

T.C.
Bilecik Şeyh Edebali Üniversitesi
İktisadi ve İdari Bilimler Fakültesi
Yönetim Bilişim Sistemleri



Twitter Veri ve Duygu Analizi

66382088944

Melih AY

ÖZET

Günümüzde veriye ulaşmak ve veri kirliliği yaratmak çok kolay, bundan dolayı her yerde gereksiz veriler olabiliyor. Bunlardan yararlı olanları alıp, bizim işimize yarayan kısımları işleyerek yapmak istediğimiz işleri ya da metin içindeki istediğimiz verilere ulaşmamızı kolaylaştıran “R Studio” var. Ben ise developer hesabı kullanarak Twitter’den “Kripto Haberleri” ile yapılan paylaşımları inceleyip metin analizi ve duygu analizi yapacağım.

1) TWITTER GELİŞTİRİCİ HESABI ALMA

Başlayabilmek için öncelikle Twitter developer hesabı açıyoruz. Tabii Twitter ilk denememde developer hesabı vermedi, bunun sebebini açıklamamı ve nerelerde kullanacağımı tam olarak belirten bir mail attıktan sonra developer hesabım açıldı. Daha sonra uygulamalar bölümünden yeni bir uygulama oluşturuyoruz ve oluşturduğumuz uygulamaya bir isim veriyoruz. Uygulamayı açarken bize token veriliyor, bunları R programlamada kullanarak Twitter ve R Studio arasındaki bağlantıyı sağlıyoruz.

Projects & Apps

Overview

Project 1

{ Products **NEW**

{ Account

Elevated

Project 1 >

V1.1 ACCESS V2 ACCESS

DEVELOPMENT APP

Twitter

Settings

QUOTA: 1 OF 3 ENVIRONMENTS

+Add App

Settings Keys and tokens

Consumer Keys

API Key and Secret ⓘ

Reveal API Key hint

Regenerate

Authentication Tokens

Bearer Token ⓘ

Generated November 30, 2021

Revoke Regenerate

Access Token and Secret ⓘ

Generated November 30, 2021

For @kemco99

Created with Read, Write, and Direct Messages permissions

Revoke Regenerate

2) GEREKLİ KÜTÜPHANELERİ İNDİRİP AKTİF HALE GETİRME

Daha sonra R studio programına girip gerekli olan paketleri indiriyoruz.

- Aşağıda görüldüğü gibi paketleri “install.packages()”komutu ile indiriyoruz.

```
1 install.packages("twitter")
2 install.packages("ROAuth")
3 install.packages("tm")
4 install.packages("RCurl")
5 install.packages("magrittr")
6 install.packages("dplyr")
7 install.packages("tidyverse")
8 install.packages("ggplot2")
9 install.packages("funModeling")
10 install.packages("lubridate")
11 install.packages("stringi")
12 install.packages("stringr")
13 install.packages("readxl")
14 install.packages("tidytext")
15 install.packages("wordcloud")
16 install.packages("RColorBrewer")
17 install.packages("tibble")
18 install.packages("tidyr")
19 install.packages("ggthemes")
20 install.packages("readr")
21 install.packages("readxl")
22 install.packages("ggpubr")
23 install.packages("formattable")
24 install.packages("ggstance")
25 install.packages("pastecs")
26 install.packages("psych")
27 install.packages("GGally")
28 install.packages("pander")
29 install.packages("rstatix")
30 install.packages("sentimentr")
31 install.packages("webshot")
32 install.packages("wordcloud2")
33 install.packages("htmlwidgets")
34 install.packages("syuzhet")
```

- Paketleri aktifleştirmek için onları "library()" komutu ile çağırıyoruz.

```
library(lubridate)
library(twitter)
library(tm)
library(ROAuth)
library(Rcurl)
library(magrittr)
library(dplyr)
library(tidyverse)
library(ggplot2)
library(funmodeling)
library(stringi)
library(stringr)
library(tm)
library(tidytext)
library(dplyr)
library(ggplot2)
library(RColorBrewer)
library(tibble)
library(tidy)
library(ggthemes)
library(readr)
library(readxl)
library(ggpubr)
library(formattable)
library(ggstance)
library(pastecs)
library(psych)
library(Gally)
library(pander)
library(rstatix)
library(sentimentr)
library(webshot)
library(wordcloud)
library(wordcloud2)
library(htmlwidgets)
library(magrittr)
library(dplyr)
```

