

TWITTER VERİLERİ ÜZERİNDE DUYGU ANALİZİ

MUSTAFA KEREM KILIÇLIOĞLU

DÖNEM PROJESİ

DANIŞMAN

Dr. Öğretim Üyesi Serpil Yılmaz

İZMİR, 2024

TEŐEKKÜR

Bu dnem projesinin hazırlanmasında gsterdiđi her trl destek ve yardımdan dolayı ok deđerli hocam Dr. Serpil Yılmaz'a en iten dileklerle teŐekkr ederim.

02 Ocak 2024

Mustafa Kerem KILILIOđLU

İÇİNDEKİLER

Sayfa No

ŞEKİL LİSTESİ.....	i
ÖZET.....	v
ABSTRACT.....	vi
1. GİRİŞ.....	1
2. TWITTER DEVELOPER HESABI . Hata! Yer işareti tanımlanmamış.	
3. VERİ TOPLAMA	4
3.1. KÜTÜPHANELER.....	4
3.2. TWEET TOPLAMA İŞLEMLERİ.....	4
3.2.1. Twitter API.....	5
3.2.2. Son 3250 Tweet.....	5
3.2.3. Filtreleme	6
4. ÇEVİRİ İŞLEMLERİ.....	8
4.1. KOD İLE ÇEVİRME	8
5. DOSYA YÜKLEME VE TWEETLERİ DÜZENLEME.....	9
5.1. DOSYA YÜKLEME	9
5.2. DÜZENLEME	10
6. SUBJECTIVITY & POLARITY.....	11
7. WORD CLOUD	13
8. NEGATİF, POZİTİF VE NÖTR TWEETLER.....	14
8.1. POZİTİF TWEETLER	14
8.2. NEGATİF TWEETLER.....	14
8.3. NÖTR TWEETLER	15
9. HARİTA.....	16
10. SONUÇLAR VE ÖNERİLER	19
11. KAYNAKÇA	20

ŞEKİL LİSTESİ

Sayfa No

Şekil 2.2.1 Sandbox-1.....	Hata! Yer işareti tanımlanmamış.
Şekil 2.2.2 Sandbox-2.....	Hata! Yer işareti tanımlanmamış.
Şekil 3.1.1 Kütüphaneler.	Hata! Yer işareti tanımlanmamış.
Şekil 3.2.1.1 Twitter API.....	5
Şekil 3.2.2.1 3250 Tweet Çekme Kodu.....	5
Şekil 3.2.2.2 3250 Tweet Çıktı.	5
Şekil 3.2.3.1 Tweet Filtreleme.....	6
Şekil 3.2.3.2 Google Sheets.....	7
Şekil 4.1.1 Kod İle Tweet Çevirme.....	8
Şekil 5.1.1 Excel Dosyasını Yükleme.	9
Şekil 5.2.1 Yüklenen Verileri Düzenleme.....	10
Şekil 5.2.2 Düzenlenmiş Tweetler.....	11
Şekil 6.1 Subjectivity & Polarity Kodları.....	12
Şekil 6.2 Subjectivity & Polarity Değerleri.....	13
Şekil 7.1 Word Cloud Oluşturma Kodları.....	14
Şekil 7.2 Word Cloud.....	14
Şekil 8.1 Polarity Değerlerinin Analizi.....	15
Şekil 8.1.1 Pozitif Tweetleri Yazdırma Kodu.....	15
Şekil 8.1.2 Pozitif Tweetlerin Yüzdesini Bulma Kodu.....	15
Şekil 8.2.1 Negatif Tweetleri Yazdırma Kodu.....	16
Şekil 8.2.2 Negatif Tweetlerin Yüzdesini Bulma Kodu.....	16
Şekil 8.3.1 Nötr Tweetleri Yazdırma Kodu.....	16
Şekil 8.3.2 Nötr Tweetlerin Yüzdesini Bulma Kodu.....	16
Şekil 9.1 Polarity & Subjectivity Haritasını Çıkarma Kodu.....	17
Şekil 9.2 Polarity & Subjectivity Haritası.....	17
Şekil 9.3 Grafikselsel Gösterim.....	18

ÖZET

TWITTER VERİLERİ ÜZERİNDE DUYGU ANALİZİ

Mustafa Kerem KILIÇLIOĞLU
İzmir Katip Çelebi Üniversitesi
Fen bilimleri Enstitüsü
Dönem Projesi
Danışman: Dr. Serpil Yılmaz
Ocak 2024, 26 sayfa

Bitirme projemde maalesef ülkemizde 6 Şubat 2023 yılında yaşanan deprem felaketi sonrası Twitter üzerinde yaşanan felaket ile ilgili atılan tweetleri, o tweetlere atılan bazı yanıt, retweet ve onun bahsedildiği tweetleri topladım. Sonrasında o tweetlere duygu analizi uyguladım.

Anahtar sözcükler: Duygu analizi, Twitter, DEPREM, 6 Şubat, Türkiye, Tweet

ABSTRACT

SENTIMENT ANALYSIS OF TWITTER DATA

Mustafa Kerem KILIÇLIOĞLU
İzmir Katip Çelebi University
Institute Of Science
Supervisor: Dr. Serpil Yılmaz
January 2024, 26 pages

In my graduation project, unfortunately, after the earthquake disaster that happened in our country on February 6, 2023, I collected the tweets about the disaster on Twitter, some responses, retweets and tweets mentioning it. Then I applied sentiment analysis to those tweets.

