**SENTIMENT ANALYS IS WITH TWITTER DATAS**

Melis Seren CEYLAN

Student

Eskibaglar Street Nural Buildings

No:40/10 Etimesgut/Ankara

+90 538 400 68 67

melisceylan49@gmail.com

Kaan BAYSAL

Student

4.cadde 124/2

Bahçelievler/Ankara

+90 537 689 04 08

kaan\_baysal@hotmail.com

Ibrahim Berat YAVAS

Student

Yazi Street Ugras Blocks 28/5

Sincan/Ankara

+90 533 682 54 79

i.beratyavas@gmail.com

Nurseli SAR

Student

Güzevler Yolaç Sokak No:5/12

Yenimahalle/Ankara

+90 553 222 31 78

nurselisarr@gmail.com

**ABSTRACT**

In this paper, we will introduce the project “Sentiment Analysis with Twitter Datas”. We will describe our steps, first datas, inputs and expected results.

**These are the programs, packages or systems that required for project build:**

* R
* Rstudio
* Twitter Dataset

More requirements could be added as the project progress. Project will be written in R language.

**Keywords**

User, tweets, input, output, expected result.

**INTRODUCTION**

We will start with introduction of the project then continue with the reasons for such a study, what can we expect as results, the benefits of results and where and how can we use them properly. Also this part will include the general understanding of projects construction steps.

Sentiment analysis is often used to help marketing and customer service teams to determine the costumer’s sentiments. Social Media is used for this. Whether the consumer is satisfied with a product reveals through these reviews. Based on this, we wanted to analyze the thoughts and feelings of our country for the general situation and news.

The project has a title as “.Sentiment Analysis with Twitter Datas” The title could be considered as a small summary of general concepts of project. In this project the data comes from Twitter. What is Twitter? Twitter was developed in American, which is an online news and social networking service. Users post and interact with messages known as "tweets". Tweets were originally restricted to 140 characters, but on November 7, 2017, this limit was doubled to 280. Users can post, like, and retweet tweets. Twitter, Inc. is based in San Francisco, California, and has more than 25 offices around the world. Twitter became one of the 10 most visited websites in 2013. As of 2018, Twitter had more than 321 million monthly active users.

There will be provided an input text which includes an twitter dataset. In the dataset there will be informations from users. In the project our first aim is to find the vital information in dataset. Then find all the necessary data for every tweet. These datas are user names, tweets, gender, birthday, hashtags. Our main focus from this point will be on users and their tweets. Our main goal from this point to perform sentiment analysis from these informations. In this project we will predict the users emotions while posting their tweets according to our datas like hashtags, words they used.

**DATA EXPLORATION**

Data analysis helps to achieve meaningful results by analyzing large volumes of data. This data is collected from a wide variety of sources, including social networks, videos, digital images, sensors and sales transaction records. The purpose of analyzing all this data is to uncover patterns and links that are not normally seen, so that they can provide valuable information about the users who created it.

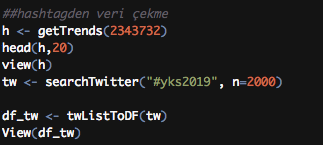
For our aim to analyse twitter users’ sentiments we reach the data from twitter. Using this dataset we can achieve hashtags, tweets which hashtags, user’s tweets, user’s personal informations and replies to other users. By using this data we can analyse sentiments of users’ for any hashtag and profile.

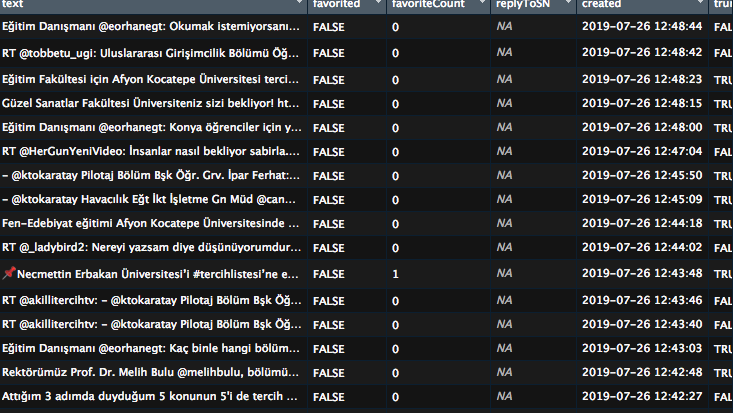
We can reach differents attribute which user’s tweet information category, they are;

1. text
2. favorited
3. favoriteCount
4. replyToSn
5. created(time)
6. truncated
7. replyToSID
8. id
9. replyToUID
10. statusSource
11. screenName
12. retweetCount
13. isRetweet
14. retweeted
15. langitude
16. latitude

We use some methods to reach the necessary data for our analyse.