3) TWITTER HESABINI R STUDIO'YA BAĞLAYIP ÇALIŞTIRMA

Şimdi de Twitter ile R studio programını bağlamak için developer hesabımızda bize verilen anahtarları kullanıyoruz.

```
# Gelistirici bilgileri
consumer_key <- "XXXXXXXXXXXXXXXXXXXX"
consumer_secret <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
access_token <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
access_secret <- "XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX"
```

Bağlantı tamamlandığında aşağıdaki bildirimi alıyoruz.

```
[1] "Using direct authentication"
```

4) SEÇİLEN HASHTAG İLE ATILAN TWEETLERİ ÇEKME

Bağlantı ayarlarını tamamladığımıza göre şimdi çekeceğimiz verilerin nereden, hangi dilden ve ne kadar çekeceğimizi belirtiyoruz. Ardından bu şekilde 21 gün boyunca verileri çektik ve uygun formata dönüştürdük.

```
# çekeceğimiz veriler |  
tweets <- searchTwitter("#kriptohaber", n=1000, locale = "tr", lang="tr")  
tweets.df <- twListToDF(tweets)  
tweet_clean <- tweets.df  
tweet_clean$text <- stri_enc_toutf8(tweet_clean$text)
```

Aşağıda gördüğümüz gibi tek çekişte 7 günlük tweet çektik.

1	1,buradan <U+0001F43B> çıkmaz ytd kendi kararınla
2	2, bu hafta vesper earn deneyimini yaşayın \$dai yatırın ve...
3	3, bitcoin aktif tekrar test edilecek karar sizin olmakla
4	4,biden fed enflasyonu dizginlemek için harekete geçece...
5	5, <U+0001F4E2> \$flux in ile ve bilgi paylaşımına ilişkin ...
6	6, <U+0001F4E2> \$flux in ile ve bilgi paylaşımına ilişkin ...
7	7, <U+0001F4E2> \$flux in ile ve bilgi paylaşımına ilişkin ...
8	8, <U+0001F4E2> \$flux in ile ve bilgi paylaşımına ilişkin ...
9	9, <U+0001F4E2> \$flux in ile ve bilgi paylaşımına ilişkin ...
10	10, <U+0001F4E2> \$flux in ile ve bilgi paylaşımına ilişkin...
11	11,<U+0001F4E2> \$flux in ile ve bilgi paylaşımına ilişkin ...
12	12, in izleme listesine aldığı küçük ve emin adımlarla büy...
13	13, yeni bir baskı yapılıyor canlı yayınlarda sürekli dile ge...
14	14, <U+0001FA99> son durum ne altcoinlerin durumu d...
15	15, <U+0001FA99> btc den hızlı bir toparlanma gelir mi ...

5) ÇEKİLEN TWEETLERİN DÜZELTİLMESİ

Daha sonra çektiğimiz tweet'lerdeki rtlerin kaldırılması, url linklerinin temizlenmesi, “#” ve “@” işaretlerinin temizlenmesi, noktalama işaretlerinin temizlenmesi, tüm harflerin küçük harfe dönüştürülmesi ve stopwords'lerin temizlenmesi için aşağıdaki komutları kullandık.

```
#####rtlerin kaldırılması
tweet_clean$text <- ifelse(str_sub(tweet_clean$text,1,2) == "RT",
                           substring(tweet_clean$text,3),
                           tweet_clean$text)

#URL linklerinin temizlenmesi
tweet_clean$text <- str_replace_all(tweet_clean$text, "http[^\s:]*", "")

#Hashtag "#" ve "@" işaretlerinin kaldırılması
tweet_clean$text <- str_replace_all(tweet_clean$text, "#\\s+", "")
tweet_clean$text <- str_replace_all(tweet_clean$text, "@\\s+", "")

#Noktalama işaretlerinin temizlenmesi
tweet_clean$text <- str_replace_all(tweet_clean$text, "[[:punct:][:blank:]]+", " ")

#Tüm harflerin küçük harfe dönüştürülmesi
tweet_clean$text <- str_to_lower(tweet_clean$text, "tr")

#Rakamların temizlenmesi
tweet_clean$text <- removeNumbers(tweet_clean$text)

#stopwordsler
liste=c(stopwords("tr"),"şey","viral","yapılır","nedir","bonus","reklam","süper","bravo","vallahı",

#Gereksiz tekrarlar ve bağlaçların temizlenmesi
tweet_clean$text = removeWords(tweet_clean$text,liste)
```

