

Comparison of Sentiment Analysis of Government of India Schemes using Tweets

V. Srividhya^{1*}, G. Raja Meenakshi²

¹ Dept of CS, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore, India

² Graduate Trainee, Cognizant Technology Solutions

*Corresponding Author: vidhyavas@gmail.com

Available online at: www.ijcsonline.org

Accepted: 18/Jun/2018, Published: 30/Jun/2018

Abstract—The pace of the monitoring of the Social media and the analysis of the social data keeps on rising high and it plays a major role in understanding the behavior of the people. Twitter, being the ninth largest social networking site in the world, is being eminent and powerful with its specialty of the short message named tweets with which people can share their opinions and also trend something worldwide with hash tags and common phrases. The Sentiment Analysis used here is to check the opinion of people related to the Government Schemes by the Central Government in the recent years with the help of Twitter Data Analysis. The tweets of the chosen schemes are classified based on the polarity and finally they are classified as positive or negative or neutral based on the opinions. This work is carried out using Digital India and Make in India tweets. Indians all over the world are sharing their ideas by tweeting and there are billions and billions of tweets tweeted every second across the world. The Sentiment Analysis is performed using R Studio. As the first step, the tweets needed for analysis are extracted with proper authentication with the help of Twitter API. The extracted tweets are cleaned by removing the stop words followed by the emotion and polarity classification. The final step is to generate the word cloud and then the comparison of the positive and the negative and the neutral tweets of the two schemes.

Keywords—Sentiment Analysis and Opinion Mining, Natural Language Processing, R - Studio

I. INTRODUCTION

Social media is the phenomenon which is emerging day to day with the help of the huge and rapid advancements in the Information Technology. Twitter is a rich resource for opinion mining and people uses this as one of the major social media to inscript their ideas. Twitter is one of the eminent Social Media compared with other sites like Facebook Twitter, Instagram, Whatsapp. Billions of tweets are shared each second across the world. To analyze the tweets, an open source approach using a set of R packages is used. A case study of Eight Indian Government Schemes shows the importance of the analysis of the user generated online opinions called tweets from Twitter. This helps in evaluating the performance of the government by monitoring the people's perspective instead of making People surveys which are time and cost effective. This phenomenon's extensiveness and the velocity with which the ideas are created and the spread about the growing need for the development of the system which detects and classify the opinions automatically as Balahur et al. [1]. To identify the sentiment expressed in the text using Natural Language Processing, Text Analytics and Computational Linguistics [2]. The sentiment analysis definitions was provided by Pang et al. also use the terms Opinion Mining, Subjectivity Analysis [3]. Recent Research in the area of Sentiment Analysis focuses on the application of

Classification of text to their polarity – positive, negative or neutral. The Opinion Mining and Subjectivity Analysis is the process to determine the polarity of opinions by humans to a specific scheme. Opinion Mining can be applied on any text form of sentiments as blogs, reviews etc. The analysis can be carried out either on document level or at sentence level. In the former case, the whole document is evaluated to determine the opinion polarity. In the latter case, the document is divided into sentences and each one is evaluated separately to determine the sentiments. To perform sentiment analysis, preprocessing is an essential task and preprocessing techniques helps to enhance the accuracy of the sentiment classification [4].

This work is carried out by R language to mine the tweets and also to implement the Opinion Mining for few Government Schemes. R Studio is a free and open source IDE for the development and deployment of R applications which can be installed on Windows, Macintosh and also Linux. The paper is organized as follows. Section I contains the introduction to the social network. Section II contains the related work of sentiment analysis. Section III gives the methodology with flow diagram. Section IV describes results and discussion. Section V concludes research work with future directions.

II. RELATED WORK

In today's digital world scenario, the effective and efficient mining of the online reviews is most needed. In the online repository, there are millions and millions of opinions and so it is difficult for the common people to follow the same. So, online web reviews mining system of Sentiment Analysis helps everyone in knowing deeper. Sentiment Analysis is a machine learning approach that identifies and categorizes the sentiments expressed as text to find the writer's idea of approach and his attitude towards the specific domain as positive or negative or neutral. Sentiment Analysis notices the polarity of the text in textual format. As it is also known as Opinion Mining, it derives the opinion of the speaker about a topic. It determines whether a piece of writing is positive, negative or neutral [5]. People's political preferences expressed online with those keenly observed over elections with uses of Social Networking Sites [6]. One can easily get a detailed and descriptive insight to predict and analyze the emotions of voters [7]. It is possible to track the real-time trends and also to capture any sudden miraculous change by keep on monitoring the conversations on the online media and get the variety of public opinion well in advance even before the announcement and the declaration of results of polls [8]. Some of the studies state that analyzing social media allows a trustworthy and efficient forecast of the final result [9]. Semantic analysis uses different approaches for the processing of text. It can be effectively taken in consideration for carrying out sentiment analysis [10]. Sentiment analysis is a research and developing a stream of Natural Language Processing method of the artificial intelligence [11]. These types of researching techniques range from document-level classification to the polarity of sentiment through words and phrases. Classifying the sentiment of Twitter messages is similar to sentence-level sentiment analysis for the limited sized tweets and the paragraph level sentiment analysis for more than one sentence [12]. Since Twitter follows micro-blogging nature, it is the best source for sentiment analysis. Major research work for sentiment analysis is carried out using tweets, but there has been some study involving the Amazon reviews dataset too. The new implementation for analyzing amazon reviews which involved detection of fake reviews, processing the genuine reviews using Apache Spark and finally rating the products [13].

