



İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİYOLOJİK VERİ BİLİMİ ANABİLİM DALI

**Türkiye’de COVID-19 Aşılması Hakkında Toplumsal
Duygusal Durumun Sosyal Medya Verileri Üzerinde Makine
Öğrenmesi Metotları Kullanılarak İncelenmesi**

Yüksek Lisans Tezi

Tuna Arda DEMİRÖZ

Ocak 2024



İSTANBUL MEDENİYET ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİYOLOJİK VERİ BİLİMİ ANABİLİM DALI

**Türkiye’de COVID-19 Aşılması Hakkında Toplumsal
Duygusal Durumun Sosyal Medya Verileri Üzerinde Makine
Öğrenmesi Metotları Kullanılarak İncelenmesi**

Yüksek Lisans Tezi

Tuna Arda DEMİRÖZ

Danışman

Prof. Dr. Hasan GÜÇLÜ

Ocak 2024

TEZ JÜRİSİ ONAYI

Tuna Arda DEMİRÖZ tarafından hazırlanan "Türkiye’de COVID-19 Aşılması Hakkında Toplumsal Duygusal Durumun Sosyal Medya Verileri Üzerinde Makine Öğrenmesi Metotları Kullanılarak İncelenmesi" başlıklı bu Yüksek Lisans Tezi, Biyolojik Veri Bilimi Anabilim Dalı’nda hazırlanmış ve jürimiz tarafından kabul edilmiştir.

JÜRİ ÜYELERİ

İMZA

Tez Danışmanı:

Unvanı Adı Soyadı Prof. Dr. Hasan GÜÇLÜ
Kurumu İstanbul Medeniyet Üniversitesi

Üyeler:

Unvanı Adı Soyadı Prof. Dr. Handan ANKARALI
Kurumu İstanbul Medeniyet Üniversitesi

Unvanı Adı Soyadı Dr. Öğr. Üyesi Emine YAYLALI
Kurumu İstanbul Teknik Üniversitesi

Tez Savunma Tarihi: 25/01/2024

BEYANLAR

Yazım ve Kaynak Gösterme Kılavuzu Beyanı

Danışmanlığımda yazılan bu tezin APA yazım ve kaynak gösterme kılavuzunda belirtilen kurallara uygun olarak yapılandırıldığı ve bu kılavuzun APA 7.0 kaynak gösterme standartlarının bu tezde tutarlı olarak uygulandığı tarafımdan incelenerek teyit edilmiştir.

Prof. Dr. Hasan GÜÇLÜ

Etik İlkelere Sadakat Beyanı

Hazırladığım bu tezin tamamen kendi çalışmam olduğunu, akademik ve etik kuralları gözeterek çalıştığımı ve her alıntıya kaynak gösterdiğimi beyan ederim.

Tuna Arda DEMİRÖZ

ÖZET

Türkiye’de COVID-19 Aşılması Hakkında Toplumsal Duygusal Durumun Sosyal Medya Verileri Üzerinde Makine Öğrenmesi Metotları Kullanılarak İncelenmesi

Tuna Arda DEMİRÖZ

Yüksek Lisans Tezi, Biyolojik Veri Bilimi Anabilim Dalı

Danışman: Prof. Dr. Hasan GÜÇLÜ

Ocak 2024

İnternet ve sosyal medya platformlarının evrimi ile birlikte bilgi üretimi, tüketimi, dolaşımı ve aktarımı, önceki dönemlere göre benzersiz bir hız kazanmıştır (Şirzad, 2022). Güncel istatistiksel ve sosyal çalışmalar ise Facebook, Twitter ve Instagram gibi sosyal medya platformlarının, insan davranışlarını etkileyen oldukça önemli verilere sahip olduğunu göstermektedir. Bu veriler, COVID-19’un yayılmasını izlemek, önlemek ve kontrol altına almak için kullanılabilir, özellikle insanların duygusal durumlarını analiz etmek ve anlamak için büyük bir kaynak sunmaktadır (Sarirete, 2021). Bu tez çalışmasında, Türkiye’de COVID-19 aşılama sürecine yönelik toplumsal duygusal tepkileri anlamak ve analiz etmek amacıyla Twitter verileri kullanılmıştır. Çalışma, aşılamanın toplumsal duygu durumunu belirli bir zaman diliminde Türkçe tivitler aracılığıyla incelemeyi ve aşılama ile ilgili düşünceleri, endişeleri, memnuniyeti ve diğer duygusal tepkileri anlamayı hedeflemiştir. Bu bağlamda, Twitter verileri istatistiksel yöntemler, makine öğrenmesi, Doğal Dil İşleme (NLP) teknikleri ve sözlük tabanlı yaklaşım kullanılarak analiz edilmiştir. Üzerine çalışılan veri seti, Şubat 2020’den Ocak 2023’e kadar dünya genelinde atılan Türkçe tivitlerden oluşmaktadır. Söz konusu dönemde aşılama ile ilgili yaklaşık 200 milyon tivit filtreledikten sonra elde edilen 38,628 tivit analiz edilmiştir.

Anahtar Kelimeler: COVID-19, Duygu Durum Analizi, Sözlük Tabanlı Yaklaşım, Doğal Dil İşleme, Twitter Verileri

ABSTRACT

Sentiment Analysis about COVID-19 Vaccination in Turkey by Using Machine Learning Methods on Social Media Data

Tuna Arda DEMİRÖZ

Master's Thesis, Department of Biological Data Science

Supervisor: Prof. Dr. Hasan GÜÇLÜ

January 2024

With the evolution of the Internet and social media platforms, the production, consumption, circulation and transfer of information has gained unprecedented speed compared to previous periods (Şirzad, 2022). Recent statistical and social studies show that social media platforms such as Facebook, Twitter and Instagram have significant data that influence human behaviour. These data offer a great resource for analysing and understanding people's emotional states, which can be used to monitor, prevent and control the spread of COVID-19 (Sarirete, 2021). This study uses Twitter data to understand and analyse public emotional reactions to the COVID-19 vaccination process in Turkey. The study aimed to examine the public sentiment of vaccination through Turkish tweets in a specific time period and to understand thoughts, concerns, satisfaction, and other emotional reactions to vaccination. In this context, Twitter data were analysed using statistical methods, machine learning, Natural Language Processing (NLP) techniques and Lexicon-based approach. The studied dataset consists of Turkish tweets posted worldwide from February 2020 to January 2023. After filtering approximately 200 million tweets related to vaccination in the period in question, 38,628 tweets were analysed.

Keywords: COVID-19, Sentiment Analysis, Lexicon Based Approach, Natural Language Processing, Twitter Data

İçindekiler

Tez Jürisi Onayı.....	iii
Beyanlar.....	iv
Özet.....	v
Abstract	vi
Giriş	1
İlgili Çalışmalar	3
Metodoloji.....	6
Veri Toplama.....	6
Veri Filtreleme ve Temizleme	6
Türkçe Tivitleri Filtreleme.....	9
CSV Formatını XLSX Formatına Çevirme	9
Tekrar Eden Tivitleri Silme.....	10
Extended Text Bilgilerini Sorgulama.....	10
Aşılı ile İlgili Tivitleri Seçme.....	10
Retweetleri Silme.....	10
"@ " ve URL ile Başlayan Kelimeleri Silme.....	11
Veri Etiketleme Süreci	11
Yönerge.....	13
Bulgular ve Analiz	14
Aşılılamaya Karşı Genel Duygu Durum.....	14
Kelime Analizi	16
Hashtag Analizi.....	20
Sözlük Tabanlı Yaklaşım Sonuçları	25
Hata Matrisi (Confusion Matrix) Analizleri	30
Aşılılamaya Karşı Duygu Durumun Zamana Göre Değişimi	34
Sonuçlar ve Tartışma	36

Sınırlılıklar.....	39
Referanslar.....	42
Ekler.....	47
Veri Filtreleme ve Temizleme Kodları.....	47
CSV Formatını XLSX Formatına Çevirme Kodu	47
Tekrar Eden Tivitleri Silme Kodu.....	48
Extended Text Bilgilerini Sorgulama Kodu.....	48
Aşı ile İlgili Tivitleri Seçme Kodu.....	49
“@” ile Mention ve URL İçeren Kelimeleri Silme Kodu.....	49
Etiketleme Yönergesi.....	51
Şekiller Listesi	56

Giriş

COVID-19, ilk olarak Çin'in Vuhan Eyaleti'nde 2019 yılı Aralık ayının sonlarında solunum yolu problemleri belirtileri (ateş, öksürük, nefes darlığı) gelişen bir grup hastada yapılan araştırmalar sonucunda 13 Ocak 2020'de tanımlanan bir virüstür ve neden olduğu hastalık da COVID-19 olarak isimlendirilmiştir (T.C. Sağlık Bakanlığı, 2023). Dünya Sağlık Örgütü (WHO), 12 Mart 2020'de COVID-19'u pandemi olarak ilan edilmiştir. Salgının yayılmasıyla birlikte pandemi hakkında haberler ve güncellemeler sosyal medyada olağanüstü bir hızla yayılmıştır. Özellikle kapanma ve karantina dönemlerinde sosyal medyanın kullanımı oldukça artmıştır. Bu nedenle, COVID-19 ile ilgili büyük miktarda veri sosyal medyada paylaşılmış ve sürekli olarak bu veriler yayılmıştır. Bu veriler pandemiye ve insanların bu pandemiyle ilgili algıları ve endişeleri hakkında önemli bilgiler sunulmuştur (Catelli et al., 2023). Başlangıçtan bu yana, bir koronavirüs hastalığı olan COVID-19 dünya çapında hızla yayılmış ve bu da çevrimiçi platformlarda tartışmaların artmasına neden olmuştur (Ahmed et al., 2020).

COVID-19 salgınının uluslararası boyutta yayılmasıyla birlikte, COVID-19 aşısıyla ilgili konular halkın ilgisini gittikçe daha çok çekmeye başlamış ve aşılama çalışmaları, salgının kontrolü ve sona erdirilmesi açısından büyük önem taşımıştır (Lyu et al., 2021). Aşılama ile bağışıklık elde edilmesinin sadece bireysel düzeyde sonuçları olmayacaktır. Aynı zamanda ve belki de daha önemlisi aşılama toplumsal düzeyde de bağışıklığın artmasını sağlayacaktır. Toplumda belli bir hastalığa karşı bağışıklığı olan insanların oranı arttıkça o hastalığın salgına dönüşme riski de azalacaktır (Akbulak & Çöl, 2022). Toplumda bağışıklığı olan insanların oranı belirli bir değerin üstüne çıktığında toplumsal bağışıklık oluşmuştur. Özellikle COVID-19 pandemisi döneminde aşı reddi ve aşıya karşı olan tereddüt toplum sağlığını ciddi şekilde tehdit eder bir boyuta ulaşmıştır

(Kader, 2019). Bu durumun başlıca sebepleri aşı güvenliği, yan etkiler, aşı hakkında eksik bilgi ve farkındalık eksikliğidir. Aşı karşıtlığı, COVID-19 aşılmasında hedeflenen oranlara ulaşılmasını engelleyen başlıca faktördür (Akbulak & Çöl, 2022). COVID-19 aşılarının geliştirilmesi ve uygulanmaya başlanması, ülkelerin pandemiyle başa çıkmak adına oluşturdukları planların en önemli parçası olmuştur. Aşılarla geniş bir erişim sağlanmasıyla birlikte, Türkiye'nin pandemiyle mücadelede arzu edilen seviyede aşılama oranlarına ulaşamadığına dair veriler ortaya atılmıştır. Bu durum, aşıya ulaşım zor olmadığı halde, toplumun genelinde aşılama konusunda belirli bir isteksizlik veya diğer faktörlerin etkisiyle gerçekleşmiş olabilme ihtimali vardır (Narmanli, 2022). Çalışmanın amacı, duygu ve düşünceleri ifade etmek için en çok kullanılan sosyal medya platformlarından biri olan Twitter'da atılmış aşı ile ilgili Türkçe tivitleri analiz ederek aşılama karşı olan duygu durum hakkında bilgi edinebilmek ve aşılama karşı duygu durumun sözlük tabanlı modeller ile çıkarımının ne denli doğru veya yanlış olabileceğini araştırmaktır.

2006 yılında kurulan ve şu anda X olarak tekrar adlandırılan Twitter, çevrim içi bir sosyal medya platformudur. Twitter kullanıcıları, metinler, fotoğraflar ve videolar içeren gönderileri, yani tivitleri paylaşabilmektedirler. Kayıtlı kullanıcılar ise tivit atabilir, tivitleri beğenebilir, tivitleri “retweet” edebilir ve diğer kayıtlı kullanıcılara doğrudan mesaj (DM) gönderebilirler (Wikipedia, 2023). Twitter'ın 2023'ün ilk yarısında Türkiye'deki abone sayısı 20,4 milyondur (In Business, 2023). Mart 2023 verilerine göre Türkiye'de Twitter kullanıcısı olduğunu belirten kişilerin %35,9'u kadın, %64,1'i erkektir. Bu kişilerin yaş gruplarına bakıldığında 18-24 yaş grubunun %32,4, 25-34 yaş grubunun %30,1, 35-44 yaş grubunun %19,8, 45-54 yaş grubunun %14,5, 55-64 yaş grubunun %11, 65 yaş ve üzeri grubunun %5,9 şeklinde olduğunu görülmektedir (Sosyal Ağ Haritası, 2023).

İlgili veri setindeki tivitler öncelikle insan gücü ile manuel olarak negatif, pozitif ve nötr sınıflarında etiketlenmiştir. Etiketli veri setinden çıkan sonuçlar analiz edilmiş ve halkın aşılamaya karşı duygu durumu değerlendirilmiştir. Bu değerlendirmelerde aşılamaya karşı halkın genel duygu durumunun ne yönde olduğu, aşılama ile ilgili atılan negatif, pozitif ve nötr duygu durum içeren tivitlerde hangi kelime ve hashtaglerin ne sıklıkla kullanıldığı incelenmiş bunlarla ilgili elde edilen sonuçlar paylaşılmış, grafik ve şekiller ile birlikte sunulmuştur. Veri seti yaklaşık üç seneye ait veri seti olduğu için aşılamaya karşı sosyal medyadaki duygu durumun zaman göre değişiminin de incelenme fırsatı oluşmuştur. Negatif, pozitif ve nötr oranlarının zamana göre değişimi, pozitif olarak etiketlenen metinlerde en sık kullanılan kelimeler, negatif olarak etiketlenen metinlerde en sık kullanılan kelimeler ve nötr olarak etiketlenen metinlerde en sık kullanılan kelimeler de analiz edilip grafik ve şekiller ile birlikte sunulmuştur. İnsan gücü ile manuel olarak etiketlenen bu veri seti sözlük tabanlı bir model ile karşılaştırılarak analiz edilmiştir. Bu analizdeki amaç Türkçe dilinde atılan bir tivit belli bir fikre karşı duygu durumun sözlük tabanlı bir model ile çalışma durumunu test edebilmektir. Çalışmada Şubat 2020 ve Ocak 2023 tarihleri arasında atılmış tivitler kullanılmıştır.

