Twitter Sentiment Analysis on Indian Government Project using R.

Pankaj Verma, Akib Mohi-Ud-Din Khanday, Syed Tanzeel Rabani, Mahmood Hussain Mir
Sanjay Jamwal

Abstract: The main motive behind this research paper is to use the power of social media to observe, examine and analyze the opinion regarding recent Indian government project as the opinion of people plays a vital role in formulating the government policies. By getting into the deeper insights on social media, one can easily analyze the behavior of people regarding various issues and policies, which was otherwise impossible using traditional sources. The case study was done on Statue of Unity. Analysis was done on one of the famous social networking sites i.e. Twitter, using R programming language. Twitter API was used to collect the primary data. Tweets were analyzed by using opinion lexicon and Emotion lexicon-based approaches. Opinion Lexicon based approach categorized the sentiment of tweets in three categories, while Emotion Lexicon based approach refined them into eight more categories. The research work done in this paper will help government to understand the emotions of people regarding their policies and will also enrich people to help them understand majority vote of people.

Keywords: Sentiment Analysis, Statue of Unity, Twitter API, Twitter data.

I. INTRODUCTION

With increase in social awareness; popularity of social networking such as twitter is increased. Twitter is mostly used for sharing information and opinions regarding the trending events [1]. Twitter is a ninth largest social network in the world and now it has 328 million of active users per month and millions of tweets are posted per day. Twitter is mostly used by celebrities, politicians, film industry personalities, sports persons. Users can tweet any information, whether text, audio or video, on which other users can react and post there opinion[2].

Today feedback requirement increases about the new product, about the government projects and schemes, executions and international talks so that with the help of these feedback companies, government and other organization can retrieve people opinion related to their products and schemes. The tweets and retweets are the best platforms for feedback and opinion retrieval system in the social network. The polarity of the tweets such that the tweets are positive or negative is quickly identified using sentiment analysis with the help of twitter data [2]. Various government schemes like Digital India, Make in India, GST, and Demonetization etc. are being Analyzed using various approaches. In this paper, a Lexicon based approach is used for analyzing the sentiments of people regarding Statue of Unity using R programming language. RStudio is used to carried out this proposed work.

1.1 Sentiment Analysis

Sentiment analysis is "the computational study of people's opinions, appraisals, and emotions towards entities, events and their attributes". Opinion is important because whenever we need to make a decision – we listen to other's opinions [3]. Sentiment analysis can be done using various subtasks like , subjectivity analysis, sentiment mining, review mining and affect analysis etc.[4].

1.1.1 Levels of sentiment analysis

II. Document-level Sentiment Analysis

In this document level sentiment analysis whole document will be analyzed and classified whether the whole opinion of documents expresses a positive or negative sentiment.

III. Sentence level Sentiment Analysis

In sentence-level sentiment analysis, we determine the polarity of the sentence, whether it is Positive, and Negative or Neutral opinion. This level attentively related to Subjectivity Classification [5], which distinguishes objective sentence and subjective sentences.

IV. Aspect Level of Sentiment Analysis

It is also known as Feature level sentiment analysis [5]. Aspect level sentiment analysis achieve fine-grained analysis. In this level, we directly look at opinion itself instead of looking to document, paragraph, sentence or phrase. This task is more interesting and more difficult too.

II. METHODOLOGY

Our methodology consists of 6 steps from A to F. Step A discusses about the extraction of tweets. Step B is about the data pre-processing to remove the noise. Step C analyses the data using lexicon-based approach. Step D classifies the tweets into different categories. Step E calculates the sentiment score and the type of polarity. The step F is used for visualization of our result.

* Correspondence Author

Pankaj Verma, MPhil Scholar, Department of Computer Sciences, BGSBU Rajouri, J&K, India. pankajvermajmu@gmail.com

Akib Mohi-Ud-Din Khanday*, Ph.D. Scholar, Department of Computer Sciences, BGSBU Rajouri, J&K, India. Email: akibkhanday@gmail.com

Syed Tanzeel Rabani. Ph.D. Scholar, Department of Computer Sciences, BGSBU Rajouri, J&K. Email:syedtanzeel@bgsbu.ac.in.

Mahmood Hussain Mir, Ph.D. scholar, department of Computer Sciences, BGSBU Rajouri, J&K. Email:mirhussain20111@gmail.com

Dr. Sanjay Jamwal, Professor, Department of Computer Sciences, BGSBU Rajouri, J&K, India. Email: sanjayjamwal@bgsbu.ac.in

A. Steps used for extracting tweets

i.   Create twitter application.
ii.   Twitter packages act as an interface to the Twitter API.
iii. For authentication ROAuth package is used.
iv.   Authentication IDs like consumer key, consumer key secret, access token, access token secret is generated.
v.   During authentication, redirection to a URL is done automatically by clicking on the authorized app, and verified using OTP.

