

*

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**TWITTER VERİSİ ÜZERİNDE COVID-19'A KARŞI ETKİLİ OLAN
İLAÇLARIN DOĞAL DİL İŞLEME KULLANILARAK ANALİZİ**

YÜKSEK LİSANS TEZİ

Orhan Talha KUM

Bilgisayar Bilimleri Anabilim Dalı

Bilgisayar Bilimleri Programı

Tez Danışmanı: Dr. Sefer BADAY

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**TWITTER VERİSİ ÜZERİNDE COVID-19'A KARŞI ETKİLİ OLAN
İLAÇLARIN DOĞAL DİL İŞLEME KULLANILARAK ANALİZİ**

YÜKSEK LİSANS TEZİ

**Orhan Talha KUM
(704191015)**

Bilgisayar Bilimleri Anabilim Dalı

Bilgisayar Bilimleri Programı

Tez Danışmanı: Dr. Sefer BADAY

İTÜ, Lisansüstü Eğitim Enstitüsü'nün 704191015 numaralı Yüksek Lisans Öğrencisi Orhan Talha KUM, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “TWITTER VERİSİ ÜZERİNDE COVID-19'A KARŞI ETKİLİ OLAN İLAÇLARIN DOĞAL DİL İŞLEME KULLANILARAK ANALİZİ” başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Dr. Sefer BADAY**
İstanbul Teknik Üniversitesi



ÖNSÖZ

Yüksek Lisans eğitimim ve tez sürecimde desteklerini, deneyimlerini benden esirgemeyen ve birlikte çalışmaktan çok memnun olduğum değerli tez danışmanım Dr. Sefer Baday’a ve bu süreçte hep benim yanımda olan ve beni destekleyen sevgili aileme çok teşekkür ederim.

Haziran 2022

Orhan Talha Kum
Bilgisayar Mühendisi



İÇİNDEKİLER

	<u>Sayfa</u>
ÖNSÖZ	vi
İÇİNDEKİLER.....	viii
KISALTMALAR.....	ix
ÇİZELGE LİSTESİ.....	x
ŞEKİL LİSTESİ.....	xi
ÖZET	xiii
SUMMARY.....	xvi
1. GİRİŞ	1
1.1 Doğal Dil İşleme Uygulamaları.....	2
1.1.1 Duygu analizi	2
1.1.1.1 Duygu analizi sınıflandırma seviyeleri	3
1.1.1.2 Hedef tabanlı duygu analizi.....	4
1.1.1.3 Hedefler için duyguları bulma.....	4
1.1.1.4 Duygu sınıflandırılmasında özellik seçimi	5
1.2 Literatür Araştırması	5
2. VERİ SETİ.....	7
3. UYGULAMA.....	8
3.1 Veri Setinin Çıkarılıp Yaklaşık Dizgi Eşleme İle İlaç İsimlerinin Aranması	9
3.1.1 Yaklaşık Dizgi Eşleme	9
3.1.2 Yaklaşık Dizgi Eşleme Uygulama Alanları	10
3.1.3 Yaklaşık Dizgi Eşleme'nin Uygulanması	11
3.2 Veri Ön İşleme.....	12
3.3 Duygu Analizinin Uygulanması	12
3.3.1 Transformers	12
3.3.2 Vader	16
3.3.3 TextBlob	18
3.4 Bilim İnsanlarına Yönelik Duygu Analizi.....	20
6. SONUÇ VE ÖNERİLER.....	23
KAYNAKLAR.....	25
ÖZGEÇMİŞ.....	35

KISALTMALAR

COVID-19	: Coronavirus Disease of 2019
API	: Application Programming Interface
DDİ	: Doğal Dil İşleme
UHEM	: Ulusal Yüksek Başarımli Hesaplama Merkezi
DNA	: Deoxyribonucleic Acid
NYSIIS	: New York State Identification and Intelligence System
NLTK	: Natural Language Toolkit
VADER	: Valence Aware Dictionary for Sentiment Reasoning
IEEE	: The Institute of Electrical and Electronics Engineers
SARS-CoV-2	: Severe Acute Respiratory Syndrome Coronavirus 2

ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 5.1 : Bilim İnsanlarına Yönelik Duygu Analizi Oranları.	20
Çizelge 5.2 : Bilim İnsanları Haricinde Duygu Analizi Oranları.	21



ŞEKİL LİSTESİ

Sayfa

Şekil 1.1 : Duygu analizi işlem aşamaları.	3
Şekil 3.1 : Metodoloji.	8
Şekil 3.2 : Csv dosya içeriği.	9
Şekil 3.3 : İlaç isminin yakшалık dizgi eşleme ile bulunması.	11
Şekil 3.4 : MongoDB dökümanları içerisinde ilaç isimleri.	12
Şekil 3.5 : Tüm ilaç isimlerinin aylara göre duygu analizi oranları.	14
Şekil 3.6 : Hydroxychloroquine için aylara göre duygu analizi oranları.	14
Şekil 3.7 : Remdesivir için aylara göre duygu analizi oranları.	15
Şekil 3.8 : Azithromycin için aylara göre duygu analizi oranları.	15
Şekil 3.9 : Tüm ilaç isimlerinin aylara göre duygu analizi oranları.	16
Şekil 3.10 : Hydroxychloroquine için aylara göre duygu analizi oranları.	17
Şekil 3.11 : Remdesivir için aylara göre duygu analizi oranları.	17
Şekil 3.12 : Azithromycin için aylara göre duygu analizi oranları.	18
Şekil 3.13 : Tüm ilaç isimlerinin aylara göre duygu analizi oranları.	18
Şekil 3.14 : Hydroxychloroquine için aylara göre duygu analizi oranları.	19
Şekil 3.15 : Remdesivir için aylara göre duygu analizi oranları.	19
Şekil 3.16 : Azithromycin için aylara göre duygu analizi oranları.	20
Şekil 6.1 : Tüm ilaçlar için duygu analizi oranları.	24
Şekil 6.2 : Tüm ilaçlar için bilim insanlarına yönelik duygu analizi oranları. ...	24
Şekil A.1 : Veri setinin çekilmesi için Python işlemleri 1. kısım	29
Şekil A.2 : Veri setinin çekilmesi için Python işlemleri 2. kısım	30
Şekil A.3 : Veri setinin çekilmesi için Python işlemleri 3. kısım	30
Şekil B.1 : İlaç ismi geçen tweetleri aramak için Python işlemleri 1. kısım.	31
Şekil B.2 : İlaç ismi geçen tweetleri aramak için Python işlemleri 2. kısım.	32
Şekil B.3 : İlaç ismi geçen tweetleri aramak için Python işlemleri 3. kısım.	33
Şekil C.1 : Duygu analizi için Python işlemleri 1. Kısım	33
Şekil C.2 : Duygu analizi için Python işlemleri 2. kısım	34
Şekil D.1 : Bilim adamlarının attığı tweetleri bulmak için Python işlemleri.	34
Şekil D.2 : Bütün ilaç isimleri için Transformers ile duygu analizi sonuçlarının yazdırılması	34



TWITTER VERİSİ ÜZERİNDE COVID-19'A KARŞI ETKİLİ OLAN İLAÇLARIN DOĞAL DİL İŞLEME KULLANILARAK ANALİZİ

ÖZET

Yeni Koronavirüs Hastalığı (COVID-19), ilk olarak Çin'in Hubei eyaletinin Wuhan şehrinde ortaya çıkıp solunum yolu belirtileri (ateş, öksürük, nefes darlığı) gösteren bir grup hastada yapılan araştırmalar sonucunda 13 Ocak 2020'de tanımlanan şiddetli akut solunum sendromu koronavirüsü 2 (SARS-CoV-2)'nin neden olduğu bir hastalıktır. Kişiden kişiye bulaşabilen bu virüsün bulaşma oranı 2020 Ocak ayında büyüme gösterdi. Dünya üzerinde çeşitli ülkelerde yaşanan virüs vakaları görülmeye başlandı ve 11 Mart 2020'de Dünya Sağlık Örgütü tarafından pandemi ilan edildi. 27 Mart 2022 itibarıyla toplam vaka sayısı 479 milyon, koronavirüs kaynaklı ölüm sayısı ise 6 milyon olmuştur. Koronavirüs tedavisi için aşı ve ilaç çalışmalarına kısa sürede başlanmış ve günümüzde aktif olarak geliştirilmeye devam etmektedir. 27 Mart 2022 itibarıyla dünya genelinde 11.2 milyar doz aşı uygulanmış olsa da an itibarıyla, COVID-19'a enfekte olan kişileri tedavi etmek için onaylanmış, güvenli ve etkili bir farmakolojik ajan bulunmamaktadır. COVID-19 tedavisinde etkili olduğu düşünülen ilaçlar insanlar tarafından kullanılmış ve kullanılmaya devam etmektedir.

Etkili olup olmadığı halen tartışma konusu olan ve COVID-19 tedavisinde kullanılan ilaçlara yönelik insanların düşüncelerini tespit etmek amaçlanmış ve bu doğrultuda sosyal medya platformlarından birisi olan Twitter tercih edilmiştir. Twitter günümüzde en yaygın olarak kullanılan sosyal medya platformlarından birisi olup akademik çalışmalarda yaygın olarak kullanılmaktadır. Twitter'ın Uygulama Programlama Arayüzü sayesinde insanların paylaşmış olduğu gönderilere ulaşabilmekte ve bu sayede veriler üzerinde akademik çalışmalar yapılabilir.

Bu çalışmada, IEEE'nin 20 Mart 2020'de COVID-19 ile alakalı paylaşılmış olan, içerisinde tweet numaralarının olduğu ve günlük olarak güncellenen veri seti kullanılmıştır. 2020 Mart ayı ile 2022 Şubat ayı arasındaki veriler çekilmiştir. Veri seti ilk olarak İstanbul Teknik Üniversitesi bünyesindeki yüksek başarılı hesaplama ve veri depolama amacıyla kullanılan Ulusal Yüksek Başarılı Hesaplama Merkezi (UHEM) platformuna aktarılmıştır. Veri seti günlük olarak yaklaşık 1-3 milyon arasında veri içerdiği için veriler parça parça aktarılmıştır. Veri seti içerisinde yalnızca tweetin numarası olduğu için tweet içeriği Twitter Uygulama Programlama Arayüzü aracılığı ile bir Python betiği kullanılarak çekilmiştir. Toplam 1.4 milyar tweet bu şekilde çekilip bir NoSQL veritabanı olan MongoDB'ye yazılmıştır. Veri seti oluştuktan sonra başka bir Python betiği aracılığıyla içerisinde COVID-19'a karşı etkili olduğu düşünülen hydroxychloroquine, chloroquine, remdesivir ve azithromycin ilaçları geçen tweetler yaklaşık dizgi eşleme metodu kullanılarak çekilmiştir. Yaklaşık dizgi eşleme metodu kullanılmasının sebebi bu metin arama metodunun ilaç isminin yanlış yazılması durumunda da eşleşme sağlamasıdır.

İçerisinde ilaç ismi geçen tweetler yeni bir veri setine aktarılmıştır. Yeni oluşan bu veri seti 3.5 milyon tweetten oluşmaktadır. Daha sonra veri seti içerisindeki tweet metinleri

iin doęal dil iřlemenin gerekli olan adımlarından birisi olan veri niřleme gerekleřtirilmiřtir. Veri niřleme ařamasında tweet ierisindeki RT, kullanıcı ismi, linkler, hashtag'in bařındaki # iřareti ve bořluklar silinmiřtir.

Veri niřleme adımından sonra ise doęal dil iřleme ktphanelerinden Transformers, Vader ve TextBlob kullanarak tweet metinleri zerinde duygu analizi uygulanmıřtır. Duygu analizi sonucunda metnin olumlu mu olumsuz mu olduęunu belirten bir duygu analizi skoru retilmiřtir. Duygu analizi skorları pozitif, ntr ve negatif olarak gruplanıp oranları aylık olarak veri grselleřtirme teknikleri kullanarak gsterilmiřtir. alıřmada bilim insanlarının ilalara olan dřncelerini tespit edebilmek amacıyla da sadece bilim insanlarına ynelik duygu analizi oranları da gsterilmiřtir.

alıřma sonucunda 2020 yılında ilalara karřı daha ntr ve negatif daęılımı az olan bir tablo grlrken 2021 Haziran ayından sonra negatif daęılımın arttıęı ve insanların ilalara olan dřncelerinin olumsuz ynde devam ettięi grlmřtr. Bilim insanlarına ynelik duygu analizi oranlarına bakıldıęında ise bilim insanları haricindeki kitleye gre daha ntr bir tablo grlmektedir.

alıřmanın sonucu olarak halen geliřtirilmeye devam eden ve pandeminin ortaya ıktıęı gnlerden beri COVID-19 tedavisine ynelik kullanılmaya devam eden ilalar hakkında insanların dřncelerinin zamanla olumsuz ynde seyir ettięi tespit edilmiřtir. Bu alıřmanın COVID-19'a etkili olan ilalar hakkında yapılacak olan akademik alıřmalara referans olacaęı dřnlmektedir.