Keywords: Sentiment analysis, Twitter, EARTHQUAKE, 6 February Turkey ,Tweet

1. GİRİŞ

Dönem projesinin amacı insanların sosyal medyadaki duygu durumlarını ölçmek ve analiz yapmaktır. Belirli konular hakkında insanların nasıl hissettiklerini ölçerek ona göre bazı çıkarımlarda bulunabiliriz. Ürünlerini kitlelere pazarlamaya çalışan bir şirket için nerelerde yanlış yaptıklarını bularak hatalarını düzeltmeyi sağlar. Belirli konularda devletlerin veya şirketlerin araştırmalarında kullanılarak insanların nabzı ölçülür ve sorunlara karşı daha etkili çözümler bulmamız sağlanır.

2. TWITTER DEVELOPER HESABI

Duygu analizi yapmamız için veri setine ihtiyacımız var. İlk olarak Twitter’da Developer Hesabı açmalıyız. Developer hesabı için Twitter bizden bazı bilgiler istiyor. Bu hesabı hangi amaçla kullanacağımız, nerelerde kullanacağımız gibi bilgileri girerek açma talebimizi gönderiyoruz. Twitter talebi inceledikten sonra eğer uygun bulursa hesabı aktifleştiriyor ve bize ücretsiz olarak kullanacağımız belirli sınırları olan Sandbox’lar veriyor.

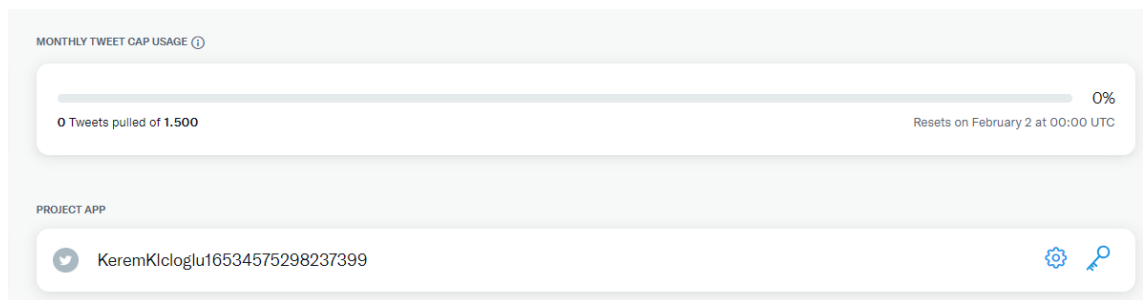
2.1. TWEETPY

Tweepy, Python ile Twitter API'sine erişmemizi sağlayan açık kaynaklı bir Python paketidir. Tweepy sayesinde Developer Hesabı’ndan alacağımız API ile belirli sınırlar altında veriler elde edebileceğiz.

2.2. SANDBOX

Twitter, Developer Hesabı’nda bize kullanmamız için Sandboxlar sunar. Bu Sandbox’lar belirli sınırlar içerisinde API’leri kullanarak projelerimizi test edebileceğimiz ortamlardır. Bu projede Twitter’ın ücretsiz olarak sunduğu Sandbox’lardan kullanacağım. Bunlar aylık 25 bin ve 5 bin tweet ile sınırlıdır. Belirli gün aralığında tweetleri alabilmemizi sağlar. Aynı zamanda Twitter bu sınırlar dışında son atılmış 3250 tweeti almamıza izin verir.

Şekil 2.2.1 Sandbox-1




Şekil 2.2.2 Sandbox-2

Free

Overview

A package that provides free access to Twitter APIs

 Your Project has Free access:
Default project-165345752982739931

1 Apps

1 environment

1 Tweets

Tweets - post up to 1500 Tweets per month

1 Cost

free

3. VERİ TOPLAMA

Twitter Developer Hesabı'nı açtıktan sonra veri toplamak için Google Colab üzerinde Python kodlama dilini kullanarak kod yazmaya başlayacağız.

3.1. Kütüphaneler

Twitter üzerindeki verileri kullanmamız için Tweep'yı import ediyoruz. Verilere duygu analizi yapabilmek, tablo olarak gösterebilmek, en çok kullanılan kelimeleri bulut halinde gösterebilmek ve bazı diğer işlemler için aşağıda gördüğümüz şekilde kütüphanelerimizi yazıyoruz.

▼ Kütüphaneler

```
[1] #Kütüphaneler
import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd

import numpy as np
import re
import matplotlib.pyplot as plt
import datetime, time
plt.style.use('fivethirtyeight')
```

Şekil 3.1.1 Kütüphaneler

3.2. Tweet Toplama İşlemleri

Tweepy'ı projemize import ettik. Bundan sonra onun kodlarını kullanarak Tweetleri çekmeye çalışacağız.

3.2.1. Twitter API

Tweetleri toplayabilmemiz için Twitter'ın bize verdiği API bilgilerini girmemiz ve hesabımıza erişim sağlamamız gerekiyor.

```
#Twitter API Bilgileri  
consumerKey = '8rFmU0YVe8QBei21GsY6d7dAi'  
consumerSecret = 'RUjBcYgIQ8y6DiLrQbNkVo57M5jrF0oAehc2LaA1JAaHe60jxI'  
accessToken = '633727850-T7pEeWJ3KD5uryV99BbQvGUCb0xTVt4IGtstZTh9'  
accessTokenSecret = 'WhPTx1HOvrpTnsorlckUqmwHVH272CeIrGdFGPvfvfJsAW'
```

Şekil 3.2.1.1 Twitter API

3.2.2. Son 3250 Tweet

Deprem ile ilgili paylaşılan son 3250 tweete göre alabiliyoruz sadece. Bu tweetleri aşağıdaki kodlar sayesinde alıyoruz.

```
tweets_list = []  
text_query = 'EARTHQUAKE'  
count = 3250  
  
for tweet in api.search(q=text_query, count=count, lang='en-en'):  
    tweets_list.append((tweet.created_at, tweet.id, tweet.text))
```

Şekil 3.2.2.1 3250 Tweet Çekme Kodu

tweets, likes, time ve ids adında dizilerimizi oluşturuyoruz. Tweepy yardımıyla 'EARTHQUAKE' hashtag'i ile atılan son 3250 tweeti alıyoruz. Her tweeti aldığında dizilere gerekli bilgileri yazdırıyoruz. En sonunda tüm dizileri bir dataframe içine atıp yazdırıyoruz.