6) EN ÇOK TEKRAR EDEN KELİMELERİ BULUT VE GRAFİK OLARAK GÖSTERME

Verilerimizi çektik, gereksiz olanları da temizledik, sırada elimizde olan verilerden bir kelime bulutu oluşturma var. Tüm işlemlerin sonucunda elimizde kalan kelimeler arasında en çok tekrar eden kelimeleri seçerek bunlardan ilk 50 tanesini kelime bulutu olarak işleyip karşımıza getiriyor.

```
# kelime bulutu oluşturma
head(tidy_tweets)

tidy_tweets %>%
  count(word, sort = TRUE) %>%
  filter(n > 100) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col() +
  xlab(NULL) +
  coord_flip() + theme_minimal() +
  ggtitle("Tweetlerde en çok kullanılan kelimeler")

library(wordcloud)
tidy_tweets %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 50))
```

7) ÇEKİLEN TWEETLERİ CSV FORMATI İLE KAYDETME

Verilerimizi çektik, gereksiz olan yerleri temizledik, kelime bulutu haline getirdik ve en son olarak bunların kayıtlarını tutabilmemiz için bir kayıt oluşturmamız gerekiyor. Bunu da R studio algılayabilmesi için “csv” formatında olması gerekiyor. Kayıt işlemini de aşağıda gördüğümüz gibi “write.csv” komutuyla yapıyoruz.

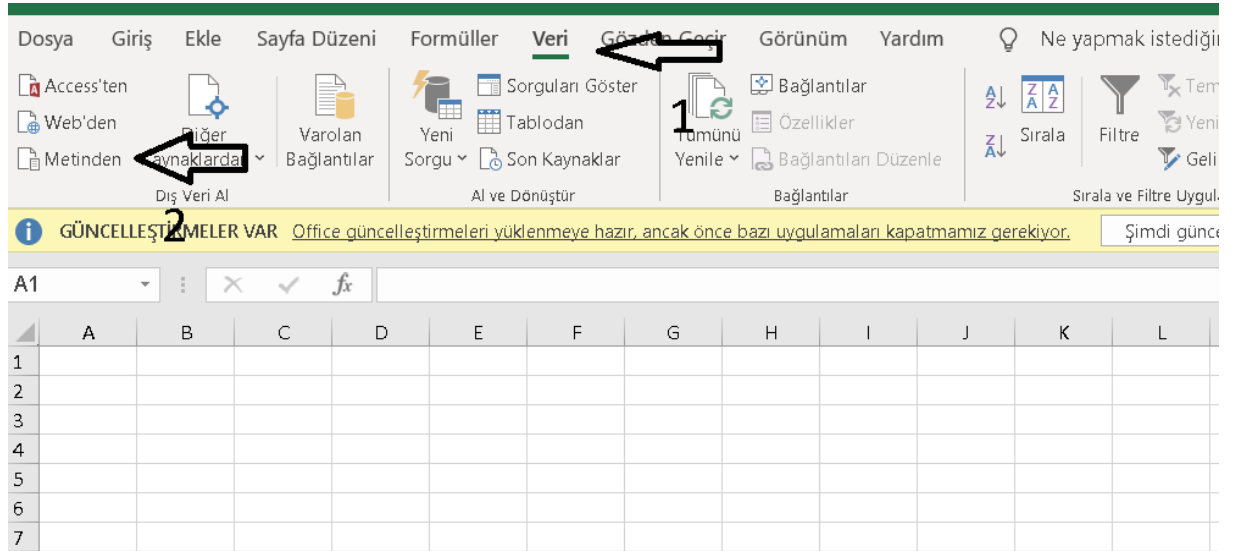
```
# Dosyayı kaydetme
```

```
write.csv(tweet_clean$text, file = "kriptohaber4.csv")
```