ANALYSIS OF DRUGS EFFECTIVE AGAINST COVID-19 USING NATURAL LANGUAGE PROCESSING ON TWITTER DATA

SUMMARY

The Novel Coronavirus Disease (COVID-19) is a disease that is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which was first identified in China's Wuhan City of Hubei province on January 13, 2020, as a result of research conducted on a group of patients who showed respiratory tract symptoms (fever, cough, breathing difficulties). The transmission rate of this virus, which can transmit human-to-human, increased in January 2020. Virus cases started to appear worldwide and pandemic was declared by The World Health Organization (WHO) on 11 March 2020. As of 27 March 2022, the pandemic had caused more than 479 million cases and 6 million confirmed deaths. Vaccine and drug studies for the treatment of coronavirus started in a short time and continue to be actively developed nowadays. Although as of 27 March 2022, 11.2 billion vaccine doses have been administered worldwide, at this time, there are no approved, safe, and effective pharmacologic agents to treat COVID-19 infected patients. The drugs believed effective in the treatment of COVID-19 were used by people and continue to be used.

It was aimed to determine the opinions of people about the drugs used in the treatment of COVID-19, whose effectiveness is still a matter of debate, and in this direction, Twitter which is one of the social media platforms was preferred. Twitter is one of the most widely used social media platforms today and it is widely used in academic studies. Thanks to the Twitter's Application Programming Interface (API), the posts shared by people can be accessed and thus academic studies can be carried out on the data.

In this study, the dataset, which was shared by the IEEE about the COVID-19 related drugs on 20 March 2020, and contains tweet ids and is updated every day, was used. Data between March 2020 and February 2022 were fetched. The dataset was first extracted to the National Center for High Performance Computing (UHEM) platform, which is used for high-performance computation and data storage within the İstanbul Technical University. The dataset was extracted as batches because it contains approximately 1-3 million data per day. Because the dataset contains only tweet ids, tweet content was fetched by using a Python script through Twitter's API. A total of 1.4 billion tweets were fetched and inserted into MongoDB which is a NoSQL database. After the dataset was created, tweets that contain hydroxychloroquine, chloroquine, remdesivir, and azithromycin, which are thought to be effective in the treatment of COVID-19, were fetched with another Python script by using approximate string matching method. The reason for using approximate string matching method is that this text search method is matched when the drug name is misspelled.

The tweets containing drug names were extracted to a new dataset. This new dataset consists of 3.5 million tweets. Then, data preprocessing, which is one of the necessary

steps in natural language processing, was applied to the tweet texts in the dataset. In the data preprocessing step, RT, username, links, # sign at the beginning of hashtags, and spaces were removed from tweets.

After data preprocessing, sentiment analysis was performed on the tweet texts by using Transformers, Vader, and TextBlob, which are natural language processing libraries. As a result of sentiment analysis, sentiment score was generated that indicates whether the tweet is positive or not. Sentiment scores were grouped as positive, neutral, and negative and their rates were shown monthly by using data visualization techniques. In the study, sentiment scores for scientists were also shown in order to determine the scientists' opinions about drugs.

While a picture with a more neutral and less negative distribution of the drugs was observed in 2020, after June 2021, it was observed that the negative distribution increased and people's opinions about the drugs continued in a negatively. When the sentiment scores for scientists are examined, a more neutral Picture is seen compared to the audience excluding scientists.

As a result of the study, it was determined that the people's opinions about the drugs that are still being developed and that have been used for the treatment of COVID-19 since the outbreak of the pandemic have a negative course over time. It is thought that this study will be a reference for academic studies about COVID-19-related drugs.



1. GİRİŞ

Yeni Koronavirüs Hastalığı (COVID-19), ilk olarak Çin'in Hubei eyaletinin Vuhan şehrinde ortaya çıkıp solunum yolu belirtileri (ateş, öksürük, nefes darlığı) gösteren bir grup hastada yapılan araştırmalar sonucunda 13 Ocak 2020'de tanımlanan şiddetli akut solunum sendromu koronavirüsü 2 (SARS-CoV-2)'nin neden olduğu bir hastalıktır [1]. Kişiden kişiye bulaşabilen bu virüsün bulaşma oranı 2020 Ocak ayında büyüme gösterdi. İlerleyen zamanlarda Avrupa, Kuzey Amerika ve Asya-Pasifik'te yer alan çeşitli ülkelerde virüs vakaları rapor edilmeye başlandı ve Dünya Sağlık Örgütü (WHO) tarafından pandemi ilan edildi [2,3]. Pandemi, kısa zamanda tüm dünyada sağlıktan ekonomiye neredeyse tüm alanlarda etkisini göstermiştir. 27 Mart 2022 itibarıyla toplam vaka sayısı 479 milyon, koronavirüs kaynaklı ölüm sayısı ise 6 milyon olmuştur [4]. Pandemi başladıktan kısa bir zaman sonra aşı ve ilaç çalışmalarına başlanmış ve günümüzde aktif olarak geliştirilmeye devam etmektedir. 27 mart 2022 itibarıyla dünya genelinde 11.2 milyar doz aşı uygulanmış olsa da an itibarıyla, COVID-19'a enfekte olan kişileri tedavi etmek için onaylanmış, güvenli ve etkili bir farmakolojik ajan bulunmamaktadır [5,6]. COVID-19 tedavisinde etkili olduğu düşünülen ilaçlar insanlar tarafından kullanılmış ve kullanılmaya devam etmektedir.

Bu çalışmada, etkili olup olmadığı halen tartışma konusu olan ve COVID-19 tedavisinde kullanılan ilaçlara yönelik insanların düşüncelerini tespit etmek amaçlanmış ve bu doğrultuda sosyal media platformlarından birisi olan Twitter tercih edilmiştir. Twitter'ın tercih edilmesinin sebebi, günümüzde en yaygın olarak kullanılan sosyal medya platformlarından birisi olması ve akademik çalışmalarda yaygın olarak kullanılmasıdır.

Koronavirüs sürecinde Twitter kullanımı, insanların koronavirüse karşı olan duygu ve düşüncelerini tespit etmek için önemli bir veri kaynağı olmuştur. Bu amaçla yapılan çalışmaların bir kısmı literatür araştırması kısmında gösterilmiştir. Koronavirüse karşı etkili olan ilaçlara yönelik insanların düşüncelerini tespit edebilmek amacıyla Doğal Dil İşleme (DDİ) teknikleri kullanılmıştır.

DDİ, yapay zekanın altında yer alan, insanların ürettiği doğal dilin işlenmesinde kullanılan önemli bir araştırma alanı olarak Twitter üzerindeki verilerin işlenmesinde pek çok çalışmada kullanılmıştır. Bu çalışmada öncelikle koronavirüsle alakalı atılan tweet'ler çekilmiştir. Ardından bu tweetler içerisinde koronavirüse karşı etkili olduğu düşünülen ilaçlar hakkında yazılmış olanlar Fuzzy Search kullanılarak filtrelenmiştir. Daha sonra DDİ teknikleri kullanılarak içerisinde koronavirüse karşı etkili olduğu düşünülen ilaç isimleri geçmiş olan tweetler analiz edilmiştir.

Bu bölümde DDİ'nin kullanım amacı ve etkili olduğu alanlar hakkında bahsedilecektir.

1.1 Doğal Dil İşleme Uygulamaları

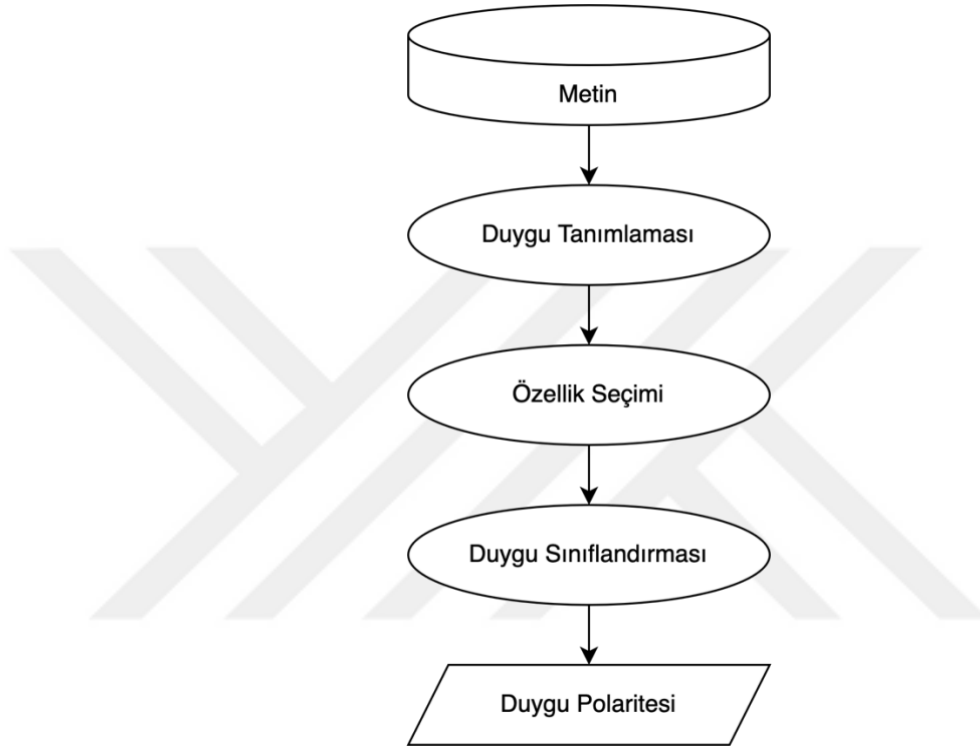
DDİ'nin pek çok farklı alanda uygulama amacı vardır. Bunlardan en sık kullanılanları aşağıda listelenmiştir.

- Makine Çevirisi: Bir doğal dili diğerine çevirmek için kullanılan bir uygulamadır.
- Otomatik Özetleme: Bir dökümandan en anlamlı bilgiyi çıkarıp dökümanı özetlemek amacıyla kullanılır.
- Duygu Analizi: Bir içerikteki duygunun olumlu, olumsuz ya da nötr olup olmadığını belirlemek amacıyla kullanılır.
- Sağlık Hizmetleri: Yüksek riskli hastaları saptamak amacıyla kullanılır.
- Metin Madenciliği: Bir metindeki yüksek nitelikteki veriyi çıkarmak amacıyla kullanılır.
- Eğitim: Gramer doğrulama, kelime düzeltme gibi alanlarda geri bildirim vermek amacıyla kullanılır.
- Tarım: Sensörlerden toplanan veriler aracılığıyla modern tarımdaki çalışma ortamı hakkında anlamlı bilgiler üretmek amacıyla kullanılır [7].

1.1.1 Duygu analizi

Duygu analizi ya da fikir madenciliği insanların fikirlerini, duygularını, değerlendirmelerini, tutumlarını ve duygularını yazı dilinden analiz eden çalışma alanıdır. DDİ'nin en aktif çalışma alanlarından birisidir ve veri madenciliği, web

madenciliği ve metin madenciliğinde yaygın olarak kullanılır [8]. Duygu analizi ve fikir madenciliği kavramları birbirlerinin yerine kullanılsa da bazı araştırmacılar az oranda kavram farklılığı olduğunu belirtmişlerdir [9]. Fikir madenciliği bir varlıkla ilgili insanların fikirlerini analiz ederken duygu analizi bir metindeki duyguyu belirleyip daha sonra analiz eder. Bu yüzden duygu analizinin hedefi fikirleri bulmak, aktardığı duyguları belirlemek ve sonra onların polaritesini sınıflandırmaktır [10].



Şekil 1.1 : Duygu analizi işlem aşamaları.

1.1.1.1 Duygu analizi sınıflandırma seviyeleri

Duygu analizinde 3 farklı seviye vardır.

- Doküman seviyesi: Bu aşamada duygu metinden çıkarılır ve tüm fikir, fikir sahibinin duygusuna bağlı olarak sınıflandırılır. Amaç, metni olumlu, olumsuz ya da nötr olarak sınıflandırmaktır.
- Cümle seviyesi: Bu aşama her cümlede ifade edilen duyguları sınıflandırmayı amaçlar. İlk adım, cümlemin objektif mi yoksa subjektif mi olduğunu belirlemektir. Eğer cümle subjektif ise cümlemin olumlu mu yoksa olumsuz mu bir fikir ürettiğine karar verir.

- Hedef seviyesi: Bu aşamada amaç fikir sahibi tarafından yorumlanan obje özelliklerini çıkarmak ve fikrin olumlu, olumsuz ya da nötr olup olmadığına karar vermektir.