III. METHODOLOGY

This work has been developed by using three phases. The first phase of the work is tweets extraction and pre-processing. The second phase is to apply Naive bayes Algorithm using package and finding polarity classification. The visualization of the result is the third phase of the work. Figure 1 shows the proposed methodology of this work.

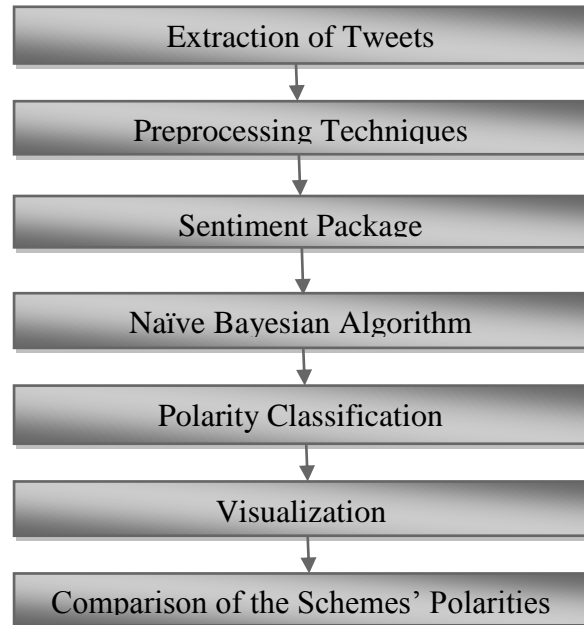


Figure 1: Proposed Methodology

A. Extraction of Tweets

The first step in mining data from Twitter is to get authenticated with the help of Twitter API. The Twitter Application can be created with the help of twitter app. When the app is created, the user is given the access to the Consumer Key and Access Token. The Consumer Secret and the Access Secret must be confidential. The package twitter in R is the interface between the R tool and the Twitter website. The package used for authentication is ROAuth. The function used to search and extract tweets is searchTwitter().

The extracted tweets are stored in DI_tweets and MI_tweets. The tweets have been extracted since the days the schemes Digital India and Make in India were launched i.e., July 1, 2015 and September 25, 2015. Totally 5000 tweets have been extracted and saved.

B. Pre-processing Techniques

The extracted tweets have been cleaned with the help of the gsub() function. The Retweets, special symbols, Web links, white spaces, punctuation, digits, HTML tags, emoticons and special symbols have been removed. All the text of the tweets are converted to lower case with the help of tolower() function. The stop words such as is, are, at, which, was, where, on etc., are removed from the extracted tweets so that one can find the sentiments easily with the root words.

C. Sentiment Package

The pre-processed tweets are the input to the functions to classify the emotions and the polarity. This can be done with the help of the package sentiment.

D. Naïve Bayesian Algorithm

Naive Bayes is a Supervised Machine Learning approach. It computes the posterior probability of a class to classify data and also Bayes theorem. In this work for the emotion and polarity classification functions, the attribute used is bytes.

E. Polarity Classification

The function classify_polarity is used to classify the tweets. After being classified for each scheme, the polarity is visualized in the form of bar graphs. The polarity is classified as either positive, negative or neutral.

F. Visualization

The polarities are visualized with the help of the ggplot() function. The word cloud is generated creating a term document matrix and the function used is comparison.cloud().

IV. RESULTS AND DISCUSSION

This work has been developed using twitter dataset. Twitter is an online social network used to send and read short messages called "tweets". The tweets used to analyze and predict the future directions by public opinion.

Figure 2 and Figure 3 shows the result of polarity classification, which has classified the tweets into positive, negative and neutral. For the proposed work 2500 tweets of two different scheme dataset has taken for polarity classification. Figure 4 and Figure 5 shows the word cloud generations of Digital India tweets and Make in India tweets.