	Tweet Id	Tweet URL	Tweet Posted Time	Tweet Content	Tweet Type	Client	Retweets Received	Likes Received	Tweet Location	Tweet Language	Use
0	"1625842348771901440"	https://twitter.com/bluwrld/status/16258423487...	15 Feb 2023 12:59:57	"128 hours buried, rescued alive, checked in L...	ReTweet	"Twitter for iPhone"	0	0		English	"1514326011835568"
1	"1625842313359400961"	https://twitter.com/PRembielak/status/16258423...	15 Feb 2023 12:59:48	"RICO from puppy to #rescuedog #USAR #PL and h...	Tweet	"Twitter Web App"	0	5		English	"760786867168108"
2	"1625842284347682817"	https://twitter.com/SadiaKh26232740/status/162...	15 Feb 2023 12:59:41	"May Allah have mercy on the Muslim Umriah, Ame...	ReTweet	"Twitter for Android"	0	0		English	"1118836582390411"
3	"1625842271286938976"	https://twitter.com/roonma/status/16258422712...	15 Feb 2023 12:59:38	"استغفرالله #TurkeySyriaEarthquake#TurkeyL...	Tweet	"Twitter for Android"	13	14		Urdu	"1430039213681033"
4	"1625842231448858625"	https://twitter.com/PiomRox/status/16258422314...	15 Feb 2023 12:59:29	"کی رنگ رنگ قرین اور #PSL8 ایک طرف #TimHorton...	ReTweet	"MedhaGayathri"	0	0		Urdu	"1503237208748474"
...
42645	"1625085212731351047"	https://twitter.com/Reading__online/status/162...	13 Feb 2023 10:51:21	"35.225 oi vekpoi σε #Toupi& #Zupia από...	ReTweet	"Twitter for Android"	1	0		Greek	"801862385195741"
42646	"1625085180229607428"	https://twitter.com/MeetOurWorld/status/162508...	13 Feb 2023 10:51:14	"CCTV footage shows people fleeing when quake ..."	Tweet	"Twitter Web App"	0	1		English	"1514401259947806"
42647	"1625085177419427843"	https://twitter.com/hdkistanbul/status/1625085...	13 Feb 2023 10:51:13	"Acil çağrı Dünyanın her yerinden bizlere gön...	Tweet	"Twitter Web App"	6	7		Turkish	"3230471"
42648	"1625085171979688888"	https://twitter.com/Maha_Bhisma/status/1625085...	13 Feb 2023 10:51:12	"Keajaiban lain di Turki: Abdul Karim, 67 ta...	ReTweet	"Twitter for Android"	6	0		in	"1605027727790125"
42649	"1625085171023351809"	https://twitter.com/linjn_/status/162508517102...	13 Feb 2023 10:51:12	"เมืองมาทกอกเมืองมาทกอกจากแผ่นดินไหว 🇹🇷🇮🇳"	ReTweet	"Twitter for iPhone"	6615	0		Thai	"770940701911547"

Şekil 3.2.2.2 3250 Tweet Çıktı

3.2.3. Filtreleme

Twitter bazı Sandbox API'leri sayesinde tweetleri belirli filtrelere göre aramamızı sağlar. Ücretsiz olarak kullanabileceğimiz Şekil 2.1 ve Şekil 2.2'deki 30-Days Sandbox'ını kullanarak 1 aylık aralıklardaki tweetleri çekebiliriz.

```
[ ] #Belli tarih aralığındaki cevapları alma
tweets = []
likes = []
time = []
ids = []
cevapverilen = []

name = 'drfahrettinkoca'

replies = tweepy.Cursor(api.search_30_day, query='to:'+name, fromDate='202112150001', toDate='202112152359', environment_name='thesis30day').items(1000)

for reply in replies:
    tweets.append(reply.text.replace('\n', ' '))
    likes.append(reply.favorite_count)
    time.append(reply.created_at)
    ids.append(reply.id)
    cevapverilen.append(reply.in_reply_to_status_id_str)

df = pd.DataFrame({'tweets':tweets, 'likes':likes, 'time':time, 'ids':ids, 'cevapverilen':cevapverilen})
df
```

Şekil 3.2.3.1 Tweet Filtreleme

Aynı işlemleri bu sefer api.search_30_day kodumuzu kullanarak gerçekleştirdik. Bunu kullandığımızda iki tarih aralığı belirtmemiz isteniyor ve bu aralık asla belirli bir sınırı aşamaz. Bu kodu kullanarak farklı farklı tarih aralıkları girerek 20bin'e yakın tweet çektim.

Her dataframe oluşturduktan sonra toplanan verileri bilgisayara ya da sanal bir ortama kaydetmemiz gerekir. Eğer kaydetmezsek tekrar dataframe oluşturduğumuzda üstüne

yazılır ya da belirli bir süre Colab'a girmezsek otomatik olarak veriler silinir. Bu yüzden dataframe'leri Google Sheets'e atmak için ayrı bir kod yazdım oradan da dilersek Excel olarak indirebiliriz.