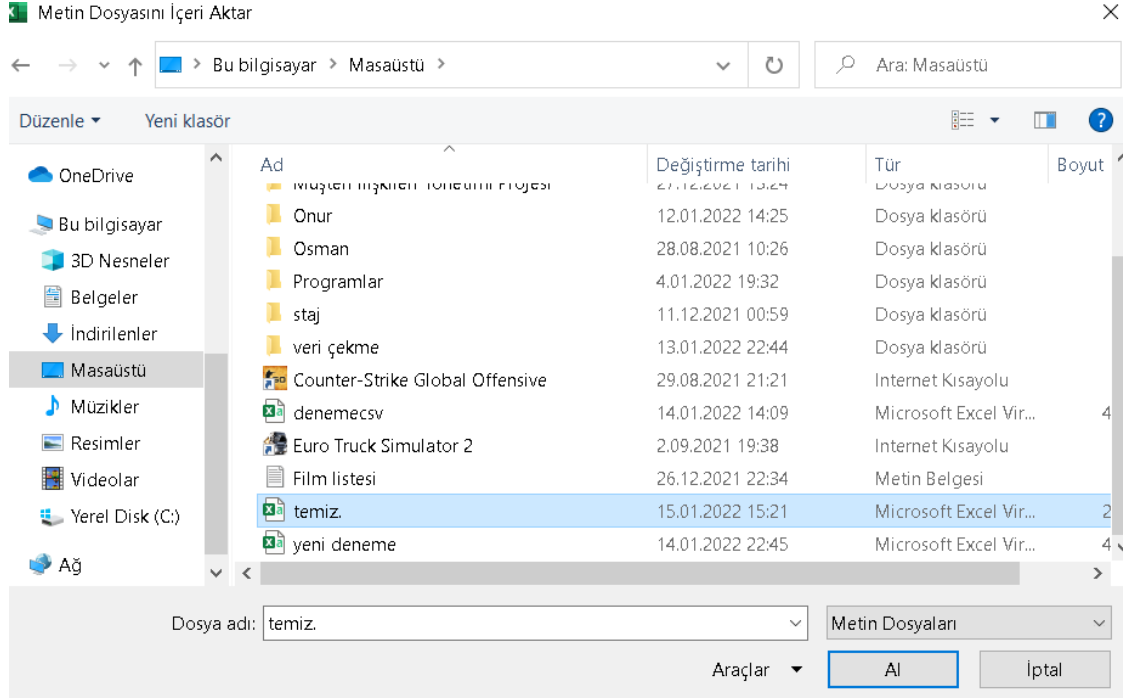
NOT: Buraya kadar yaptığımız işlemler şu an için pek anlam ifade etmiyor çünkü gündem her gün değişmekte bu yüzden de ilk çektiğimiz veriler ile son çektiğimiz veriler birbirinden farklı olacak çünkü her kelime bulutu o gün ile alakalı bilgi vereceğinden asıl kelime bulutu hepsini birleştirdiğimiz zaman ortaya çıkacaktı. Diğer bir yandan da gereksiz kelimeleri azalttığımız için kelime fazlalığından biraz da olsa kurtulmuş olduk.

8) KAYDEDİLEN CSV DOSYALARINI BİRLEŞTİRME

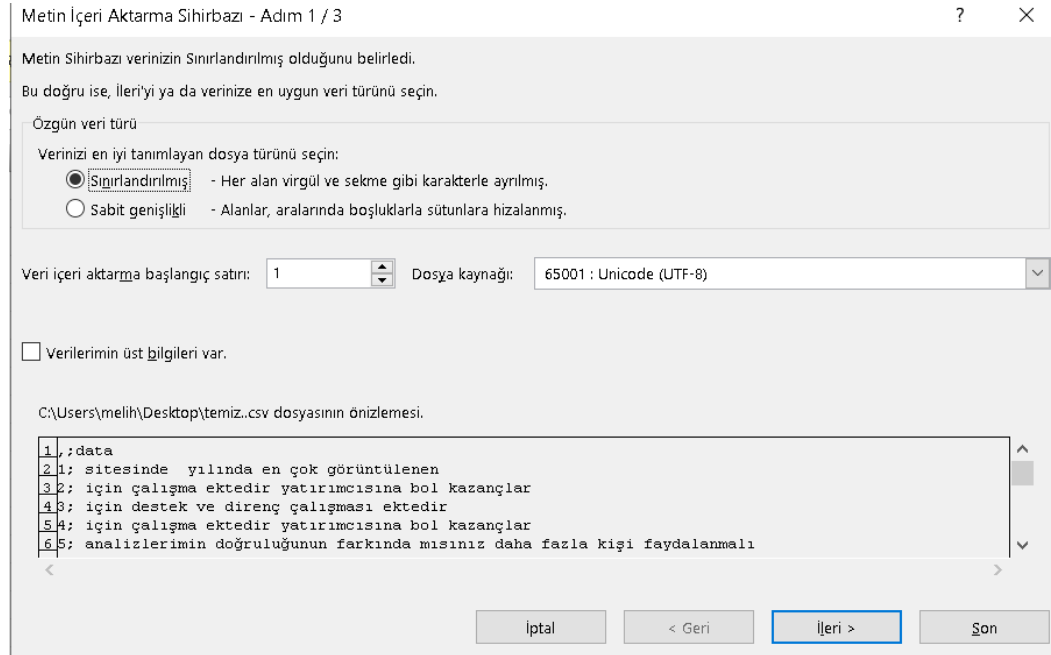
21 gün boyunca çektiğimiz verileri tek bir excell dosyasında birleştirdikten sonra, yeni bir excell dosyası açıp verileri r studio'da çalışabilmesi için bir takım değişiklik gerekiyor. Önce açtığımız excell dosyasına giriyoruz, ardından üst kısımda bulunan veri sekmesine basıyoruz ve sonra metinden seçeneğini seçiyoruz.