İlgili Çalışmalar

Pandemi dönemindeki toplumun aşılamaya karşı olan duygu durumu ile ilgili aşı kararlarını etkileyen faktörler, aşı karşıtlığı ve insanların aşılamaya karşı bakışı hakkında birçok çalışma mevcuttur. Yılmaz ve arkadaşlarının yaptığı çalışmada Türkiye’de insanların COVID-19 aşısına bakışını değerlendirmek için 1293 katılımcı ile Aralık 2020’de çevrim içi anket yapılmıştır. Katılımcıların %41,2’si COVID-19 aşısına olumlu yaklaşırken %20,9’u COVID-19 aşısına olumsuz yaklaşmış, kalan %37,9’u kararsız olduklarını belirtmişlerdir. Kararsız yaklaşanların %75,9’u ise COVID-19 aşısının yeni bir aşı olduğunu ve bu aşının yan etkilerinin korkutucu

olduğunu belirtmişlerdir (Yılmaz et al., 2021). Çağlar ve Selvi, Twitter verilerini kullanarak COVID-19 aşılı (Pfizer/BioNTech, Sinopharm, Sinovac, Moderna, Oxford/AstraZeneca, Covaxin ve Sputnik V) hakkında atılan tivitleri zamansal ve mekansal bağlamdaki dolaşımını veri madenciliği teknikleriyle analiz etmiştir. Bu analizler sonucunda afet ve kriz yönetimleri konuları üzerine önemli çıkarımlar yapabilmeyi hedeflemişlerdir. Çiçen, aşı kararını etkileyen faktörleri incelerken aşıya ücret ödenmesi ve aşı olanlara ücret ödenmesi durumlarının aşılama karşı olan duygu durumu etkileyip etkilemeyeceğini araştırabilmek için 327 katılımcı ile 2022 yılının Mayıs ayında bir anket araştırması yapmıştır. Bu grubun %33'ü aşının ücretli olması durumunda aşılama karşı negatif bir duygu durum sergileyeceklerini belirtmişlerdir. Grubun %95'i ise aşılama karşılığında ücret ödenmesi durumunda aşılama karşı negatif bir duygu durum sergileyeceklerini belirtmişlerdir (Çiçen, 2023). Bu sonuçlar insanların aşıya olan güveninin oldukça düşük olduğunu ortaya çıkarmaktadır. Aşılama sonucunda kişiye ücret ödenmesi durumunda kişinin aşıya duygu durumunun negatif olduğu düşünülmektedir. Bunun yerine kişinin ücret ödediği bir aşıya karşı duygu durumu daha pozitif yöndedir. Aşılama konusundaki duygu durum analizi üzerine ise özellikle son dönemlerde birçok çalışma mevcuttur. Bu çalışmalar, küresel bir pandemi olan COVID-19'un ortaya çıkması ve yayılması sürecinde, aşı geliştirme çabalarının hızla yoğunlaştığı ve uygulandığı dönemden oldukça etkilenmiştir. İnsanların COVID-19 pandemisindeki kapanma, karantina ve kısmi kısıtlamalar döneminde sosyal medya kullanımında büyük ölçüde farklılık olmuştur (OKTAY, n.d.). Pandeminin ilk aylarında virüs hakkında bilgi eksikliği nedeniyle aşı çalışmalarına kıyasla virüse yakalananların nasıl tedavi edileceği hakkında atılmış tivitler daha fazladır (Kurt, 2021). Fakat veri setinden yola çıkarak aşılama ile ilgili atılan tivitlerin oranının gittikçe arttığını gözlemlenebilmektedir. Daha önce yapılan çalışmalarda öncelikle bu büyüklükte bir veri seti ve zaman dilimi üzerinde bir çalışmaya

rastlanmamıştır. Yapılan çalışmalarda genellikle kelime ve içerik bazlı analizler yapılmıştır. Narmanlı, çalışmasında Türkiye’de aşı kararsızlığı yaşayan Twitter kullanıcılarının tivitlerinde aşı ile ilgili birlikte hangi konuların tartışıldığını, aşuya karşı güvensizliğinin nedeninin neler olarak ifade edildiğini ve aşuya karşı kararsızlık yaşayan kullanıcıların tivitlerinde hangi yanlış bilgilerin öne çıktığını infodemi bağlamında değerlendirmek için 20 Eylül-29 Eylül 2021 tarihleri arasında “aşı” hashtagiyle atılan 2,349 tivit içeriği çözümlemesi yöntemiyle analiz etmiştir. Aşıya karşı negatif duygu durum içerisinde olan kullanıcıların tivitlerinin tematik analizleri sonucunda 2,349 tivitinin %35,4’ünde aktörler, kurumlar ve uygulamalardan bahsedilmiştir. Bu temaları %30,8 ile aşı güvenilirliği izlemiştir (Narmanlı, 2022). Sütçü ve arkadaşlarının yaptığı çalışmada COVID-19 aşısı ile ilgili atılan 1,258 tivit incelenmiştir. Bu tivitlerde en sık kullanılan kelimeler analiz edilmiş ve tivitlerin temaları içerik analizi yöntemleri ile belirlenmiştir. Toplam 1,258 tivitinin bu temalara göre dağılımları belirlenmiş ve en çok tivit atılan temanın %32,8 ile pandeminin gerçek olduğuna inanmayan söylemler kategorisine dahil olduğu belirlenmiştir. Bu çalışmada tema analizine ek olarak yapılan kelime sıklığı analizine göre 1,258 tivitinin 1,012’sinde “Plandemi” kelimesi geçmiştir. Bu kelimeleri sırasıyla “Biyontechyanetki”, “Maske” ve “Covid” kelimeleri 67, 66 ve 62 kez ile takip etmiştir (Sütçü et al., n.d.). Aslan, çalışmasında Twitter API üzerinden toplanan, Haziran 2021 ve Eylül 2021 arasında iki aylık bir süreçte aşılama ile ilgili atılan 176,345 tivit üzerinde BiGRU-CNN tabanlı bir derin öğrenme modeliyle analizler gerçekleştirmiştir. CNN, yani Evrişimli Sinir Ağı, bir metni birden fazla cümleye bölerek cümle düzeyinde bir içerik ifadesi yaratabilmek için her cümleden farklı taneciklere sahip n-gram bilgilerini çıkarmıştır. BiGRU, yani Çift-yönlü Tekrarlayan Birim ise, cümlelerdeki bağlamsal ve anlamsal özellikleri çıkarmış ve buna dikkat puanı ekleyerek bir sonuç elde etmiştir (Yan et al., 2021). Çeşitli veri temizleme aşamalarından geçmiş 176,345 tivitinin her birinin Textblob kütüphanesi kullanılarak polaritesi ve

öznelliği hesaplanmıştır. Polarite değerlerini -1, 0 ve 1 (negatif, pozitif ve nötr) şeklinde değerlendiren çalışma 176,345 tivitın %49,6'sının nötr, %34,9'unun pozitif ve %15,5'inin ise negatif polariteye sahip olduđu sonucuna ulaşmıştır (Aslan, 2022). Lyu ve arkadaşlarının yaptığı çalışmada konu modelleme teknikleri kullanılarak 2020 Mart ayından itibaren geçen bir senelik zaman diliminde atılmış 1,499,421 tivitlik bir veri setinde zaman içinde deđişen duygu durumları anlamlandırılmaya çalışılmıştır. Bu anlamlandırmalar sonucunda halkın aşılama gelişmeleriyle birlikte aşılama karşı duygu durumu ve endişenin ne denli deđiştii hakkında bilgi edinmeye odaklanmıştır. Sarirete ise çalışmasında pandemi sürecinde insanların aşılama karşı olan duygu durumunu tivitler üzerinden analiz etmek için bir çalışma gerçekleştirmiştir. Bu çalışmada duygu durum analizi (sentiment analysis) ve doğal dil işleme (natural language processing) tekniklerini birlikte kullanmayı hedeflemiştir. #COVIDVaccine hashtagine sahip tivitler üzerinde bir araştırma yapan çalışma sonuçlarda pozitif ve negatif duygu durumların bulunduđunu ve negatif duygu durumun daha baskın olduđunu belirtmiştir.

Metodoloji

Veri Toplama

Tivitler Şubat 2020 ve Ocak 2023 tarihleri arasında belirli aylarda Twitter API kullanılarak çekilmiştir. Bu veriler toplanırken "SARS-CoV-2", "COVID", "coronavirus", "nCov2019" anahtar kelimeleri ile filtreleme yapılmıştır.

Veri Filtreleme ve Temizleme

Bu 200 milyona yakın tivit üzerinde Python kodları ile temizleme ve filtreleme işlemleri gerçekleştirilmiştir.

- Veri setinde her bir tivitın hangi dilde atıldıđı belirten “lang” isimli bir sütün bulunmaktadır. Bu etiketler aracılıđıyla, toplamda 200 milyona yakın tivit ierisinden sadece Trke dilinde yazılmıř olanlar tespit edilmiř ve sadece bu dildeki veriler analiz iin filtrelenmiřtir.
- Trke dilinde atılan tivitlere ulařtıktan sonra, tekrar eden metinler ayıklanmıřtır. rneđin, birbirleriyle tamamen rtřen  metin bulunması durumunda, bu metinlerden yalnızca biri veri setinde korunmuř; diđer iki metin ise veri setinden ıkarılmıřtır.
- Ařılamaya karřı duygu durumu analiz etmek amacıyla, ilk olarak tivit ieriklerinde frekansı en yksek olarak yer alan kelimeler belirlenmiřtir. Bu kelimeler arasından ařılamayla ilgili olanları belirlemek adına en sık tekrar eden kelimelerin ařılama ile iliřki analizi gerekleřtirilmiřtir. alıřılacak veri setinin, sadece ařılama konusunu ieren tivitleri kapsamasını sađlamak amacıyla, bu kelimeleri iermeyen tivitlerin eleme iřlemleri gerekleřtirilmiřtir. Ardından, geriye kalan veriler zerinde temizleme iřlemlerine devam edilmiřtir.
- Ařılamayla ilgili Trke tivitleri analiz etme srecinde, birbirini duygu durum olarak tekrar eden fakat aynı metin olmayan tivitleri elemek iin elde edilen veri setinde yer alan retweet edilmiř ieriklerin temizlenmesi amalanmıřtır (Lyu et al., 2021). Bu temizleme iřlemini gerekleřtirmek iin Excel zerinde kullanılan bir yntem takip edilmiřtir. Retweet edilmiř tweetlerin “@RT” ile bařladıđı tespit edilmiř ve tivitler veri setinden silmek amalanmıřtır.

Bu adımlardan sonra 200 milyona yakın tivit, ařılama hakkında atılmıř 38,628 tivite dřrlmřtir. Bu filtreleme ve temizle iřlemlerinin ařamaları Veri Filtreleme ve Temizleme

bölümünde ve bu işlemleri kodları ise Veri Filtreleme ve Temizleme Kodları bölümünde detaylı olarak gösterilmiştir.

Veri filtreleme sürecinde temel olarak yedi ana adım esas alınmıştır. İlk olarak, büyük veri setleri içerisindeki Türkçe metinlerin belirlenmesi için bir filtreleme gerçekleştirilmiştir. Ardından, "text" sütununda tekrar eden metinlerin varlığı kontrol edilmiş ve bu durumda sadece bir satır bırakılarak diğer tekrarlanan satırlar silinmiştir. Sadece aşılama ile ilgili metinlere ulaşabilmek için, aşılama ile ilgili kelimelerin bulunduğu metinler seçilerek, aşılama ile ilgili kelimeler içermeyen metinler veri setinden çıkarılmıştır. Veri setinde tekrar eden duygu durumları elemek ve daha kararlı bir veri seti üzerinde çalışabilmek için retweet edilmiş metinler Excel'deki fonksiyonlar ile tespit edilip silinmiştir. Son olarak metinlerdeki okuma kolaylığını artırabilmek ve metinlerdeki kullanıcı adlarının ve paylaşılan linklerin veri gizliliği gereği kaldırılması için "@" ile başlayan kelimeler ve URL içeren kelimeler (http: ile başlayan URL'ler) veri setinden silinmiştir. Bu aşamada metin tamamı değil sadece metin içerisinde bahsi geçen içerikler silinmiştir. Bu adımlar, veri setinin temizlenmesini ve sadece projenin odaklandığı konuyla ilgili metinlerin veri setine dahil olmasını amaçlamaktadır. Türkçe metinleri belirleme, tekrarlanan içerikleri ortadan kaldırma ve aşılama ile ilgili olmayan metinleri eleme süreçleri, veri setinin daha odaklı, tutarlı ve anlamlı olmasına katkı sağlamaktadır. Veri filtreleme aşamaları Şekil 1'de detaylı şekilde gösterilmiştir.



Şekil 1 Veri Filtreleme Aşamaları

Türkçe Tivitleri Filtreleme

Veri setimizdeki metinler birçok dilden toplanmıştır. Fakat veri içerisinde Türkçe olan bilgiler “lang” sütununda “tr” bilgisi ile tutulmaktadır. Elimizdeki veriler orijinal ham JSON formatından çevrilmiş CSV (virgülle ayrılmış değerler) formatındadır. Geliştirilen Python kodu, bu dosyalardan sadece “lang” sütunu değeri “tr” olan satırları içeren yeni bir CSV formatında dosya oluşturmaktadır.

CSV Formatını XLSX Formatına Çevirme

Türkçe dilindeki verilerin .csv formatında filtrelenmesi, Python programlama dili kullanılarak daha etkili bir şekilde gerçekleştirilebilmekte ve bu sürecin ardından elde edilen .csv dosyalarının .xlsx formatına dönüştürülmesinde de yine Python kodlarından yararlanılmaktadır. Bu yöntemler, veri analizi ve manipülasyonu için Python'un sağladığı hız ve esneklik avantajlarından faydalanarak işlemlerin daha hızlı ve verimli bir şekilde tamamlanmasını sağlamaktadır.

Tekrar Eden Tivitleri Silme

Tivitlerdeki analizi yanılmamak adına, birden fazla metinden sadece bir tanesini veri setinde koruma amacıyla Python kodu kullanarak bir filtreleme işlemi gerçekleştirilmiştir. Bu yöntem, analiz sürecinde çeşitli metin kaynaklarının etkileşimini azaltarak, sonuçların tek bir metin örneği üzerinden daha odaklı ve güvenilir olmasını sağlamaktadır.

Extended Text Bilgilerini Sorgulama

Veri setinde “text” sütununun bazı metinlerinde Twitter API’den kaynaklı yarıda kesilmeler mevcuttur. Bu metinler eğer yarıda kesilmiş ise “extended_tweet.full_text” satırındaki değer dolu görünmektedir. Eğer bu değer doluysa “extended_tweet.full_text” satırındaki değeri alınıp “text” sütunundaki metnin yerine konulacağı bir Python kodu oluşturmuştur.

Aşı ile İlgili Tivitleri Seçme

Çalışmada sadece aşılama ile ilgili duygu durumların ölçülebilmesi için, sadece aşılama ilgili kelimeleri içeren metinler veri setinde korunmuş; aşıyla ilgili olmayan metinler ise veri setinden silinmiştir. Bu işlem, Python kodu kullanılarak gerçekleştirilmiştir, bu sayede analizin odak noktası olan aşılama ile ilgili duygu durum içeriklerin belirlenmesi sağlanmıştır.

Retweetleri Silme

Excel üzerinde yapılan alfabetik filtre ya da “@RT” ile başlayanları seç fonksiyonuyla retweet edilmiş metinler tespit edilebilir durumdadır. Retweet olan metinlerin başında “@RT” ifadesi bulunmaktadır. Retweet içeren metinler veri setinden silinmiştir.

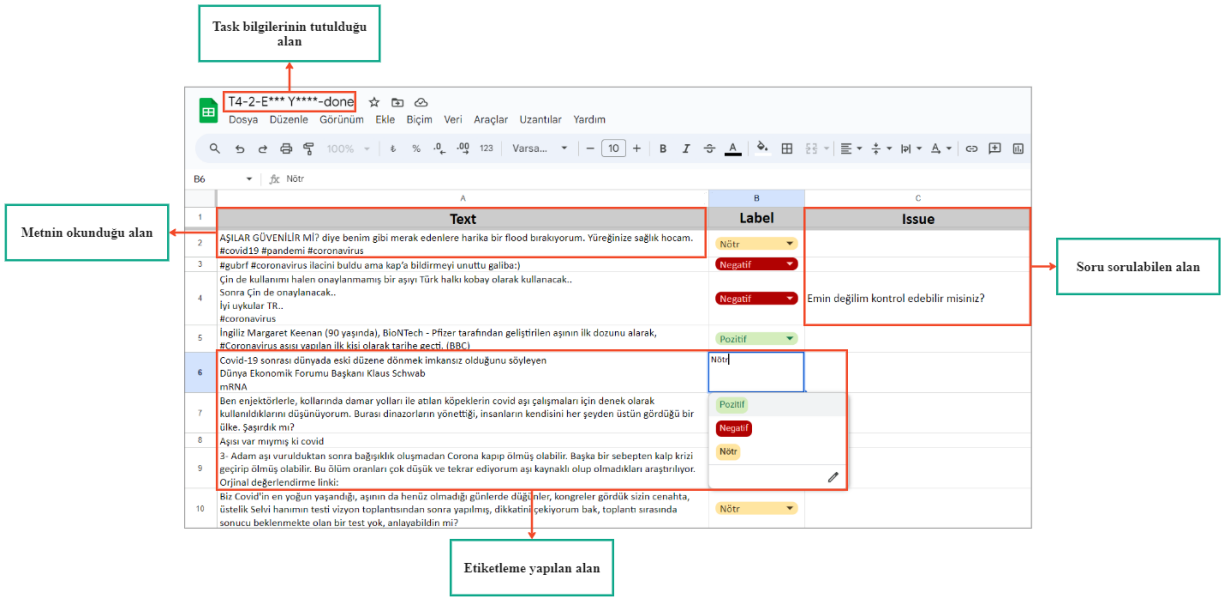
"@" ve URL ile Başlayan Kelimeleri Silme

Tivitlerin içerisinde fazla sayıda "@" ile kullanılan mention ve ilgili bir websitesine yönlendirmek için URL'ler mevcuttur. Etiketleme aşamasında buradaki bilgiler metnin duygu durumunu belirleyen bilgiler vermediğine karar verilmiştir. URL'ler özel ve bilgi vermeyen kodlardır. Mentionlarda ise ("@" ile başlayan kelimeler) sadece Twitter kullanıcı isimleri mevcuttur. Bu kelimeleri metin içerisinden silmek hem etiketleme hem analiz sürecini kolaylaştırmıştır. İlgili kelimeler Python kodu kullanarak metinlerin içerisinden kaldırılmıştır. Metnin orijinal yapısına zarar verilmemiştir.