Google Sheets İşlemleri

```
[ ] #Googler Sheets / Doğrulama
!pip install --upgrade --quiet gspread
from google.colab import auth

auth.authenticate_user()

import gspread
from oauth2client.client import GoogleCredentials

gc = gspread.authorize(GoogleCredentials.get_application_default())
```

```
[ ] #Googler Sheets / Kaydetme

from gspread_dataframe import get_as_dataframe, set_with_dataframe

sh = gc.create('Tweetler')
worksheet = gc.open('Tweetler').sheet1

set_with_dataframe(worksheet, df2)
```

Şekil 3.2.3.2 Google Sheets

İlk kodu çalıştırdığımızda Google Drive'a erişim izni istenir erişim izni verdikten sonra Karşımıza çıkan key'i kodun altında çıkacak olan ekrana girmemiz gerekir. Onu girdikten sonra 2. Kod çalıştırılır bu sayede dataframe tweetler adı altında sheets dosyası olarak Google Drive'a kaydedilir.

4. ÇEVİRİ İŞLEMLERİ

Duygu analizi yapmamızı sağlayan kodlar sadece İngilizce kelimeler üzerinde analiz yapıyor. O yüzden tweetleri İngilizceye çevirmemiz gerekiyor.

4.1. Kod İle Çevirme

Colab üzerinden çeviri yapmamızı sağlayan kütüphaneyi eklememiz gerekiyor. Onu yaptıktan sonra dataframe üzerindeki her tweete tek tek çeviri yapacak bir fonksiyon yazmalıyız. Bu çevirileri yeni bir dizi oluşturup içine atacağız ve duygu analizine başlayacağız (Renotte 2022).

```
[ ] %%capture
!pip install git+https://github.com/neuml/txtai#egg=txtai[pipeline]

[ ] %%capture
from txtai.pipeline import Translation
#Çeviri modeli oluşturma
translate = Translation()

#Tweetleri İngilizceye çevirip ekrana yazdır
def tweetCevir(yazi):
    return translate(yazi, 'en')

df['englishtweets'] = df['tweets'].apply(tweetCevir)
df
```

Şekil 4.1.1 Kod İle Tweet Çevirme

5. DOSYA YÜKLEME VE TWEETLERİ DÜZENLEME

Google Sheets üzerinden düzenlediğimiz verilerin hepsini bir bütün olarak tekrar Colab'a yüklememiz gerekiyor. Bunları yükledikten sonra tweetlerin içinde duygu analizinde sıkıntı çıkarabilecek şeyleri çıkartmamız gerekiyor. Linkleri, @ işareti ile bahsetmeleri, # (hashtag) işaretini ve RT yazılarını kod yardımıyla çıkarmalıyız.

5.1. Dosya Yükleme

```
#Google Colab'a veri yükleme
from google.colab import files
uploaded = files.upload()

#Excel dosyasını dataframe'e atıp yazdırma
df = pd.read_excel('Tweetler.xlsx', sheet_name= 0)
df
```

Şekil 5.1.1 Excel Dosyasını Yükleme

İlk kodu çalıştırdıktan sonra bilgisayardan dosya seçilir ve yükleme yapılır. İkinci kod ile dataframe'e yüklenen excel dosyası aktarılır. Dosya sistemde kalıcı değildir. Colab üzerinde belirli süre hiçbir şey yapılmadığında silinir.

5.2. Düzenleme

```
#Tweetlerden bahsetme, rt, link, hashtag gibi şeyleri çıkarma
def tweetDuzenle(yazi):
    yazi = re.sub(r'[A-Za-z0-9]+', '', yazi) #mentionları kaldırma
    yazi = re.sub(r'#', '', yazi) #hashtag sembolünü kaldırma
    yazi = re.sub(r'RT[\s]+', '', yazi) #RT yazısını kaldırma
    yazi = re.sub(r'https?:\/\/\S+', '', yazi) #Linkleri kaldırma

    return yazi

df['englishtweets'] = df['englishtweets'].apply(tweetDuzenle)
dfyanitlar['englishtweets'] = df['englishtweets'].apply(tweetDuzenle)
df
```

Şekil 5.2.1 Yüklenen Verileri Düzenleme

Oluşturduğumuz fonksiyonla tüm tweetler düzenlenir.

	Tweet Id	Tweet URL	Tweet Posted Time	Tweet Content	Tweet Type	Client	R R
0	"1625842348771901440"	https://twitter.com/bluwrld/status/16258423487...	15 Feb 2023 12:59:57	"128 hours buried, rescued alive, checked in t...	ReTweet	"Twitter for iPhone"	
1	"1625842313359400961"	https://twitter.com/PRembielak/status/16258423...	15 Feb 2023 12:59:48	"RICO from puppy to #rescuedog #USAR #PL and h...	Tweet	"Twitter Web App"	
2	"1625842284347682817"	https://twitter.com/SadiaKh26232740/status/162...	15 Feb 2023 12:59:41	"May Allah have mercy on the Muslim Ummah, Ame...	ReTweet	"Twitter for Android"	
3	"1625842271286398976"	https://twitter.com/roonma/status/16258422712...	15 Feb 2023 12:59:38	"استغفرالله ن#TurkeySyriaEarthquakeIn#Turkey\...	Tweet	"Twitter for Android"	
4	"1625842231448858625"	https://twitter.com/PlomRox/status/16258422314...	15 Feb 2023 12:59:29	"كى رنگا رنگ تريب اور #PSL8 اک طرف" #TimHorton...	ReTweet	"MedhaGayathri"	
...
42645	"1625085212731351047"	https://twitter.com/Reading__online/status/162...	13 Feb 2023 10:51:21	"35.225 oi vεkpoi σε #Touptkia & #Eupia ατό...	ReTweet	"Twitter for Android"	
42646	"1625085180229607428"	https://twitter.com/MeetOurWorld/status/162508...	13 Feb 2023 10:51:14	"CCTV footage shows people fleeing when quake ..."	Tweet	"Twitter Web App"	
42647	"1625085177419427843"	https://twitter.com/hdkistanbul/status/1625085...	13 Feb 2023 10:51:13	"Acil çağrı! Dünyanın her yerinden bizlere gön...	Tweet	"Twitter Web App"	
42648	"1625085171979685888"	https://twitter.com/Maha_Bhisma/status/1625085...	13 Feb 2023 10:51:12	"Keajaiban lain di Turki; Abdul Karim, 67 ta...	ReTweet	"Twitter for Android"	
42649	"1625085171023351809"	https://twitter.com/linjn_/status/162508517102...	13 Feb 2023 10:51:12	"น้องแมวกลอน้องหมาหลังจากรอดจากแผ่นดินไหว 🙏🙏🙏"	ReTweet	"Twitter for iPhone"	