1.1.1.2 Hedef tabanlı duygu analizi

Hedef tabanlı duygu analizi, verilen hedef varlıkların farklı yönlerini tanımlama ve her ifade edilen fikrin duygu polaritesini tahmin etmeye dayalıdır. Bu, 2 ayrı aşamada gerçekleştirilir.

- Hedef çıkarma: Bir cümledeki bütün hedef terimlerini tanımlamakla ilgilidir.
- Hedef duygu sınıflandırması: Farklı hedeflerdeki fikirlerin olumlu, olumsuz ya da nötr olup olmadığını belirler. Bunu, duygu sözlükleri sayesinde gerçekleştirir. Olumlu ya da olumsuz olarak sınıflandırılan kelimelere sahip olan farklı tipte duygu sözlükleri mevcuttur. Örnek olarak: The General SInquirer, LIWC, MPQA Subjective Cues Lexion, Bing Liu's Opinion Lexicon ve SentiWordNet verilebilir. Online sözlükler yeterli kelime içermedikleri veya metinde geçen alanla ilgili olmadığı için yeterli olmayabilir. Bu durumda yeni bir duygu sözlüğü oluşturulabilir. Bu, yarı-gözetimli sözlükler sayesinde yapılabilir.

1.1.1.3 Hedefler için duyguları bulma

Hedefler için duyguları bulma 4 ayrı aşamaya bölünebilir [11].

- Metin çıkarımı: Metinden cümleleri ve ifadeler çıkarılır.
- Duygu sınıflandırılması: Her çıkarılan cümle ve terim için için olumlu, olumsuz ya da nötr olup olmadığına karar vermek için bir duygu sınıflandırıcısı çalıştırılır.
- Hedef çıkarıcı: Bir duygu ifade eden hedefler için hedef terim çıkarıcı uygulamalar ve tanımlanan her hedef için duygu polaritesini belirler.
- Hedef polarite birleşmesi: Hedefler için duyguları gruplandırır ve nihai özeti üretir (örn. Treatment: positive, covid: negative).

1.1.1.4 Duygu sınıflandırılmasında özellik seçimi

Duygu sınıflandırma, duygu analizinde önemli aşamalardan birisidir. Duygu sınıflandırmadaki ilk adım metin özelliklerini seçme ve çıkarmadır. Mevcut özelliklerden bazıları aşağıdaki gibidir [12]:

- Terimlerin varlığı ve sıklığı: Bu özellikler tekil kelimeler ya da kelime n-gramları ve onların sıklık sayılarıdır. Kelimelere ikili (binary) ağırlık verilir ya da özelliklerin bağıl önemliliklerini belirlemek için terim frekans ağırlıkları kullanılır [13].
- Sözcük türleri: Fikirler için önemli bir gösterge olduğu için sıfatlar bulunur.
- Fikir kelimeleri ve cümleleri: İyi, kötü, çirkin ve nefret gibi yaygın olarak fikirleri belirtmek için kullanılan kelimelerdir.
- Olumsuzlamalar: Olumsuz kelimelerin ortaya çıkması fikir yönelimini değiştirebilir.

1.2 Literatür Araştırması

Daha önce koronavirüse karşı etkili olan ilaçlar hakkında atılan tweetler için yapılan duygu analizi çalışmaları incelenmiştir.

[14] numaralı çalışmada 28 Şubat 2020 ile 22 Mayıs 2020 arasında içerisinde hydroxychloroquine, remdevisir and convalescent plasma bahsedilmiş olan tweetler çekilip tweet sayısı, günlük yapılan retweet sayısı ve duygu analizi yapılmıştır. Çalışma sonucunda Amerika kongre üyeleri tarafından atılan 72 tweet'i duygu analizleri ile 3 partiye bölerek gösterilmiştir. Çalışmada kısıtlı bir zaman aralığına bakılmış ve sonuç kısmında sadece kongre üyelerinin attığı tweet'lerin duygu analizi paylaşılmıştır.

[15] numaralı çalışmada COVID-19 ile alakalı atılan 424 milyon tweet çekilmiş. Bunların arasından içerisinde ilaç geçenler filtrelenmiş. Sonuç olarak en çok bahsedilen tweetler frekans değerleri de verilerek paylaşılmış. Bu çalışmada herhangi bir duygu analizi yapılmayıp gelecek çalışmalar kısmında duygu analizi uygulanacağından bahsedilmiştir.

[16] numaralı çalışmada 1 Mart 2020 ile 5 Haziran 2020 arasındaki COVID-19 ile alakalı paylaşılan 84.04M tweet çekilmiştir. Çalışma sonucunda ülkelere göre ve

#WorkFromHome, #SocialDistancing gibi hashtagler kullanılarak atılan tweetler için duygu analizi yapılmıştır. Çalışmada içinde ilaç isimleri geçen tweetler için bir duygu analizi yapılmamıştır.

[17] numaralı çalışmada 13 Aralık 2019 ile 9 Mart 2020 arasında koronavirüs ile alakalı spesifik hashtagler kullanılan tweetler çekilmiş. Bu çalışma sonucunda duygu analizi ve konu modelleme uygulanmıştır ancak ilaç ile alakalı olan tweetler için bir çalışma yapılmadığı görülmüştür.



2. VERİ SETİ

Çalışmada kullanılan veri seti IEEE'nin koronavirüslerle alakalı paylaşılan tweetler veri setinden çekilmiştir [18].

Bu veri seti, içerisinde tweet numarası ve TextBlob duygu analizi skoru bulunan ve günlük olarak paylaşılan csv dosyalarından oluşmaktadır. Veri setinin oluşumunda, 90'dan fazla farklı anahtar kelime ve pandemiye referans eden yaygın kullanılan hashtag'ler kullanılarak koronavirüsle alakalı atılan tweet'ler için Gerçek zamanlı Twitter beslemesi izlenmiştir. Veri seti 20 Mart 2020 tarihinde oluşmaya başlanmıştır.

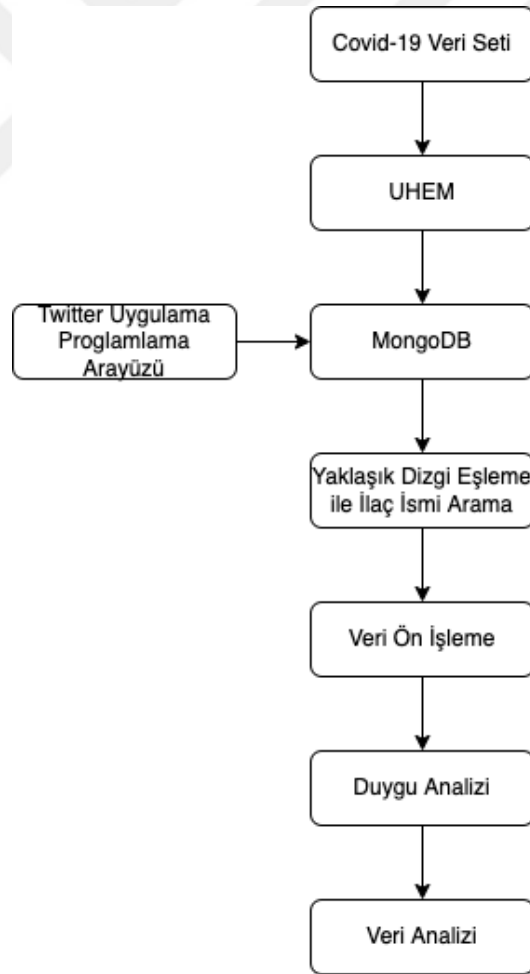
1 csv dosyasının içerisinde yaklaşık olarak 1-3 milyon arasında tweet bulunmaktadır. Çalışmada toplam 700 csv dosyası kullanılmıştır. Dosyalar IEEE üzerinde günlük ve 100'lük toplu zip dosyaları halinde tutulmaktadır.

Çalışmada dosyalar 100'lük zip dosyaları halinde UHEM (Ulusal Yüksek Başarımlı Hesaplama Merkezi) sistemine yüklenmiştir.

Twitter'ın veri ile ilgili olan politika kısıtlamaları sebebiyle tweet içeriği veri seti içerisinde paylaşılmamıştır. Bu nedenle Twitter'ın Uygulama Programlama Arayüzü'ne erişmeyi sağlayan bir Python kütüphanesi olan Tweepy aracılığıyla tweetler tweet id kullanılarak çekilmiştir.

3. UYGULAMA

Çalışmada ilk olarak veri seti UHEM sistemine aktarılmıştır. Aktarılan veriler Twitter Uygulama Programlama Arayüzü kullanılarak özümlenip MongoDB veritabanına yazılmıştır. Daha sonra yaklaşık dizgi eşleme yöntemi kullanılarak içerisinde ilaç ismi geçen tweetler bir csv dosyasına aktarılmıştır. Bu dosyadaki veriler üzerinde daha sonra veri ön işleme yapıp duygu analizi metodları uygulanmıştır. Çalışmanın sonunda veriler grafikler ile analiz edilip çıkarım yapılmıştır. Aşağıdaki şekilde çalışmada uygulanan metodoloji gözükmemektedir.



Şekil 3.1 : Metodoloji.

3.1 Veri Setinin Çıkarılıp Yaklaşık Dizgi Eşleme İle İlaç İsimlerinin Aranması

Veri seti gerçek zamanlı olarak beslendiği için dosyalar aralıklarla tekrar çekilmiştir. Veri seti UHEM sistemine yüklendikten sonra bir Python betiği ile Tweepy kütüphanesi kullanılarak tweet içeriği çekilip MongoDB veri tabanına yazılmıştır. Veri tabanı olarak MongoDB kullanılmasının sebebi MongoDB'nin yüksek hacimli veri depolama için tasarlanmış bir NoSQL veri tabanı olmasıdır.

Tweetler MongoDB'ye kaydedildikten sonra bir Python betiği kullanılarak içerisinde ilaç ismi geçen dökümanlar yaklaşık dizgi eşleme metodu ile çekilip bir csv dosyasına aktarılmıştır.

dateTime	_id	sentiment_score	created_at	text	user_id	user_description	drugName	vader_score	retweeted_st
2020-03-19 19:52:38	1240727902976716802	0.00	2020-03-19 19:52:38	RT @INKs_Shirt: Waiting for a hataraku saibou e...	1240727902976716802	are	hydroxychloroquine	0.0000	I
2020-03-19 19:52:53	1240727966633558016	0.35	2020-03-19 19:52:53	@StreamAJ Is it true that chloroquine is used ...	1240727966633558016	NaN	chloroquine	0.4215	F
2020-03-19 19:53:04	1240728013035184137	-0.70	2020-03-19 19:53:04	RT @PatPenn2: Bill Hemmer, on Fox, just had a D...	1240728013035184137	Happily married to my best friend, Marine Offi...	chloroquine	-0.4939	I
2020-03-19 19:53:09	1240728034178666496	0.75	2020-03-19 19:53:09	RT @BDUTT: Just want to say that our Indian do...	1240728034178666496	Love travel and music	chloroquine	0.6908	.
2020-03-19 19:53:14	1240728056173559809	0.40	2020-03-19 19:53:14	Now that it has been confirmed that chloroquin...	1240728056173559809	stay focused and be guided 🍌 Food addict 🍌 Foo...	chloroquine	0.8822	F

Şekil 3.2 : Csv dosya içeriği.

3.1.1 Yaklaşık Dizgi Eşleme

Dizgi eşleme bir metindeki bütün karakter örüntü oluşumlarının bulunması problemidir [19].

Yaklaşık dizgi eşleme ya da bulanık arama ise bir örüntü ile tam olarak eşleşmek yerine yaklaşık olarak eşleşen örüntüleri bulma tekniğidir. Matematiksel olarak problem 2 ayrı metin $T [1...n]$ ve $P [1...m]$ ve bir uzaklık fonksiyonu $D (x [i...j], y [a...b])$, $x [i...j]$ ve $y [a...b]$ x ve y 'nin alt dizgisi olarak tanımlanabilir [20]. $D (x [i...j], y [a...b])$, $x [i...j]$ 'u $y [a...b]$ 'ye çevirmek için gereken minimum maliyeti hesaplar. x 'i y 'ye çevirmek için 3 işlem vardır.

1. Yerine Koyma: x 'deki bir karakterin alınıp eşleştirmek için y ile değiştirilmesidir.
2. Ekleme: y 'de aynı pozisyondaki bir karakterle eşleştirmek için x 'e bir karakterin eklenmesi işlemidir.
3. Silme: Eklemenin tersi olarak x 'deki bir karakterin silinmesi işlemidir.

Yaklaşık dizgi eşleme için nihai girdi kabul edilebilir en yüksek hata k ile ifade edilir.

