. Biontech için Duygu Analizi sonuçları (1.1.2023-23.02.2023).



Şekil 18. Aşılar hakkında yıllara göre duygu değişimi



Şekil 19. Aşılar hakkında yıllara göre duygu değişimi



Şekil 20. Aşılar hakkında yıllara göre duygu değişimi

Sonuçlar ve gelecek çalışma

Yaklaşık iki yıl ülkemizde ve tüm dünyada etkili olan COVID-19 pandemisi tüm insanlığın sağlığını tehdit etmiş önemli sayıda can kayıplarına sebep olmuştur. İnsanlar üzerinde olumsuz izler bırakan COVID-19 virüsüne karşı geliştirilen aşılar hastalığı hafif atlabilmek için koruyucu çözüm olarak görülmektedir. Ancak aşı karşıtı tutumlar ve aşı uygulamalarından sonra

gelişen sorunlar ve yan etkiler insanları tedirgin etmektedir. Bu noktadan hareketle, bu çalışmada COVID-19 aşılardan Sinovac ve Biontech özelinde, metin madenciliği teknikleri ile duygu analizi gerçekleştirilmiştir. Bu amaçla aşı hakkındaki olumlu ya da olumsuz tecrübelerini ve düşüncelerini Twitter üzerinden paylaşan insanların tweetleri elde edilerek duygu analizi gerçekleştirilmiştir. Analiz sonuçları incelendiğinde COVID-19 aşılarnın toplumda çoğunlukla olumlu yönde etki yaptığı görülmüştür.

Öte yandan, analiz sonuçlarına göre aşı ile ilgili olumlu duygular fazla olsa da, nüfusun bir kısmının endişe ve korku duyması, aşılarn insanlar üzerindeki etkisinin olumsuz taraflarını da yansıtmıştır. Bu çalışmanın farklı aşı türleri ya da medikal tedavi yöntemleri için de uyarlanabileceği ve bu sayede insanlar için yol gösterici olabileceği düşünülmektedir. Gelecek çalışmada, makine öğrenmesi algoritmaları ile duygu sınıflandırması yapılarak tweetler üzerine duygu analizi gerçekleştirilip sonuçları bu çalışma ile karşılaştırılacaktır.

Kaynaklar

- [1] “Instagram,” [Online]. Available: <https://www.instagram.com/>.
- [2] “Facebook,” [Online]. Available: <https://www.facebook.com/>.
- [3] “Twitter,” [Online]. Available: <https://twitter.com/>.
- [4] R. Feldman and J. Sanger, “The Text Mining Handbook”, Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press, 2006.
- [5] R. Dehkharghani, Y. Saygin, B. Yanikoglu and K. Oflazer, “SentiTurkNet: a Turkish polarity lexicon for sentiment analysis”, Lang Resources & Evaluation, 50:667–685, 2016.
- [6] S. Baccianella, A. Esuli, F. Sebastiani, “SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining”, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, May, Valletta, Malta, 2010.
- [7] E. Cambria, D. Olsher, D. Rajagopal, “VSenticNet 3: A Common and Common-Sense Knowledge Base for Cognition-Driven Sentiment Analysis”, July, Conference: AAAI, 2014.
- [8] S. M. Mohammad and P. D. Turney, “Crowdsourcing a word-emotion association lexicon”, *Computational Intelligence*, 29(3), 436–465. 2013.
- [9] M. Özcelik, N. B. Arıcan, Ö. Bakay, E. Sarıış, G. N. Bayazit, Ö. Ergelen ve T. O. Yıldız, “HisNet: A Polarity Lexicon based on WordNet for Emotion Analysis”, In *Proceedings of the 11th Global Wordnet Conference*, pages 157–165, University of South Africa (UNISA). Global Wordnet Association, 2021.
- [10] S. S. Sharma and G. Dutta, “SentiDraw: Using star ratings of reviews to develop domain specific sentiment lexicon for polarity determination”, *Information Processing and Management* 58 (2021) 102412, 2021.