Şekil 5.2.2 Düzenlenmiş Tweetler

Şekil 5.2.2’de düzenleme öncesi tweetleri ve düzenleme sonrası tweetleri görebiliriz. Tweetleri düzenlediğimize göre duygu analizi için veri üzerinde çalışmaya başlayabiliriz.

6. SUBJECTIVITY & POLARITY

Python ile TextBlob kütüphanesi yardımıyla tweetlerin Subjectivity ve Polarity değerlerini bulacağız (Shah 2020). Subjectivity ve Polarity'nin Türkçe karşılıkları Öznellik ve Kutupluktur. Kodlayacağımız fonksiyonlarla bu değerleri bulup duygu analizi için kullanmaya başlayacağız. Bu değerleri kullandığımız dataframe içinde 'polarity' ve 'subjectivity' adında iki sütun açarak göstereceğiz.

```
#Subjektifliği elde etme (subjectivity)
def getSubjectivity(yazi):
    return TextBlob(yazi).sentiment.subjectivity

#Polariteyi elde (polarity)
def getPolarity(yazi):
    return TextBlob(yazi).sentiment.polarity

#Subjectivity ve Polarity'yi dataframe üzerinde göster
df['subjectivity'] = df['englishtweets'].apply(getSubjectivity)
df['polarity'] = df['englishtweets'].apply(getPolarity)

##dfYanitlar['subjectivity'] = df['englishtweets'].apply(getSubjectivity)
##dfYanitlar['polarity'] = df['englishtweets'].apply(getPolarity)

df
```

Şekil 6.1 Subjectivity & Polarity Kodları

Tweet Language	Name	Username	User Bio	Verified or Non-Verified	Profile URL	User Followers	User Following	User Account Creation Date	subjectivity	polarity
English	"\ B.Roy 🇮🇳"	bluwrid	"outspoken & wise 🇮🇳"	Non-Verified	https://twitter.com/bluwrid	30	50	13-Apr-2022 19:34:17	0.533333	0.366667
English	"Przemo Rembielak"	PRembielak	"56, Poland, Open-Minded, Anti Political Corre..."	Non-Verified	https://twitter.com/PRembielak	272	731	03-Aug-2016 10:38:18	0.000000	0.000000
English	"Sadiah Khan"	SadiaKh26232740	"	Non-Verified	https://twitter.com/SadiaKh26232740	239	520	18-Apr-2019 11:20:03	0.000000	0.000000
Urdu	"Genuine Trendy News"	roonma	"Let's Up to Date with Trends."	Non-Verified	https://twitter.com/roonma	4	13	24-Aug-2021 05:28:51	0.000000	0.000000
Urdu	"PloM (🇮🇳🇮🇳)"	PiomRox	"your Pom"	Non-Verified	https://twitter.com/PiomRox	89	1583	14-Mar-2022 05:11:07	0.000000	0.000000
...
Greek	"Reading__online"	Reading__online	"📰📰 news and more"	Non-Verified	https://twitter.com/Reading__online	15	486	24-Nov-2016 18:57:45	0.000000	0.000000
English	"Our World"	MeetOurWorld	"Our World, Our Stories."	Non-Verified	https://twitter.com/MeetOurWorld	160	160	14-Apr-2022 00:33:02	0.000000	0.000000
Turkish	"HDK İstanbul"	hdkistanbul	"Halkların Demokratik Kongresi İstanbul İl Mec..."	Non-Verified	https://twitter.com/hdkistanbul	2744	330	03-May-2015 12:33:56	0.000000	0.000000
in	"Sentanu Putra"	Maha_Bhisma	"#GakFollowPorno Dua akun sudah disuspend. Mat..."	Non-Verified	https://twitter.com/Maha_Bhisma	981	1088	20-Dec-2022 02:30:29	0.000000	0.000000
Thai	"jelly"	linjn_	"	Non-Verified	https://twitter.com/linjn_	28	146	31-Aug-2016 11:06:01	0.000000	0.000000

Şekil 6.2 Subjectivity & Polarity Değerleri

Şekil 6.2'deki kırmızı alanda subjectivity ve polarity değerlerimizi göre görebiliriz.

7. WORD CLOUD

Word Cloud bir veri setinde en çok hangi kelimelerin geçtiğini gösterir. Bizim veri setimizde en çok hangi kelime kullanılmış bunu öğrenmek için Word Cloud kullandım.

```
#Word cloud oluşturma

tumKelimeler = ' '.join( [twts for twts in df['englishtweets']] )
wordCloud = WordCloud(width = 1800, height=900, random_state = 21, max_font_size = 200).generate(tumKelimeler)

plt.imshow(wordCloud, interpolation = "bilinear")
plt.axis('off')
plt.show()
```

Şekil 7.1 Word Cloud Oluşturma Kodları



Şekil 7.2 Word Cloud

Şekil 7.1'deki kodları kullanarak Şekil 7.2'deki Word Cloud'ı oluşturdum. 6 Şubat depremi sonrası tweetleriyle ilgilendiğimiz için doğal olarak en çok EARTHQUAKE (deprem), Turkey(Türkiye), Buried(gömü/toprak altı), Rescued(kurtarıldı) kelimeler kullanılmış.