B. Data Pre-processing

The data we extracted is raw tweets. It contains unnecessary data that is not required for analysis. Now before sentiment analysis, we need to be remove these unnecessary data from extracted twitter dataset such as HTML links, emoticons, punctuations '@', stopwords(i.e. is, at the, on), RT, numbers, white spaces are removed, stemDocument(i.e. "running" will be changed to its root form run), convert tweets to lower case such that the resultant data set hold only valuable information for the analysis[8].

C. Lexical Analysis

The Lexical analysis is carried out with the help of lexicon-based approach. In this analysis we used opinion lexicon dictionary created by Hui and Lui in 2006. This database contains 2006 positive and 4783 negative opinion words. This database is loaded in our working directory of RStudio and each word in the tweets are compared with the words in this database that we have loaded and then sentiment is predicted by using this [4].

D. Classification

In building a classifier model, the pre-processed tweets were scored and classified by polarity (positive or negative) and categorized into eight different types of

emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust). To accomplish this, we used the R package titled "Syuzhet" and from this package we used NRC emotion-based dictionary. Hui and Lui database of sentiment words is used to score the tweets for detecting polarity.

E. Calculating sentiment score

Hui Lui lexicon is used for calculating the sentiment score of every tweet with the help of scoring function.

$$S = \Sigma P - \Sigma N$$

Where S = Sentiment Score.
    P = Positive Words.
    N = Negative Words.
    a) Polarity types

(i) *Positive polarity*-When the number of positive words is greater than negative words.
(ii) *Negative polarity*- When number of negative words are more than positive words.
(iii) *Neutral polarity*- When number of positive and negative words are equal or if there is no existence of any opinion words.

F. Visualization

Sentiment analysis is graphically represented by using "RColorBrewer", and "ggplot2" in RStudio. In our work bar charts are used for representing outcomes of the sentiment analysis.

III.   STATUE OF UNITY: A CASE STUDY.

Statue of Unity is the world's tallest statue built in the memory of great Indian activist Sardar Vallabhbhai Patel (1875-1950) who played a vital role in the independence movement of India and was its first Home minister. This world's tallest statue stands at the height of 182 meters in the state of Gujrat. The worth of Rs.2,989 crore was used to make this Statue. There is a vast number of opinions shared by people regarding making of Statue of Unity on various social networking sites. In order to observe the opinion about the Statue of Unity, Twitter information has been collected and has been analyzed to see the reaction of people regarding it.

A. Collecting Statue of Unity related Tweets

For extracting tweets from twitter firstly authentication is being done using Twitter API application. After creating the twitter application, we got access to Twitter APIs credentials for Twitter Authentication server.
Consumer_Key<-"yyyyyyyyyyyyyyyyyyyyyyyyy".
Consumer_Secret_Key<-"yyyyyyyyyyyyyyyyyyyyyyyyy"
Access_Token_Key<-"yyyyyyyyyyyyyyyyyyyyyyy".
Access_Secret_Key<-"yyyyyyyyyyyyyyyyyyyyyyy".
setup_twitter_oauth(consumer_key = consumerKey , consumer_secret = consumerSecret , access_token = accesstoken , access_secret = accesssecret).

B. Access twitter data sets.

After authentication with Twitter API, we extracted tweets related to Statue of Unity. We used searchTwitter () function to access the data. In this work we first extracted 11000 tweets to classify them on the basis of polarity classification. Then in second shift, we extracted 19000 tweets and classified them on the basis of emotion-based sentiment analysis technique.
SOU_tweets = searchTwitter("statue of unity", n=19000, lang="en").
SOU_tweets = searchTwitter("statue of unity", n=19000, since = "2017-01-01", lang="en").

C. Classification of Tweets.

1. Classification by polarity.

On the refined dataset polarity operation is being applied with bigram features. The polarity function is used for generating the sentiment scores for each tweet. The Sentiment Score of each tweet can be positive or negative on the basis of opinions of public [2]. These are represented by bar chart in figure 1.
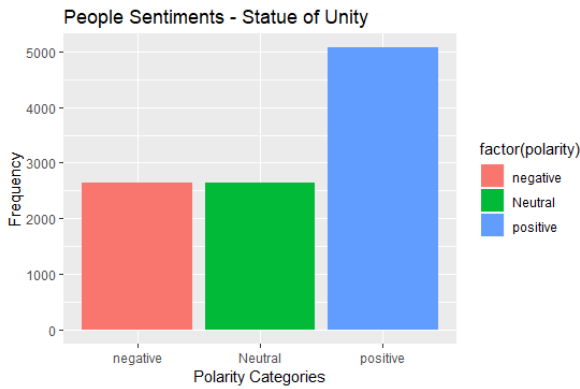
Fig. 1. Classification by polarity-based approach using sentiment score.

