Sentiment Analysis of Twitter Data Using Statistical Analysing Tool R Studio

** Radhika Goyal*
*** Meenal Kakkar*

Abstract

Today, social media is considered as the best platform for the exchange of data and views on a large scale. With increase in number of users, the data generated on daily basis is increasing rapidly. This data is useful in analyzing people's opinions about the ongoing trends and activities occurring across the world. The data generated on these platforms is analyzed and this is known as social media data analysis. This type of data analysis is also known as opinion mining where the views and thoughts of people are extracted and processed, which then helps in knowing the current mindset of people regarding various activities. To accomplish the need of knowing people's mindset, the popular social media site Twitter can be used.

Twitter is an online micro - blogging service that facilitates delivery and interpretation of 140 posts known as "tweets". It is one of the platforms for gathering large and diverse data, and from this gathered data it is easy to analyze a person's sentiments. This is known as sentiment analysis.

Sentiment analysis uses the idea of data analysis to extract emotions, comments, reviews, etc. from different social media platforms. The paper contains tweets that are extracted using the hashtag IndiaStrikeBack (#IndiaStrikeBack) for the duration of ten days when people were emotional about the response given back by the Indian armed forces against the attack. The paper then analyses the thoughts and feelings of various users of Twitter behind their tweets about the action taken by the Indian armed forces in response to the Pulwama attack. For doing the analysis of the extracted tweets, a robust tool R was used.

R is a statistical tool that is user-friendly and easy to use. It is used for performing data analysis of the data that has been extracted from social media platform. With the help of packages such as Twitter, the extraction of tweets from Twitter and doing its analysis becomes easier. The extracted data is stored in Comma Separated Values (csv) files and is further represented in the form of bar charts. The opinions of the people are categorized into two different sentiments, namely, (positive and negative). There is a classification of users' emotion into 8 categories, namely, anger, anticipation, disgust, fear, joy, sadness, surprise, and trust that are taken from NRC sentiment dictionary. The feelings transmitted by sarcasm and irony are not very well served by the instrument for analyzing feelings.

Keywords : #IndiaStrikeBack, sentiment analysis, social media, Twitter

I. INTRODUCTION

With the growing advent of technology, there is major growth in social networking which has connected people around the globe over a network. This growth has led to a change in the way many people express their views, feelings and opinions, as well as the platforms they use.

Social networking helped people in sharing their views, opinions and thoughts resulting in generation of huge amount of data that has further been used for analysts for determining the current mindsets of people around the world. These networking sites provide the analysts large amount of data that needs to be analyzed and meaningful information can be extracted.

Manuscript Received: July 24, 2019; Revised: August 17, 2019; Accepted: August 20, 2019. Date of Publication: October 5, 2019.
* R. Goyal is a student with MCA Department, International School of Informatics and Management, Jaipur, Rajasthan, India – 302 001. (email:radka16398@gmail.com)
** M.Kakkar is Assistant Professor, MCA Department, International School of Informatics and Management Jaipur, Rajasthan, India – 302 001.(email : meenal.kakkar@icfia.org)
DOI : 10.17010/ijcs/2019/v4/i5/149458

One such social media platform is Twitter which is considered one of the most popular social networking services that facilitate the delivery and interpretation of small posts known as "tweets". The tweets are a way through which people express their emotions, views, and opinions which are used by these companies and organizations to analyze their product or service quality and work on their product/service accordingly. Tweets are considered as a type of data, and with growing number of tweets, the rate of growth of data is very high due to which data analytics has become a buzzword. One of its applications is sentiment analysis, which is a research area. Sentiment analysis uses the idea of data analysis to extract emotions from different social platforms, comments, reviews, etc.

II. SENTIMENT ANALYSIS

Analyzing emotions or opinions from statements or sentences and classifying them on the basis of sentiments they convey using text is known as sentiment analysis.

Sentiment analysis provides a way through which unstructured information is transformed into structured information having views of the public.

The sentiments can be categorized on the basis of the classes, where classes can be classified into two class classifications (positive, negative) or three class classifications (positive, neutral, negative).

Different approaches such as the lexicon approach and machine learning approach have been been used. The lexicon approach uses the scoring technique to evaluate sentiments, whereas the machine learning approach uses some models on the training data set.