Veri Etiketleme Süreci

Filtreden geçirilmiş 38,628 tivit etiketleyici takımına dahil edilen 30 farklı kişi tarafından etiketlenmiştir. Bu kişiler öncelikle bir test aşamasından geçirilmiştir. Kişilere yönerge iletildikten sonra test için seçilmiş 50 metin içeren bir görev tanımlanmıştır. Öncesinde doğru cevapları belirlenen bu metinlerde etiketleyici adaylarından %90 ve üzeri başarı beklenmiştir. Hazırlanan veriler 500 adet metin içeren görevler halinde etiketleyici takım ile paylaşılmıştır. Kişiler etiketlemelerini tamamladığını bir denetleyici 500 adet metinden rastgele 50 metni değerlendirmiş, kişinin hata oranı %10 veya üzerindeyse görev sahibi kişiye veri setini tekrar değerlendirmesi için ek süre ve geri dönüşler vermiştir. Sonrasında ise verinin kontrol edilmesi için 3 farklı kontrolcü her bir görevin %60'ını tekrar kontrol etmiştir. Önemli bir not olarak, etiketleyiciler metinleri işlerken tarih, tivit atan kişileri Twitter'daki takipçi sayısı, tivitın atıldığı tarih veya tivitı atan kişilerin kullanıcı isimleri gibi dışsal bilgileri görmemişlerdir. Bu bağımsızlık, etiketleme sürecinin tarafsız ve objektif bir şekilde gerçekleşmesini sağlamıştır. Projede etiketleme işlemleri için kullanılacak platform seçimi, mevcut etiketleyici takımın aşına olduğu bir ortamın tercih edilmesi ilkesine dayanmıştır. Etiketleyici takımın daha önceki metin etiketleme çalışmalarının büyük bir

kısmı, Google Drive üzerinde bulunan Google Sheets üzerinden gerçekleştirilmiştir. Etiketleme yapılacak platform seçimi tamamlandıktan sonra Google Sheets üzerinden görev dağıtımları yapılmıştır. Etiketleyici takım ile iletişim ise Discord uygulaması üzerinden kurulmuştur. Bu platformlar, etiketleyici ekibin kullanımına zaten aşina oldukları araçlar olduğu için seçilmiştir. Dağıtılan görevler üzerinde her bir etiketleyicinin karşılaştığı üç ana sütun bulunmaktadır. Bu sütunlar şunlardır: "text" (metin), "label" (etiket), ve "issue" (sorun). Etiketleyiciler, "text" sütunundaki metni okuduktan sonra daha önceden Google Sheets üzerinde belirlenmiş olan "label" sütunundan, Google Sheets içerisinde bulunan veri doğrulama fonksiyonu ile belirlenmiş "negatif", "pozitif" ve "nötr" seçenekleri arasından tercihlerini yapmışlardır. Etiketleyici takım, "issue" sütununu, metinle ilgili sorularını denetleyicilere iletmek üzere kullanılmıştır. Etiketleme arayüzünün örnek bir görseli Şekil 2’de gösterilmiştir. Bu proje Co-one Arge ve Teknoloji A.Ş. ile yürütülmüştür.



Şekil 2 Etiketleme Arayüzü Örnek Görsel

Veri etiketleme sürecinin tüm aşamaları Şekil 3’de detaylı bir şekilde gösterilmiştir.



Şekil 3 Veri Etiketleme Süreci Akışı

Yönerge

Projedeki amacı, görev özetini ve etiketleme detaylarını etiketleyici takıma aktarabilmek amacıyla bir etiketleme yönergesi oluşturulmuştur. Bu yönerge oluşturulurken projeyi gerçekleştirmedeki amaç aktarılmıştır. Proje için bir etiketleme yönergesi oluşturulmadan önce yönergede kullanılmak üzere uygun örnek metinlerin bulunması, etiketleyici takımın incelemesi için örnek bir veri setinin oluşturulması ve etiketleme sırasında karşılaşılabilecek uç durumların önceden belirlenebilmesi

için 446 metin etiketlenmiştir. Bu örnek 446 metnin içerisinde her bir kategori için açıklamalarıyla birlikte dörder metin örneği verilmiştir. Yine etiketleyici takımın projeyi daha iyi şekilde anlayabilmesi için 446 etiketli metin içerisinde seçilen 200 metin, yönerge içerisine bir Google Drive linki olarak yerleştirilmiştir. Kişilerin yönergeyi okuyup anladıktan sonra bu örnek etiketli veri setini incelemeleri tavsiye edilmiştir. Bunun amacı, etiketleyicilerin süreci daha iyi kavramalarına ve etiketleme işlemlerini başarıyla gerçekleştirmelerine katkıda bulunmaktır. Yönergenin görselleri Etiketleme Yönergesi bölümünde bulunmaktadır.

Bulgular ve Analiz

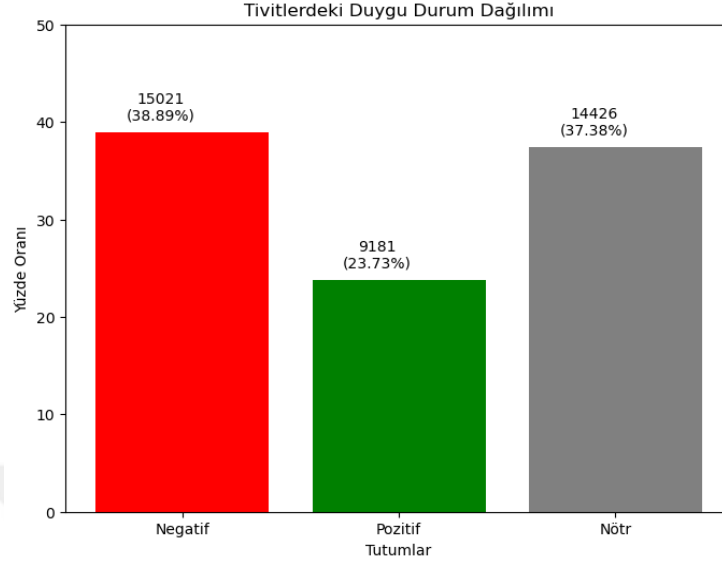
Aşılılamaya Karşı Genel Duygu Durum

Aşılılamaya yönelik duygu durumu ölçmek amacıyla incelenen 38,628 tivit içinde, en yaygın duygu durumun negatif olduğu gözlemlenmiştir. Bu bağlamda, 15,021 tivit aşılılamaya karşı negatif bir duygu durum yansıtmaktadır. Pozitif duygu durumlar ise 9,181 tivit içinde belirgin bir varlık göstermektedir. Ayrıca, aşılılamaya karşı nötr bir duygu durum sergileyen tivit sayısı ise 14,426 olarak tespit edilmiştir.

Sayısal verilere dayalı olarak elde edilen oranlar ise şu şekildedir:

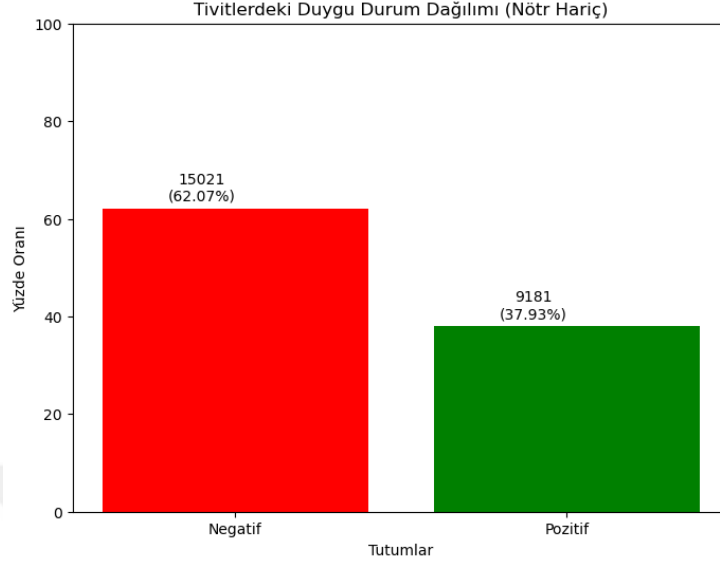
- Negatif duygu durum oranı: %38,89
- Pozitif duygu durum oranı: %23,73
- Nötr duygu durum oranı: %37,38

Aşılılamaya karşı genel duygu durum yüzdeleri ve sayıları Şekil 4'de gösterilmiştir.



Şekil 4 Tivitlerdeki Genel Duygu Durum Dağılımı

Aşılamaya yönelik duygu durumları içeren 38,628 tivit içinde, pozitif duygu durumlar 9,181 tivit ile temsil edilirken, negatif duygu durumlar ise 15,021 tivit ile öne çıkmaktadır. Bu veriler ışığında, pozitif ve negatif duygu durumların birbirine oranı incelendiğinde, negatif duygu durumların pozitif duygu durumlara göre yaklaşık olarak 1,64 kat daha fazla olduğu gözlemlenmiştir. Sadece pozitif ve negatif duygu durumların yüzdeleri ve sayıları Şekil 5 ile gösterilmiştir.



Şekil 5 Tivitlerdeki Genel Duygu Durum Dağılımı (Nötr Hariç)

Kelime Analizi

Veri setinin analizleri sonucunda bir tivit veri seti içeriğine kelime bazlı bir analiz yapıldığında kullanıcının duygu durumu hakkında çeşitli ipuçları elde edilebildiği çıkarımı yapılmıştır. Kullanıcıların belli bir konudaki duygu durumunu belirten kelimeler genellikle tivitinin duygu durumunu da yansıtmaktadır. Bu nedenle, bu çalışmada tivit içeriklerinde yapılan kelime bazlı analizler, kullanıcının aşılama konusundaki duygu durumunu anlamak ve doğrulamak için önemli bir araç niteliğindedir. Bu analiz içerisinde veri setinde genel olarak kullanılan, pozitif duygu durum olarak değerlendirilen ve negatif duygu durum olarak değerlendirilen veri setlerindeki en sık kullanılan kelimeler incelenmiştir.

Veri seti genel olarak değerlendirildiğinde en sık tekrar edilen kelimeler sırasıyla şu şekildedir;

- “aşı”: 11,389
- “covid”: 7,398
- “doz”: 2,582

- “kalp”: 1,986
- “mRNA”: 1,541
- “ilaç”: 1,386
- “yan”: 1,348
- “bağışıklık”: 1,109 kez tekrar etmiştir.

Veri filtreleme aşamalarında “aşı”, “ilaç” ve “bağışıklık” kelimeleri aranarak filtrelemeler yapılmıştır. Fakat “kalp” ve “yan” kelimelerinin sıklıkla tekrar etmesi veri setindeki negatif duygu durum içeren tivitlerin yüzde olarak fazla çıkmasını destekler niteliktedir. “kalp” ve “yan” kelimeleriyle birlikte sıklıkla kullanılacak “krizi” 740 kez, “etki” kelimesi 390 kez tekrar etmiştir. Bu da veri setinde sıklıkla aşılama sürecinin yan etkilerinden ve bu yan etkilerden en tehlikelisi olan kalp krizinden bahsedildiğini kanıtlar niteliktedir.

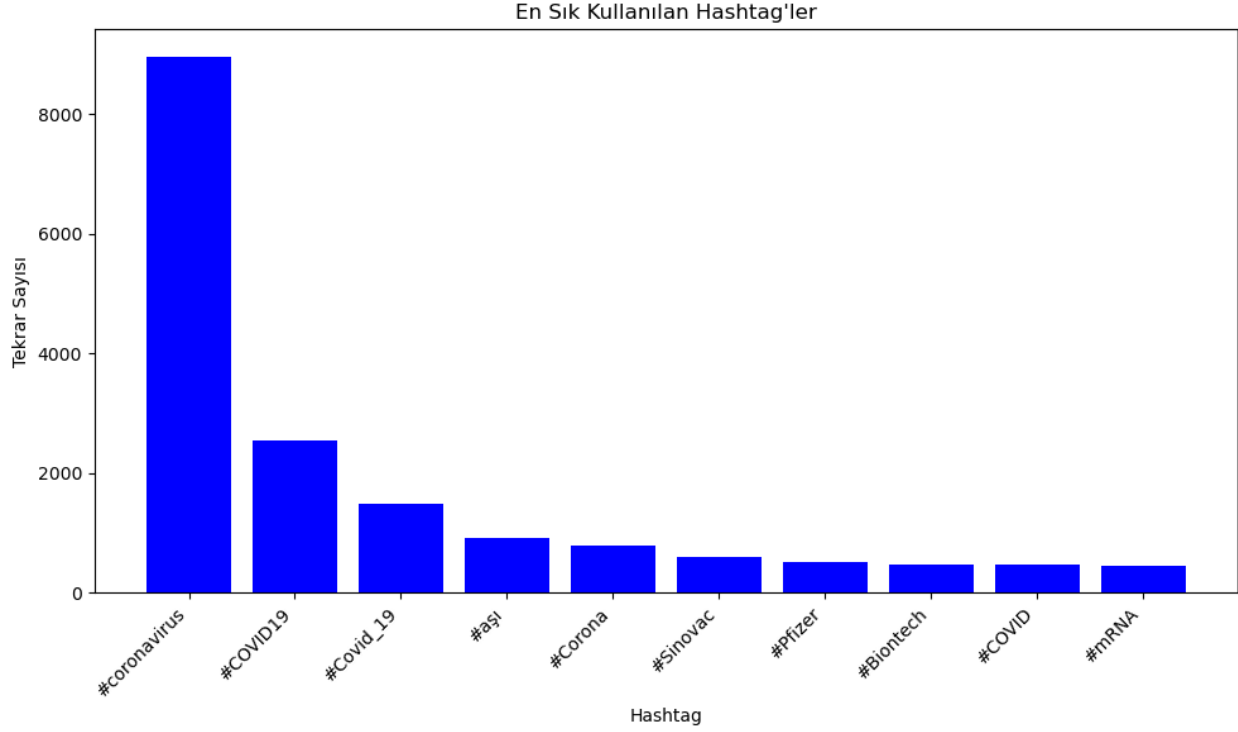
Negatif olarak değerlendirilen duygu durumlarda en sık tekrar edilen ve negatif duygu durum ile ilişkilendirilebilen kelimeler sırasıyla şu şekildedir;

- “kalp”: 1,176
- “mRNA”: 1,020
- “yan”: 800
- “krizi”: 502
- “Pfizer”: 492
- “olmadım”: 405
- “grip”: 355
- “hasta”: 326 kez tekrar etmiştir.

Genel veri setinde sıklıkla tekrar eden “kalp”, “krizi” ve “yan” kelimeleri negatif olarak değerlendirilen tivit setinde de sıklıkla tekrar edilmiştir. Bu kelimelere ek olarak “grip” ve “hasta” gibi aşılama karşı duygu durum ile negatif ilişkilendirilebilecek kelimeler sıklıkla kullanılmıştır. En sık tekrar eden kelimelere bakıldığında “olmadım” kelimesi de karşımıza çıkmaktadır. Bu

- “grip”: 189
- “ilaç”: 174
- “hafif”: 171
- “Biontech”: 169 kez tekrar etmiştir.