8. NEGATİF, POZİTİF VE NÖTR TWEETLER

Bulduğumuz polarity değerlerine göre tweetimizin pozitif mi, negatif mi yoksa nötr mü olduğunu bulacağız. Tweet eğer polarity 0'dan düşükse negatif, 0 ise nötr, 0'dan yüksekse pozitif olarak hesaplanacak. Bir fonksiyon oluşturup dataframe'deki her polarity değerini tek tek işleme koyacağız. Dataframe içinde 'Analysis' adında yeni bir sütun açıp bulduğumuz değerleri oraya yazdıracağız.

Tweet Language	Username	User Bio	Verified or Non-Verified	Profile URL	User Followers	User Following	User Account Creation Date	subjectivity	polarity	Analysis
English	bluwrid	"outsoken & wise"	Non-Verified	https://twitter.com/bluwrid	30	50	13-Apr-2022 19:34:17	0.533333	0.366667	Pozitif
English	PRembielak	"56, Poland, Open-Minded, Anti Political Corre..."	Non-Verified	https://twitter.com/PRembielak	272	731	03-Aug-2016 10:38:18	0.000000	0.000000	Nötr
English	Sadiakh26232740	"	Non-Verified	https://twitter.com/Sadiakh26232740	239	520	18-Apr-2019 11:20:03	0.000000	0.000000	Nötr
Urdu	roonma	"Let's Up to Date with Trends."	Non-Verified	https://twitter.com/roonma	4	13	24-Aug-2021 05:28:51	0.000000	0.000000	Nötr
Urdu	PiomRox	"your Pom"	Non-Verified	https://twitter.com/PiomRox	89	1583	14-Mar-2022 05:11:07	0.000000	0.000000	Nötr
...
Greek	Reading__online	"📰📰 news and more"	Non-Verified	https://twitter.com/Reading__online	15	486	24-Nov-2016 18:57:45	0.000000	0.000000	Nötr
English	MeetOurWorld	"Our World, Our Stories."	Non-Verified	https://twitter.com/MeetOurWorld	160	160	14-Apr-2022 00:33:02	0.000000	0.000000	Nötr
Turkish	hdkistanbul	"Halkların Demokratik Kongresi İstanbul İl Mec..."	Non-Verified	https://twitter.com/hdkistanbul	2744	330	03-May-2015 12:33:56	0.000000	0.000000	Nötr
in	Maha_Bhisma	"#GakFollowPorno Dua akun sudah disuspend. Mat..."	Non-Verified	https://twitter.com/Maha_Bhisma	981	1088	20-Dec-2022 02:30:29	0.000000	0.000000	Nötr
Thai	linjn_	"	Non-Verified	https://twitter.com/linjn_	28	146	31-Aug-2016 11:06:01	0.000000	0.000000	Nötr

Şekil 8.1 Polarity Değerlerinin Analizi

8.1. POZİTİF TWEETLER

Tüm tweetler arasından eğer pozitif tweetleri yazdırmak istersek aşağıdaki kodu kullanabiliriz.

```
#Pozitif tweetleri yazdır
j=1
sortedDF = df.sort_values(by = ['polarity'])
for i in range(0, sortedDF.shape[0]):
    if( sortedDF['Analysis'][i] == 'Pozitif'):
        print(str(j) + ' ' + sortedDF['englishtweets'][i])
        print()
        j = j + 1
```

Şekil 8.1.1 Pozitif Tweetleri Yazdırma Kodu

Pozitif tweetlerin yüzdesini de bu şekilde bulabiliriz. Tüm tweetlerin yüzde 49.7'ü pozitif.

```
# Pozitif tweetlerin yüzdesini al
ptweets = df[df.Analysis == 'Pozitif']
ptweets = ptweets['englishtweets']

round((ptweets.shape[0] / df.shape[0]) * 100, 1)
```

29.3

Şekil 8.1.2 Pozitif Tweetlerin Yüzdesini Bulma Kodu

8.2. NEGATİF TWEETLER

Tüm tweetler arasından eğer negatif tweetleri yazdırmak istersek aşağıdaki kodu kullanabiliriz.

```
#Negatif tweetleri yazdır
j=1
sortedDF = df.sort_values(by = ['polarity'], ascending = 'False')
for i in range(0, sortedDF.shape[0]):
    if( sortedDF['Analysis'][i] == 'Negatif'):
        print(str(j) + ' ' + sortedDF['englishtweets'][i])
        print()
        j = j + 1
```

Şekil 8.2.1 Negatif Tweetleri Yazdırma Kodu

Negatif tweetlerin yüzdesini de bu şekilde bulabiliriz. Tüm tweetlerin yüzde 9.1'i negatif.

Şekil 8.2.2 Negatif Tweetlerin Yüzdesini Bulma Kodu

```
# Negatif tweetlerin yüzdesini al
ntweets = df[df.Analysis == 'Negatif']
ntweets = ntweets['englishtweets']

round((ntweets.shape[0] / df.shape[0]) * 100, 1)
```

14.1

8.3. NÖTR TWEETLER

Tüm tweetler arasından eğer nötr tweetleri yazdırmak istersek aşağıdaki kodu kullanabiliriz.

```
#Nötr tweetleri yazdır
j=1
sortedDF = df.sort_values(by = ['polarity'], ascending = 'False')
for i in range(0, sortedDF.shape[0]):
    if( sortedDF['Analysis'][i] == 'Nötr'):
        print(str(j) + ' ' + sortedDF['englishtweets'][i])
        print()
        j = j + 1
```