For using the lexicon approach, different lexicon analyzers are available out of which NRC is the most commonly used. The NRC emotion lexicon analyzer provides the classification of user's sentiment into eight emotions and two sentiments.

III. METHODOLOGY

For performing the sentiment analysis, the following steps are performed, that is, from the extraction of tweets from Twitter to pre-processing of tweets forthe analysis and generating the graphical representation.

A. Twitter Application Creation

Extraction of tweets is only possible after the creation of Twitter's developer account that allows the permission of tweet extraction by providing four security keys to the user.

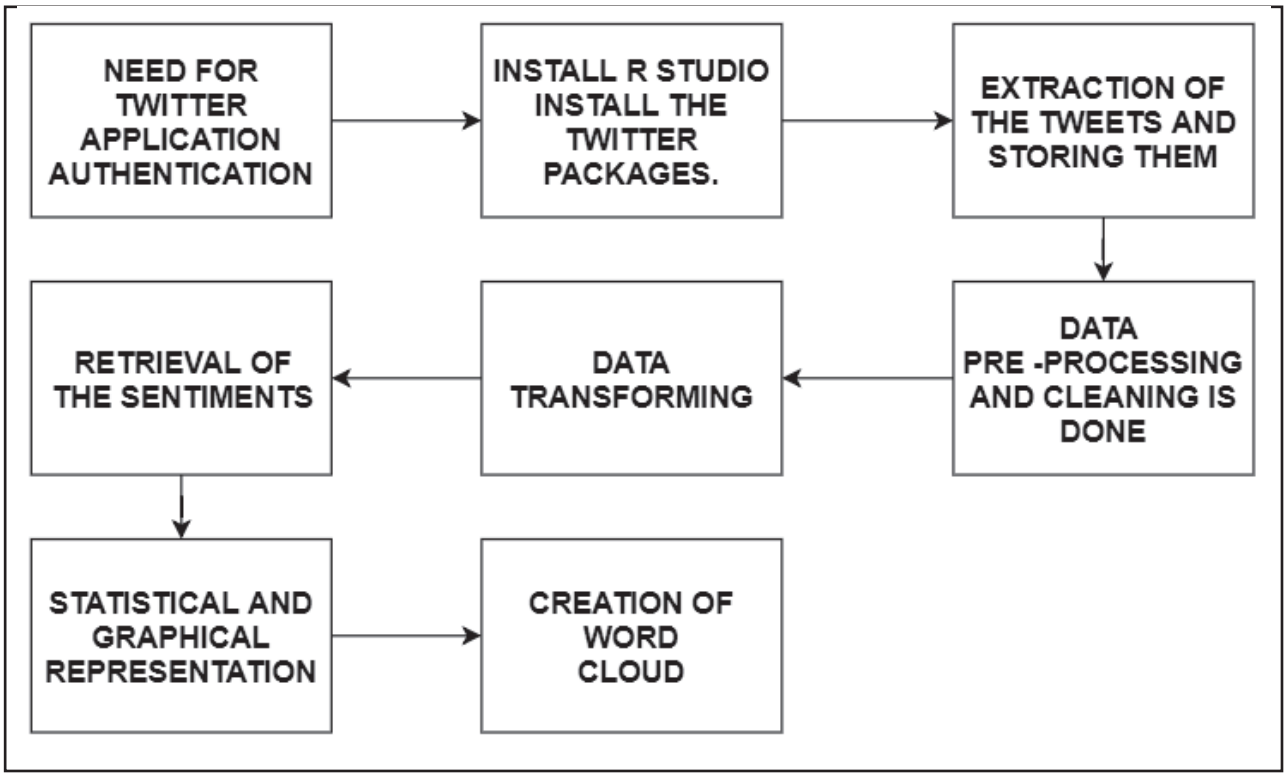

Fig. 1. Flowchart of Experiment

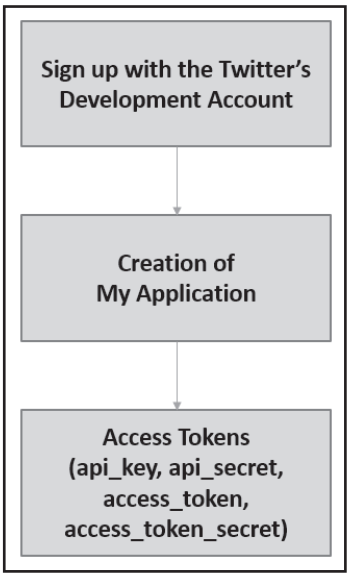

Fig. 2. Process of Twitter Application Creation

B. Using R-Studio

Installing and loading the packages in R-Studio that are required for performing the sentiment analysis. Packages that were used are:

- Library(twitterR): R based Twitter client
- Library(tm): Text mining package
- Library(syuzhet): Extract sentiment and sentiment-derived plots arcs from text.
- Library(ggplot2): Data visualizations using the grammar of graphics.
- Library(scales): Scale functions for visualizations.
- Library(wordcloud): Used for creating word clouds.
- Library(RColorBrewer): Color Brewer palettes.

C. Tweets Extraction

After successful authentication the extraction of tweets takes place using hashtag.

D. Pre-processing and Cleaning of Tweets

The syntactical correction of tweets takes place making the available data more readable and also reducing the ambiguity in extraction.

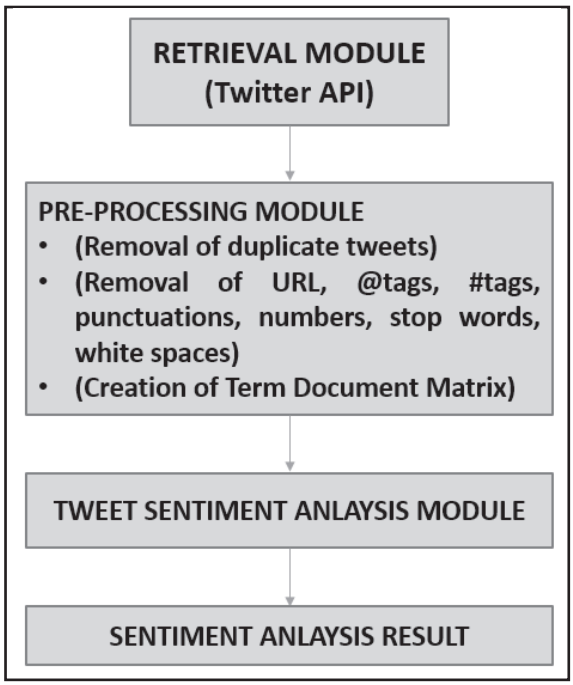

Fig. 3. Process of Data Cleaning

E. Data Transformation

Data after cleaning is arranged in data frame and matrix is created for performing some operations.

F. Retrieval of Sentiments

The main task of extracting the emotions is done from the tokenized words.

G. Visualizing the Results

This is done for making the analysis easy and observing the sentiments derived.

IV. DATA EXTRACTION

Data extraction means collection of the data and creating data sets. Here, data set refers to the collection of tweets that have been extracted from Twitter that have the hashtag #IndiaStrikesBack. This hashtag came into existence after the Balakot airstrike conducted by India in the early morning on February 26, 2019.

V. WORD CLOUD FORMATION

The #IndiaStrikesBack dataset status according to public opinion on Twitter is visualized here. The visualizations are done using the bar graphs, and with the use of RColorBrewer package we can play with the