Daha sonrası problem $D(x[i...j], y[a...b]) \leq k$ koşulunu sağlayan bir $P[i...j]$ 'nin hesaplanmasıdır.

Uzaklık fonksiyonlarında en yaygın olarak Hamming uzaklık ve Levenshtein mesafe algoritmaları kullanılır [21,22].

Yaklaşık dizgi eşleme algoritmaları 2 ayrı kategoriye ayrılabilir [23]:

- Filtreleme tabanlı algoritmalar: Bu algoritmalar 2 aşamadan oluşur. İlk aşamada olası örüntü oluşumlarının yerleri tespit edilir. İkinci aşamada ise ilk aşamada belirlenen tüm konumlar doğrulanır. Bu algoritmalar en kötü durum senaryolarına karşı yetersizdir [24].
- Geri iz sürme tabanlı algoritmalar: Bu algoritmalar genel olarak dizgi eşleme algoritmalarının bir türevidir. Bu yaklaşımda mevcut bir dizgi eşleme algoritması yaklaşık dizgi eşleme algoritmasına dönüştürülür.

3.1.2 Yaklaşık Dizgi Eşleme Uygulama Alanları

Yaklaşık dizgi eşleme, ilk olarak 1960 ve 1970 yılları arasında araştırmalara konu olmuştur. Bu yıllarda yaklaşık dizgi eşlemenin çözüme kavuşturduğu konular hala en güncel uygulama alanları olarak varlığını sürdürmektedir. Aşağıda bu uygulama alanları listelenmiştir [25].

- Hesaplamalı Biyoloji: DNA ve protein sıralaması belirli harflerden oluşan uzun metinler olarak düşünülebilir. Bu sıralama canlıların genetik kodlarını temsil eder. Bu sıralamadaki belirli DNA zincirlerini bulmak yaklaşık dize eşleşmeyi gerektirir.

- Sinyal İşleme: Konuşma tanıma için ses sinyallerinin metine iletiminde kullanılır. Resim, ses ve video verilerinde arama yapmak amacıyla da kullanılır.
- Metin Çıkarımı: Yazılı bir metindeki hatalı kelimeleri düzeltme amacıyla uygulanır.
- Diğer Alanlar: Bulanık aramanın kullanıldığı alanlar her geçen gün artmaktadır. El yazısı tanıma, görüntü sıkıştırma, veri madenciliği, optik okuma sistemi, dosya sıkıştırma gibi pek çok farklı alanda kullanılmaktadır.

3.1.3 Yaklaşık Dizgi Eşleme'nin Uygulanması

Yaklaşık Dizgi Eşlemenin çalışmada uygulanmasında Python'ın fuzzy kütüphanesi kullanılmıştır [26]. MongoDB üzerinde yeni bir koleksiyon oluşturularak aranacak kelimelerin orijinal ve metaphone değerleri fonetik uygulamalar sayesinde yazılmıştır. Fuzzy kütüphanesinin içinde yer alan, fonetik algoritmalarından Double-Metaphone ve NYSIIS (New York State Identification and Intelligence System) kullanılmıştır. Daha sonra içerisinde ilaç isimlerinden hydroxychloroquine, chloroquine, azithromycin, remdesivir geçen tweetler oluşturulan yeni koleksiyon kullanılarak aranmıştır. Bu sayede ilaç ismi tam olarak doğru bir şekilde yazılmayan tweetler de bir ilaç ismi ile eşleştirilmiş oldu. Aşağıda örnek olarak doğru yazılmayan bir ilaç isminin başarılı bir şekilde eşleştirildiği görülmektedir.

```
In [19]: data.query("_id==1280057785279889408").iloc[0].text
Out[19]: 'Get a clue LIBERAL AMERICANS, hydroxychloroquine works!!!!\nAfrica has been taking it for years. https://t.co/wB5insqbvU'
```

```
In [15]: data.query("_id==1280057785279889408").iloc[0].drugName
Out[15]: 'hydroxychloroquine'
```

Şekil 3.3 : İlaç isminin yaklaşık dizgi eşleme ile bulunması.

```
{
  "_id" : ObjectId("6255f3b4eaa55d71bfd9ef53"),
  "keyword" : BinData(0,"SFRSSw=="),
  "original" : "hydroxychloroquine",
  "item" : 1
}
{
  "_id" : ObjectId("6255f3b4eaa55d71bfd9ef54"),
  "keyword" : "HADRAXACHLARAGAN",
  "original" : "hydroxychloroquine",
  "item" : 1
}
{
  "_id" : ObjectId("6255f3b4eaa55d71bfd9ef55"),
  "keyword" : BinData(0,"S0xSSw=="),
  "original" : "chloroquine",
  "item" : 1
}
{
  "_id" : ObjectId("6255f3b4eaa55d71bfd9ef56"),
  "keyword" : "CHLARAGAN",
  "original" : "chloroquine",
  "item" : 1
}
```

Şekil 3.4 : MongoDB dökümanları içerisinde ilaç isimleri.

3.2 Veri Önileme

Makine öğrenmesinde en önemli basamaklardan birisi olan veri önileme ham veriyi kullanışlı ve etkili formata çevirmek için kullanılan bir veri madenciliği tekniğidir. Çalışmada tweet metni içerisinde yer alan RT, kullanıcı ismi, linkler, hashtag'in başındaki # işareti ve boşluklar silinmiştir.

3.3 Duygu Analizinin Uygulanması

İçerisinde ilaç isimleri geçen tweetler bir csv dosyasına aktarıldıktan sonra Python'ın pandas kütüphanesi aracılığıyla Jupyter Notebook platformu kullanılarak okundu. 3 ayrı doğal dil işleme kütüphanesi ile tweetler üzerinde duygu analizi gerçekleştirildi. Duygu analizi skorları veri seti üzerinde TextBlob, NLTK Vader ve Transformers için ayrı ayrı sütunlara yazılmıştır. Yazılan bu değerler daha sonra veri seti üzerinde aylara göre gruplanarak analiz edilmiştir. Aşağıda 3 kütüphanenin sonuçları 3 ayrı ilaç için ayrı ayrı incelenmiştir.

3.3.1 Transformers

Hugging Face şirketinin bir kütüphanesi olan Transformers, görüntü, metin ve ses gibi işlemler yapmamız için bize önceden eğitilmiş gelişmiş modeller sağlar [27]. Önceden

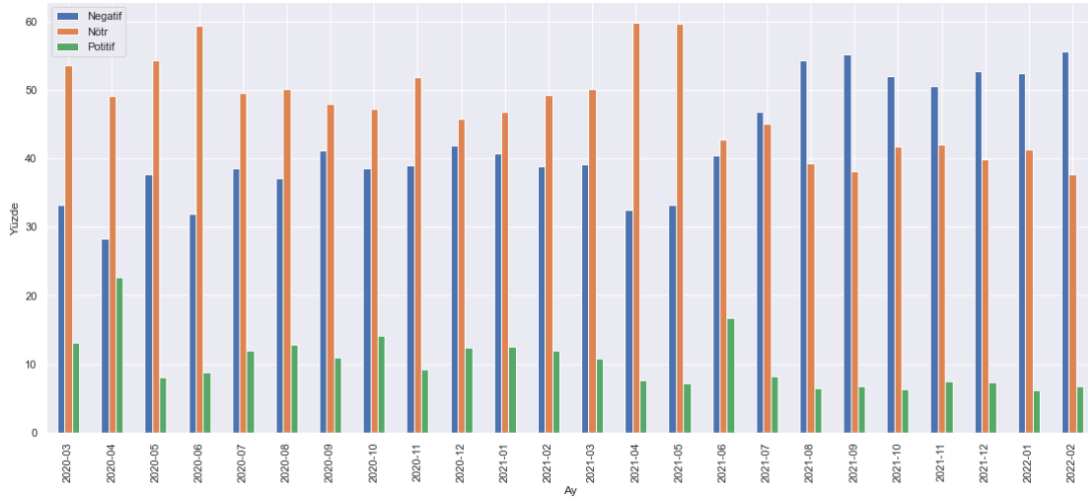
eđitilmiş modelleri kullanmak hesaplama maliyetlerini ve modeli bařtan eđitmek iin gerekli olan zamanı azaltır. Modeller farklı amalarla kullanılabilir:

- Metin: Metin sınıflandırma, bilgi ıkarımı, soru cevaplama, zetleme, eviri ve metin retimi.
- Grsel: Grnt sınıflandırma, nesne takibi ve grnt bltleme.
- Ses: Konuřma tanıma ve ses sınıflandırma.
- ok modlu: Optik karakter tanıma, taranan dkmanlardan bilgi ıkarımı, video sınıflandırma ve grsel soru cevaplama.

Transformers ktphanesi en popler derin đrenme ktphanelerinden PyTorch, TensorFlow ve JAX'a destek vermektedir.

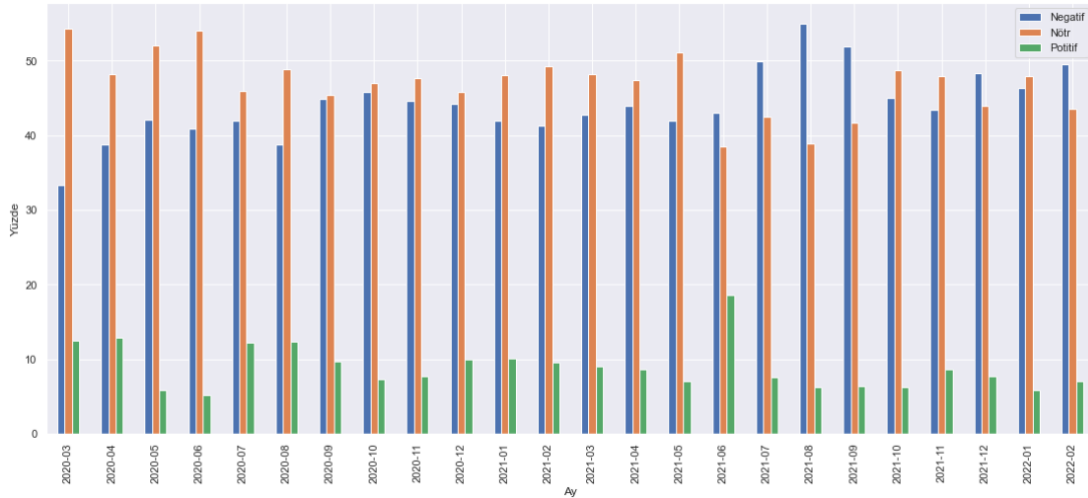
Transformers ile duygu analizi hesaplanırken ktphane ierisinde nceden eđitilmiş olan BERTweet modeli kullanılmıştır. BERTweet, 850 milyon İngilizce tweet kullanılarak geliřtirilmiş ilk geniř kapsamlı dil modelidir [28]. 2019 yılında geliřtirilen BERT ile aynı mimariyi kullanır ve RoBERTa n eđitim prosedr ile eđitilmiştir. Rakipleri olan RoBERTa ve XLM-R ve SOTA modellerine gre 3 ayrı tweet DDİ uygulamalarından metin parası etiketleme, varlık ismi tanıma ve metin sınıflandırmada daha iyi performans gstermiştir. BERTweet 850 milyon tweeti 2 ayrı kaynaktan ekmiştir. İlk kaynak 2012 Ocak ayı ile 2019 Ađustos ayı arasındaki 845 milyon İngilizce tweetten oluřurken ikinci kaynak 2020 Ocak ayından 2020 Mart ayına kadar Covid-19 ile alakalı atılan 5 milyon tweetten oluřmaktadır. BERTweet PyTorch'un bir ktphanesi olan fairseq ve transformers ktphaneleri ierisinde yer almaktadır.

Ařađıda Transformers ktphanesi kullanılarak duygu analizi skorları incelenmiştir.



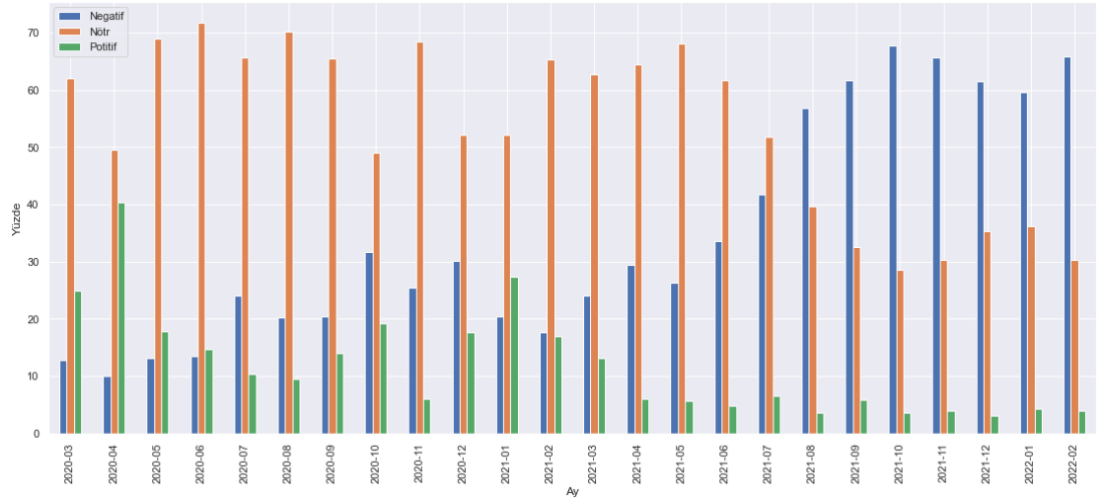
Şekil 3.5 : Tüm ilaç isimlerinin aylara göre duygu analizi oranları.