Şekil 8.3.1 Nötr Tweetleri Yazdırma Kodu

Nötr tweetlerin yüzdesini de bu şekilde bulabiliriz. Tüm tweetlerin yüzde 41.2'si nötr.

```
# Nötr tweetlerin yüzdesini al
ntweets = df[df.Analysis == 'Nötr']
ntweets = ntweets['englishtweets']

round((ntweets.shape[0] / df.shape[0]) * 100, 1)

56.6
```

Şekil 8.3.2 Nötr Tweetlerin Yüzdesini Bulma Kodu

9. HARİTA

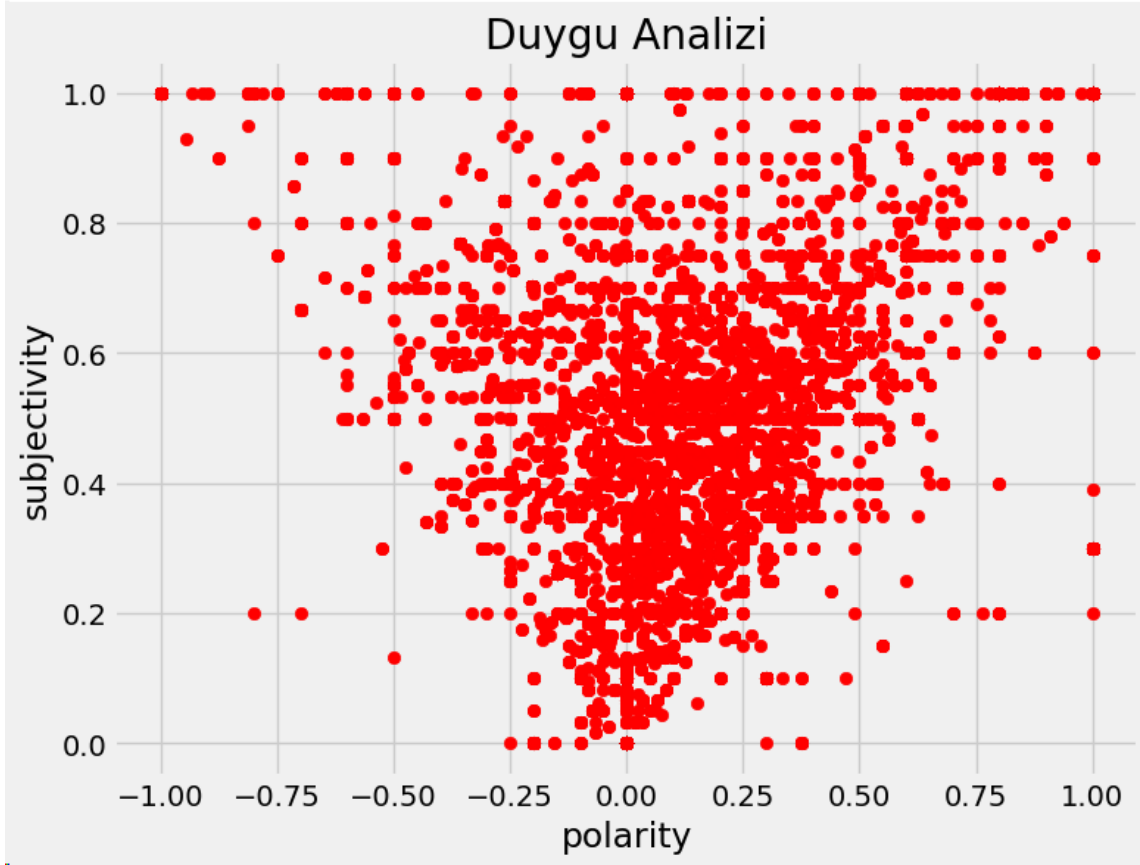
Tüm tweetlerin polarity ve subjectivity değerlerini daha iyi görebilmemiz için harita oluşturacağız. Haritada hangi değerlerde yoğunluk olduğunu daha rahat görebiliriz.

Aşağıdaki kodlarla bu haritayı oluşturabiliriz.

```
#Polarity ve Subjectivity haritasını çıkar
plt.figure(figsize=(8,6))
for i in range(0, df.shape[0]):
    plt.scatter(df['polarity'][i], df['subjectivity'][i], color='Red')

plt.title('Duygu Analizi')
plt.xlabel('polarity')
plt.ylabel('subjectivity')
plt.show()
```