Şimdi ise çekeceğimiz veriyi seçiyoruz ve al seçeneğine basıyoruz.



Ardından ilk çıkan sayfada bir şey işaretlemeden ileri diyoruz.



İkinci sayfada diğer seçeneğine basıp yanındaki kutucuğa “işaretini” koyuyoruz bu ayrılacak olan yerleri “ işareti ile ayırmaya yarıyor.

Metin İçeri Aktarma Sihirbazı - Adım 2 / 3

Bu ekran verilerinizin içerdiği ayrıncıları ayarlamanıza olanak verir. Aşağıdaki önizlemede metninizin ne şekilde etkilendiğini görebilirsiniz.

Ayrıncılar

☒ Sekme
☐ Noktalı Virgül
☐ Virgül
☐ Boşluk
☒ Diğer: " "

☐ Ardışık ayrıncıları tek olarak işle

Metin niteleyicisi: " " v

Veri önizleme

```
;;data
1; sitesinde yılında en çok görüntülenen
2; için çalışma ektedir yatırımcısına bol kazançlar
3; için destek ve direnç çalışması ektedir
4; için çalışma ektedir yatırımcısına bol kazançlar
5; analizlerimin doğruluğunun farkında mısınız daha fazla kişi faydalanmalı
```

İptal < Geri İleri > Son

3. sayfada da bir değişiklik yapmadan son seçeneğine basıyoruz ve işlemimizi bitiriyoruz.

Metin İçeri Aktarma Sihirbazı - Adım 3 / 3

Bu ekran sütunları seçmenizi ve Veri Biçimini belirlemenizi sağlar.

Sütun veri biçimi

☒ Genel
☐ Metin
☐ Tarih: GAY v
☐ Sütunu içeri aktarma (atla)

'Genel', sayısal değerleri sayılara, tarih değerlerini tarihlere ve diğer tüm değerleri metne dönüştürür.

Gelişmiş...

Veri önizleme

```
Genel
;;data
1; sitesinde yılında en çok görüntülenen
2; için çalışma ektedir yatırımcısına bol kazançlar
3; için destek ve direnç çalışması ektedir
4; için çalışma ektedir yatırımcısına bol kazançlar
5; analizlerimin doğruluğunun farkında mısınız daha fazla kişi faydalanmalı
```

İptal < Geri İleri > Son

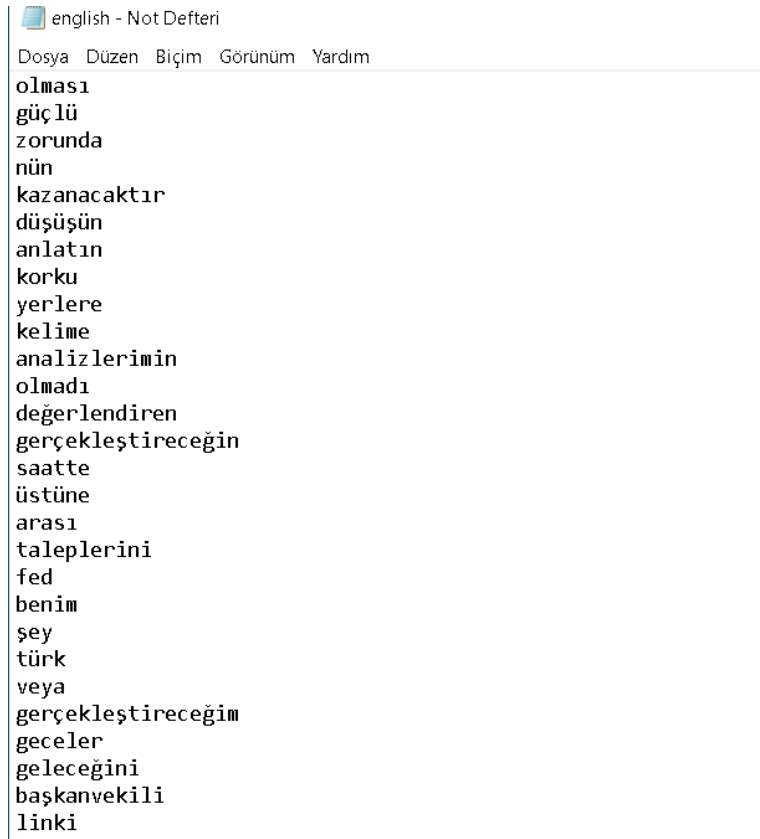
9) BİRLEŞTİRDİĞİMİZ VERİLERİ R STUDIO İLE ÇALIŞTIRMA, VERİYİ İNCELEYİP GEREKSİZ KELİMELERİ ÇIKARTMA