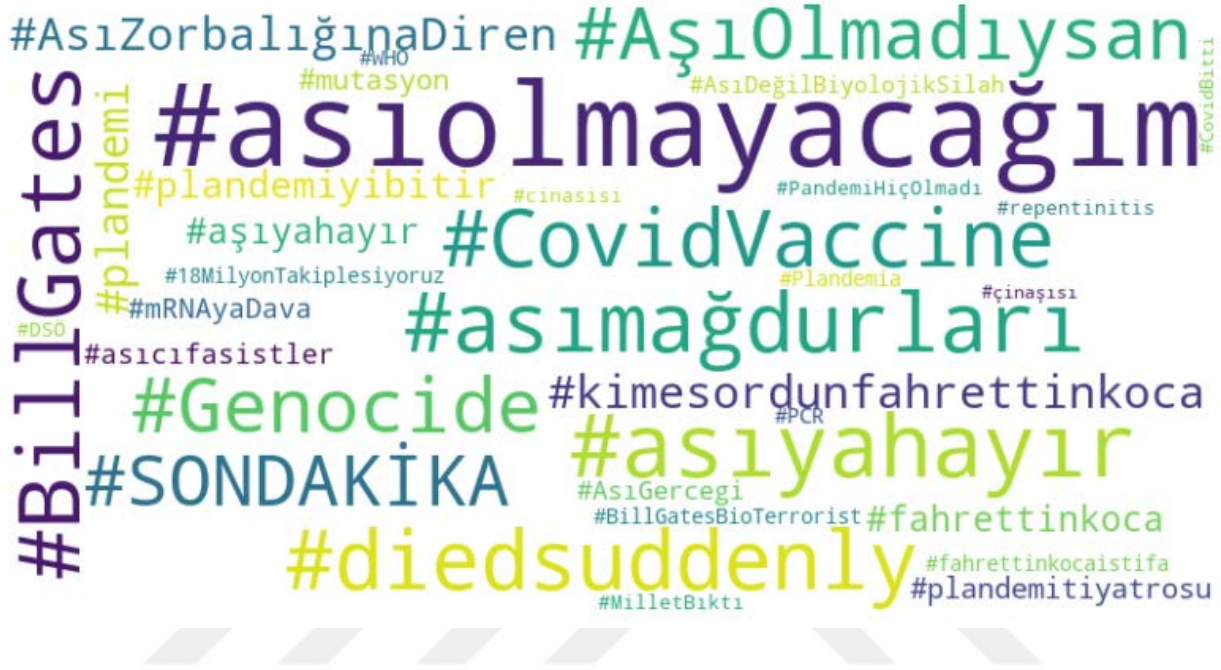
“kalp”, “yan”, “Pfizer” ve “grip” kelimeler negatif olarak değerlendirilen duygu durum veri setinde de sıklıkla kullanılan kelimeler arasındadır. Bu kelimeler pozitif veri setinde daha az tekrar edilmesine rağmen hatırı sayılır bir sayıya sahiptir. “oldum”, “etkili”, “hafif” ve “ilaç” gibi kelimeler ise aşılamaya karşı duygu durumun pozitif değerlendirildiği veri setimizdeki sonuçları doğrular nitelikte bilgiler vermektedir. Ek olarak, pozitif duygu durum veri setinde “Pfizer”, “biontech”, “Biontech” ve “Biontech” kelimeleri toplamda 740 kez tekrar edilirken “Moderna”, “sinovac” ve “Sinovac” kelimeleri toplamda 298 kez tekrar edilmiştir. “Pfizer” ve “Biontech” kelimelerinin sıklıkla kullanılması da daha çok Pfizer Biontech aşısının olumlu değerlendirildiğini kanıtlar niteliktedir. Pozitif duygu durum olarak etiketlenen veri setinde sıklıkla tekrar eden kelimeler kelime bulutu ile Şekil 7’de gösterilmiştir.



Şekil 8 Tüm Veri Seti için En Sık Tekrar Eden Hashtagler

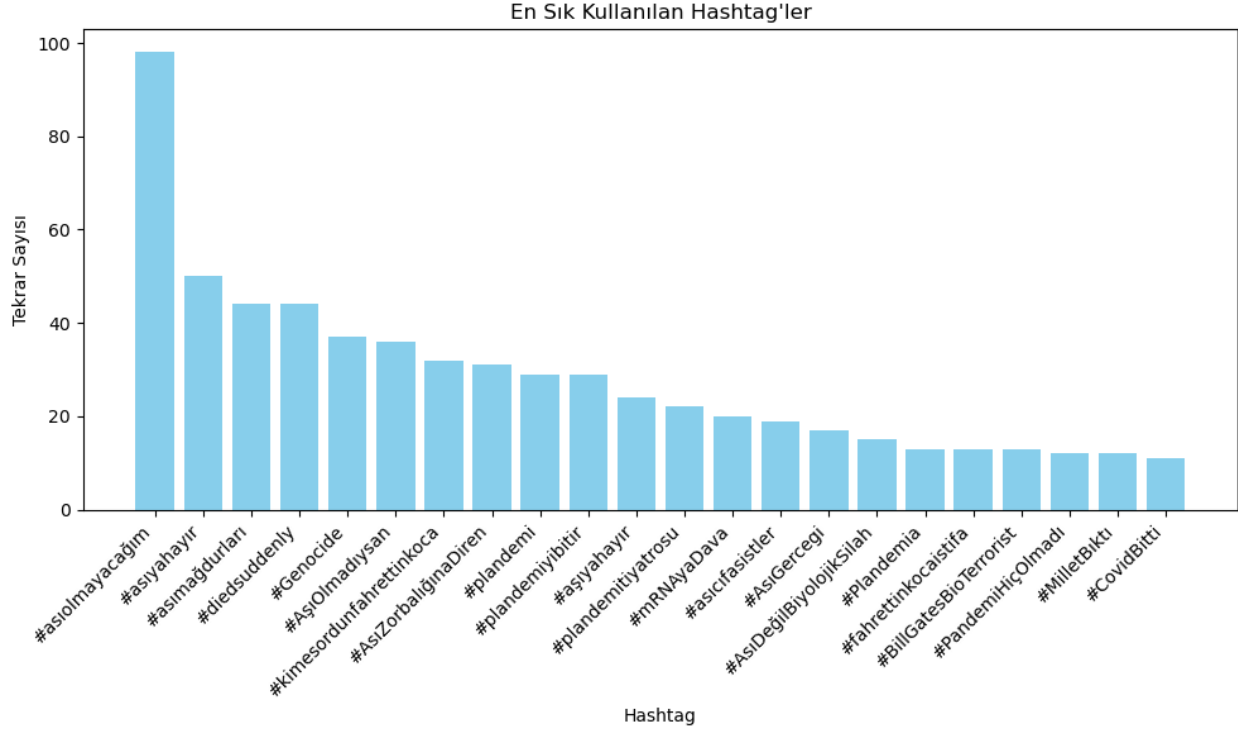
“coronavirus”, “COVID19”, “Covid_19”, “Corona” ve “COVID” hashtagleri en sık tekrar eden 10 hashtag arasında kendine yer bulmaktadır. Bu durum veri setinin toplanış şekli ile oldukça ilgilidir. En sık tekrar eden ilk 10 hashtag içerisinde “aşı” 914 kez, “Sinovac” 598 kez, “Pfizer” 500 kez, “Biontech” ise 463 kez tekrar etmektedir. Verinin filtreleme yöntemlerinde aşılama ile ilgili metinleri bulabilmek için kelime bazlı filtrelemeler yapılmıştır. Bu hashtaglerin sık tekrar etmesinin sebeplerinden biri de budur. Aşılamaya karşı pozitif ve negatif duygu durum içeren metinlerdeki hashtagler için analiz yapıldığında dikkat çeken sonuçlar ile karşılaşılmaktadır. Veri setinin genelinde en sık kullanılan 10 hashtagdeki gibi genel kavramlar elendiğinde elde edilen sonuçlar ve sık tekrar eden hashtagler hakkında bazı analizler şu şekildedir; Negatif duygu durum içeren metinlerde en sık kullanılan hashtag’lerin içeriği genel olarak aşı karşıtlığı ve aşılama

karşı duyulan şüphe ile ilgili olduğu görüşmüştür. Negatif duygu durum veri seti için en sık kullanılan hashtagler kelime bulutu ile Şekil 9’da gösterilmiştir.



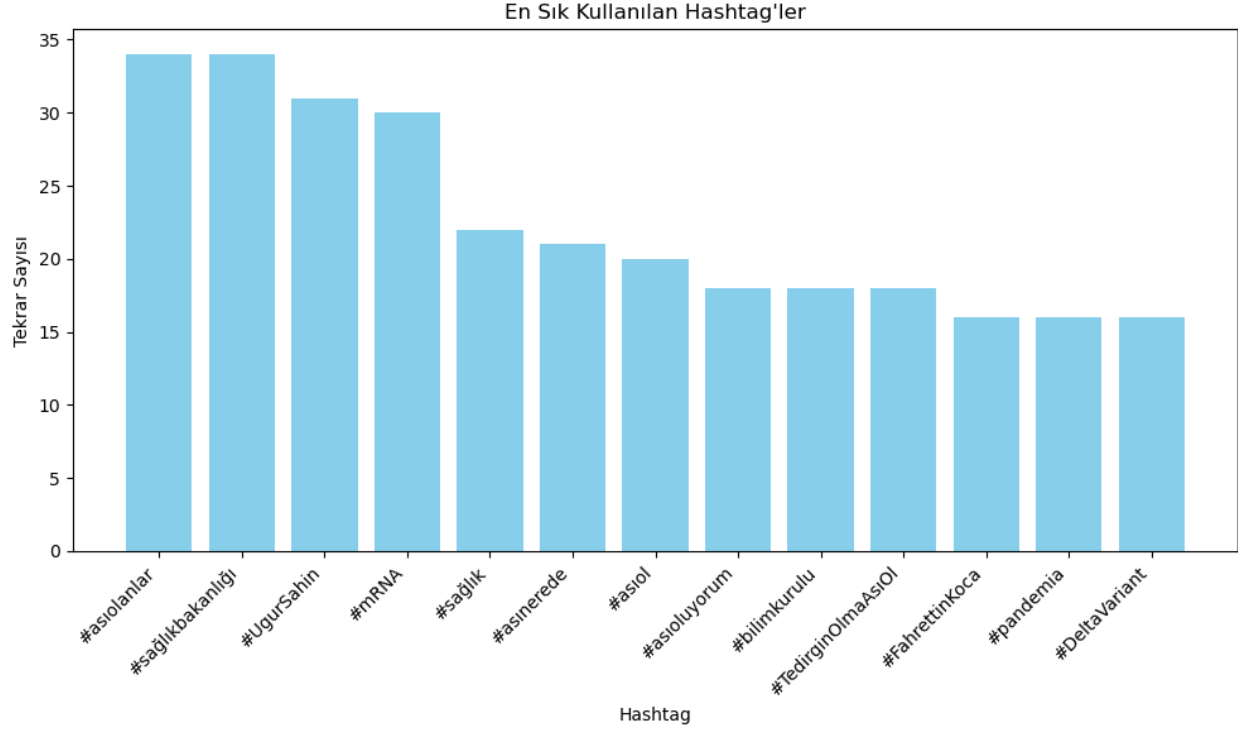
Şekil 9 Negatif Duygu Durum Veri Seti için Kelime Bulutu (En Sık Tekrar Eden Hashtagler)

Bariz olarak aşı karşıtlığı içeren hashtagler ve tekrar sayıları ise Şekil 10’da gösterilmiştir. #asıolmayacağım, #asımağdurları, #plandemi, #plandemiyatrosu, #mRNAYaDava ve #CovidBitti gibi aşı karşıtı hashtagler negatif duygu durum içeren tivitlerde sıklıkla kullanılmıştır. Özellikle “Plandemi” hashtagi bu dönemde oldukça popüler olmuş ve tivitlerde sıklıkla kullanılmıştır. Çalışılan veri seti Türkçe olmasına rağmen negatif duygu durum içeren veri setinde en sık kullanılan hashtagler arasında İngilizce hashtagler de kullanıldığı görülmüştür. #diedsuddenly, #Genocide ve #BillGatesBioTerrosist hashtagleri de negatif duygu durum içeren hashtagler arasındadır. Bu da etiketli veri setinin tutarlılığını gösterir durumdadır.



Şekil 10 Negatif Duygu Durum Veri Seti için En Sık Tekrar Eden Hashtagler

Pozitif duygu durum içeren metinlerde en sık kullanılan hashtaglerin içeriği genel olarak aşılama sürecini olumlu değerlendiren hashtagler ile ilgili olmuştur. Pozitif duygu durum veri seti için en sık kullanılan hashtagler kelime bulutu ile Şekil 11’de gösterilmiştir.



Şekil 12 Pozitif Duygu Durum Veri Seti için En Sık Tekrar Eden Hashtagler

Sözlük Tabanlı Yaklaşım Sonuçları

Birçok çalışmada, sözlük tabanlı bir yaklaşım kullanarak duygu durum analizi gerçekleştirilmiştir. Bu çalışmada, açık kaynaklı bir kaynaktan elde ettiğimiz Türkçe duygu durum analizi setinden yardım alarak oluşturulan model, insan gücü ile manuel olarak etiketlenen veri setiyle karşılaştırılmıştır. Bu karşılaştırmanın amacı, özellikle duygu durum analizi modelinin aşılama karşı duygu durumunda ne kadar doğru sonuçlar üretebileceğini test etmektir. Sözlük tabanlı modeli oluşturmak için Chen ve Skinea'nın Building Sentiment Lexicons for All Major Languages çalışmasının veri setinden yardım alınmıştır. Çalışmanın veri setinde Türkçe dili özelinde bir duygu durum analizi için 1,519 negatif kelime 981 pozitif kelime bulunmaktadır. Bu veri seti içeriğinde çalışmanın daha sağlıklı sonuçlar verebilmesi için sadece bir manipülasyon yapılması

gerekmiştir. “Aşı” kelimesi negatif veri setine ait bir kelimedir ve durumun aşılama üzerine geliştirilen modelde belli önyargılar yaratacağı düşünülmüştür ve bu kelime veri setinden çıkarılmıştır. Bu pozitif ve negatif kelimeler kullanılarak, insan gücüyle manuel olarak etiketlenmiş veri setindeki her bir tivitın negatif, pozitif veya nötr kategorilerinden hangisine daha uygun olabileceğini anlayabilmek için basit bir model kullanılmıştır.

Bir tivit üzerinden örnek vermek gerekirse;

“İngiltere’de bu özgür grup, aşı noktasının önünde covid aşılarını protesto ediyor ve ‘aşı’ kurbanı olduğu iddia edilenlerin çok sayıda fotoğrafını yayınlıyor...”

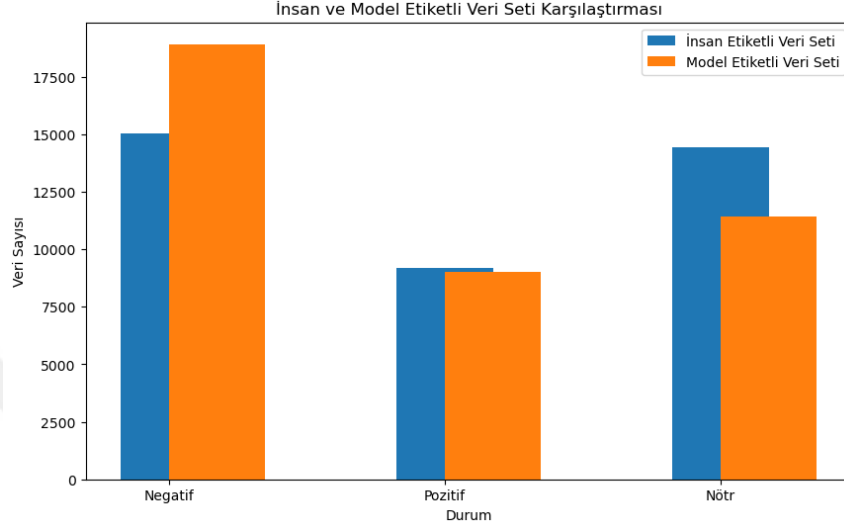
Kırmızı ile altı çizilen 3 kelime negatif değerli, yeşil ile çizilen kelime pozitif değerlidir. Tivit içerisindeki diğer kelimeler duygu durum analizi veri setinde bulunmayan kelimelerdir. Burada -3+1=-2 hesabıyla tivitın duygu durum analiz sonucu negatif çıkmaktadır.

Bir başka örnek ile göstermek gerekirse;

“Sağlık Bakanı Fahrettin Koca, “Şuan itibariyle 1. doz aşısını olan sağlık çalışanı sayısı 100.000’i geçti. Türkiye aşı programı uygulamakta güçlü bir alt yapıya sahiptir. Bu güce inanan, destek veren herkese teşekkür ederim” dedi.”

Kırmızı ile altı kırmızı ile çizilen 1 kelime negatif değerli, yeşil ile çizilen 6 kelime pozitif değerlidir. Tivit içerisindeki diğer kelimeler duygu durum analizi veri setinde bulunmayan kelimelerdir. Burada -1+6=5 hesabıyla tivitın duygu durum analiz sonucu pozitif çıkmaktadır. Bu formül ile veri seti analiz edilmiş, çıkan sonuçlar eğer artı bir değer ise “pozitif”, eksi bir değer ise “negatif” ve değer sıfır ise “nötr” olarak değerlendirilmiştir. Bu değerlendirmelerin sonucunda bulunan pozitif, negatif ve nötr dağılımı insan gücü ile manuel olarak etiketlenmiş veri setinde

benzerdir. İnsan gücüyle manuel olarak etiketlenmiş veri seti ve sözlük tabanlı model tarafından etiketlenmiş veri setinin sınıflara sayısal dağılımı Şekil 13’te gösterilmiştir.