* To access hashtags:

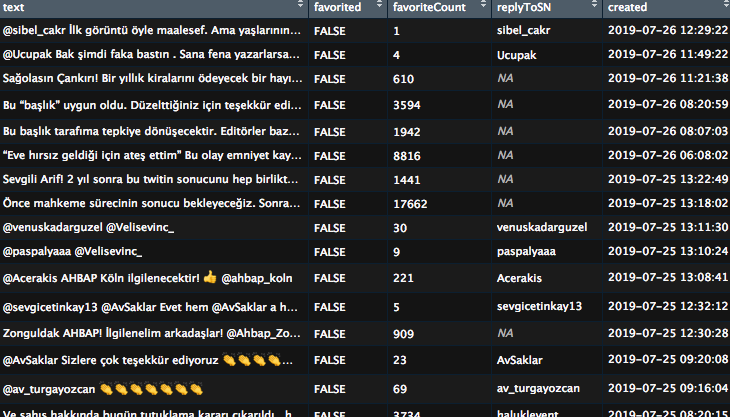




In first line, we get trend hashtags in ‘Ankara’ with getTrends() function. In that function we use Ankara’s ID ,which is ‘2343732’. Trend hashtags that we found are saved to “h” variable. With this trend hashtags, we can choose a hashtag, get tweets which include that hashtag. Then datas is transformed to list and save it to “df\_tw”. We can see datas in a folder by using view() method.

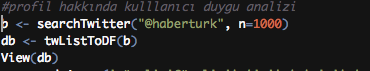
* To access user’s tweets:

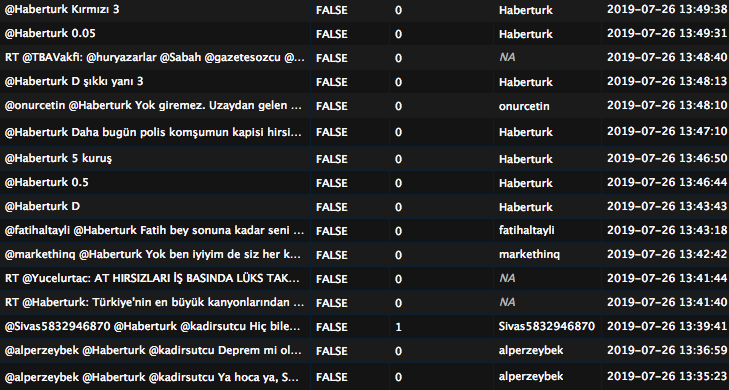




userTimeLine() function helps us to get users tweets, tweets’ favorite counts, retweet counts and users’ personal informations etc. In first image, we get Haluk Levent’s tweets and we save datas to a variable. Then we can see datas in a folder by using View() method.

* To access to replies and mentions to a profile:



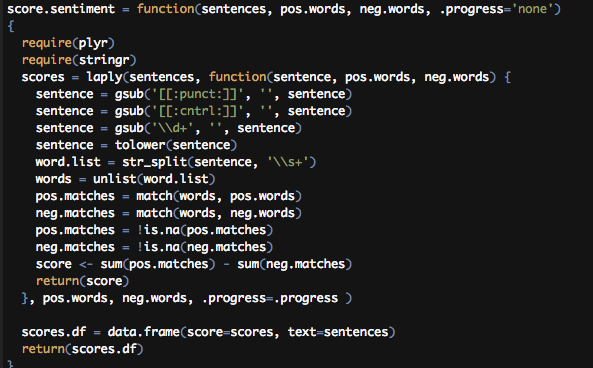


In this case, we can reach some tweets that written by some users about “Haberturk”. In that tweets, there are some positive and negative opinions.

**ANALYSIS METHODS**

After reaching the necessary data for our analyse. We use some methods to analyse sentiments of users’ who tweeted to hashtags and profiles. We also analyse a user’s emotions with the help of this datas and methods.

* For sentiment analysis:



Score.sentiment function is taken 4 variables. They are sentences, positive words, negative words and progress type. Sentences are tweets’ texts, positive and negative words are decided before starting sentiment analysis and sended to function. Then sentences are cleaned up by gsub() function. After then sentences are splitted into words and added to a list. That list can be used for compare words to the dictionaries of positive & negative terms. Lastly, scores are created.