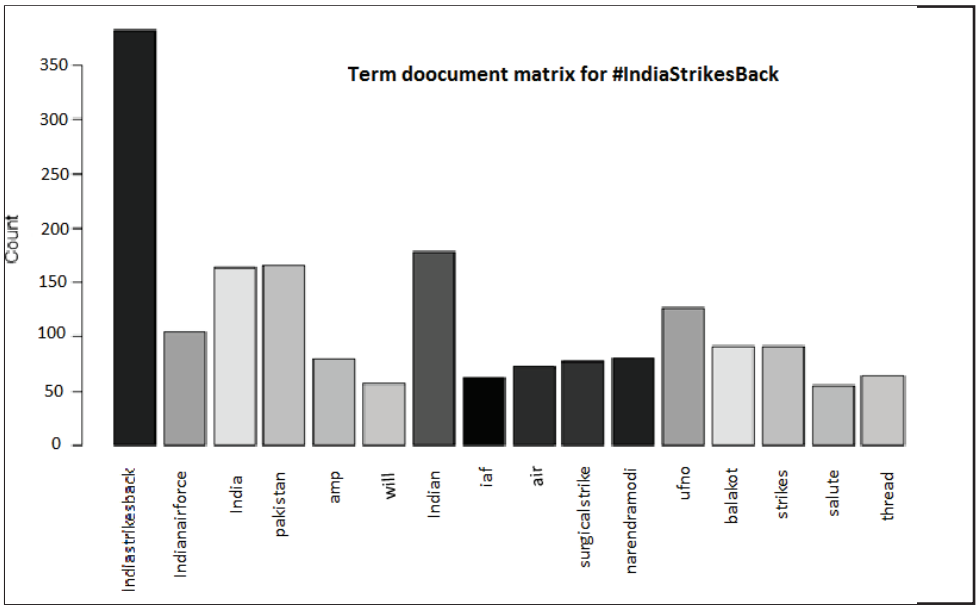

Fig. 4. TDM Graph for Most Used Words

colors in the graph.

Fig 4. depicts the most frequently used words in tweets. The word Indiastrikeback has the highest occurrence followed by the words Indian and India. Other words like Balakot, indianairforce are also used quite frequently.

The word cloud in Fig 5 shows frequently used words in the tweets are Indiastrikesback, Indian, Indianairforce and so on. The variation in size and color of words depicts the frequency like the word Indiastrikesback has higher frequency than the word Indian, whereas the remaining words show the responses and reactions of people over Twitter.

VI. THE ANALYSIS

The main aim of this study was to study and analyze the sentiments of people after India struck back in Balakot

Fig. 5. Word Cloud for Words Used Frequently

on February 26. This analysis comprised of 8 emotions and 3 sentiments.

The NRC emotion lexicon is used to calculate emotions and sentiments since it is the most widely used library for getting sentiments of people.

Bar graph is used to depict sentiments behind the tweets.

