International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2494-9150 Vol-02, Issue 07, Oct 2016.

SENTIMENTAL ANALYSIS OF #BREXIT USING R-TOOL

[1]SHIV KUMAR GOEL, [2]SURAJ KUTTERI, [3]SACHITH POOJARY, [4]HITESH PATIL

[1]Deputy HOD, [2,3,4]Student, [1,2,3,4]MCA Dept, VESIT, Mumbai, Maharashtra, India.

[1]shivkumar.goel@ves.ac.in, [2]suraj.kutteri@ves.ac.in, [3]sachith.poojary@ves.ac.in, [4]hitesh.patil@ves.ac.in

Abstract - **World is changing rapidly and getting adapted to the changes. People opinion in all the things happening in the world around, do matter a lot as its the main source of knowledge that will provide abundant of information that can be made beneficial not only to that person but also to others if that source of opinion is being properly studied and utilized. People reaction the events, topics and happening of the world is completely different that is formed by their experience and effect. And now-a-days people have started expressing their personal views and emotions. People opinion can be known from newspapers, blogs and websites on recent happenings. These sources help us to collect massive amount of data and give an insight about Big-data and its applications in Indian-political scenarios, manufacturing of certain products, ordering of certain medicines depending on the season etc. Emotions expressed usually vary from one person to another. These emotions can correspond to a wide range of intensities that vary from very mild to strong. Analyzing the emotions require an adequate processing and understanding of these expressions.**

Keywords — Microblogging, Tweeter, GGPlots, Histogram, Pie-Chart, R-Programming, Word Cloud, Lexicon approach etc.

I. INTRODUCTION

Microblogging websites have evolved to become a source of abundant information. This is due to nature of microblogs on which people post real time messages about their opinions on a variety of topics, discuss current issues, complain, and express positive sentiment for products they use in daily life. In fact, companies manufacturing such products have started to poll these microblogs to get a sense of general sentiment for their product. This polling helps them to know about the popularity of the product, to gain information about the knowledge of pricing people's opinion. Many times these companies study user reactions and reply to users on microblogs. One challenge is to build technology to detect and summarize an overall sentiment.

Twitter is one of the "micro-blogging" social networking website that has a large and rapidly growing user base. The aim of this paper is to collect tweets using a Twitter API on keywords #BREXIT. This paper will determine the sentiment orientation of the tweets. Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. The analysis will determine the tweet status updates (which cannot exceed 140 characters) considering the sentiments of the people that reflects positive opinion, negative opinion or neutral opinion on the behalf of the person who tweeted. This paper carried out analysis using LEXICON Approach.

© 2016, IJREAM All Rights Reserved.

II. THE STUDY

Lexicon based approaches are quite popular in the sentiment analysis domain. These approaches involve tokenization of a particular corpus of text into unigrams, which are then assigned a polarity score. The aggregated sum of these scores determines the sentiment behind the text. It is generally classified as positive, negative or neutral depending on the calculated score. The flowchart describing a general lexicon based approach of Twitter data is shown Figure 1. Flowchart of a general lexicon based approach. Our study considers three lexicon based approaches relevant to our domain. They are explained in detail in the following sub-sections(Fig1).

III. METHODS

The Different phases involved in our approach was

1} Planning And Data-Collection.

2} Data Pre-Processing.

3}Data Loading

4}Data Probing and development of Dashboard.

5}Cloud Hosting.

6}Documentation.

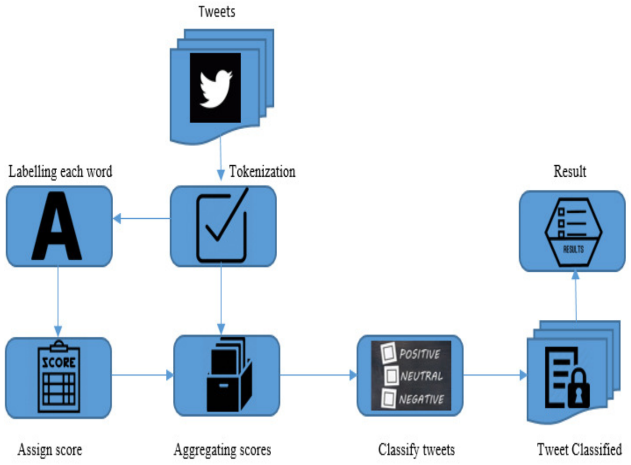


Figure 1 Flowchart of a general lexicon based approach

1. Planning And Data-Collection.

The data was collected and the total of 905 tweets related to #Brexit were collected. The data was gathered using Search API officially provided by Twitter was utilized. The Search API allows developers to look up tweets containing a specific word or a phrase. One of the constraints imposed by Twitter in the Search API was that it could produce only 1500 tweets at a time. The R programming language was used to carry out our sentiment analysis experiment. The "twitteR" package available for the R environment was used for extract the tweets from Twitter.