1.1 Classification by sentiment score.

The SentiStrength Scale is used for measuring the polarity(Sentiment) score for positive as well as negative tweets in which +1 means weak Positive(-1 for weak Negative) and +5 means Strong Positive(-5 means Strong Negative)[2]. The sentiment score is a more specific numerical illustration of the Opinions of people. Figure 2 shows that the sentiment score on Statue of Unity tweets. Based on this score and data extracted, Table 1. below summarizes the polarity and the corresponding percentage.
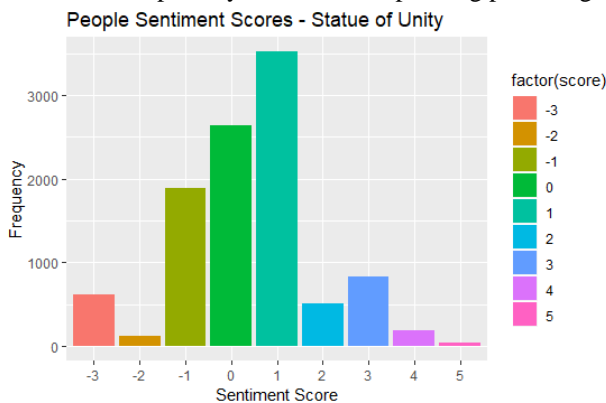

Fig.2 Sentiment score on Statue of Unity Tweet

Table 1. Polarity count on the basis of sentiment score

Lexicon Based	Positive	Negative	Neutral
Count	5490	2750	2750
Percentage	49.90	25.04	25.04

1.2 Classification by Emotions

Classification by emotion of pre-processed tweets is done by using NRC emotion-based dictionary from package named "Syuzhet" in R library. In this approach the presence of eight different emotion and their corresponding valence (positive or negative) are calculated in pre-processed data set [7]. The Table 2. below Summarizes the classification into 8 emotions.
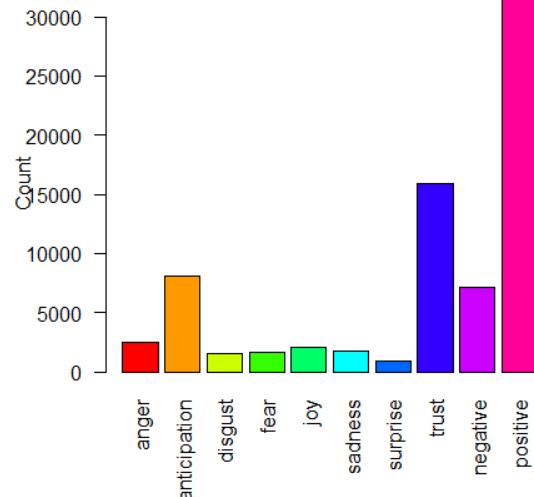

Fig 3. Classification by emotion on Statue of Unity tweets.

Table 2: Sentiment score for each emotion.

Emotion-Based	Count	Percentage
Anger	3117	3.17%
Anticipation	9409	9.58%
Disgust	1872	1.90%
Fear	1948	1.98%
Joy	2397	2.44%
Sadness	2326	2.36%
Surprise	1086	1.10%
Trust	21216	21.61%

1.3 Comparative Analysis

On Comparing the Polarity and Emotion based approaches, it was analyzed that former provides the general sentiment while as later provides the in-depth analysis by refining the polarity into eight different emotional categories. The results showed that the Polarity based method showed 49.90% of tweets as positive, 25.04% as negative and 25.04% as neutral while as Emotion based method refined the sentiment of analyzed tweets into Eight categories having total positive count of 44.75% and total negative count of 11.06%. It is concluded from our analysis that majority of people support the government initiative in making the "Statue of Unity". Table 3. below summarizes the comparison between these two discussed approaches.

Table3: Comparative Analysis of Polarity and Emotion Based in terms of Percentage (%).