Tüm işlemleri hallettiğimize göre şimdi dosyamızı çağırmaya geldi. Aşağıda ilk satırda dosyamızı çağırdık, str() komutu ile dosyamızın özetini gördük.

```
deneme <- read.table(file.choose(),header=T, sep = ";", encoding= "UTF-8" )  
str(deneme)
```

Daha sonra ise verileri çekerken yaptığımız gibi stopwords ve gereksiz kelimeleri temizledik. Stopwords'lerin çok tekrar edenlerini yanına yazdım, çoğunluğunu "english.dat" dosyasına yazdım. Hala kalkmayanlar oldu bunları da excell dosyasından siliyoruz.

```
#Gereksiz kelimeler  
#stopwordsler  
liste=c(stopwords("en"),"çok","bonus","yatırım","arkadaşlar")  
  
#Gereksiz tekrarlar ve bağlaçların temizlenmesi  
deneme$data = removewords(deneme$data,liste)
```



Gereksiz kelimeleri de sildik ve ardından wordclud oluşturdum.

```
#kelime bulutu  
wordcloud(deneme$data,min.freq=4,scale=c(3,1,5),max.words=60)  
|
```

KELİME BULUTUM



KELİME BULUTUNDA ÇIKAN KELİMELERİN ANLAMLARI

Direnç: Fiyatlardaki yükselişin durmasının beklendiği seviye olarak açıklanabilir. Direnç seviyelerinde finansal ürünlerdeki alışın durması veya yavaşlaması, satıcıların bu seviyeden tepki vermeleri veya kar realizasyonu yapmaları olarak yorumlanır.

Coin: Kripto para veya coin, bir değişim aracı olarak çalışmak üzere, işlemleri güvence altına almak için matematiğin bir dalı olan kriptografi kullanılarak tasarlanmış bir dijital unsurdur.

Short: Short ya da diğer adıyla kısa pozisyon bir finansal varlığı fiyatının düşeceği beklentisi ile satışa çıkarmak anlamına gelir. Short işlemdeyse sahip olmadığınız bir varlığı ödünç (borç) alarak işlem yaparsınız. Kısaca özetlemek gerekirse short işlem bir varlığa sahip olmadan satabilmek demektir.

Metaverse: Metaverse kelimesi, "meta" (ötesinde anlam) ve "evren" önekinin bir portmantosudur; terim tipik olarak algılanan bir sanal evrene bağlı kalıcı, paylaşılan, 3D sanal alanlardan oluşan, gelecekteki internet yinelemesi kavramını tanımlamak için kullanılır.

Kripto: Türk Dil Kurumu kapsamında bakıldığında İngilizce olarak Kripto kelimesi 'gizli ve görünmez' biçiminde tabir edilmektedir.

Airdrop: Airdrop, bir kripto para belirteci veya paranın, genellikle ücretsiz olarak çok sayıda dijital cüzdana dağıtılmasıdır. Airdrop'lar öncelikle dikkat çekmenin ve yeni takipçiler kazanmanın bir yolu olarak uygulanmakta, bu da daha büyük bir kullanıcı tabanı ve daha geniş bir para kullanımı ile sonuçlanmaktadır.

Toparlanma: Kripto piyasasında belli olaylar sonrasında ya da belirli dönemlerde ayı piyasası denilen döneme giriyor ve bunun sonucunda piyasa değerleri aşağı düşmekte. Düşüş bittiğinde, piyasa yükselmeye başladığı döneme denir.

Kademeli: Kripto'da piyasa çok hızlı olduğu için alım satımlar bazen kademeli olarak alınır, satılır. Kademe ise belirli bir düzey belirlemek ve piyasa oraya geldiğinde oradan alım satım yaparak bu şekilde birkaç düzey belirleyip gerçekleştirmek olarak tanımlanabilir.