* For sentiment analysis with a user’s tweets:



Positive and negative words can be seen at first and second line. User’s tweets are gotten and added into a variable. Sentiment analysis is done with tweets and positive - negative words.

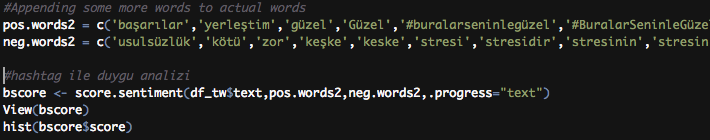
* For sentiment analysis with users’ replies to a profile:



Users on Twitter are able to comment on each other. We were able to evaluate the Tweets, that are tweeted to haberturk, according to the words we divided into two groups that we identified positively and negatively.

* For sentiment analysis with hashtag:





We put our result to bscore to show our result in histogram. We use positive-words and negative-words which help us to find the emotions in score.sentiment function.

* For day and time analysis of a user’s tweets:

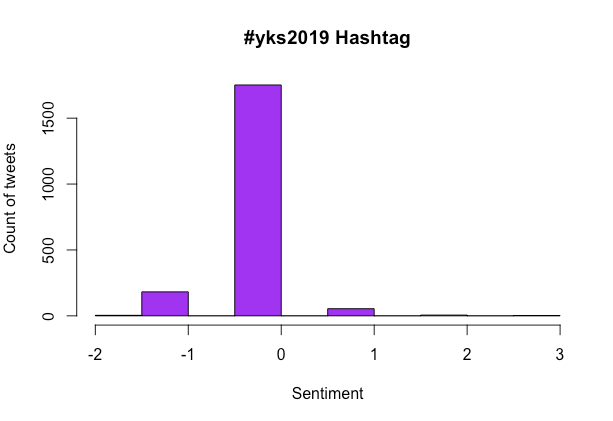


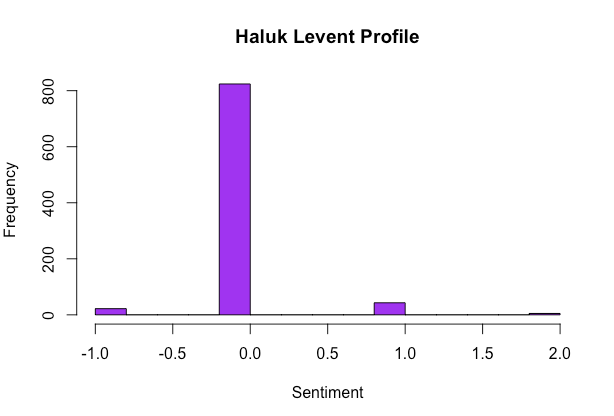
Usage time range can be analyzed also. In this case, one of a chosen user’s tweets can be taken and created a histogram. This histogram shows the day of tweets. Date and time can be analyzed by using tweets’ created time. In this way, we can analyze which days the user tweeted in general.

**RESULTS**

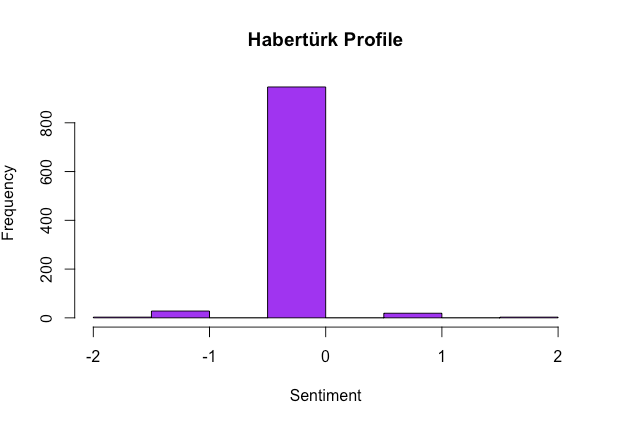
We showed the results of the analysis on the histogram.

In histograms, we can see the emotion analysis of tweets with hashtags thrown in. We can see the distribution of a person's tweets in hours. Then we can analyze the emotions of the comments made on a person's account. We can see the detailed information of the histograms. We have displayed on the bottom of the histograms.