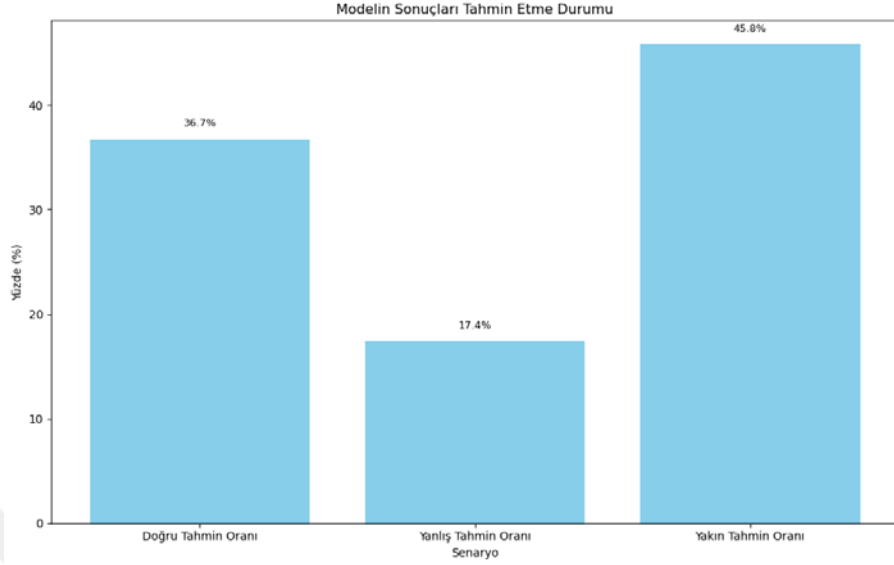
Şekil 13 İnsan Etiketli Veri Seti ve Model Etiketli Veri Seti Sayısal Karşılaştırması

İnsan gücü ile manuel olarak etiketlenen veri seti ile modelin etiketlediği veri seti arasındaki negatif, pozitif ve nötr dağılımına bakıldığında modelin “pozitif” duygu durum oranını neredeyse tahmin edebildiği görülmektedir. İnsan gücü ile manuel olarak etiketlenen veri setinin pozitif duygu durum dağılımı %23,77 iken modelin tahmini %22,88’dir. İnsan gücü ile etiketlenen veri setinin negatif duygu durum dağılımı %38,88 iken modelin tahmini %48,07’dir. Nötr dağılımda ise insan etiketli veri setinin negatif duygu durum dağılımı %37,35 iken modelin tahmini %29,05’dir. Karşılaştırılmaların sapma payı değerlerine bakıldığında pozitif duygu durum için değer %3,82 çıkmaktadır. Bu değer, negatif duygu durum için %21,12, nötr duygu durum için %24,97’dir. Sapma payının hesaplandığı formül Şekil 14’te gösterilmiştir.

$$SapmaPayı = \frac{|Değer_1 - Değer_2|}{\frac{Değer_1 + Değer_2}{2}} \times 100$$

Şekil 14 Sapma Payı Formülü

Veri setindeki insan gücü ile manuel etiketlenen veri setinin sonuçları ile modelin sonuçları teker teker karşılaştırıldığında 38,626 tivitın 14,183'ünü model doğru tahmin etmektedir. Yani model %36,7 oranında başarıya sahiptir. Modelin sonucu doğru bilme senaryoları pozitif ise pozitif demesi, negatif ise negatif demesi ve nötr ise nötr demesi ile üç senaryodan oluşmaktadır. Modelin insan etiketli veri setinde negatif olan bir duygu duruma pozitif tahmini yaptığı ve insan etiketli veri setinde pozitif olan bir duygu duruma negatif tahmini yaptığı tivit sayısı, yani modelin tamamen yanıldığı tivit sayısı 38,626 tivitte 6,724'dir. Modelin tamamen yanılma oranı %17,4'tür. Modelin yakın tahmin yaptığı yani insan etiketli veri setinde pozitif ya da negatif olan bir duygu duruma nötr tahmini yaptığı sayı 17,719'dur. Yani modelin yakın tahmin etme oranı %45,8'dir. Modelin insan etiketli veri setinin sonuçlarını tahmin edebilme durumu Şekil 15'te detaylı bir biçimde gösterilmiştir.



Şekil 15 Modelin İnsan Etiketli Veri Setindeki Yüzdesele Tahmin Sonuçları

Sonuçlar incelendiğinde, modelin etiketli veri setinde gerçekleştirdiği tahminlerin başarı düzeyinin yeterli olmayabileceği sonucuna ulaşılmaktadır. Duygu durum analizinde, sözlük bazlı analizlerin kullanıldığı durumlarda, modelin kapsayıcılık derecesinin artırılması gerektiğini işaret etmektedir. Bu durum, gelecekteki çalışmalarda kullanılacak modeller için önemli bir gelişim alanını ifade etmektedir. Özellikle etiketlenmiş veri setleri, duygu durum analizi gibi karmaşık tahmin analizlerinde model performansının artırılmasına yönelik değerli bir kaynak oluşturmaktadır. Bu tür veri setleri, gelecekteki çalışmalarda daha güvenilir ve etkili modellerin geliştirilmesine katkı sağlamalıdır. Bu bağlamda, duygu durum analizi gibi tahmini zor analizlerde, kelime seviyesinde analizlere odaklanan modellerin daha sağlıklı sonuçlar elde etmesini sağlamak için kapsamlı bir model geliştirme stratejisinin benimsenmesi önemlidir. Bu çalışmada, duygu durum analizi alanındaki araştırmalara katkıda bulunmak üzere tasarlanmıştır ve elde edilen sonuçlar, gelecekteki çalışmaların bu alandaki modellerin etkinliğini artırmaya odaklanması için önemli ipuçları sunmaktadır.

Hata Matrisi (Confusion Matrix) Analizleri

Etiketli veri setinden alınan pozitif ve negatif duygu durum içeren sonuçlar ile sözlük tabanlı modelden alınan pozitif ve negatif durum içeren sonuçlar hata matrisi kapsamında değerlendirilmiştir. Veri setinde olan nötr sınıfını daha doğru sonuçlar elde edebilmek için işlemler içerisinde katılmamıştır. Nötr sınıfını işlemlere katılmadığında 15,644 karşılaştırılabilir değer elde edilmiştir. Bu da hata matrisindeki sonuçlara ulaşabilmek için yeterli bir sayı olmuştur.

Makine öğreniminde, modelin belirli bir görev hakkında öğrenim kazanabilmesi için etiketli veri setleri kullanılabilir. Modelin performansını değerlendirebilmek için gerçek sonuçlarla karşılaştırmak gerekir (Ersöz & Çınar, 2021). Hata matrisi (Confusion Table) bu karşılaştırma için kullanılan bir yöntemdir. Bu çalışmada, hata matrisi tablosu gerçek sınıf ve modelin tahmin ettiği sınıflar arasındaki doğruluğu ölçmektedir ve yapılan hata türleri hakkında da bilgi vermektedir. Görselde her bir sütun model tarafından tahmin edilen sınıfı, her bir satırda etiketli veri setinin sonuçlarını, yani gerçek sınıfları göstermektedir.

Sonuçlar dört kategori ile belirtilir;

Gerçek Pozitif (TP – True Positive): Modelin tahmini ve gerçek tahminin her ikisi de pozitifdir (İnsanın pozitif duygu durum olarak değerlendirdiği bir metni modelin de pozitif değerlendirmesi).

Gerçek Negatif (TN – True Negative): Modelin tahmini ve gerçek tahminin her ikisi de negatiftir (İnsanın negatif duygu durum olarak değerlendirdiği bir metni modelin de negatif değerlendirmesi).

Yanlış Pozitif (FP – False Positive): Modelin tahmini pozitifdir, fakat gerçek değer tahmini negatiftir (İnsanın negatif duygu durum olarak değerlendirdiği bir metni modelin pozitif değerlendirmesi).

Tip I hata içerir. Tip I hata gerçek tahminin ilişki olmadığını (negatif) belirttiği senaryoya modelin ilişki var (pozitif) deme olasılığını belirtir. Tip I hata testin hata payı veya anlamlılık düzeyi olarak da bilinir. Modelin başarılı sayılabilmesi için Tip I hata oranının en fazla %5 olması gerekmektedir. Çalışmadaki modelin Tip I hata oranı %27'dir. Bu orana bakıldığında model başarılı sayılmamaktadır.

Yanlış Negatif (FN – False Negative): Modelin tahmini negatiftir, fakat gerçek değer tahmini pozitifdir (İnsanın pozitif duygu durum olarak değerlendirdiği bir metni modelin negatif değerlendirmesi).

Tip II hata içerir. Tip II hata gerçek tahminin ilişki olduğunu (pozitif) belirttiği senaryoya modelin ilişki yok (negatif) deme olasılığını belirtir. Tip II hata testin gücü veya anlamlılık olasılığı olarak da bilinir. Modelin başarılı sayılabilmesi için Tip II hata oranının en fazla %20 olması gerekmektedir. Çalışmadaki modelin Tip II hata oranı %16'dır. Bu orana bakıldığında testin, yani modelin güçlü olduğu görülmektedir.

Modelin hata matrisi detayları Şekil 16'da gösterilmiştir.

		Tahmin Edilen Sınıf	
		Pozitif	Negatif
Gerçek Sınıf	Pozitif	3130 TP	2454 FN Type II Error
	Negatif	4270 FP Type I Error	5790 TN

Şekil 16 Modelin Hata Matrisi (Confusion Matrix) Tablosu

Doğruluk (Accuracy): Doğruluk oranı modelin başarı seviyesini ölçmek için sıklıkla kullanılan bir metriktir. Fakat doğruluk oranı modelin başarı seviyesini ölçebilmek için tek başına yeterli bir metrik değildir. Doğruluk oranı modelin doğru tahmin ettiği durumların toplam duruma oranıdır. Modelin doğruluk oranı Doğruluk formülü ile $(3130+5790)/(3130+2454+4270+5790)$ şeklinde hesaplandığında sonuç %57,02 çıkmaktadır. Modelde pozitif ve negatif olmak üzere iki kategori olduğunu bilinmektedir, bu sebeple %57,02 oranı bize başarı seviyesinin yüksek olmadığını göstermektedir.

$$\text{Doğruluk (Accuracy)} = \frac{TP + TN}{TP + FP + TN + FN}$$

Kesinlik (Precision): Kesinlik oranı modelin pozitif olarak tahmin ettiği değerlerin ne kadarının gerçek pozitif göstermektedir. Yani modelin tahmin ettiği pozitif sınıfların gerçekte ne kadar pozitif değerine ait olduğu hakkında bilgi verir. Tip 1 hatanın oranı ile yakından ilgilidir. Kesinlik değerinin yüksek olması modelin başarısı ile ilgilidir. Modelin kesinlik oranı kesinlik formülü ile

3130/(3130+4270) şeklinde hesaplandığında sonuç %42,29 çıkmaktadır. Modelin başarı seviyesi kesinlik oranına bakıldığında da yüksek değildir.

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

Duyarlılık (Recall): Duyarlılık oranı modelin pozitif olarak tahmin etmesi gereken değerlerin ne kadarını ne kadarını pozitif tahmin edebildiğini göstermektedir. Bir diğer deyişle pozitif durumların ne kadar başarılı tahmin edildiğinin sonucunu belirtir. Modelin başarılı sayılabilmesi için mümkün olduğunca yüksek olması gerekmektedir. Modelin duyarlılık oranı duyarlılık formülü ile 3130/(3130+2454) şeklinde hesaplandığında %56,05 çıkmaktadır. Modelin başarı seviyesi duyarlılık bağlamında da yüksek bir seviyede değildir.

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

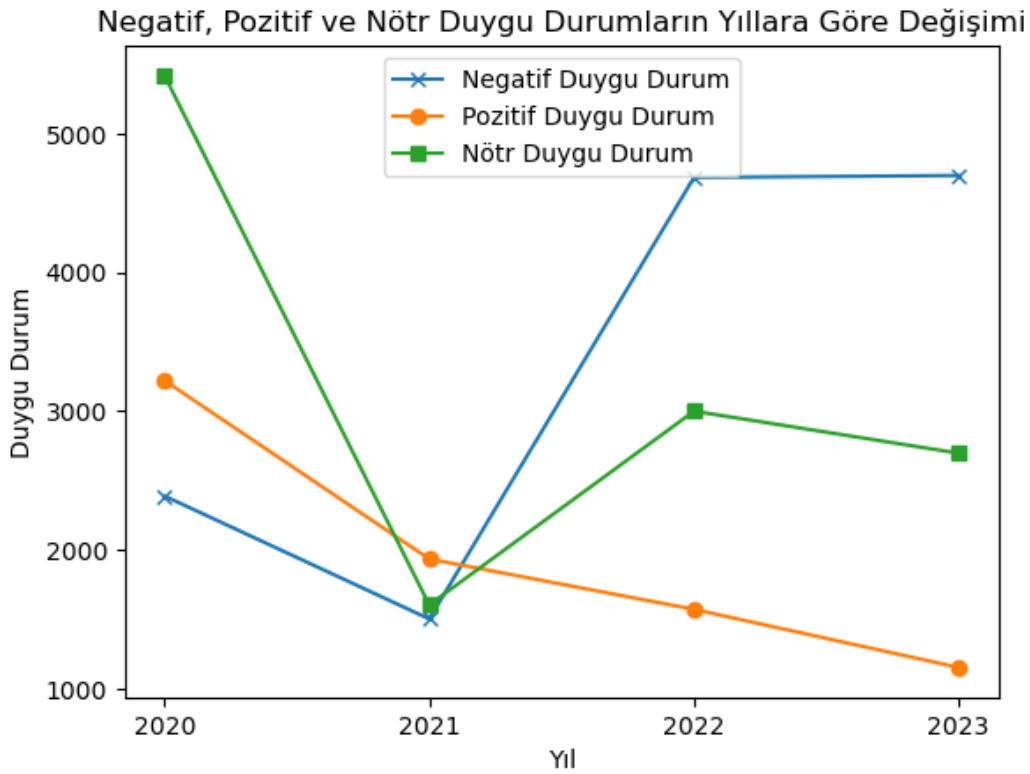
F1 Skoru: F1 skoru kesinlik ve duyarlılık oranlarının daha dengeli bir ortalamasıdır. Dengeli bir ortalama vermesinin sebebi uç durumları engelleyebilmek içindir. Sadece ortalamalarını aldığı bir senaryoda kesinlik değeri 0 ve duyarlılık oranı 1 olursa F1 skor %50 oranına sahiptir sonucu çıkar ve bu sonuç modelin başarı seviyesi konusunda araştırmacıyı yanıltıcı bir sonuca götürür. F1 skoru 1 değerine yaklaştıkça modelin başarılı, 0 değerine yaklaştıkça modelin başarısız olduğunu göstermektedir. Modelin F1 skoru formül ile $2 * (0,429 * 0,560) / (0,429 + 0,560)$ şeklinde hesaplandığında sonuç %48,5 çıkmaktadır. Modelin başarı seviyesinin düşük olduğunu gösteren bir değerdir.

$$F1 Skoru (F1 Score) = 2 * \frac{Kesinlik (Precision) * Duyarlılık (Recall)}{Kesinlik (Precision) + Duyarlılık (Recall)}$$

Hata matrisi sonuçlarını analiz edildiğinde sözlük tabanlı modeli insan gücü ile manuel etiketlenmiş veri seti ile karşılaştırdığımızda sözlük tabanlı modelin insanların aşlamaya karşı duygu durumunu başarıyla tahmin edemediği görülmektedir.