2. Data Pr-Processing Phase:

Data Transformation/Cleansing

Data Transformation / Cleansing is an important component of the data mining process. It involves recognition, removal of errors and inconsistency to improve the quality of the dataset prior to the process of analysis. The tweets were cleaned from irrelevant data to improve their quality. In our case, we observed that there were certain elements that did not provide any information and hence, had to be removed before processing. The elements were as follows:

1) Links: People generally have a tendency to attach documents (images, blogs, videos, web direction etc.) along with their tweets. These links or URLs had to be eliminated since they were of no use to our analysis.

2) Mentions: Mentions are used in Twitter to reply, acknowledge or start a conversation. Mentions are always written using "@" sign followed by the username. These mentions do not contain any relevant information thus, are removed.

3) Removing Punctuations and other miscellaneous data: Punctuations marks like quotes (""), commas (,) and semicolons (:) do not have any significant role in our analysis and hence, were removed from all the tweets present in our dataset.

3. Data Loading

The resultant .csv file is loaded into R-Studio which was further use for the analysis.

4. Data Probing and Development of Dashboard

a) Creating Bag of Words

The simplest and most widely used lexicon based approach is the baseline approach (also called "Bag of Words Approach"). In this method, there are two dictionaries – that of the positively tagged words and negatively tagged words. After tokenization, each individual word of the tweet is searched within those dictionaries, and depending upon the location of the word, it is assigned a polarity score.

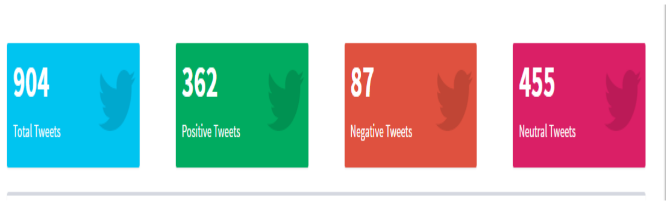

904 Total Tweets
362 Positive Tweets
87 Negative Tweets
455 Neutral Tweets

Figure 2

Consider a tweet from our dataset: “Great things can be accomplished with ease when you have the best team in the world”. At the end of data loading, the text ready for analysis is – “great things can be accomplished with ease when you have the best team in the world”.

Following the technique explained above, each of the following words- “great”, “accomplished”, “ease” and “best” are given a sentiment score of +1 since they are present in the positive words dictionary. On aggregation, the total polarity score of +4 is obtained, indicating that the sentiment behind the tweet is positive.

1) Scoring: If the individual token is found in the positive words dictionary, it is assigned a +1 polarity score value, if present in the negative words dictionary, a score of -1 and lastly, if not present in any of them, a score of 0 is assigned.

2) Aggregation: The total sum of the scores of each word present in the text is calculated and the on the basis of the final polarity value, the tweet can be categorized as positive, neutral or negative.

#TO SORT THE DATA AS PER SENTIMENTS

score.sentiment <- function(sentences, pos.words, neg.words, .progress='none')

require(plyr)

On the basis of above method the tweets on #Brexit was analysed and the result obtained is shown in FIG 2.

b) Creating Word Cloud to analyze the Tweets

A word cloud is a graphical representation of frequently used words in a collection of text files. The height of each word in this picture is an indication of frequency of occurrence of the word in the entire text.

So using this, the context in which the word was used and the frequency can be analyzed which will help to analyze the exact sentiment of the people (as shown in FIG 3,).

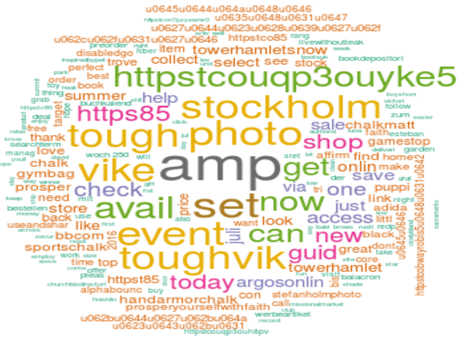


Figure 3

c) Creating Histogram/GGPlot/ Pie-Chart for the sentiments of the tweets

Histogram and GGPlot will help to make the comparative study of the sentiments on day to day basis. Pie-Chart provides us the exact percentage of people in or not in the favor of Brexit. So graphical representation in form of Histogram and Pie-Charts(as shown in FIG 4 to 7).