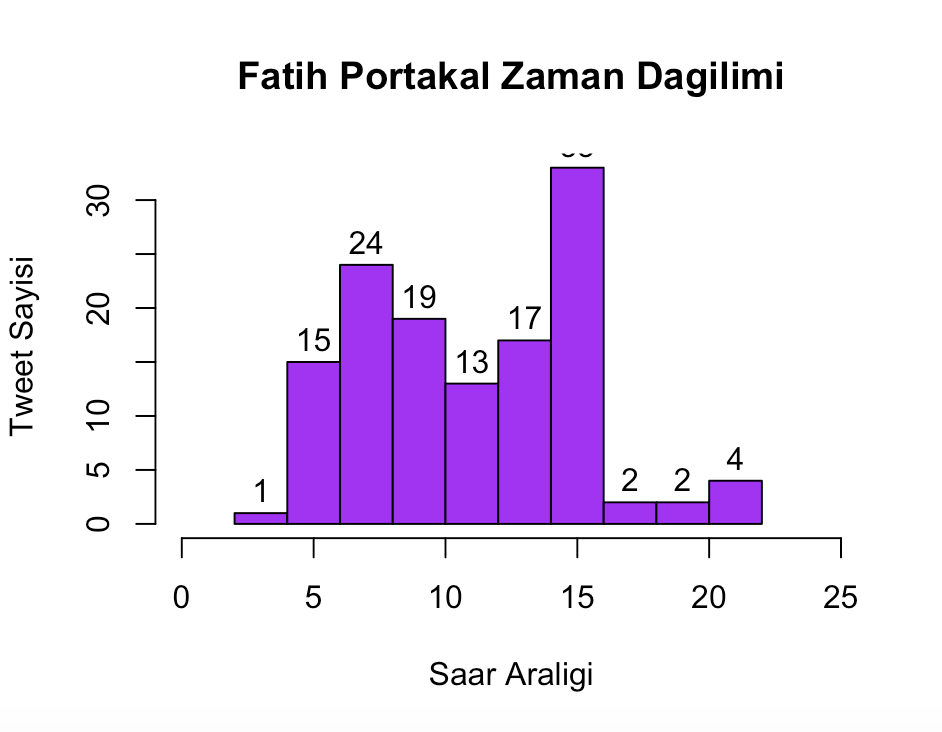
**Figure 1 :** In this histogram, tweets’ sentiments can be seen. 163 of tweets’ that included “#yks2019” have negative emotions. At the same time 54 of the tweets have positive emotions. Rest of the tweets , that are 1771 pieces, are neutral emotions. Rest of the 1771 tweets are retweets and some of them are about some news. Because of these, they are not voted.

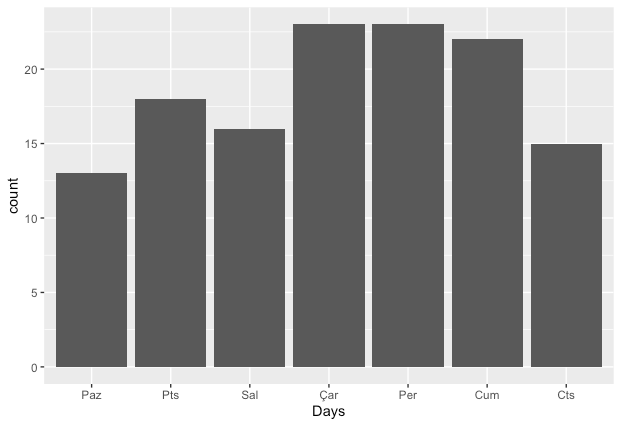


**Figure 2 :** In this histogram, we find the sentiments of Haluk Levent. 111 tweets of Haluk Levet’s tweets are positive. 27 of his tweets are negative and the rest are neutral.



**Figure 3 :** In this graphic, we can see the emotions of users’ who mentioned to Haberturk’s profile.

**Figure 4 :** In this histogram, the distribution of a person's tweets in hours can be seen.We reviewed 2000 tweets for this histogram.We divided 12 pieces a day. The values we see in the histogram are 16 times the twits scored at that time of day.



**Figure 5 :** In this graphic, we can see the distribution of Fatih Portakal’s tweet per day.

**CONCLUSION**

The use of social media is not only grow up with individual accounts. It was also spread with corporate social media accounts. In this way social media is not only used for entertainment. It also used for advertising, promotion and improve the quality of product service.

**REFERENCES**

1. <http://www.wikizero.biz/index.php?q=aHR0cHM6Ly90ci53aWtpcGVkaWEub3JnL3dpa2kvVHdpdHRlcg>
2. <https://www.google.com.tr/amp/s/www.r-bloggers.com/string-functions-in-r/amp/>
3. <https://www.dummies.com/programming/r/how-to-split-strings-in-r/>
4. <https://twitter.com/Haberturk>
5. <https://twitter.com/hashtag/yks2019?src=hashtag_click>
6. <https://twitter.com/haluklevent>
7. <https://www.udemy.com/veri-bilimi-ve-makine-ogrenmesi-egitimi/>