Ethereum: Ethereum ilk olarak Kuzey Amerika Bitcoin Konferansında, Ethereum kurucusu Vitalik Buterin tarafından tanıtımı yapılan bir sistemdir. Genel olarak bir alt coin statüsünde görülse de Ethereum esasında blockchain teknolojisinin gelişmesini ve daha fazla alanda kullanılmasını amaçlayan bir yenilikçi bir sistemdir.

Vadeli: Vadeli işlem, işlemde olan bir kripto para için önceden belirlenmiş gelecek fiyattan alım emri ya da satım emri verilerek işlem yapılmasına denir. Geleneksel vadeli işlemlerde, önceden belirlenmiş bir son kullanma tarihi ve ödeme tarihi bulunmaktadır.

Token: Token'lar, Blockchain ekosisteminde dijital olarak transfer edilebilen varlıklardır. Bir token; değeri, hizmeti veya bir ürünü temsil edebilir.

Binance: Binance, çeşitli kripto para birimlerinin ticareti için bir platform sağlayan bir kripto para borsasıdır. Ocak 2018 itibarıyla Binance, işlem hacmi açısından dünyanın en büyük kripto para borsası olmuştur.

Kapanış: Kripto paralarda alım satım işlemleri günün her saatinde gerçekleşse de trader'lar bir kripto para biriminin nasıl performans gösterdiğini ölçme şekli, günlük 24 saatlik bir zaman dilimidir ve buna "günlük kapanış" ismi verilmektedir.

Mexc: Meksika Borsası, Meksiko, Meksika'da bulunan bir borsadır. São Paulo merkezli BM&F Bovespa'dan sonra Latin Amerika'daki en büyük ikinci borsadır. Aynı zamanda Amerika'daki en büyük beşinci borsadır.

Avax: Avalanche (AVAX); merkeziyetsiz finans uygulamalarını, finansal varlıkları, alım satım ve diğer hizmetlerinin piyasaya sürülmesi içinde gereken altyapısı sağlayan bir platformdur. Her türden katılımcının, her türden değer varlığını oluşturmaya veya alım satım işlemleri yapmasına olanak sağlar.

Cüzdan: Kripto cüzdanları, özel anahtarlarınızı saklayıp kripto paranızı güvende tutar ve erişilebilir bir şekilde saklar. Ayrıca Bitcoin ve Ethereum gibi kripto para birimleri göndermenizi, almanızı ve harcamanızı sağlar.

Trade: Borsada yatırım araçlarını alıp satan kişilere trader denilmektedir. Aslında her alış satış işlemi trade işlemidir. Fakat büyük ve sürekli alış satış yapan yatırımcılara trader desek daha doğru olur.

Long: Genellikle uzun süreli yatırımları çağırıştırır. Bir kripto para biriminin fiyatının, faaliyette bulunduğunuz zaman dilimine bağlı olarak bir süreliğine yükselmek üzere olduğunu hissettiğinizde uzun süre devam etmek ilginizi çekebilir.

Cardano: Cardano, akıllı sözleşmeler için halka açık bir blockchain platformu çalıştırmayı amaçlayan bir kripto para birimi ağı ve açık kaynaklı bir projedir. Cardano'nun dahili kripto para birimine ADA denir.

Listeleme: Kripto piyasası sürekli gelişerek büyümekte, coinlerin bu piyasaya girebilmeleri için belirli kriterler var bunları karşılayanlar ise piyasaya girmeye hak kazanır ve borsa artık kendisinin satışa çıkaracağı coinleri liste halinde paylaşır, buna listeleme denir.

Shib: Shiba Inu, aynı zamanda Shiba Token olarak da bilinir, Ağustos 2020'de "Ryoshi" olarak bilinen anonim bir kişi tarafından oluşturulan merkezi olmayan bir kripto para birimidir. "DOGE Killer", olarak adlandırılan kripto para, Dogecoin'den sonra piyasaya sürülmüştür.

Blockchain: Blockchain yani blok zinciri teknolojisi de transfer ve ver güvenliğini sağlamak için kullanılan teknolojilerdir. Blok zinciri, geleneksel veri tabanları gibi değil de işlemlerin arka arkaya listelediği dijital bir kayıt defteri olarak çalışır. Veriler giriş tarihine göre genellikle arka arkaya sıralanır.

Startup: Amerika'daki silikon vadisinden çıkmış olan Startup, daha sonra tüm dünyaya yayılmış olan bir tür iş akımıdır. Bir problem ya da soruna yönelik çözümler üretmek suretiyle daha sonra bu çözümleri uygulayarak hızla gelişen şirketlerdir.