Yukarıdaki şekilde içerisinde tüm ilaç isimleri geçen tweetler dahil edilmiş şekilde oranların dağılımı görülmektedir. İlk aylarda ilaçlar hakkında nötr bir dağılım olduğu görülmektedir. 2021 Temmuz ayından sonra dağılımın negatife kaydığı ve bu şekilde devam ettiği görülmektedir.



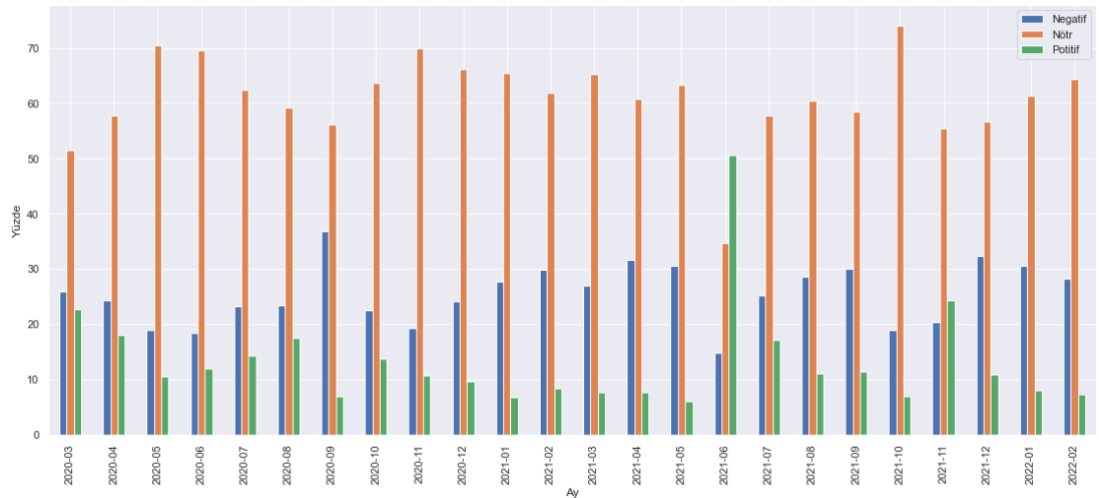
Şekil 3.6 : Hydroxychloroquine için aylara göre duygu analizi oranları.

Hydroxychloroquine’de de benzer şekilde ilk aylarda nötr bir dağılım görünürken, 2021 Temmuz ayından sonra negatif değerlerin ön plana çıktığı görülmektedir.



Şekil 3.7 : Remdesivir için aylara göre duygu analizi oranları.

Remdesivir’de ilk aylarda pozitif oranların negatifine de geçtiği görülmektedir. Benzer şekilde 2021 Ağustos ayından sonra negatif dağılımın arttığı ve pozitif değerlerin çok düşük kaldığı görülmektedir.



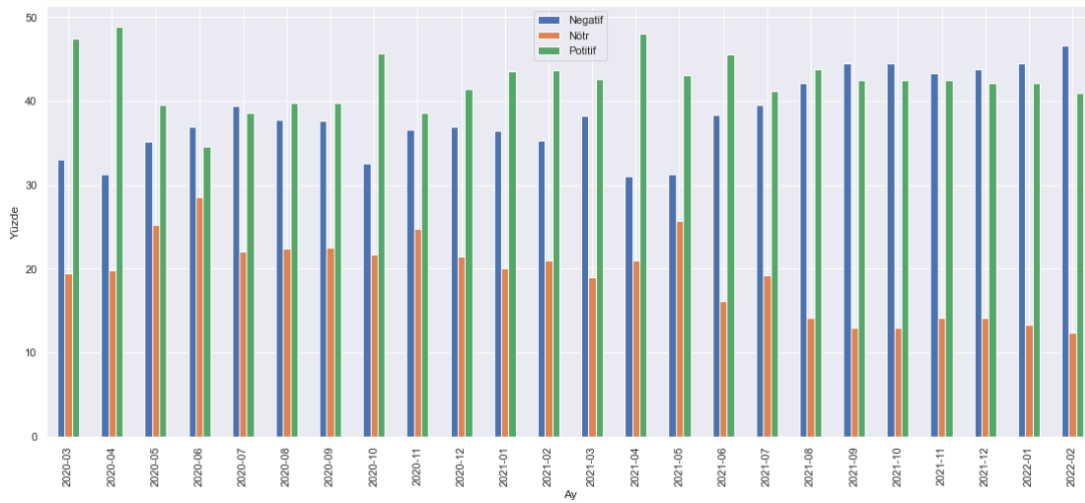
Şekil 3.8 : Azithromycin için aylara göre duygu analizi oranları.

Azithromycin’de genel olarak nötr değerlerin ağırlıkta olduğu görülmektedir. Ancak 2021 Haziran ayında pozitif dağılımın negatif ve nötr değerleri de geçip yükseldiği görülmektedir. 2021 Haziran ayında Azithromycin’in bu kadar yüksek çıkmasının sebebi incelediğinde 10 Haziran tarihinde yapılan bir haberde yüksek dozda Hydroxychloroquine ve Azithromycin alımının Covid-19 hastalarının hayatta kalma oranını %200 artırdığı geçilmekte ve bu haberin pozitif sonuçlar üzerinde etkili olduğu anlaşılmaktadır [29].

3.3.2 Vader

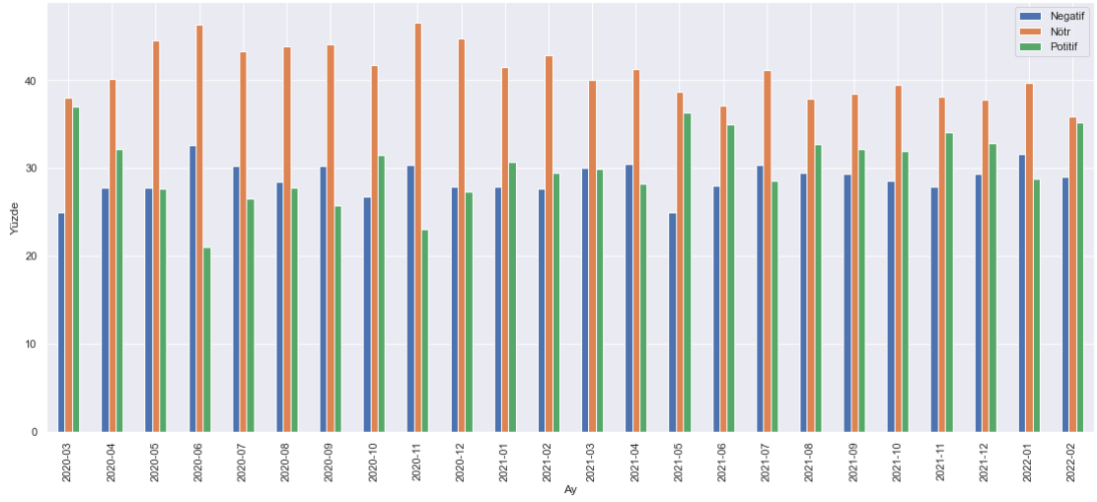
Vader (Valence Aware Dictionary for Sentiment Reasoning) bir sözlük ve kural tabanlı duygu analizi aracıdır [30]. Vader sözlüğü kendisini geleneksel duygu sözlüklerinden birisi olan LIWC (Linguistic Inquiry and Word Count)'e göre daha çok sosyal medya içeriklerine hassas olmasıyla ayırır. Vader metnin polaritesini hesaplayıp geriye olumlu, olumsuz, negatif, nötr skorlarını ve bunların normalize edilmiş hali olan bileşik bir skor döner.

Aşağıda vader kütüphanesi kullanılarak ortaya çıkan sonuçlar görülmektedir.



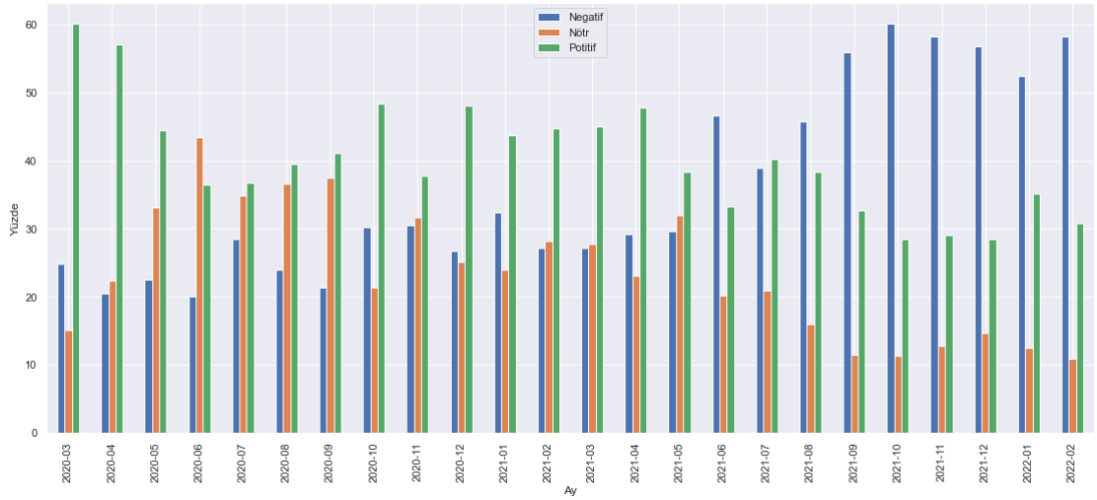
Şekil 3.9 : Tüm ilaç isimlerinin aylara göre duygu analizi oranları.

Vader'da Transformers'a benzer bir şekilde negatif dağılımın 2021 sonlarına doğru arttığı görülmektedir. Transformers'dan farklı şekilde ilk aylarda pozitif oranların nötr değerleri geçtiği görülmektedir. Bunun sebebi, vader kütüphanesinin Transformers'a göre bileşik bir skor dönmesi ve duygu sınıflandırmasının bu skor üzerinden yapılmasıdır. Vader bileşik skorunda -0.05'ten küçük değerler negatif, -0.05 ile 0.05 arasında olan değerler nötr ve 0.05'ten büyük değerler ise pozitif olarak işaretlenmiştir.



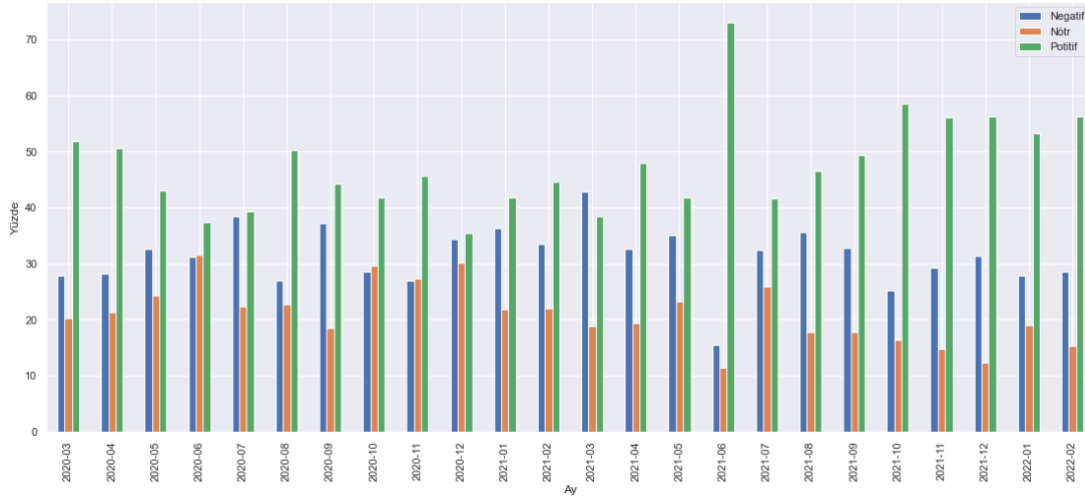
Şekil 3.10 : Hydroxychloroquine için aylara göre duygu analizi oranları.