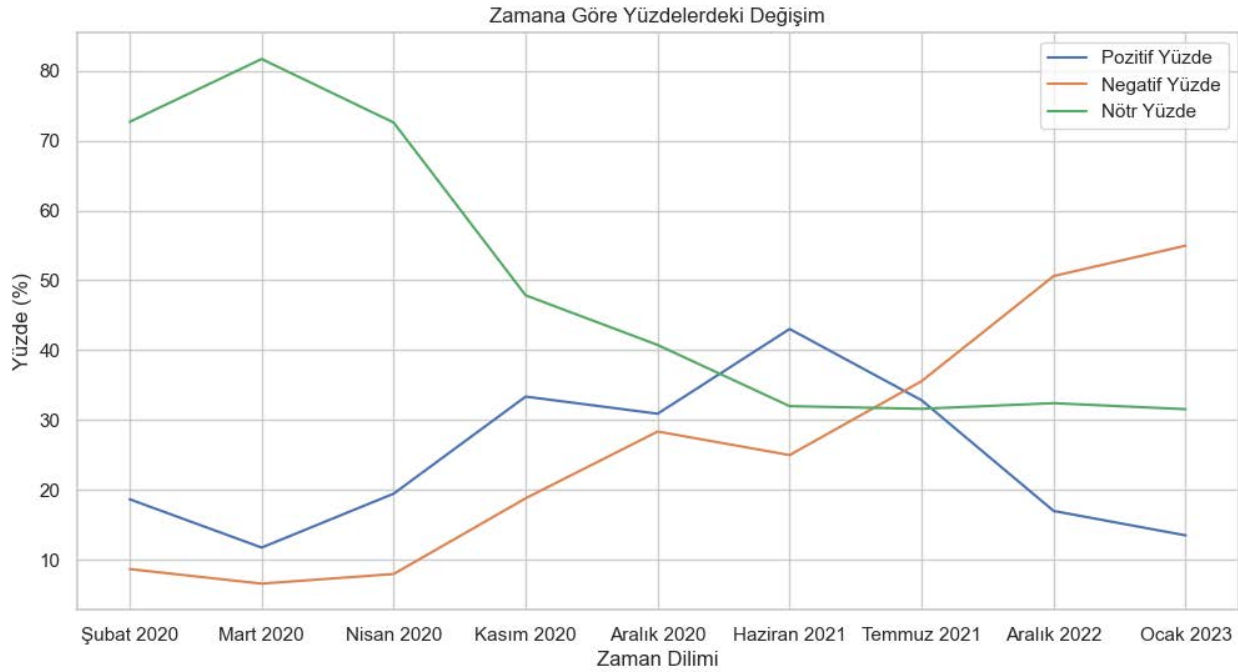
Aşlamaya Karşı Duygu Durumun Zamana Göre Değişimi

Etiketlenen 33,863 veride aşlamaya karşı duygu durumun zamana göre değişimi incelenmiştir. 2020 yılına ait 11,026, 2021 yılına ait 5,039, 2022 yılına ait 9,253 ve 2023 yılına ait 8,545 etiketli veri bulunmaktadır. Yıllara göre, sayı olarak değişimine bakıldığında pozitif duygu durumun 2020 başlarında negatife göre daha fazla olduğu fakat 2021 yılında eşitlenen sayıların 2022 ve 2023 yıllarında negatif duygu durumun lehine değiştiğini görülmektedir. Negatif, pozitif ve nötr duygu durumların yıllara göre sayısal değişimleri Şekil 17’de gösterilmiştir.



Şekil 17 Negatif, Pozitif ve Nötr Duygu Durumların Yıllara Göre Değişimi

Pozitif ve negatif duygu durum yüzdelerinin zamana göre değişimi incelendiğinde negatif duygu durumun gittikçe yükselen bir grafiğe sahip olduğu görülmektedir. Pandeminin başlarından itibaren değerlendirilen süreçte kullanıcıların aşılamaya karşı negatif duygu durumu giderek artmıştır. Aşılamaya karşı pozitif duygu durumun ise pandeminin başlarından itibaren 2021 yılı ortasına kadar artarak devam ettiğini fakat 2021 yılı ortalarından itibaren bir düşüş yaşadığını görülmektedir. Genel veri setinde aşılamaya karşı negatif duygu durum %38,89, aşılamaya karşı pozitif duygu durum %23,73 olmasına rağmen negatif duygu durum yüzdesi pozitif duygu durumu yüzdesini ancak 2021 yılının ortalarında geçebilmiştir.



Şekil 18 Negatif, Pozitif ve Nötr Duygu Durumun Zamana Göre Yüzdesele Değişimi

Aşılamaya karşı negatif, pozitif ve nötr duygu durum yüzdeleri tek tek incelendiğinde negatif duygu durumun pandeminin başından 2023 başına kadar belirgin bir artış eğiliminde olduğunu, aşılamanın bu süreçte artmasının negatif duygu durumu çok fazla değiştirmedeği görülmektedir.

Aşılama sürecinde ve sonrasında kaygı duyulan yan etkilerin (kalp krizi vs.) bu durumda etkili olduğu tahminler arasındadır. Pozitif duygu durumun ise 2021 yılının ortalarına kadar belirgin bir artış gösterdiği fakat bu zaman diliminden 2023 başlarına kadar keskin bir düşüş yaşadığı görülebilmektedir. Aşılama karşı nötr bir duygu durum içerisinde olmanın yüzdesel olarak pandeminin başından itibaren düşüş gösterdiği görülmektedir. Nötr duygu durumun yüzdesel oranı pandeminin başlarında yüksek seviyelerdeyken bu oran 2021 yılında azalmıştır. Bunun sebebini pandeminin başlarında, özellikle aşı çalışmalarının henüz başlangıç seviyesinde olduğunu ve daha çok COVID-19 geçirenlerin hastalıktan korunma ve hastalık ile nasıl mücadele edeceğine dair tivitler atmaya yöneldiğini varsayarak açıklanabilir durumdadır (Kurt, 2021).

Sonuçlar ve Tartışma

Bu çalışmanın temel amaçlarından biri Türkçe dilinde bir duygu durum analizi yapılarak literatüre katkıda bulunmaktır. 38,622 tivit in insan gücüyle manuel olarak etiketlendiği veri setinin analizleri Türkçe dili özelinde aşılama karşı olan duygu durumu hakkında çeşitli bilgiler vermektedir. Bunun yanında bu çalışma büyük hacimde etiketli veri sağlayarak gelecekte duygu durum analizi (sentiment analysis) ve aşılama konusunda yapılacak çalışmalara destek olması planlanmıştır.

Tez çalışmasının başlığında makine öğrenmesi yöntemlerinin kullanılarak verilerin incelendiği belirtilmesine rağmen, sözlük tabanlı yaklaşımın sonuçlarında istenen bulguların elde edilememesi dikkate değerdir. Araştırmanın ilerleyen aşamalarında, makine öğrenmesi alanına doğrudan katkı yapmak yerine, etiketlenen veri setinin gelecekteki çalışmalara ışık tutması hedeflenmektedir. Bu bağlamda, çalışma, makine öğrenmesi yöntemlerinin kullanımını vurgulamak yerine, etiketlenmiş veri setinin daha geniş bir anlayış ve uygulama alanı için bir temel oluşturma amacını gütmektedir. Bu doğrultuda 38,626 etiketli metnin bulunduğu veri seti aperta.ulakbim.gov.tr adresindeki arşive

01.02.2024 tarihinde yüklenmiştir. Açık erişime sahip olan veri setine aperta.ulakbim.gov.tr/record/263524 linki ile erişilebilir durumdadır. Ayrıca aperta.ulakbim.gov.tr adresindeki “Arama” alanında 263524 ile yapılan aramada yine veri seti karşımıza çıkmaktadır.

Aşılama karşı duygu durum analizi üzerine birçok çalışma mevcuttur. Literatürün, bu çalışma ile benzer ve farklı yönleri olduğu görülmektedir. Sütçü ve arkadaşlarının yaptığı çalışmada COVID-19 aşısı ile ilgili 11-18 Şubat 2021 tarihleri arasında atılan 1,258 tivit incelenmiştir. Twitter API Key ile elde edilen veriler ön işleme sırasında “tokenizasyon”, “filtreleme” ve “dönüştürme” aşamalarından geçirilmiştir. Tokenizasyon aşamasında, metinler anlamlılık düzeylerine göre ayrılmıştır. Filtreleme aşamasında metinlerde bulunan gereksiz kelime ve noktalama işaretleri kaldırılmıştır. Dönüştürme aşamasında ise “URL” ve “HTML” bilgileri metinlerden silinmiştir. Filtreleme aşamaları kısmi benzerlikler göstermektedir. Sonrasında veri seti üzerinde en sık tekrar eden kelimeler ve temalar incelenmiştir. En sık tekrar eden kelimelerin genellikle aşılama karşı negatif duygu durum içerdiği belirlenmiş ve bu yönüyle de benzer sonuçlar elde edilmiştir. En sık tekrar kelime “Plandemi” olarak öne çıkmıştır. Bu kelime çalışmamızda negatif duygu durum olarak belirtilen küme içerisinde en sık tekrar eden kelimelerden biridir. Çalışma aşılama karşı atılan tivitlerin temasını incelediğinde oranı en yüksek olan tema %32,8 oranı ile “Pandeminin Gerçek Olduğuna İnanmayan Söylemler” temasıdır. Bu çalışmada aşılama karşı genel duygu durum, hashtag ve kelime analizleri değerlendirildiğinde bu tema çalışması ile ortak paydada bulunduğu çıkarılan sonuçlar arasındadır. Lyu ve arkadaşlarının yaptığı çalışmada ise R yazılım metoduyla 11 Mart 2020 (Dünya Sağlık Örgütü’nün COVID-19’u pandemi ilan edildiği gün) ve 31 Ocak 2021 tarihleri arasında 583,499 farklı kullanıcıdan elde edilen 1,499,421 tivit üzerinde konu modelleme (topic modeling) teknikleri kullanılarak aşılama karşı duygu durum analizi yapılmıştır. Çalışmadaki amaç, sosyal

medyada COVID-19 aşılama süreci ile ilgili konuları ve duygu durumları belirlemek ve sürü bağışıklığına ulaşma sürecini etkileyebilecek kamuoyu algılarını, endişelerini ve duygu durumlarını daha iyi anlayabilmektir. Sonuçlar doğrultusunda süreç içerisinde değişen konular ayırt edilmeye çalışılmıştır. Veri seti belli kelimeler ile filtreleme yapılarak elde edilmiştir (“CoronavirusPandemic”, “covid19”, “2019ncov”). Veri toplama yöntemi bu çalışma ile benzerlikler göstermektedir, fakat veri toplanan ortam farklıdır. Sonrasında, veri seti sekiz adımlı bir ön süreç ile filtelenmiştir. Bu filtrelemelerden bazıları bu çalışma benzerlik göstermektedir. İngilizce dilinden verilere ulaşmak, aşılama ile ilgili tivitlere ulaşmak için kelime filtreleri yapmak ve tekrar eden tivitleri kaldırmak bunlardan birkaçıdır. Çalışmadan bir senelik süreç içerisinde aşılama çalışmaları ile ilgili gelişmelerin olduğu özel tarihlerde atılan tivitlerdeki duygu durum değişimi analiz edilmiştir. İngilizce diline özel duygu durum analizi (sentiment analysis) için modeller kullanan çalışma öfke (anger), beklenti (anticipation), iğrenme (disgust), korku (fear), sevinç (joy), üzüntü (sadness), şaşkınlık (surprise) ve güven (trust) duygularının zamansal değişimini incelemiştir. Bu duygularda aşılama ile ilgili gelişmelerin olduğu tarihlerde ne gibi değişiklik olduğunu öğrenebilmek çalışmanın bir diğer amacıdır. Örneğin Pfizer firmasının aşının etkililik oranının %90'ın üzerinde olduğunu açıkladığı günde negatif bir duygu durum olan korkunun (fear) keskin bir düşüş yaşadığı gözlemlenmiştir. Yine negatif bir duygu olan üzüntü (sadness) düşüş yaşamış, pozitif bir duygu olarak nitelendirilen sevinç (joy) oranında artış yaşanmıştır. Aslan ise çalışmasında Haziran 2021 ve Eylül 2021 tarihleri arasında atılan 176,345 tivit üzerinde bir çalışma gerçekleştirmiştir. Veriler Twitter API üzerinden MAXQDA nitel veri analizi yöntemi ile toplanmıştır. Veri toplama aşamasında hedef tivitlere ulaşabilmek için “covid19aşı”, “biontech”, “Sinovac”, “aşıhayatkurtarı”, “aşıyanetkileri” gibi terimler kullanılmıştır. Veriler toplandıktan sonra ön işleme aşamalarından geçirilmiştir. Bu aşamalardan

bazıları bu çalışma ile benzerlik göstermektedir. Tekrarlanan tivitleri veri setinden temizler, tivitin retweet olduğunu ifade eden RT içeren tivileri temizler ve metin içlerinden “@” içeren kelimeleri temizlemektedir. Bu çalışmadan farklı olarak tivitler içerisindeki bütün kelimeler ek almış halleri eklerinden ayrılarak tek bir kök halinde toplanmıştır. Bu işleme kök çözümüleme (lemmatization) adı verilir. Örneğin “aşılar” kelimesi “aşı” olarak güncellenmiştir. İnsan eliyle etiketleme yapmak için fazla büyük bir veri setine sahip olan çalışma ön işleme aşamaları sonucunda elde ettiği Türkçe kelimeleri Python’ın Googletrans kütüphanesini kullanarak İngilizceye çevirmektedir. Tüm bu aşamalardan sonra yine Python’ın NLTK ve TextBlob kütüphanelerini kullanan çalışma her bir tivitin öznellik ve polarite değerlerini çıkarmaktadır. Polarite değerlerine bakıldığında tivitlerin her biri için pozitif, negatif ve nötr olmak üzere sonuçları çıkartılmıştır. 176,345 tivitinin %49,6’sının nötr, %34,9’unun pozitif ve %15,5’inin ise negatif polariteye sahip olduğu sonucuna ulaşmıştır. Bu sonuç çalışmadaki genel duygu durum dağılımından farklı bir sonuç gösterse de, tez için kullanılan veri setinde 2021 Haziran ve 2021 Temmuz verilerinin negatif, pozitif ve nötr değerlerine bakıldığında pozitif duygu durumunun insan gücü ile manuel olarak etiketlenmiş veri setinde de negatiften fazla olduğu görülmektedir.

Sınırlılıklar

Bu çalışmada, yalnızca Türkçe dilinde atılan tivitler incelenmiştir. Dolayısıyla, dünya genelinde aşılama karşı duygu durumunu ve tutumunu ölçen kapsamlı bir analiz olmaması, bu çalışmanın sınırlılıkları arasında yer almaktadır. Araştırmanın odaklandığı dil ve coğrafi sınırlamalar, genel bir uluslararası perspektifi içermekte zorluklar yaratmaktadır. Bu bağlamda, benzer duygu durumu ve tutumları farklı kültürler ve diller arasında karşılaştırmak için daha geniş bir veri setinin ele alınması, gelecekteki çalışmalar için anlamlı olabilir. Çalışma sırasında, Twitter API aracılığıyla çekilen tivitler için kullanılan anahtar kelimeler içermeyen ancak COVID-19 aşısıyla ilgili olan

tivitlerin gözden kaçırılmış olma olasılığı bulunmaktadır. Bu, araştırmanın odak noktası olan anahtar kelimelerin dışında kalan ancak aşırıya dair önemli içerikler içeren tivitleri kapsam dışı bırakma riskini taşımaktadır. Bu durum, çalışmanın kapsamını genişletmek ve daha kapsamlı bir perspektif sunmak amacıyla gelecekteki benzer çalışmalarda göz önünde bulundurulmalıdır. Sadece tek bir sosyal medya platformu olan Twitter'ın kullanılması, yine sınırlılıklar arasında değerlendirilebilmektedir. Bu durum, araştırmanın geniş bir sosyal medya kullanıcı kitlesini temsil etme kapasitesini sınırlayabilir ve farklı platformlarda var olan çeşitli görüş ve etkileşimleri dışlayabilir. Çalışmada kullanılan veri seti, üç senelik geniş bir zaman dilimini kapsasa da, yıllara göre dağılımında belirgin bir dengesizlik bulunmaktadır. Örneğin, veri setinde 2020 yılına ait 11,026, 2021 yılına ait 5,039, 2022 yılına ait 9,253 ve 2023 yılına ait 8,545 tweet bulunmaktadır. Ayrıca, aylara göre dağılım da değişiklik göstermektedir. Örneğin, 2021 yılının Temmuz ayından 2022 yılının Aralık ayına kadar herhangi bir veri bulunmamaktadır. Bu durum, çalışmanın zaman açısından bazı sınırlılıklarla karşılaştığını ve bu durumu göz önünde bulundurmanın gerekliliğini ortaya koymaktadır. Etiketli veri seti ve sözlük tabanlı model arasındaki karşılaştırmalar için gerçekleştirilen analizlerde, Türkçe diline özgü daha kapsamlı bir veri seti kullanılması daha faydalı olabilir sonucuna varılabilmektedir. Model oluşturmak için kullanılan veri setinde toplamda 2500'e yakın pozitif ve negatif kelime bulunmaktadır. Bu, sözlük tabanlı modelin başarı düzeyini artırabilir ve daha gerçekçi sonuçlar elde etmemize yardımcı olabilir. Türkçe dilinde duygu analizi yapabilen bazı kütüphaneler olsa da (örneğin Zemberek vb.), bu kütüphaneleri veri seti üzerinde değerlendirdiğimizde istenilen sonuçları alınamamıştır. Türkçe, duygu analizi yapmak için daha fazla etiketli veriye ihtiyaç duyan bir dil olarak öne çıkmaktadır. Bu nedenle, Türkçe dilinde duygu analizi yapmak, çalışmanın sınırlılıklarından biridir. Ek olarak kullanılan sözlük tabanlı modelde negatif, pozitif ve nötr sınıflarını belirlemek için setteki pozitif ve negatif