10) DUYGU ANALİZİ

Çektiğimiz tweetlerin üzerinde pozitif veya negatif olup olmadığını belirlemeye duygu analizi denir.

POLARİTE ANALİZİ

En basit duygu sözlüğünü oluşturan bu testte skorlar -1 ile 1 arasında değerler almaktadır. Kelime ve cümle bazında yapılan sınıflandırmalarda negatif ve pozitif olmak iki kategori bulunmaktadır.

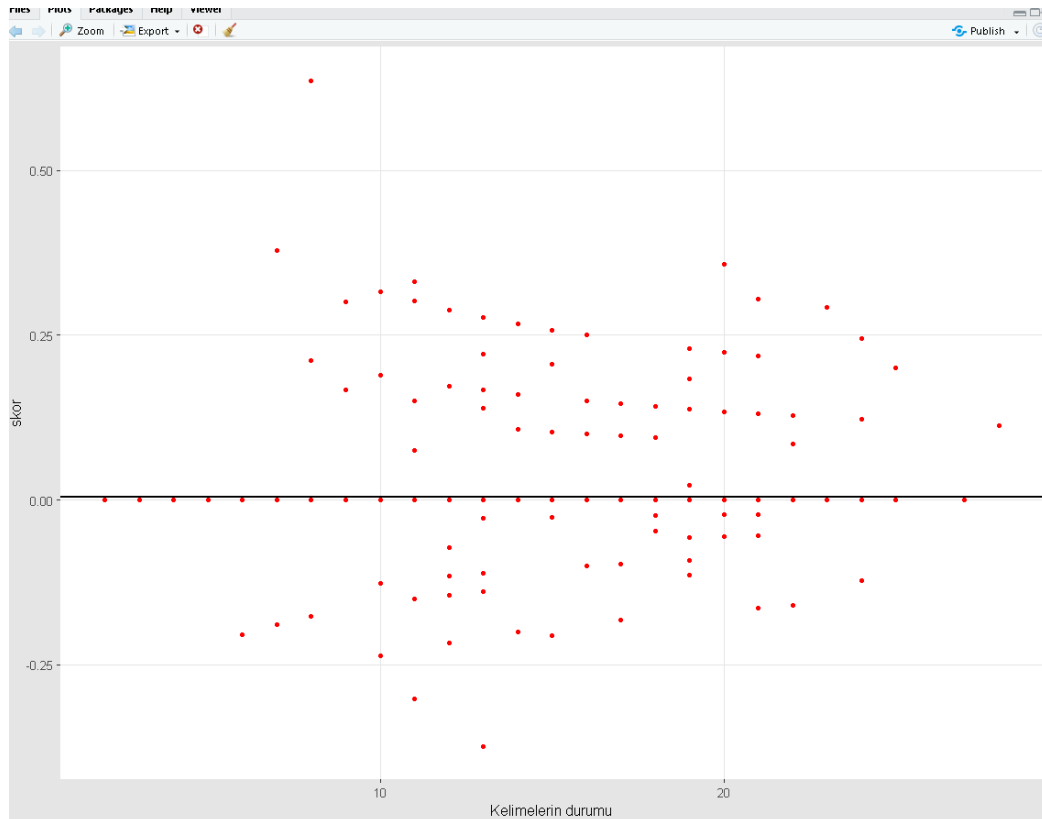
```
# Polarite
polarite<-spellcheck(deneme$data$word)

tablo<-cbind(deneme$data, polarite[,c(3,4)])

ggplot(tablo, aes(word_count, sentiment))+
  geom_point(color="red")+
  geom_hline(yintercept = mean(tablo$sentiment), color= "black", size=1)+
  labs(y ="skor", x = "kelimelerin durumu") +
  theme_igray()+
  labs()+
  theme(plot.caption = element_text(hjust = 0, face = "italic"))

stat.desc(polarite$sentiment, basic=T) %>% pander()
```

Kodları çalıştırdığımızda kelimeleri duygu durumuna göre pozitif veya negatiflik durumuna göre 1 ve -1 arasında konumlandırır.



Çıkan sonuç göre 0'ın çok az üzeri oldu. Yani yaptığımız analizde çekilen tweetlerde ki kelimelerin ağırlığı pozitif yönde oldu.

Analizim çıkan cevap;

mean	SE.mean	CI.mean.0.95	var	std.dev	coef.var
0.003567	0.0009339	0.001831	0.002263	0.04757	13.34

Polarite analizimin sayısal konumu 0,003567 noktasıdır.