Hydroxychloroquine'e bakıldığında nötr dağılımın önde geldiği görülmektedir. Tüm ilaç isimlerine nazaran hydroxychloroquine'de pozitif değerlerin yüksek olduğu görülmektedir.



Şekil 3.11 : Remdesivir için aylara göre duygu analizi oranları.

Remdesivir'de 2021 Haziran ayından sonra negatif oranların arttığı ve nötr değerlerin düştüğü görülmektedir.



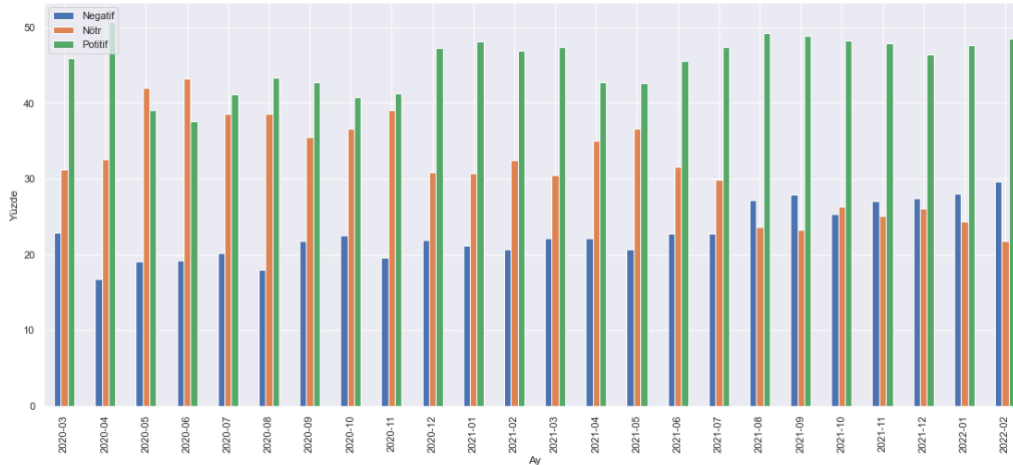
Şekil 3.12 : Azithromycin için aylara göre duygu analizi oranları.

Azithromycin’de genel olarak pozitif dağılımın önde olduğu görülmektedir. Haziran ayında %70’in üzerinde pozitif oran olmasının sebebi daha önce Transformers kısmında bahsedilen haber sebebiyledir.

3.3.3 TextBlob

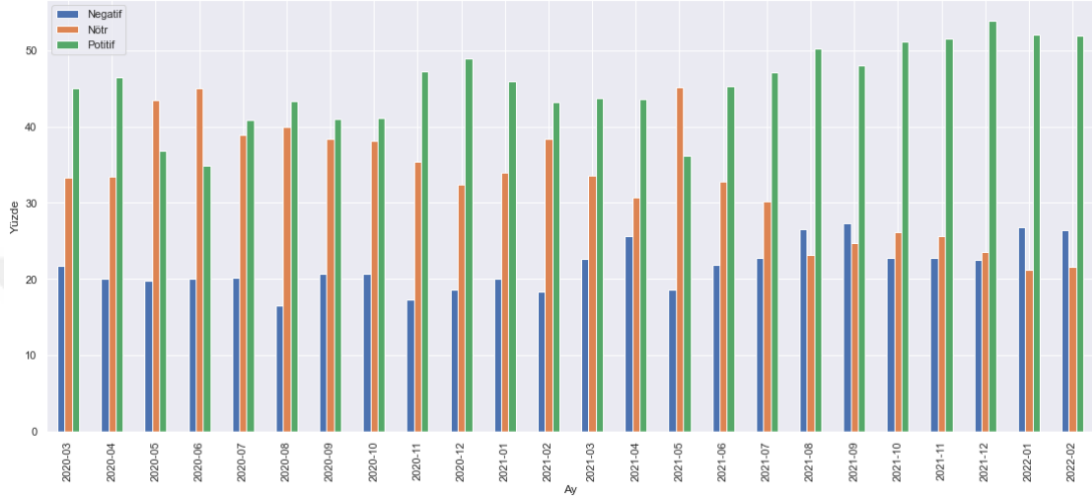
TextBlob metin formatındaki verileri analiz etmek için kullanılan bir Python kütüphanesidir [31]. Sınıflandırma, duygu analizi ve konuşma tanıma gibi DDİ işlemleri için bir uygulama programlama arayüzü sağlar. NLTK (Natural Language Toolkit) çatısı üzerinde çalışır. Uygulama programlama arayüzü geriye polarite ve öznellik seviyelerini döner. Polarite negatif bir duygu belirten -1 ile pozitif bir duyguyu belirten +1 arasında bir değer alır. Öznellik ise bir bilginin gerçeğe mi dayalı yoksa kişisel bir fikir mi olduğunu belirten 0 ile 1 arasında bir değer alır.

Aşağıda TextBlob kullanılarak aylara göre duygu analizi dağılımları görülmektedir.



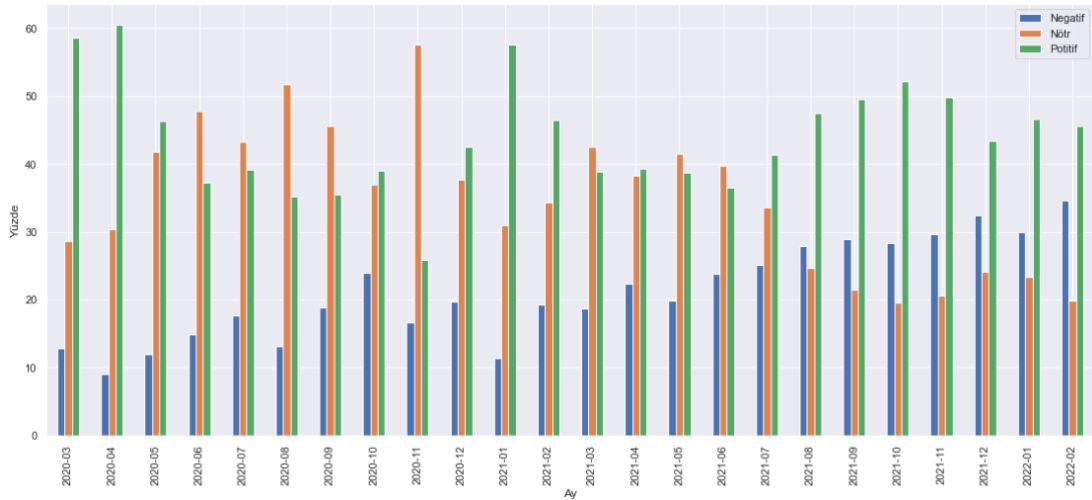
Şekil 3.13 : Tüm ilaç isimlerinin aylara göre duygu analizi oranları.

TextBlob’da Transformers ve Vader’a göre pozitif dağılımların yüksek çıktığı görülmektedir. Bunun sebebi TextBlob’un Vader gibi geriye bir polarite dönmesidir. Ancak Vader’dan farklı olarak burada 0’ın altındaki değerler negatif, 0 nötr, 0’dan büyükler ise pozitif olarak sınıflandırılmıştır. İlk aylarda nötr değerlerin yüksek ve negatif değerlerin düşük olduğu görülürken 2021 Temmuz ayından sonra negatif değerlerin nötrleri geçtiği görülmektedir.



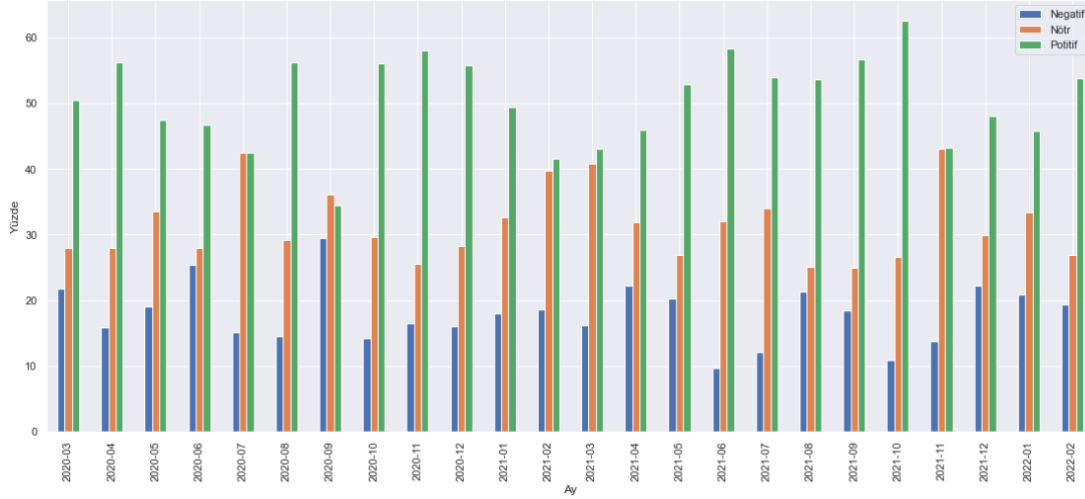
Şekil 3.14 : Hydroxychloroquine için aylara göre duygu analizi oranları.

Hydroxychloroquine’de de benzer şekilde ilk aylarda nötr bir dağılımın baskın olduğu görülürken 2021 Ağustos ayından sonra negatif dağılımın arttığı görülmektedir. TextBlob’da nötr değerlerin yüksek çıkması incelediğinde Vader ve Transformers’da pozitif olarak bulunan bazı tweetlerin TextBlob’da 0 olarak bulunduğu görülmüştür.



Şekil 3.15 : Remdesivir için aylara göre duygu analizi oranları.

Remdesivir’de de ilk aylarda nötr değerlerin yoğunlukta olduğu ve negatif dağılımın düşük olduğu görülürken 2021 Haziran ayından sonra nötr değerlerde azalış, negatif ve pozitiflerde artış olduğu görülmektedir.



Şekil 3.16 : Azithromycin için aylara göre duygu analizi oranları.

TextBlob’da Azithromycin için genel olarak pozitif dağılımın önde olduğu görülmektedir.

3.4 Bilim İnsanlarına Yönelik Duygu Analizi

Bilim insanlarının COVID-19’a karşı etkili olan ilaçlar hakkındaki düşüncelerini değerlendirmek için kullanıcı biyografisine bakılarak yeni bir veri seti oluşturuldu. Kullanıcı biyografisi içerisinde phd, prof, md, scientist, university, lab, doctor, db ,postdoc, science, biomedical, surgeon, biologist, biochemist, microbiologist, researcher kelimeleri geçen tweetler yeni bir veri setine aktarıldı. Daha sonra bu veri seti üzerinde duygu analizi oranları incelendi. Aşağıda Transformers kütüphanesi kullanılarak 3 ilaç için duygu analizi oranları bilim insanları ve bilim insanları haricindeki veri seti için tablo halinde gösterilmiştir.

Çizelge 5.1 : Bilim İnsanlarına Yönelik Duygu Analizi Oranları.

İlaç İsmi	Pozitif Oranı (%)	Negatif Oranı (%)	Nötr Oranı (%)
Tüm İlaçlar	11.22	33.15	55.61
Hydroxychloroquine	6.58	38.12	55.29
Remdesivir	15.84	23.70	60.45
Azithromycin	14.02	22.50	63.46

Çizelge 5.2 : Bilim İnsanları Haricinde Duygu Analizi Oranları.

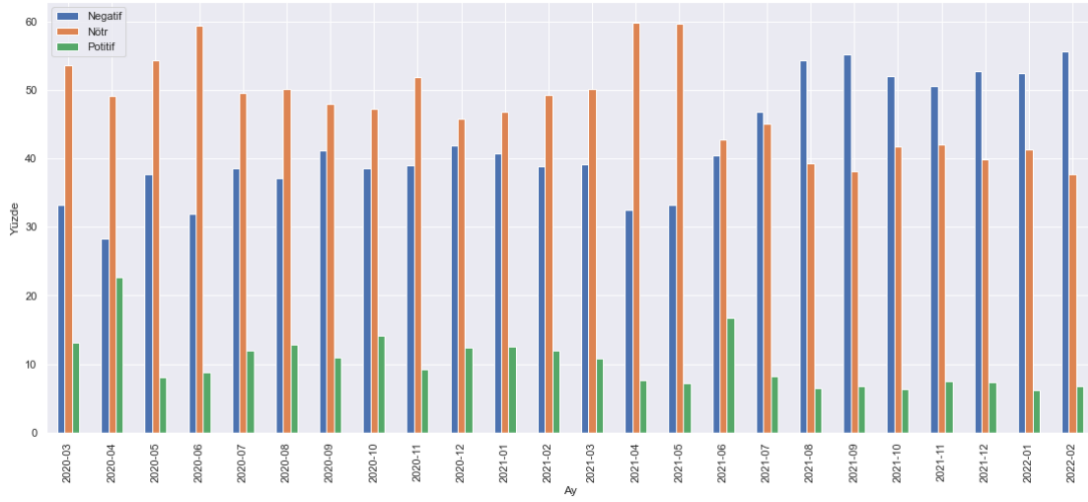
İlaç İsmi	Pozitif Oranı (%)	Negatif Oranı (%)	Nötr Oranı (%)
Tüm İlaçlar	10.28	40.59	49.12
Hydroxychloroquine	8.78	41.97	49.23
Remdesivir	11.99	32.70	55.29
Azithromycin	13.62	24.67	61.69

Çizelgelerdeki oranlara bakıldığında bilim insanlarındaki negatif dağılımın düşük olduğu ve nötr dağılımın ağırlıkta olduğu görülmektedir. Tüm ilaçlar baz alındığında bilim insanlarında yaklaşık %7 daha az negatif dağılımın olduğu görülmektedir. Buradaki %1 fark pozitif oranlara, %6 fark ise nötr dağılıma eklenmiştir.