kelimelerin her biri 1 puan olarak değerlendirilmiştir. Örneğin negatif kelime seti içerisindeki “yanlış”, “hasta”, “tehdit” ve “tehlike” gibi bariz negatif kelimelerin modele etkisi sadece 1 puan iken, “sıradan”, “fakir”, “nesne” ve “kukla” gibi negatifliği diğerleri kadar bariz olmayan kelimeler de model çıktısı değerlendirilirken 1 puanlık değere sahiptir. Bu durum, model çıktılarındaki analizlerin değerlendirilmesinde bazı sınırlamalar olduğunu göstermektedir. Çalışmada veri filtreleme ve temizleme aşamalarında ek adımlar eklemek mümkündür. Kelime analizi ve sözlük tabanlı analizlerde daha sağlıklı veriler elde etmek için tüm veri setindeki büyük harfleri küçük harflere çevirme işlemi yapılabilir. Örneğin, kelime analizi kısmında sık tekrar eden kelimeleri tespit etme aşamasında 'Aşı' ve 'aşı' kelimelerini aynı değer içerisinde elde edebilirdi. Benzer şekilde, sözlük tabanlı analizlerde bu yöntemin kullanılması sonuçlarda potansiyel değişikliklere neden olabilirdi. Bu nedenle, büyük harfleri küçük harflere çevirme işleminin uygulanmaması, çalışmanın sınırlılıkları arasında yer almaktadır. Yine veri filtreleme ve temizleme aşamalarında, veri setindeki her kelimenin ek almış farklı formlarını ayırmayarak tek bir kök halinde birleştirilmemesi, yani kök çözümleme (lemmatization) işleminin uygulanmaması, çalışmanın kelime analizi ve sözlük tabanlı analizlerde bazı kısıtlamalara neden olabileceği konusunda dikkate alınmalıdır. Bu durum, özellikle kelime analizi bölümünde sık kullanılan kelimelerin farklı formlarını tanıma ve analiz etme yeteneğini sınırlayabilmektedir. Aynı şekilde, sözlük tabanlı analizlerde de kelime köklerini kullanmamış olmak, analiz sonuçlarının eksik veya yanıltıcı olma olasılığını artırabilir. Bu nedenle, kök çözümleme işlemi eksikliği, çalışmanın metodolojik sınırlılıkları arasında göz önünde bulundurulmalıdır.

Referanslar

Dinh, Ly, and Nikolaus Parulian. "COVID -19 Pandemic and Information Diffusion Analysis on Twitter." *Proceedings of the Association for Information Science and Technology* 57, no. 1 (October 2020): e252. <https://doi.org/10.1002/pra2.252>.

Çağlar, Burak, and Hüseyin Zahit Selvi. "Veri Madenciliği Uygulamalarının Web Tabanlı Mekânsal Görsel Analitik Ortamda Sunumu: COVID-19 Aşı Tweet'leri Örneği." *Afyon Kocatepe University Journal of Sciences and Engineering* 23, no. 2 (May 3, 2023): 417–26. <https://doi.org/10.35414/akufemubid.1206851>.

Catelli, Rosario, Serena Pelosi, Carmela Comito, Clara Pizzuti, and Massimo Esposito. "Lexicon-Based Sentiment Analysis to Detect Opinions and Attitude towards COVID-19 Vaccines on Twitter in Italy." *Computers in Biology and Medicine* 158 (May 2023): 106876. <https://doi.org/10.1016/j.combiomed.2023.106876>.

Ahmed, Wasim, Josep Vidal-Alaball, Joseph Downing, and Francesc López Seguí. "COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data." *Journal of Medical Internet Research* 22, no. 5 (May 6, 2020): e19458. <https://doi.org/10.2196/19458>.

Lyu, Joanne Chen, Eileen Le Han, and Garving K Luli. “COVID-19 Vaccine–Related Discussion on Twitter: Topic Modeling and Sentiment Analysis.” *Journal of Medical Internet Research* 23, no. 6 (June 29, 2021): e24435. <https://doi.org/10.2196/24435>.

Akbulak, Mert Aykut, and Meltem Çöl. “DÜNYADA VE TÜRKİYE’DE AŞILANMA TUTUMU VE COVID-19 AŞILARINA BAKIŞ.” *ESTÜDAM Halk Sağlığı Dergisi* 7, no. 3 (October 20, 2022): 531–40. <https://doi.org/10.35232/estudamhsd.1093147>.

OKTAY, F. (n.d.). PANDEMİ DÖNEMİNDE SOSYAL MEDYA KULLANIMI. *INTERNATIONAL CONGRESS OF EURASIAN SOCIAL SCIENCES-4*, 215.

Kurt, E. V. (2021). ONLINE PANDEMİ: COVID-19 PANDEMİSİNE İLİŞKİN TWITTER’DA PAYLAŞILAN SAĞLIK BİLGİLERİNİN TEMATİK ANALİZİ. *Gümüşhane Üniversitesi İletişim Fakültesi Elektronik Dergisi*.

Narmanli, Didem. “Aşı Kararsızlığı Bağlamında Aşı Tartışmaları: Twitter’da Kovid-19 Örneği.” *TRT Akademi* 7, no. 14 (January 30, 2022): 28–57. <https://doi.org/10.37679/trta.1013435>.

ŞiRzad, Nefise. “Kovid-19 Sürecinde Aşı Kararsızlığı: Aşı Karşıtı Tweetlere İlişkin Bir Analiz.” *TRT Akademi* 7, no. 14 (January 30, 2022): 58–81. <https://doi.org/10.37679/trta.1010334>.

Sütçü, Kübra, Burak Tekerek, and Gökçen Özler. “AŞI KARŞITI TWITTER PAYLAŞIMLARININ METİN MADENCİLİĞİ VE İÇERİK ANALİZİ YÖNTEMİYLE İNCELENMESİ,” n.d.

Yılmaz, Hatice İlke, Başak Turğut, Göksu Çitlak, Oğulcan Mert, Bilge Parali, Muhammed Engin, Aylin Aktaş, and Orhan Alimoğlu. “Türkiye’de İnsanların COVID-19 Aşısına Bakışı.” *Dicle Tıp Dergisi* 48, no. 3 (September 1, 2021): 583–94. <https://doi.org/10.5798/dicletip.988080>.

Narmanli, Didem. “Aşı Kararsızlığı Bağlamında Aşı Tartışmaları: Twitter’da Kovid-19 Örneği.” *TRT Akademi* 7, no. 14 (January 30, 2022): 28–57. <https://doi.org/10.37679/trta.1013435>.

Ersöz, Filiz, and Yasemin Çinar. “Veri Madenciliği ve Makine Öğrenimi Yaklaşımlarının Karşılaştırılması: Tekstil Sektöründe bir Uygulama.” *European Journal of Science and Technology*, December 10, 2021. <https://doi.org/10.31590/ejosat.1035124>.

T.C. Sağlık Bakanlığı [T.C. Sağlık Bakanlığı COVID-19 Bilgilendirme Platformu] (Director). (n.d.). T.C. Sağlık Bakanlığı COVID-19 Bilgilendirme Platformu. <https://covid19.saglik.gov.tr/TR-66300/covid-19-nedir-.html>

Chen, Yanqing, and Steven Skiena. “Building Sentiment Lexicons for All Major Languages.” In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 383–89. Baltimore, Maryland: Association for Computational Linguistics, 2014. <https://doi.org/10.3115/v1/P14-2063>.

Kader, Çiğdem. “AŞI KARŞITLIĞI: AŞI KARARSIZLIĞI ve AŞI REDDİ - ANTI-VACCINATION: VACCINE HESITANCY and REFUSAL.” *ESTÜDAM Halk Sağlığı Dergisi* 4, no. 3 (October 18, 2019): 377–88. <https://doi.org/10.35232/estudamhsd.590304>.

Yan, Wei, Lifan Zhou, Zhengjiang Qian, Le Xiao, and Haixia Zhu. “Sentiment Analysis of Student Texts Using the CNN-BiGRU-AT Model.” Edited by Tongguang Ni. *Scientific Programming* 2021 (October 22, 2021): 1–9. <https://doi.org/10.1155/2021/8405623>.

Aslan, Serpil. “BiGRU-CNN Tabanlı Derin Öğrenme Modeliyle Türkiye’deki Covid-19 Aşılarına Yönelik Twitter Duygu Analizi.” *International Journal of Pure and Applied Sciences* 8, no. 2 (December 31, 2022): 312–30. <https://doi.org/10.29132/ijpas.1087486>.

Çiçen, Yıldırım Beyazıt. “Türkiye’de COVID-19 Aşı Kararlarını Etkileyen Faktörlerin İncelenmesi.” *Fiscaoeconomia* 7, no. 3 (September 18, 2023): 2516–38. <https://doi.org/10.25295/fsecon.1320531>.

Twitter - Vikipedi. (2009, April 18). Vikipedi, Özgür Ansiklopedi. Retrieved January 7, 2024, from <https://tr.wikipedia.org/wiki/Twitter>

Raporlar | Sosyal ağ haritası. (2021). SOSYAL AĞ HARİTASI. Retrieved January 8, 2024, from <https://sosyalagharitasi.gov.tr/report?year=2023>

İşte Türkiye 'de bu yıl kullanıcı sayısını en fazla artıran sosyal medya platformu. (n.d.). Retrieved January 8, 2024, from <https://www.inbusiness.com.tr/sectorler/is-dunyasi/2023/08/16/iste-turkiyede-bu-yil-kullanici-sayisini-en-fazla-artiran-sosyal-medya-platformu>

Turk.Net. (2023, July). Hashtag Nedir? Ne İşe Yarar Ve Nasıl Yapılır? Retrieved January 13, 2024, from <https://turk.net/blog/hashtag-nedir-ne-ise-yarar-ve-nasil-yapilir/>

Öğündür, G. (2021, December 12). Doğruluk (Accuracy) , Kesinlik(Precision) , Duyarlılık(Recall) ya da F1 Score ? *Medium*. <https://medium.com/@gulcanogundur/do%C4%9Fruluk-accuracy-kesinlik-precision-duyarl%C4%B1%C4%B1k-recall-ya-da-f1-score-300c925feb38>

Jayaswal, V. (2021, December 15). Performance Metrics: Confusion matrix, Precision, Recall, and F1 Score. *Medium*. <https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262>

Demiröz, T. A. (2024, February 1). Türkiye\de COVID-19 Aşılması Hakkında Toplumsal Duygusal Durumun Sosyal Medya Verileri Üzerinde Makine Öğrenmesi Metotları Kullanılarak İncelenmesi. Retrieved from <https://aperta.ulakbim.gov.tr/record/263524>

Ekler

Veri Filtreleme ve Temizleme Kodları

CSV Formatını XLSX Formatına Çevirme Kodu

Gereken kütüphaneleri indirilir.

```
import os
import pandas as pd

def filter_and_save_csv(input_folder, output_file):
    # input_folder: CSV dosyalarının bulunduğu klasör bulunur.
    # output_file: Sonuçların kaydedileceği CSV dosyası bulunur.

    # Ardından boş bir DataFrame oluşturulur.
    result_df = pd.DataFrame()

    # Input klasöründeki her CSV dosyası kontrol edilir.
    for file_name in os.listdir(input_folder):
        if file_name.endswith(".csv"):
            file_path = os.path.join(input_folder, file_name)
            # CSV dosyasını okuyoruz.
            df = pd.read_csv(file_path)
            # "lang" sütunu "tr" olanlar seçilir.
            tr_rows = df[df['lang'] == 'tr']
            # Sonuçlar ana DataFrame'e eklenir.
            result_df = pd.concat([result_df, tr_rows])

    # Sonuçlar yeni bir CSV dosyasına kaydedilir.
    result_df.to_csv(output_file, index=False)

# Yazdırma örneği bu şekilde çıkar.
input_folder = "/path/to/your/folder"
output_file = "/path/to/your/folder"
filter_and_save_csv(input_folder, output_file)
```

Tekrar Eden Tivitleri Silme Kodu

Gereken kütüphaneleri indirilir.

import pandas **as** pd

def tekrar_edenleri_sil(veri_seti, sutun_adi, tekrar_limiti=1):

try:

Excel dosyası okunur.

df = pd.read_excel(veri_seti)

'text' sütununda tekrar edenler sayılır ve limit kontrol edilir.

tekrar_sayilari = df['text'].value_counts()

tekrarli_satirlar = tekrar_sayilari[tekrar_sayilari > tekrar_limiti].index

Tekrar eden satırlar silinir.

df = df[~df['text'].isin(tekrarli_satirlar)]

Yeni veri seti kaydedilir.

df.to_excel("yeni_veri_seti.xlsx", index=False)

print("Tekrar edenler silindi, yeni veri seti oluşturuldu.")

except Exception **as** e:

print(f"Bir hata oluştu: {e}")

Excel dosyasının yolu belirtilir.

veri_seti_yolu = "/path/to/your/folder"

Fonksiyonu varsayılan olarak tekrar limiti 1, yani bir tane bırakacak şekilde çağrılır.

tekrar_edenleri_sil(veri_seti_yolu, 'text', tekrar_limiti=1)

Extended Text Bilgilerini Sorgulama Kodu

Gereken kütüphaneler indirilir.

import pandas **as** pd

Excel dosyası okunur.

df = pd.read_excel("/path/to/your/folder")

İlgili sütunlardaki işlemi gerçekleştirilir.

for index, row **in** df.iterrows():

```
if pd.isna(row['extended_tweet.full_text']):
    df.at[index, 'text'] = row['extended_tweet.full_text']

# Sonuç yeni bir Excel dosyasına yazdırılır.
df.to_excel("/path/to/your/folder", index=False)
```

Aşı ile İlgili Tivitleri Seçme Kodu

```
# Gereken kütüphaneleri indirilir.
import pandas as pd
# Excel dosyası okunur (pandas kullanarak).
excel_file = "/path/to/your/folder"
# Excel dosyasının adı ve yolu buraya eklenir.
data = pd.read_excel(excel_file)
# Korunmak istenen kelimeler liste olarak belirtilir.
korunacak_kelimeler = [
    " asi ", " AŞI ", " AŞI", " aşı ", ...
]
# Sadece belirli kelimeleri içeren satırlar seçilir.
filtered_data = data[data['text'].str.contains('|'.join(korunacak_kelimeler), case=False, na=False)]
# Filtrelenmiş veri yeni bir Excel dosyasına kaydedilir.
output_excel = "/path/to/your/folder"
# Kaydedilmek istenen Excel dosyasının adı ve yolu eklenir.
filtered_data.to_excel(output_excel, index=False)
```

“@” ile Mention ve URL İçeren Kelimeleri Silme Kodu

```
# Gereken kütüphaneleri indirilir.
```

```

import pandas as pd
def remove_at_words(text):
    # "@" ve "https:" ile başlayan kelimeleri bulan regex deseni kullanılır.
    pattern = r"s*https://\w+\s*" ve pattern = r"s*@ \w+\s*"
    # Regex desenine uyan kelimeleri boşluk ile değiştirme işlemi gerçekleştirilir.
    result = re.sub(pattern, ' ', text)
    # Başlangıçtaki ve sondaki boşluklar kaldırılır.
    return result.strip()
# Excel dosyası okunur.
input_file_path = "/path/to/your/folder"
output_file_path = "/path/to/your/folder"
# Excel dosyasındaki veri yüklenir.
df = pd.read_excel(input_file_path)
# Belirtilen sütunlardaki metin verilerine uygulanır.
columns_to_process = ['text']
# İlgili sütun adları güncellenir.
for column in columns_to_process:
    df[column] = df[column].apply(remove_at_words)
# Sonuç yeni bir Excel dosyasına kaydedilir.
df.to_excel(output_file_path, index=False)
print(f"İşlem tamamlandı. Sonuçlar '{output_file_path}' dosyasına kaydedildi.")