- [11] R. Dehkharghani, "SentiFars: A Persian Polarity Lexicon for Sentiment Analysis", *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 19, 2, Article 21, 12 pages, September, 2019.
- [12] "Emotion Ontology", [Online]. Available: <https://biportal.bioontology.org/ontologies/MFOEM>.
- [13] M. Dragoni, S. Poria, E. Cambria, "OntoSentNet: A Commonsense Ontology for Sentiment Analysis", *IEEE Intelligent Systems*, May/June, p.77-85, 2018.
- [14] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", In: *Proceedings of the ACL-02 conference on empirical methods in natural language processing-Volume 10* (pp. 79–86), Association for Computational Linguistics, 2002.
- [15] A. Xie, Agarwal, B. Vovsha, I. O. Rambow and R. Passonneau, "Sentiment analysis of Twitter data", In: *Proceedings of the workshop on languages in social media Association for Computational Linguistics*, (pp. 30–38), 2011.
- [16] S. M. Başarslan ve F. Kayaalp, "Sentiment Analysis with Machine Learning Methods on Social Media", *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal Regular Issue*, Vol. 9, N. 3 (2020), 5-15, 2020.
- [17] F. Fang Yao, Y. Wang, "Domain-specific sentiment analysis for tweets during hurricanes (DSSA-H): A domain-adversarial neural-network-based approach", *Computers, Environment and Urban Systems* 83 (2020) 101522, 2020.
- [18] N. Mukhtar, A. M. Khan, N. Chiragh, "Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains", *Telematics and Informatics*, 35 (2018) 2173–2183, 2018.
- [19] D. Michailidis, N. Stylianou and I. Vlahavas, "Real Time Location Based Sentiment Analysis on Twitter - The AirSent System", In *SETN '18: 10th Hellenic Conference on Artificial Intelligence*, July 9–15, 2018, Rio Patras, Greece. ACM, New York, NY, USA, Article 4, 4 pages. <https://doi.org/10.1145/3200947.3201052>, 2018.
- [20] S. Saran, L. Singla & P. Singh, "Twitter analytics for integrated research in biodiversity", *Asian conference on remote sensing*. In *Proceedings of the 40th Asian Conference on Remote Sensing ACRS*, 2019.
- [21] M. Albayrak, K. Topal ve V. Altuntaş, "Sosyal Medya Üzerinde Veri Analizi: Twitter", *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, C.22, Kayfor 15 Özel Sayısı, s.1991-1998, 2017.
- [22] M. G. Almatar, H. S. Alazmi, L. Li and E. A. Fox, "Applying GIS and Text Mining Methods to Twitter Data to Explore the Spatiotemporal Patterns of Topics of Interest in Kuwait", *ISPRS Int. J. Geo-Inf.*, 9, 702; 2020, doi:10.3390/ijgi9120702.
- [23] Z Song and J. Xia, "Spatial and Temporal Sentiment Analysis of Twitter data", In: *Capineri, C, Haklay, M, Huang, H, Antoniou, V, Kettunen, J, Ostermann, F and Purves, R. (eds.) European Handbook of Crowdsourced Geographic Information*, Pp.205–221, 2016, London: Ubiquity Press. DOI: <http://dx.doi.org/10.5334/bax.p>. License: CC-BY 4.0
- [24] M. Häberle, M. Werner and X. X. Zhua, "Geospatial text-mining from Twitter – a feature space analysis with a view toward building classification in urban regions", *European Journal of Remote Sensing* 2019, Vol. 52, No. S2,2–11, <https://doi.org/10.1080/22797254.2019.1586451>, 2019.
- [25] S. Alowaidi, M. Saleh, O. Abulnaja, "Semantic Sentiment Analysis of Arabic Texts", (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 2. 2017.
- [26] B. Resch, F. Usländer and C. Havas, "Combining machine learning topic models and spatiotemporal analysis of social media data for disaster footprint and damage assessment", *Cartography and Geographic Information Science*, 45:4, 362-376, 2018, DOI: 10.1080/15230406.2017.1356242.
- [27] L. Tavoschi, F. Quattrone, E. D'Andrea, P. Ducange, M. Vabanesi, F. Marcelloni and L. P. Lopalco, "Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy", *Human Vaccines & Immunotherapeutics*, 16:5, 1062-1069, 2020, DOI: 10.1080/21645515.2020.1714311.
- [28] L. Nemes and A. Kiss, "Social media sentiment analysis based on COVID-19", *Journal of Information and Telecommunication*, 5:1, 1-15, DOI: 10.1080/24751839.2020.1790793, 2021.
- [29] A. Beşkirli, E. Gülbandılar ve E. Dağ, "Metin Madenciliği Yöntemleri ile Twitter Verilerinden Bilgi Keşfi", *ESTUDAM Bilişim Dergisi*, Cilt 2, Sayı 1, 21-25, 2021.
- [30] Çakmak, E. T., & Oğuzlar, A. (2022). Sosyal Medyada Duygu Analizi: COVID-19 Sürecinde 5G Algısı. *International Journal of Social Inquiry*, 15(1), 55-68.
- [31] Aygün, I., Kaya, B., & Kaya, M. (2021). Aspect based twitter sentiment analysis on vaccination and vaccine types in covid-19 pandemic with deep learning. *IEEE Journal of Biomedical and Health Informatics*, 26(5), 2360-2369.
- [32] C. J. Lyu, L. E. Han, K. G. Luli, "COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis", *J Med Internet Res* 2021, 23(6), e24435, 2021.
- [33] S. Liu and J. Liu, "Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis", *Vaccine* 39 (2021) 5499–5505, 2021.
- [34] Marcec, R., & Likic, R. (2022). Using twitter for sentiment analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 vaccines. *Postgraduate medical journal*, 98(1161), 544-550.