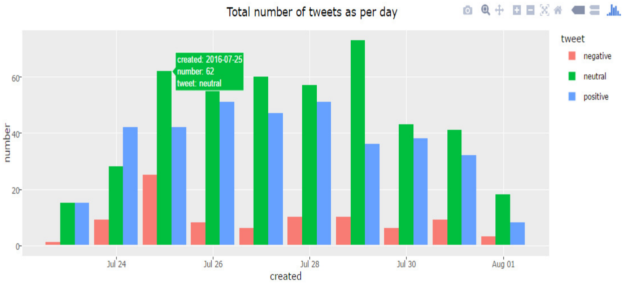


Fig. The histogram above showcases the distribution of tweets across Positive, Negative and Neural polarity per day.

Figure 4

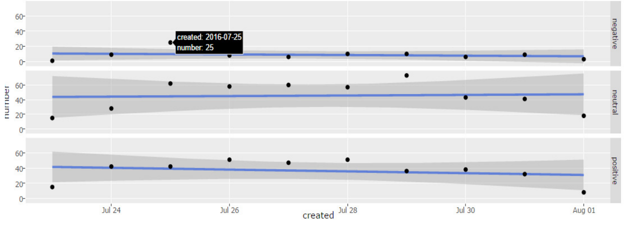


Fig. This plot shows the distribution of positive,negative and neutral sentiments as per day. For each tweet a net score of positive,negative and neural sentiments are computed and this plot shows the distribution of scores.

Figure 5

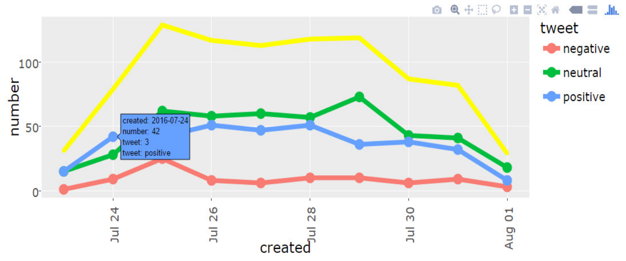


Fig. The plot above showcases the distribution of tweets across Positive, Negative and Neural polarity per day. The yellow line indicates summary statistics