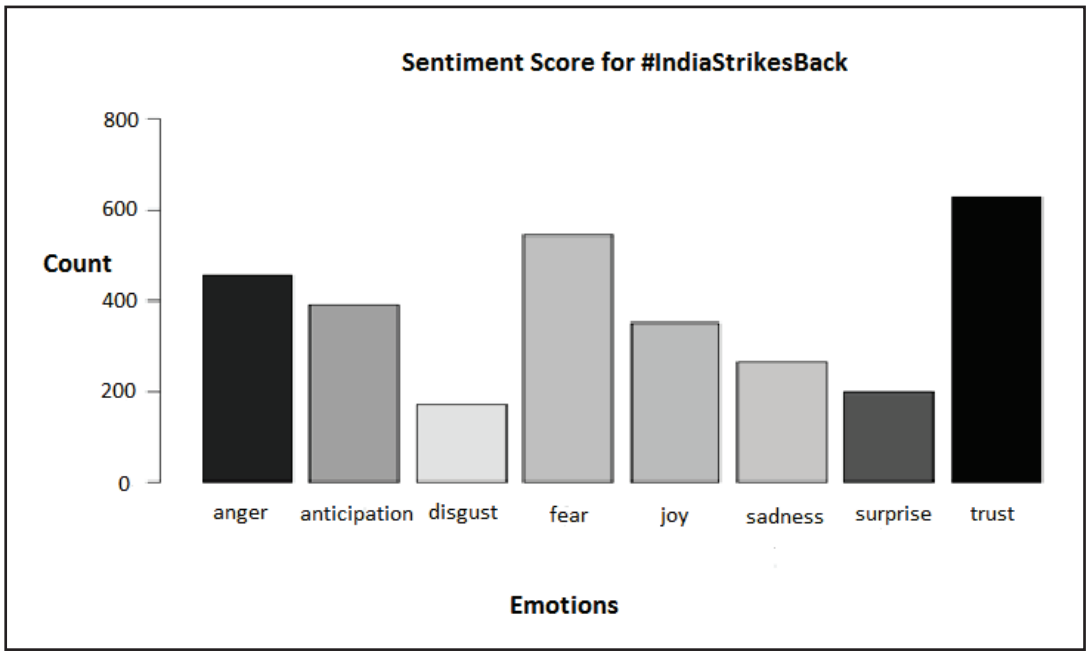

Fig. 6. Bar Graph Showing Sentiment Analysis

Fig.5 depicts a high level of trust that shows the confidence of people on the Indian armed forces and Indian soldiers for their bravery. 'Fear' depicts the fear in the minds of people for their safety and fear of losing their loved ones during the strike back. The get_nrc sentiment function from the package syuzhet was used which compared all the tokenized words with the word sentiment.EmoLexhad a large number of words with different emotions. The graph shows the mixed opinions of people around the world over the twitter.

VII. LIMITATIONS AND CHALLENGES

- Irony, humor, sarcasm present in the sentiments are not served well with the present sentiment analysis tool.
- Emotions that are communicated through emoticons are not properly interpreted.
- Any use of abbreviations, slang words or some local languages extracted words stand nowhere in the sentiment dictionary.
- Number of tweets used are 5000 which is not much compared to the overall tweets posted by people across the world.
- The use of mixed language words that are not translated makes it difficult for the analysis.

VIII. CONCLUSION

From the experiment it was concluded that out of 1154 tweets, more than 500 tweets were positive and the remaining were a combination of neutral and negative tweets. This shows that a majority of the citizens of India and people across the world reacted positively to the airstrikes conducted by India and support the Indian army and government in their brave decision.

The analysis further shows that the application of analytics to know about the sentiment of people with development in technology, and doing further research in text mining and analysis.

REFERENCES

[1] S. Singh and S. S. Choudhary, "Social media data analysis: Twitter sentimentalanalysis using R language,"*International Journal of Advances in*

Electronics and Computer Science, vol. *4*, no. *11*, 2017.
[2] J. Spencer and G. Uchyigit, "Sentimentor: Sentiment analysis ofTwitter Data,"School of Computing, Engineering and Mathametics, University of Brighton, 2012.
[3] I. Keka and B. Çiço, "Statistical treatment for trend detection and analyzing of electrical load using programming languageR," in *4th Mediterranean Conference on Embedded Computing*, *Montenegro,*2015. Doi: 10.1109/MECO.2015.7181906
[4] T. D. Chung, R. Ibrahim, S. M. Hassan, and N. S. Rosli, "Fastapproach for automatic data retrieval using R programminglanguage," in *2nd IEEE International Symposium on Robotics and Manufacturing Automation(ROMA)*, 2016.
[5] B. Gupta, M. Negi, K. Vishwakarma, G. Rawat, and P. Badhani, "Study of Twitter sentiment analysis using machine learning algorithms on Python," *International Journal of Computer Applications*,vol. *165*, no. *9*, pp. 29 – 34, 2017.
[6] V. V. Raniand K. S. Rani, "Twitter streaming and analysis through R," *Indian Journal of Science and Technology*, vol. *9*, no. *45*, 2016. Doi: 10.17485/ijst/2016/v9i45/97914
[7] T. R. Komati, S. B. Allamsetty, and C. V. Pinnamaraju, "Implementation of sentiment analysis on Twitter data," *International Journal of Pure and Applied Mathematics*, vol. 116, no. *5*, pp. 69-74, 2017.
[8] S. Singh and S. S. Choudhary, "Social media data analysis: Twitter sentimentalanalysis using R language," *International Journal of Advances in Electronics and Computer Science*, vol. *4*, no. *11*,pp. 2394-2835, 2017.

About the Authors

Radhika Goyal is pursuing Masters in Computer Applications from International School of Informatics and Management, Technical Campus, Jaipur. Her interest is in database mining. She is doing specialization in Databases System in her post-graduation course.

Meenal Kakkar is Assistant Professor (Computer Science) with International School of Informatics and Management Jaipur. She has completed her MCA (Masters of Computer Application) from Rajasthan Institute of Engineering and Technology, Jaipur. She has five years of experience which includes industry experience. Her research areas are Cloud Computing and Machine Learning.