- [35] Å. F. Nielsen, “A new anew: evaluation of a word list for sentiment analysis in microblogs”, CEUR Workshop Proc 2011;718:93–8, 2011.
- [36] C. Villavicencio, J. J. X. Macrohon, X. A. Inbaraj, J. H. Jeng, J. G. Hsieh, “Twitter Sentiment Analysis towards COVID-19 Vaccines in the Philippines Using Naïve Bayes”, *Information*, 12, 204, 2021, <https://doi.org/10.3390/info12050204>.
- [37] Ansari, M. T. J., & Khan, N. A. (2021). Worldwide COVID-19 Vaccines Sentiment Analysis Through Twitter Content. *Electronic Journal of General Medicine*, 18(6).
- [38] Çılgın, C., Gökçen, H., & Gökşen, Y. (2022). Twitter’da COVID-19 aşılara karşı kamu duyarlılığının çoğunluk oylama sınıflandırıcısı temelli makine öğrenmesi ile duygu analizi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 38(2), 1093-1104.
- [39] Aslan, S. (2022). BiGRU-CNN Tabanlı Derin Öğrenme Modeliyle Türkiye’deki Covid-19 Aşılarına Yönelik Twitter Duygu Analizi. *International Journal of Pure and Applied Sciences*, 8(2), 312-330.
- [40] Mermer, G., & Özsezer, G. (2023). Discussions About COVID-19 Vaccination on Twitter in Turkey: Sentiment Analysis. *Disaster Medicine and Public Health Preparedness*, 17, e266.
- [41] “Covid19,” [Online]. Available: <https://covid19.saglik.gov.tr/>.
- [42] “WHO,” [Online]. Available: <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---1-september-2023>
- [43] T. Jo, “Text Mining Concepts, Implementation, and Big Data Challenge”, *Studies in Big Data*, Springer, Volume 45, ISBN 978-3-319-91814-3, <https://doi.org/10.1007/978-3-319-91815-0.2019>.
- [44] X. Wu, X. Zhu, G. Wu and W. Ding, “Data Mining with Big Data”, *IEEE Transactions On Knowledge And Data Engineering*, Vol. 26, No. 1, 2014.
- [45] G. Salton, C.S. Yang, “On the specification of term values in automatic indexing”. *J. Doc.* 29, 351–372, 1973.
- [46] “Wordnet,” [Online]. Available: <https://wordnet.princeton.edu/>.
- [47] “Orange,” [Online]. Available: <https://orangedatamining.com/>.
- [48] J. Demšar and B. Zupan, “Orange: Data Mining Fruitful And Fun”, *Proceedings of the 15th International Multiconference, Information Society-IS*, Volume A, 2012.
- [49] M. Hu and B. Liu, “Mining and Summarizing Customer Reviews”, *KDD’04*, August 22–25, Seattle, Washington, USA, 2004.
- [50] C. J. Hutto, E. Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text”, *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, Michigan, USA, June 1–4, 2014.