Figure 6

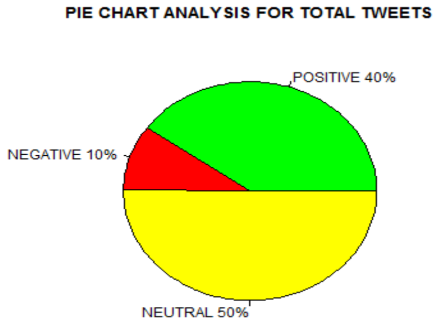
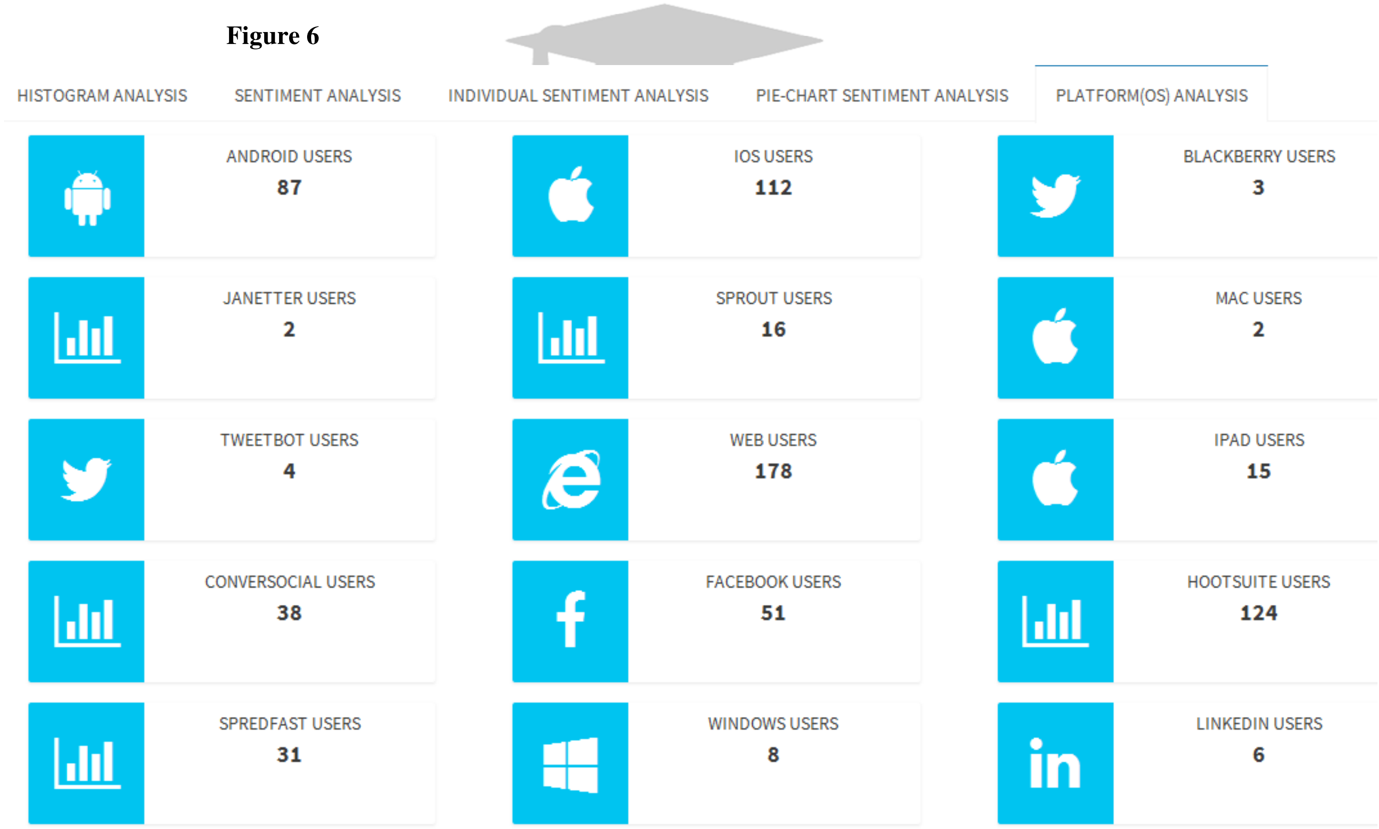


Figure 5



ANDROID USERS 87	IOS USERS 112	BLACKBERRY USERS 3
JANETTER USERS 2	SPROUT USERS 16	MAC USERS 2
TWEETBOT USERS 4	WEB USERS 178	IPAD USERS 15
CONVERSOCIAL USERS 38	FACEBOOK USERS 51	HOOTSUITE USERS 124
SPREDFAST USERS 31	WINDOWS USERS 8	LINKEDIN USERS 6

Figure 8

d) Creating Platform(OS) analysis

On the basis of the data collected we were able to analyze the platforms used and mostly preferred by the people. This will also help to analyze the people's preference over the mobile, tabs, PC's(as shown in FIG 8).

IV. CONCLUSION

So the above study shows how the opinion of the people based on topic is different from each other. It also shows how this opinion studies can be helpful for different studies. It will also help the government to understand about their decision implementation reactions of the folks, their progress, their advantages and disadvantages. It is also helping us to analyze the market about people's likes and dislikes, their preferences, their knowledge etc. Our main aim was to demonstrate how the R-Programming can be used to analyze the sentiment and how it would be fruitful in decision making.

REFERENCES

[1] Chapter 10: Text Mining, in book R and Data Mining: Examples and Case Studies http://www.rdatamining.com/docs/RDataMining.pdf

[2] R Reference Card for Data Mining http://www.rdatamining.com/docs/R-refcard-data-mining.pdf

[3] Free online courses and documents http://www.rdatamining.com/resources/

[4] RDataMining Group on LinkedIn (12,000+ members) http://group.rdatamining.com

[5] https://www.r-bloggers.com/sentiment-analysis-with-machine-learning-in-r/

[6] https://www.r-bloggers.com/sentiment-analysis-on-donald-trump-using-r-and-tableau/

[7] "Data Analysis on Current Affairs Using R" by Manjunath Mulimani, Nireeksha V Shetty, Nikshitha Shetty, Renita Maria Lolo, Rajashree L Dept. Of Computer Science, Sahyadri College of Engineering & Management, Karnataka, India.