Methods	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Negative	Positive	Neutral
Polarity based	NA	NA	NA	NA	NA	NA	NA	NA	25	50	25
Emotion Based	3.17	9.58	1.90	1.98	2.44	2.36	1.10	21.61	11.06	44.75	NA

Table title above the table: *Polarity and Emotion Analysis of Tweets in Percentage*

IV. CONCLUSION AND FUTURE WORK

This research work analyses the emotion of people regarding the government project "Statue of Unity" using social media analytics and various sentiment analysis techniques. Analysis was done using Polarity and Emotion Based approaches. In Polarity based approach we calculated sentiment scores and counted a number of positive, negative and neutral tweets about this project. Emotion based approach further refines the already analyzed polarity. On comparing these two methods, Emotion based method reflects in detail the emotion of people about this project. The results showed that most of the people appreciate and support this project. This type of sentiment analysis can be performed for evaluating other government schemes and projects and monitoring the growth of the schemes from people's perspective. This work can be further extended to analyze the large amount of data regarding various other government schemes. This will help govt. agencies to understand opinion of people regarding their polices and will also help people to understand the majority vote.

REFERENCES

1. K. Arun, A. Srinagesh, and M. Ramesh, "Twitter Sentiment Analysis on Demonetization an Initiative Government of India," Int. J. Recent Trends Eng. Res., vol. 4, no. 6, pp. 252–258, 2017.
2. D. Suganthi and A. Geetha, "Twitter Sentiment Analysis on GST tweets using R tool," Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol., vol. 2, no. 5, pp. 793–796, 2017.
3. M. Edison and A. Aloysius, "Concepts and Methods of Sentiment Analysis," pp. 16288–16296, 2016.
4. R. Wagh and P. Punde, "Survey on Sentiment Analysis using Twitter Dataset," Proc. 2nd Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2018, no. Iceca, pp. 208–211, 2018.
5. B. Naiknaware, S. Kawathekar, and S. Deshmukh, "Sentiment Analysis of Indian Government schemes using Twitter datasets," Int. Conf. Recent Adv. Comput. Sci. Eng. Technol. IOSR J. Comput. Eng., pp. 70–78.
6. S. Singh and S. S. Choudhary, "Social Media Data Analysis:Twitter Sentimental Analysis Using R Language," Int. J. Adv. Electron. Comput. Sci., vol. 4, no. 11, pp. 13–17, 2017.
7. K. Ganesan and A. P. Patil, "Predictive Analysis of Tweets on Goods and Services Tax(GST) in India using Machine Learning," Int. J. Eng. Technol. Manag. Appl. Sci. www.ijetmas.com, vol. 5, no. 8, pp. 2349–4476, 2017.
8. A. Rashid, N. Anwer, M. Iqbal, and M. Sher, "A Survey Paper : Areas , Techniques and Challenges of Opinion Mining," vol. 10, no. 6, pp. 18–31, 2013.
9. M. Govindarajan and M. Romina, "A Survey of Classification Methods and Applications for Sentiment Analysis," pp. 11–15, 2013.
10. K. Huda, "Classification Technique for Sentiment Analysis of Twitter Data," vol. 8, no. 5, pp. 2551–2555, 2017.
11. Syed Tanzeel, and Qamar Rayees "Quantifying Suicidal Ideation on Social media" [Unpublished]

AUTHORS PROFILE

Pankaj Verma: The Author has done Bachelors in Science snd MCA from Jammu University. He is currently pursing M. Phil in computer science from BGSBU Rajouri. His research Interests are about Machine Learning, Data analysis

Akib Mohi-Ud-Din Khanday: The Author has done his Bachelors and Masters in University of Kashmir and IUST Awantipora respectively. He is currently pursuing Ph.D. in Computer Sciences from BGSBU Rajouri. His research interests are Machine learning and Data Analytics.

Syed Tanzeel Raban: The Author has done BCA from University of Kashmir, MCA from BGSBU Rajouri and M.Phil. in Computer Sciences from Bharathidasan University, Tamil Nadu. He is Currently Pursuing Ph.D. in Computer Sciences from BGSBU Rajouri. His Research interests are Social Network analysis, Machine learning, Deep Learning, Software Metrics

Mahmood Hussain Mir: The Author has done BCA from Kashmir University, MCA from BGSBU Rajouri, M.Phil. from Bharathidasan University. He is Currently Pursuing Ph.D. from BGSBU Rajouri. His research interests are IOT, Machine learning and data Analytics

Dr. Sanjay Jamwal:The Author is Currently Sr. Assistant Professor in the department of Computer Sciences BGSBU Rajouri. He has done MCA from University of Jammu, M. Phil from Annamalai University Tamil Nadu and Ph. D from University of Jammu. He has almost 20 years off experience in teaching and Research. His research Interests are Machine Learning, Data Analytics and Networking.