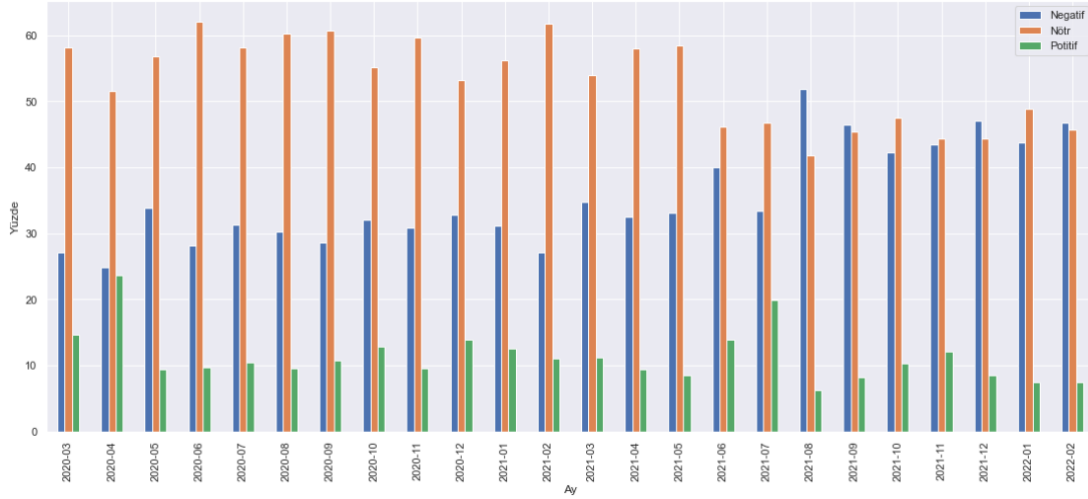


6. SONUÇ VE ÖNERİLER

Bu çalışma kapsamında 2019 yılının sonunda ortaya çıkan ve 27 Mart 2022 itibarıyla 479 milyon vaka ve 6 milyon ölüme neden olan COVID-19 için geliştirilen ve Covid-19'a etkili olup olmadığı halen tartışma konusu olan ilaç çalışmaları hakkında sosyal medya üzerinde duygu analizi uygulanarak insanların bu ilaçlara karşı düşünceleri tespit edilmiştir. En yaygın kullanılan sosyal medya platformlarından birisi olan Twitter üzerinde COVID-19 ile alakalı paylaşılan 2020 Mart ayı ile 2022 Şubat ayı arasındaki 1.4 milyar tweet Twitter Uygulama Programlama Arayüzü kullanılarak çekilmiştir. Çekilen bu tweetler içerisinde COVID-19'a etkili olduğu düşünülen 3 ilaç ismi geçenler bir veri setine aktarılmıştır. İlaç isimleri aranırken, metin aramada birebir eşleştirme olmadan, birbirine yakın kelimelerin eşleşmesine imkan veren yaklaşık dizgi eşleme metodu uygulanmıştır. Bu sayede, içerisinde yazılırken hata yapılma olasılığı olan ilaç isimleri geçen tweetlerin de veri setine eklenmesi sağlanmıştır. Veri seti çekildikten sonra tweetler içerisindeki özel karakterlerin ve linklerin temizlenmesi için veri ön işleme yapılmıştır. Daha sonra veri seti üzerinde Python'ın doğal dil işleme kütüphanelerinden Transformers, Vader ve TextBlob kullanılarak duygu analizi uygulanmıştır. Duygu analizi uygulandıktan sonra COVID-19'a etkili olduğu düşünülen Hydroxychloroquine, Remdesivir ve Azithromycin ilaçları için aylık olarak duygu analizi oranları veri görselleştirme teknikleri kullanılarak gösterilmiştir. Ayrıca bilim insanlarının bu ilaçlara olan düşüncelerini tespit edebilmek için kullanıcı biyografisine bakılarak bilim insanlarına yönelik duygu analizi oranları da gösterilmiştir. Sonuç olarak, 3 ayrı doğal dil işleme kütüphanesinin verdiği sonuçlar da göz önünde bulundurularak 2020 Mart ayından 2021 Haziran ayına kadar daha nötr bir tutum varken 2021 Haziran ayından sonra daha olumsuz düşünceler olduğu anlaşılmaktadır. Bilim insanlarının ilaçlara olan düşüncelerinde ise daha nötr bir dağılım olduğu görülmektedir. Aşağıda Transformers kütüphanesi kullanılarak aylık olarak yüzde cinsinde duygu analizi oranları bilim insanları ve bilim insanları haricindekiler için gösterilmiştir.



Şekil 6.1 : Tüm ilaçlar için duygu analizi oranları.



Şekil 6.2 : Tüm ilaçlar için bilim insanlarına yönelik duygu analizi oranları.

Yukarıdaki şekillerde de görüldüğü üzere ilk aylarda daha nötr bir dağılım varken 2021 Haziran ayından sonra olumsuz düşünceler artmıştır. Bilim insanlarının ilaçlara olan düşüncelerinin ise daha nötr olduğu görülmektedir.

Bu çalışma, COVID-19'a karşı geliştirilmeye devam eden ve insanların COVID-19 tedavisinde kullandıkları ilaçların etkisini göstermekte ve bu alanda yapılacak olan araştırmalara referans olacağı düşünülmektedir.

KAYNAKLAR

- [1] **Url-1** <<https://covid19.saglik.gov.tr/TR-66300/covid-19-nedir-.html>>, erişim tarihi 27.03.2022.
- [2] **Url-2** <<https://www.cdc.gov/coronavirus/2019-ncov/locations-confirmed-cases.html#map>>, erişim tarihi 27.03.2022.
- [3] **World Health Organization.** (2020). *Coronavirus disease 2019 (COVID-19) Situation Report – 51* [PowerPoint slides]. Erişim <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf>
- [4] **Weekly epidemiological update on COVID-19 - 29 March 2022.** (2022). Erişim: 10 Nisan 2022, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---29-march-2022>.
- [5] **Url-3** <<https://ourworldindata.org/covid-vaccinations>>, erişim tarihi 27.03.2022.
- [6] **Tarighi, P., Eftekhari, S., Chizari, M., Sabernavaei, M., Jafari, D., & Mirzabeigi, P.** (2021). A review of potential suggested drugs for coronavirus disease (COVID-19) treatment. *European Journal of Pharmacology*, 895, 173890.
- [7] **Ba, P., Nandhini, J. M., & Tc, G.** (2021). An Analysis of the Applications of Natural Language Processing in Various Sectors.
- [8] **Liu, B.** (2012). (2020). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- [9] **Tsytsarau, M., & Palpanas, T.** (2012). Survey on mining subjective data on the web. *Data Mining and Knowledge Discovery*, 24(3), 478-514.
- [10] **Medhat, W., Hassan, A., & Korashy, H.** (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.
- [11] **Katrekar, A., & AVP, B. D. A.** (2005). An introduction to sentiment analysis. *GlobalLogic Inc.*
- [12] **Aggarwal, C. C.** (2015). Mining text data. In *Data mining* (pp. 429-455). Springer, Cham.
- [13] **Mejova, Y., & Srinivasan, P.** (2011). Exploring feature definition and selection for sentiment classifiers. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 5, No. 1, pp. 546-549).
- [14] **Hamamsy, T., & Bonneau, R.** (2020). Twitter activity about treatments during the COVID-19 pandemic: case studies of remdesivir, hydroxychloroquine, and convalescent plasma. *medRxiv*.
- [15] **Tekumalla, R., & Banda, J. M.** (2020). Characterizing drug mentions in COVID-19 Twitter Chatter, *arXiv preprint arXiv:2007.10276*.

- [16] **Sharma, K., Seo, S., Meng, C., Rambhatla, S., & Liu, Y.** (2020). Covid-19 on social media: Analyzing misinformation in twitter conversations. *arXiv preprint arXiv:2003.12309*.
- [17] **Boon-Itt, S., & Skunkan, Y.** (2020). Public perception of the COVID-19 pandemic on Twitter: sentiment analysis and topic modeling study. *JMIR Public Health and Surveillance*, 6(4), e21978.
- [18] **Url-4** <<https://ieee-dataport.org/open-access/coronavirus-covid-19-tweets-dataset>>, erişim tarihi 18.12.2020.
- [19] **Rasool, A., Tiwari, A., Singla, G., & Khare, N.** (2012). String matching methodologies: A comparative analysis. *REM (Text)*, 234567(11), 3.
- [20] **Kumar, N., Bibhu, V., Islam, M., & Bhardwaj, S.** (2010). Approximate string matching algorithm. *International Journal on Computer Science and Engineering*, 2(3), 641-644.
- [21] **Sankoff, D., & Mainville, S.** (1983). Common subsequences and monotone subsequences. *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison*, 363-365.
- [22] **Levenshtein, V.** (1965). Binary codes capable of correcting spurious insertions and deletion of ones. *Problems of information Transmission*, 1(1), 8-17.
- [23] **Hakak, S. I., Kamsin, A., Shivakumara, P., Gilkar, G. A., Khan, W. Z., & Imran, M.** (2019). Exact string matching algorithms: Survey, issues, and future research directions. *IEEE access*, 7, 69614-69637.
- [24] **Kucherov, G., Salikhov, K., & Tsur, D.** (2016). Approximate string matching using a bidirectional index. *Theoretical Computer Science*, 638, 145-158.
- [25] **Navarro, G.** (2001). A guided tour to approximate string matching. *ACM computing surveys (CSUR)*, 33(1), 31-88.
- [26] **Url-5** <<https://github.com/yougov/fuzzy>>, erişim tarihi 01.04.2022.
- [27] **Url-6** < <https://huggingface.co/docs/transformers/index>>, erişim tarihi 01.04.2022.
- [28] **Nguyen, D. Q., Vu, T., & Nguyen, A. T.** (2020). BERTweet: A pre-trained language model for English Tweets. *arXiv preprint arXiv:2005.10200*.
- [29] **OAN Newsroom.** (2021, 10 Haziran). Researchers Find Higher Doses Of Hydroxychloroquine, Azithromycin Increase COVID Patient Survival By 200 Percent. *One America News Network*. Erişim adresi <https://www.oann.com/researchers-find-higher-doses-of-hydroxychloroquine-azithromycin-increase-covid-patient-survival-by-200-percent/>
- [30] **Hutto, C., & Gilbert, E.** (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).
- [31] **Url-7** <<https://textblob.readthedocs.io/en/dev/>>, erişim tarihi 01.04.2022.

- [32] **Lamsal, R.** (2021). Design and analysis of a large-scale COVID-19 tweets dataset. *Applied Intelligence*, 51(5), 2790-2804.
- [33] **Pérez, J. M., Giudici, J. C., & Luque, F.** (2021). pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks. *arXiv preprint arXiv:2106.09462*.