```

Etiketleme Yönergesi

“Aşılama” ile İlgili Atılan Tivitler Üzerinde Duygu Durum Etiketleme Projesi

Tuna Arda Demiröz · demiroztuna@gmail.com

AMAÇ

Bu etiketleme projesinde amacımız bir yüksek lisans tezine yardımcı olmak amacıyla pandemi döneminde “aşı” ile ilgili atılmış tweetlerden aşya karşı tutumun duygu durumu hakkında bir çıkarım yapabilmektir.

GÖREV ÖZETİ

Veri setimiz “aşı” hakkında atılmış tivitlerden oluşmaktadır. Her bir metni dikkatli bir şekilde okuduktan sonra metni paylaşan kişinin aşya karşı olan tutumunu “negatif”, “pozitif” ve “nötr” seçenekleriyle değerlendirmeniz gerekecektir. Tivitler pandeminin farklı dönemlerinde atılmıştır. Bu sebeple pandeminin başındaki aşı geliştirme süreci, aşılama süreci ve aşılama süreci sonrası dönemleri için farklı metinlerle karşılaşabilirsiniz.

Bu dönemler için baz alabileceğiniz örnekler;

- **Aşı Geliştirme Süreci:** Geliştirilen veya geliştirilecek olan aşilar hakkında verilen bilgiler, haberler ve duygu durum belirten tivitler hakkında tutum çıkarımı yapabilirsiniz.
- **Aşılama Süreci:** Aşılama süreci hakkında devlet politikaları, haberler veya aşılama sonrası oluşan yan etkiler hakkındaki duygu durum belirten tivitler hakkında tutum çıkarımı yapabilirsiniz.
- **Aşılama Süreci Sonrası:** Aşılama üzerinden uzun bir süre geçse de insanlar aşılama hakkında tivitler atmaya devam etmiştir.



Bu dönemler hakkında etiketleme **yapılmayacaktır**. Sadece bilgi verme amaçlıdır.

YÖNERGE İÇERİĞİ

Yönergede örnek etiketleme örnekleri ve etiketlemelerin nerede ve nasıl yapılacağı ile ilgili bilgiler ilerleyen sayfalarda verilecektir. Fakat daha öğretici olması adına aşağıdaki linki kopyalayıp bir başka sekmede yapıştırdıktan sonra 200 verilik bir örnek etiketli veri seti göreceksiniz. Yönergenin tamamını inceledikten sonra bu örnek veri setini incelemeniz rica olunur.

Örnek Veri Seti Linki



https://docs.google.com/spreadsheets/d/fwbkLcWnNqL_UGyxMaJrmokNGWwnj8iBu50AWbTr4kgw/edit?usp=sharing

Şekil 19 Etiketleme Yönergesi (Sayfa 1)

ÖRNEK "NEGATİF" METİNLER

Negatif

40 yaşında sporcu
Tam doz covid 🦠🦠🦠
Oluşan pıhtılar
Şaşırdık mı ?!?!

Yorum:

Burada aşılama sonrası bir durum
"negatif" olarak değerlendirilmiştir.
Pıhtı buradaki odak noktamız.

Negatif

Fauci covid aşıları nedeniyle yüzbinlerce kişinin
ölümünden ve sakatlanmasından yargılanacak hocam
darısı başınıza...🙏

Yorum:

Yine aşılama sonrası genel ama
olumsuz bir tutum. "negatif" olarak
etiketlenmelidir.

Negatif

Açıkça maddi çıkarı olan kişilerin Covid ve aşı gibi önemli
konularda bilimsel tartışmayı etkilemesi endişe verici

Yorum:

Bir olumsuz gibi görünme de
aşılama şüpheli bir yaklaşım söz
konusu burada. "negatif" olarak
etiketlenmelidir.

Negatif

Covid Aşısını. bakanlık resmi yazı dilindeki hukuki ismi
"Deneysel tedavi"ymiş 😊

Yorum:

Bu metindeki gibi ünlem veya emoji
ile tutumu belli eden "negatif"
metinler mevcuttur.



"Negatif" etiketlemeleri diğer etiketlemeye göre daha kolay anlaşılabilir. Aşılama ile ilgili endişeler, yanlış inanışlar, aşılama ile ilgili negatif haberler bu etikete dahil olmalıdır.

ÖRNEK "POZİTİF" METİNLER

Pozitif

En acısı şu,Ece seçkin cahil bir insan değil (mesela Yıldız Tilbe'nin boş konuşması çok mühim değil),ciddi bir üniversite bitirmiş bir insan.Aşıya bağlı yan etki yaşadıysa gerçekten üzgünüm, ama milyonların hayatını kurtaran bir mucizenin üstü kişisel olaylarla çizilemez

Yorum:

Tivit genellikle olumsuz bir tutuma sahip olsa bile altı çizili cümle aşıya karşı tutumun "pozitif" olduğunu gösterir.

Pozitif

Çevremde onlarca kişi en az üç doz biontech aşısı vuruldu. Çoğu aşıdan sonra covid geçirdi. Hepsisi hafif atlattı. Hiçbirinde olumsuz herhangi bir şey gelişmedi.

Yorum:

Tivit aşı sonrası döneme aittir. Aşıların olumsuz bir yanı olmadığı işaret eden ibareler mevcuttur.

Pozitif

Sağlık Emekçilerine İlk Aşılar Yapılmaya Başlandı
#sağlık #asinerede #asiolanlar #CoronaVac #coronavirus

Yorum:

Burada olumlu haber toplumun aşıya karşı tutumunu motive ediyor. Ayrıca #asinerede tiviti aşıya olan inancın da bir göstergesidir.

Pozitif

Ankara Üniversitesi Rektörü Prof. Dr. Ünüvar da aşı oldu
#ankara #ankarauniversitesi #coronavirus #COVID19

Yorum:

İnsanları aşılama teşvik eden bir haber olarak değerlendirip olumlu olarak işaretlenmelidir.



"Pozitif" etiketlemeleri yaparken kişilerin aşılama tutumlarına ek olarak aşılama ile ilgili teşvik edici ve aşılama ile ilgili olumlu haberler de dikkate alınmalıdır.

Şekil 21 Etiketleme Yönergesi (Sayfa 3)

ÖRNEK "NÖTR" METİNLER

Nötr

Bu kadar şiddetli ateş ,boğaz ve kemik ağrısını en son covid olduğumda yaşamıştım. Umarım covid pozitif değildir 😞🙄

Yorum:

Burada covid19 ile ilgili belirtiler olumsuz bir tutum ile belirtilmiştir fakat aşı ile bir ilgisi yoktu. Bu sebeple "nötr" olarak etiketlenmelidir.

Nötr

Hocam Merhaba, Covid 19 aşısı vurulmaya devam edilmesi gerekir mi? İyi çalışmalar dilerim. 😊

Yorum:

Burada tahminimiz bir uzmana soru yöneltilmiş. Fakat aşılamaya ile ilgili tutumun pozitif veya negatif olduğunu anlamamız güç. "nötr" uygun cevap olacaktır burada.

Nötr

*07.07.2018 doğumlu Yiğit Aras
6 aylıkken SMA tanısı kondu
4 doz Spinraza aldı, 5. Dozu alamadı
2 defa kalbi durdu, covid atlattı ve sedyeyle ambulansa taşınırken kalça kemiği kırıldı. Yaşadığı bunca zorluğa rağmen Yiğit Aras hala hayata tutunmaya devam ediyor.*

Yorum:

Aşılamaya ile ilgili bilgi verilse de aşılamaya karşı tutum hakkında herhangi bir çıkarım yapılamıyor. "nötr" olarak etiketlenmesi uygundur.

Nötr

*#fahrettinkoca canlı yayında aşığı vurdurdu ve aşığıya tepkisi 🤔🤔
#coronavirus*

Yorum:

Bu gibi bir başka bağlantıya yönlendiren tвитler mevcuttur. Burada "tepki" olarak bahsedilen duruma pozitif ya da negatif diyemiyoruz. Doğru seçenek "nötr".



"Pozitif" ya da "Negatif" tutum olarak belirleyemediğimiz tвитleri "Nötr" olarak etiketlemek uygundur. Eğer + veya - bir tarafa yakın fakat emin değilsek "Nötr" güvenli olan seçenektir.



Bazı metinler ise aşılamadan tamamen bağımsızdır. Bunlar filtrelemelerden kaçan bazı metinlerdir. Bunlar genellikle "Nötr" etiketi alırlar.

Şekil 22 Etiketleme Yönergesi (Sayfa 4)

NASIL ETİKETLEME YAPILIR?

Sizlere verilecek bağlantı linkleri ile Google Sheets üzerinden görseldeki gibi bir sayfaya yönlendirileceksiniz. "A" sütununda metinler, "B" sütununda ise etiketlemeleriniz olacak.

	A	B
1	text	label
2	İğincir biz de allecek maske ve aşız ve dahi mesafesiz eski güzel normallere göre geçirdik bu 3 senemizi ve hasta bile olmadık. Ancak etrafında covid olup ağır geçirenlerin hemen hemen hepsi aşı ve maskeli korunanlardı	Negatif
3	CoronaVac aşısı sağlık çalışanlarına uygulanmaya başladı #coronavirüs #COVID19	Pozitif
4	Mahkumlara bakmak sizin sorumluluğunda, ceza evi ağır kalabalık, grip mevsimi başladı, herkes hasta, bir şeyler yap artık! Kapalı mahkumlar neden covid izni alamıyor? Hayatları epi derecede değerli değil mi 100yılAflı ŞanınaYakışır #AflatTurkiyem	Nötr
5	Peki 3 doz aşı olmayıp covid olarak bağışıklık kazansaydınız ya da kazanmasaydınız bu yaşadığınız sorunları yapamayacağımız nereden biliyorsunuz?	Negatif
6	En acısı şu,Ece seçkin cahil bir insan değil (mesela Yıldız Tilbe'nin boş konuşması çok mühim değil),ciddi bir Üniversite bitirmiş bir insan.Aşya bağlı yan etki yaşadıysa gerçekten üzgünüm, ama milyonların hayatını kurtaran bir mucizenin üstü kişisel olaylarla çözülemez	Pozitif

"B" sütununda alana klavye tuşu ile geçtiğinizde "enter" a basarsanız etiketleme kategorilerini göreceksiniz. Mouse ile oka tıklarsanız yine aynı şekilde kategoriler karşınıza çıkacaktır.

	A	B
1	text	label
2	İğincir biz de allecek maske ve aşız ve dahi mesafesiz eski güzel normallere göre geçirdik bu 3 senemizi ve hasta bile olmadık. Ancak etrafında covid olup ağır geçirenlerin hemen hemen hepsi aşı ve maskeli korunanlardı	
3	CoronaVac aşısı sağlık çalışanlarına uygulanmaya başladı #coronavirüs #COVID19	Pozitif
4	Mahkumlara bakmak sizin sorumluluğunda, ceza evi ağır kalabalık, grip mevsimi başladı, herkes hasta, bir şeyler yap artık! Kapalı mahkumlar neden covid izni alamıyor? Hayatları epi derecede değerli değil mi 100yılAflı ŞanınaYakışır #AflatTurkiyem	Negatif
5	Peki 3 doz aşı olmayıp covid olarak bağışıklık kazansaydınız ya da kazanmasaydınız bu yaşadığınız sorunları yapamayacağımız nereden biliyorsunuz?	Nötr
6	En acısı şu,Ece seçkin cahil bir insan değil (mesela Yıldız Tilbe'nin boş konuşması çok mühim değil),ciddi bir Üniversite bitirmiş bir insan.Aşya bağlı yan etki yaşadıysa gerçekten üzgünüm, ama milyonların hayatını kurtaran bir mucizenin üstü kişisel olaylarla çözülemez	
7	Adam, Fauci'nin yargılanmasını boşuna istemiyormuş... Nerede ise diğer tarafı boyluyormuş!	
	"İki fena değil. Ancak takviye (booster) beni yerlere serdi. Aşız ilen olduğum COVID'den çok daha berbattı"	

Emin olmadığınız veya fikir almak istediğiniz durumlarda "C" sütununa yorum bırakabilirsiniz. Bu yorumlara cevap yine aynı yerden verilecektir.

	A	B	C
1	text	label	isim
2	İğincir biz de allecek maske ve aşız ve dahi mesafesiz eski güzel normallere göre geçirdik bu 3 senemizi ve hasta bile olmadık. Ancak etrafında covid olup ağır geçirenlerin hemen hemen hepsi aşı ve maskeli korunanlardı	Negatif	
3	CoronaVac aşısı sağlık çalışanlarına uygulanmaya başladı #coronavirüs #COVID19	Pozitif	
4	Mahkumlara bakmak sizin sorumluluğunda, ceza evi ağır kalabalık, grip mevsimi başladı, herkes hasta, bir şeyler yap artık! Kapalı mahkumlar neden covid izni alamıyor? Hayatları epi derecede değerli değil mi 100yılAflı ŞanınaYakışır #AflatTurkiyem	Nötr	Nötr geldinde iğarettedim falan emni değilim. Kontrol edebilir misiniz?
5	Peki 3 doz aşı olmayıp covid olarak bağışıklık kazansaydınız ya da kazanmasaydınız bu yaşadığınız sorunları yapamayacağımız nereden biliyorsunuz?	Negatif	
6	En acısı şu,Ece seçkin cahil bir insan değil (mesela Yıldız Tilbe'nin boş konuşması çok mühim değil),ciddi bir Üniversite bitirmiş bir insan.Aşya bağlı yan etki yaşadıysa gerçekten üzgünüm, ama milyonların hayatını kurtaran bir mucizenin üstü kişisel olaylarla çözülemez	Pozitif	
7	Adam, Fauci'nin yargılanmasını boşuna istemiyormuş... Nerede ise diğer tarafı boyluyormuş!	Negatif	
8	"İki fena değil. Ancak takviye (booster) beni yerlere serdi. Aşız ilen olduğum COVID'den çok daha berbattı"	Negatif	
9	Fauci covid ayları nedeniyle yüzümlerize kişinin ölümünden ve sakatlanmasından yargılanacak hicam damı bağtıza. .	Negatif	
10	retis, serin pudri'n'i beklemeydin? bakurdu o aşısı sana? rus aşısı daha çok severdin hem emirsin?	Negatif	

Şekil 23 Etiketleme Yönergesi (Sayfa 5)

Şekiller Listesi

Şekil 1 Veri Filtreleme Aşamaları	9
Şekil 2 Etiketleme Arayüzü Örnek Görsel.....	12
Şekil 3 Veri Etiketleme Süreci Akışı.....	13
Şekil 4 Tivitlerdeki Genel Duygu Durum Dağılımı.....	15
Şekil 5 Tivitlerdeki Genel Duygu Durum Dağılımı (Nötr Hariç).....	16
Şekil 6 Negatif Duygu Durum Veri Seti için Kelime Bulutu (En Sık Tekrar Eden Kelimeler).....	18
Şekil 7 Pozitif Duygu Durum Veri Seti için Kelime Bulutu (En Sık Tekrar Eden Kelimeler).....	20
Şekil 8 Tüm Veri Seti için En Sık Tekrar Eden Hashtagler.....	21
Şekil 9 Negatif Duygu Durum Veri Seti için Kelime Bulutu (En Sık Tekrar Eden Hashtagler).....	22
Şekil 10 Negatif Duygu Durum Veri Seti için En Sık Tekrar Eden Hashtagler.....	23
Şekil 11 Pozitif Duygu Durum Veri Seti için Kelime Bulutu (En Sık Tekrar Eden Hashtagler).....	24
Şekil 12 Pozitif Duygu Durum Veri Seti için En Sık Tekrar Eden Hashtagler.....	25
Şekil 13 İnsan Etiketli Veri Seti ve Model Etiketli Veri Seti Sayısal Karşılaştırması.....	27
Şekil 14 Sapma Payı Formülü.....	28
Şekil 15 Modelin İnsan Etiketli Veri Setindeki Yüzdesel Tahmin Sonuçları.....	29
Şekil 16 Modelin Hata Matrisi (Confusion Matrix) Tablosu.....	32
Şekil 17 Negatif, Pozitif ve Nötr Duygu Durumların Yıllara Göre Değişimi.....	34
Şekil 18 Negatif, Pozitif ve Nötr Duygu Durumun Zamana Göre Yüzdesel Değişimi.....	35
Şekil 19 Etiketleme Yönergesi (Sayfa 1).....	51
Şekil 20 Etiketleme Yönergesi (Sayfa 2).....	52
Şekil 21 Etiketleme Yönergesi (Sayfa 3).....	53
Şekil 22 Etiketleme Yönergesi (Sayfa 4).....	54
Şekil 23 Etiketleme Yönergesi (Sayfa 5).....	55