Şekil 9.1 Polarity & Subjectivity Haritasını Çıkarma Kodu

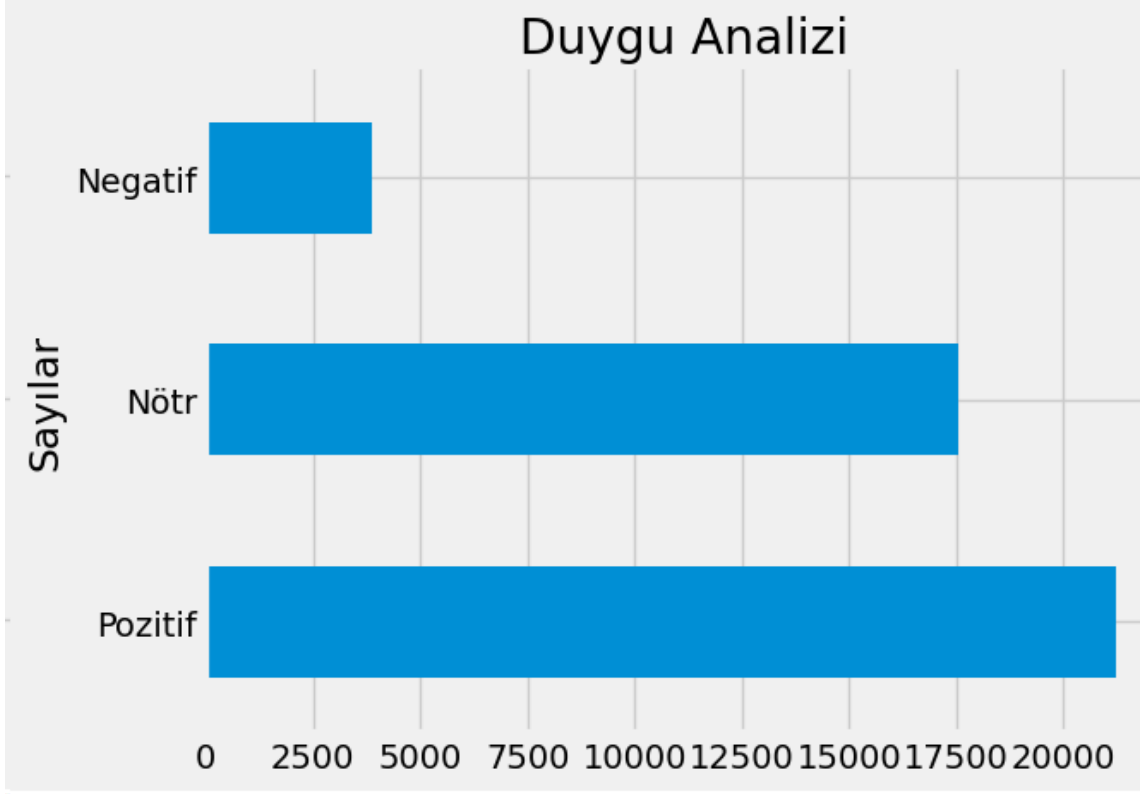


Bu kodu çalıştırdıktan sonra tüm tweetleri işleme 10-15 dakika sürecektir. 10-15 dakika çalıştırdıktan sonra oluşturulan harita aşağıdaki gibidir.

Şekil 9.2 Polarity & Subjectivity Haritası

Tweet sayılarını daha iyi görmek için şu şekilde bir grafikte de gösterebiliriz.

Tweetlerin yarısından fazlasının duygu analizi nötr olarak ölçülmüş sonrasında sırasıyla en çok pozitif ve negatif tweet bulunuyor.



Şekil 9.3 Grafikselsel Gösterim

Nötr yani polarity değeri 0 olan tweetler diğerlerine göre çok daha fazladır. Bazı araştırmalara göre nötr duygusal değerdeki verilerden öğreneceğimiz şeyler diğerlerine göre daha azdır. O yüzden nötr değerler önemsizdir. Duygu analizi yaparken ilk olarak tarafsızlığın ölçülmesi gerekir. Eğer subjectivity yani öznellik az ise o verinin daha az bilgilendirici olduğu düşünülür. Bu sebeple ilk olarak öznelliğe daha sonra polariteye bakılır (Vryniotis 2013).

Moshe Koppel ve Jonathan Schler (Koppel 2006) 2006 yılında yaptığı araştırmalarda nötr değerlerin önemsiz görülmemesi gerektiğini göstermiştir. Her polarite probleminde olduğu gibi, üç kategorinin tanımlanması gerektiğini (pozitif, negatif ve nötr) ve nötr kategorinin tanıtılmasının genel doğruluğu bile iyileştirebileceğini öne sürmüşlerdir.

10.SONUÇLAR VE ÖNERİLER

Tweetlerin duygu analizini yaptım ve bunun sonucunda pozitif, negatif ve nötr tweetleri tespit ettim. Pozitif tweetler her ikisine göre daha fazla çıktı sonrasında nötr ve en az da negatif çıktı. Duygu analizini yapmak için her kelimenin anlamına göre verilmiş belirli polarity değerlerini kullandık. Kendi fikrimce bu sonuçlar bizim bazı sorunları anlamamızı sağlayabilir ama çok doğru bir sonuca ulaşamayız. İnsanlar sosyal medyada %100 doğru dil bilgisiyle yazmadığı için herkes birçok kısaltma kullanıyor ya da kelimeleri, cümleleri yanlış yazıyor. Kendimiz bile bazen anlayamadığımız bir şeyi basit bir algoritma ile çalışsan bu sistemin anlamasını bekleyemeyiz. Bazı insanlar kötü bir şeyi ironi yaparak güzel bir dille anlatabilir ya da tam tersi kötü şeyler yazarak güzel bir şey söyleyebilir. Bu tarz anlamları bu algoritmanın çıkarması çok zordur. Çok daha doğru bir sonuca ulaşılması için Doğal Dil İşleme ve Derin Öğrenme kullanılarak büyük verilerle yapay zekanın eğitilmesi gerekir. Böyle yaptığımızda yapay zeka güzel anlayıp cümlenin duygusunu daha iyi analiz edebilir.

11.KAYNAKÇA

- [1] V. Vryniotis, «The importance of Neutral Class in Sentiment Analysis,» DatumBox, [Çevrimiçi]. Available: <https://blog.datumbox.com/the-importance-of-neutral-class-in-sentiment-analysis/>.
- [2] J. S. Moshe Koppel, «The importance of neutral examples for learning sentiment,» *Computational Intelligence*, no. 22, 2006.
- [3] P. Shah, «Sentiment Analysis using TextBlob,» towards data science, 27 Haziran 2020. [Çevrimiçi]. Available: <https://towardsdatascience.com/my-absolute-go-to-for-sentiment-analysis-textblob-3ac3a11d524>.
- [4] N. Renotte, «Nicholas Renotte - Youtube,» [Çevrimiçi]. Available: <https://www.youtube.com/channel/UCHXa4OpASJEwrHrLeIzw7Yg>.