EKLER

EK A: Veri setinin çekilmesi için Python işlemleri

```
import pandas as pd
import numpy as np
import os
import pymongo
from pymongo import MongoClient
from pymongo import UpdateOne
import tweepy
import logging
import time
import sys
from os import listdir
from os.path import isfile, join
import re
import configparser

databaseConfig = configparser.RawConfigParser()
databaseConfig.read('/okyanus/users/okum/corona_tweets/corona_talha/database.properties')

PATH=databaseConfig.get('PathSection', 'PATH')+"dr_corona_tweets_1"
ERROR_PATH=databaseConfig.get('PathSection', 'ERROR_PATH')
DONE_PATH=databaseConfig.get('PathSection', 'DONE_PATH')
NOT_INSERTED_PATH=databaseConfig.get('PathSection', 'NOT_INSERTED_PATH')

fileList = [f for f in listdir(PATH) if isfile(join(PATH, f))]

def getLocalClient():
    return MongoClient(databaseConfig.get('DatabaseSection', 'database.host'), 27017)

client=getLocalClient()

db = client.twitter
tweetsColl=db.tweets

auth = tweepy.OAuthHandler(databaseConfig.get('ApiSection_1', 'consumer.key'),databaseConfig.get('ApiSection_1',
'consumer.secret'))
auth.set_access_token(databaseConfig.get('ApiSection_1', 'access.token'),databaseConfig.get('ApiSection_1',
'access.secret'))

api = tweepy.API(auth,wait_on_rate_limit=True, wait_on_rate_limit_notify=True,retry_count=10, retry_delay=5,
retry_errors=set([503]))

root_logger= logging.getLogger()
root_logger.setLevel(logging.INFO)
handler = logging.FileHandler("corona_tweets_1.log", 'w', 'utf-8')
handler.setFormatter(logging.Formatter('%(asctime)s %(message)s'))
root_logger.addHandler(handler)
```

Şekil A.1 : Veri setinin çekilmesi için Python işlemleri 1. kısım

```

for file in fileList:
    if file.endswith('.csv'):
        splittedStrs=file.split(".")
        fileName=splittedStrs[len(splittedStrs)-2]
        print("{}".format(fileName))
        DATAPATH=PATH+'/'+file
        data=pd.read_csv(DATAPATH,sep=",",header=None,dtype={0: 'str'})
        batch_size = 100
        try:
            for i in range(0, len(data), batch_size):
                startMillis = float(round(time.time() * 1000))
                batched_data=data[i:i+batch_size]
                tweet_ids=batched_data[0].tolist()
                tweetIdSentimentScoreMap={batched_data[0][index]:batched_data[1][index] for index,row in
batched_data.iterrows()}
                tweet_statuses=[]
                try:
                    tweet_statuses = api.statuses_lookup(tweet_ids,tweet_mode="extended")
                except Exception as e:
                    logging.info("exception occurred while fetching tweets! batch skipped::fileName::{}".format(fileName))
                    appendToNotInserted(batched_data)
                    continue
                tweetDocumentList=[]
                if len(tweet_statuses)<1:
                    logging.info("tweet_statuses is empty for tweets!::fileName::{}::tweet ids::
{}".format(fileName,tweet_ids))
                for tweet in tweet_statuses:
                    if tweet is not None:
                        tweetId = tweet.id
                        sentimentScore = tweetIdSentimentScoreMap[str(tweetId)]
                        possibly_sensitive=False
                        if hasattr(tweet, 'possibly_sensitive'):
                            possibly_sensitive=tweet.possibly_sensitive
                        isRetweeted=retweetStatusExists(tweet)
                        original_tweet_id=None
                        original_tweet_user_id=None
                        if isRetweeted:
                            original_tweet_id=tweet.retweeted_status.id
                            original_tweet_user_id=tweet.retweeted_status.user.id

```

Şekil A.2 : Veri setinin çekilmesi için Python işlemleri 2. kısım

```

tweet_doc={
    "_id":tweetId,
    "sentiment_score":sentimentScore,
    "created_at":tweet.created_at,
    "text":tweet.full_text,
    "user_id":tweet.id,
    "user_name":tweet.user.name,
    "user_screenname":tweet.user.screen_name,
    "user_location":tweet.user.location,
    "user_description":tweet.user.description,
    "followers_count":tweet.user.followers_count,
    "friends_count":tweet.user.friends_count,
    "listed_count":tweet.user.listed_count,
    "user_createdat":tweet.user.created_at,
    "user_favouritescount":tweet.user.favourites_count,
    "user_verified":tweet.user.verified,
    "user_statusescount":tweet.user.statuses_count,
    "retweet_count":tweet.retweet_count,
    "favorite_count":tweet.favorite_count,
    "possibly_sensitive" : possibly_sensitive,
    "retweeted_status": isRetweeted,
    "original_tweet_id":original_tweet_id,
    "original_tweet_user_id":original_tweet_user_id
}
tweetDocumentList.append(tweet_doc)
logging.info("tweet inserted to list successfully::fileName::{}::tweetId::{}::sentimentScore::
{}".format(fileName,tweetId,sentimentScore))
else:
    logging.info("tweet is null::fileName::{}".format(fileName))
    if len(tweetDocumentList) > 0:
        currentTimeMillis=float(round(time.time() * 1000))
        upserts=[ UpdateOne({'_id':x['_id']}), {'$set':x}, upsert=True) for x in tweetDocumentList]
        result = tweetsColl.bulk_write(upserts)
        logging.info("tweets saved successfully::fileName::{}::length::{}::time::
{}".format(fileName,len(upserts),currentTimeMillis-startMillis))
        logging.info("All tweets processed. File is moving under done::fileName::{}".format(fileName))
        os.rename(DATAPATH, DONE_PATH+file)
    except Exception as e:
        logging.info("exception occurred! file is moving under error::fileName::{}::error::{}".format(fileName,e))
        os.rename(DATAPATH, ERROR_PATH+file)

```

Şekil A.3 : Veri setinin çekilmesi için Python işlemleri 3. kısım

EK B: İlaç ismi geçen tweetlerin yaklaşık dizgi eşleme metodu ile bulunması için Python işlemleri

```
import pandas as pd
import numpy as np
import os
import pymongo
from pymongo import MongoClient
import tweepy
import logging
import time
import fuzzy
import nltk
from nltk.corpus import stopwords
import csv
import re
import configparser

databaseConfig = configparser.RawConfigParser()
databaseConfig.read('/okyanus/users/okum/corona_tweets/corona_talha/database.properties')

client = MongoClient(databaseConfig.get('DatabaseSection', 'database.host'), 27017)

root_logger= logging.getLogger()
root_logger.setLevel(logging.INFO)
handler = logging.FileHandler("drug_search.log", 'w', 'utf-8')
handler.setFormatter(logging.Formatter('%(asctime)s %(message)s'))
root_logger.addHandler(handler)

logging.info("drug_search started")

db = client.twitter
tweetsColl=db.tweets
searchTermsColl=db.searchTerms

searchTermsColl.drop()

logging.info("searchTerms collection dropped")

terms=["hydroxychloroquine","chloroquine","azithromycin","remdesivir"]
dmeta = fuzzy.DMetaphone()

startMillis = float(round(time.time() * 1000))
```

Şekil B.1 : İlaç ismi geçen tweetleri aramak için Python işlemleri 1. kısım

```

def insertSearchTerms():
    for word in terms:
        if len(word) <= 2:
            continue
        fuzzy_terms = []
        fuzzy_terms.append(dmeta(word)[0])
        fuzzy_terms.append(fuzzy.nysiis(word))
        for term in fuzzy_terms:
            searchTermsColl.insert({
                "keyword": term,
                "original": word,
                "item": 1
            })
    insertSearchTerms()
    nltk.download('stopwords')
    stop_words = set(stopwords.words('english'))
def findFuzzyTerms(tweetText):
    search_words = tweetText.split(" ")
    fuzzy_terms = []
    results=[]
    for word in search_words:
        if word in stop_words:
            continue
        fuzzy_terms.append(dmeta(word.encode('utf-8'))[0])
        fuzzy_terms.append(fuzzy.nysiis(word))
        try:
            results = searchTermsColl.find(
                {"$or": [
                    {"keyword": {"$in": fuzzy_terms}},
                    {"original": {
                        "$regex": '^' + re.escape(tweetText), "$options": "i"}}
                ]})
        except:
            logging.info("exception occurred in find query!::tweetText::
{:fuzzyterms".format(re.escape(tweetText),fuzzy_terms))
        return results
def yield_rows(cursor, chunk_size):
    chunk = []
    for i, row in enumerate(cursor):
        if i % chunk_size == 0 and i > 0:
            yield chunk
            del chunk[:]
        chunk.append(row)
    yield chunk

```

Şekil B.2 : İlaç ismi geçen tweetleri aramak için Python işlemleri 2. kısım


```

data['clean_text'] = data['text'].apply(text_preprocessing)

sentiment_analysis = pipeline('sentiment-analysis',model="finiteautomata/bertweet-base-sentiment-analysis")

analyzer = SentimentIntensityAnalyzer()

data["original_tweet_id"].fillna(0, inplace = True)
data["original_tweet_user_id"].fillna(0, inplace = True)

data['original_tweet_id'] = data['original_tweet_id'].astype(int)
data['original_tweet_user_id'] = data['original_tweet_user_id'].astype(int)

logging.info("hugging face started")
startMillis = float(round(time.time() * 1000))

data_temp_hugging_face=data.copy()

for index,val in data_temp_hugging_face.iterrows():
    try:
        cleanTweetText=val["clean_text"]
        pol_score=analyzer.polarity_scores(cleanTweetText)
        data_temp_hugging_face.loc[index, 'vader_score'] = pol_score['compound']

        sentiment = sentiment_analysis(cleanTweetText)
        data_temp_hugging_face.loc[index, 'hugging_face_label'] = sentiment[0]['label']
        data_temp_hugging_face.loc[index, 'hugging_face_score'] = sentiment[0]['score']
        print(cleanTweetText)
        logging.info("hugging face sentiment analysis calculated::index::{}".format(index))
    except:
        logging.exception("Error occurred while sentiment analysis::index::{}".format(index))

currentTimeMillis=float(round(time.time() * 1000))
logging.info("hugging face finished::time::{}".format(currentTimeMillis-startMillis))
currentTimeMillis=float(round(time.time() * 1000))
logging.info("hugging face csv export started")
data_temp_hugging_face.to_csv('hugging_face_dataframe.csv',index=False)
currentTimeMillis=float(round(time.time() * 1000))
logging.info("hugging face csv export finished::time::{}".format(currentTimeMillis-startMillis))

```

Şekil C.2 : Duygu analizi için Python işlemleri 2. kısım

EK D: Duygu analizi sonuçlarının yorumlanması için Python işlemleri

```

data_user_desc=pd.DataFrame(data_temp, columns=col_names + ['dateTimeStr'])
user_description_keywords=["phd","prof","md","scientist", "university", "lab", "doctor", "dr",
"postdoc", "science", "biomedical", "surgeon", "biologist", "biochemist", "microbiologist", "researcher"]

def user_filter(row):
    if any(word.upper() in set(str(row['user_description']).upper().split()) for word in user_description_keywords):
        return True
    else:
        return False

data_user_df=data_user_desc[['user_description']].apply(user_filter, axis=1)
data_user_desc=data_user_desc[data_user_df]

user_desc_non_retweeted=data_user_desc[data_user_desc['_id'].isin(non_retweeted_data['_id'])]

data_temp_discarded=data_temp[~data_temp['_id'].isin(data_user_desc['_id'])]

data_temp_ordered=data_temp_discarded.reset_index(inplace=False).sort_values('dateTime')

non_retweeted_ordered=data_temp_ordered[data_temp_ordered['_id'].isin(non_retweeted_data['_id'])]

data_user_desc_ordered=data_user_desc.reset_index(inplace=False).sort_values('dateTimeStr')
user_desc_non_retweeted_ordered=user_desc_non_retweeted.reset_index(drop=True,
inplace=False).sort_values('dateTimeStr')

```

Şekil D.1 : Bilim adamlarının attığı tweetleri bulmak için Python işlemleri

```

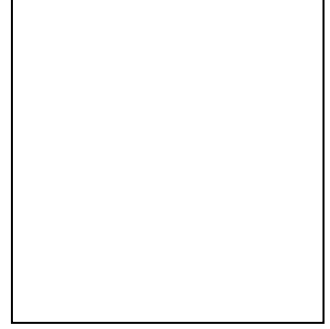
non_retweeted_ordered_group = non_retweeted_ordered.groupby(['dateTimeStr',
'hugging_face_label']).agg({'hugging_face_score': 'count'})
non_retweeted_ordered_pcts = non_retweeted_ordered_group.groupby(level=0).apply(lambda x:100 * x / float(x.sum()))

non_retweeted_ordered_pcts_plot=non_retweeted_ordered_pcts.unstack().plot(kind='bar')
non_retweeted_ordered_pcts_plot.set(xlabel='Ay', ylabel='Yüzde')
non_retweeted_ordered_pcts_plot.legend(labels=["Negatif", "Nötr", "Potitif"])

```

Şekil D.2 : Bütün ilaç isimleri için Transformers ile duygu analizi sonuçlarının yazdırılması

ÖZGEÇMİŞ



Ad-Soyad : Orhan Talha Kum

ÖĞRENİM DURUMU:

- **Lisans** : 2018, İstanbul Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği

MESLEKİ DENEYİM VE ÖDÜLLER:

- Yazılım Mühendisi, 2018 – hâlen, Netmera
- İstanbul Üniversitesi Bilgisayar Mühendisliği Bölüm Birincisi