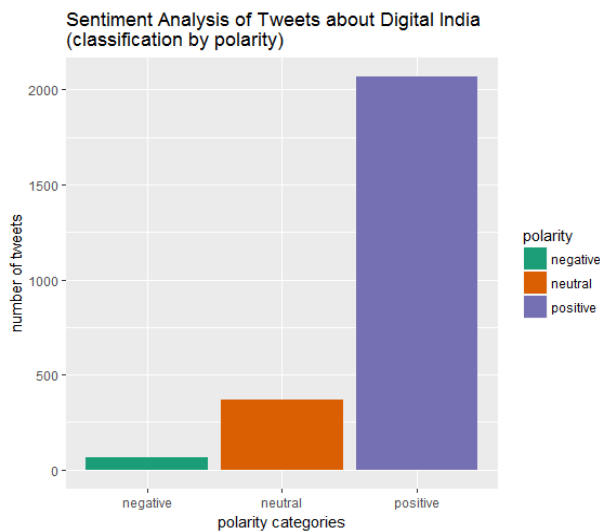


Figure 2 Polarity Classification of Digital India tweets

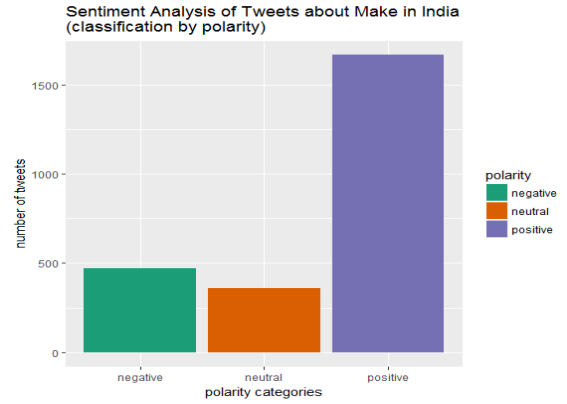


Figure 3 Polarity Classification of Make in India tweets

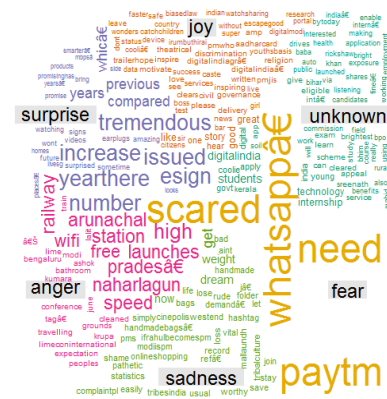


Figure 4 Word cloud of Digital India tweets

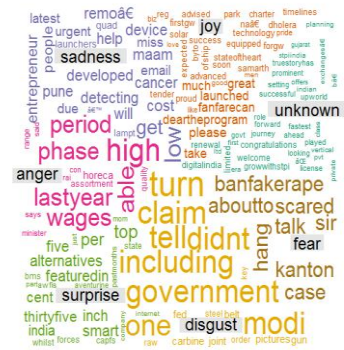


Figure 5 Word cloud of Make in India tweets

(i) Comparison of Schemes' Polarities

As the tweets were classified based on Subjectivity Lexicon, the tweets which have more number of positive words are classified as positive and with more negative words are neutral and the tweets which neither falls under positive nor negative are classified as neutral tweets. After being classified separately, both the polarity results of Digital India and Make in India are compared and the conclusions are arrived.

The following table 1 represents the result of Polarity of two different schemes – Digital India and Make in India. The Figure 6 shows the comparison chart between Digital India and Make in India tweets with positive, negative and neutral tweets. From the result it is concluded that compared with Make in India scheme, Digital India scheme has more positive sentiment tweets.

Table 1 Result of Polarity of Twitter Datasets

Trend Name	Positive sentiments	Negative Sentiments	Neutral Sentiments	Tweets used
#Digital India	2050	350	100	2500
#MakeIn India	1700	475	325	2500

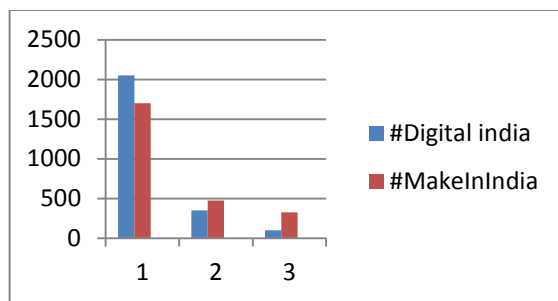


Figure 6 Comparison of Polarity of Two Schemes

V. CONCLUSION

Sentiment Analysis is one of the effective and efficient techniques for analyzing the behavior of the people. The twitter data, commonly known as tweets are the samples for the different opinions of different people. Sentiment Analysis helps to find the different perspective and aspects of the people's mind and thoughts with which the polarity of the schemes can be found and it helps in predicting the implementation of the various Government schemes effectively and also to take decisions for the future schemes. In this paper, Polarity Classification is performed such as obtaining positive, negative and neutral tweets for particular Government Schemes such as Digital India and Make in India. From this people and Government can find out the different opinion behind the declared scheme. The manual survey from people is tiresome as it is costly and time consuming. So, this Sentiment Analysis is very handy and effective for the evaluation of the Government scheme and also to monitor them on a regular basis of its growth among the people and also to know the people's perspective.

REFERENCES

[1] A. Balahur, G. Jacquet, "Sentiment analysis meets social media – Challenges and solutions of the field in view of the current information sharing context". Information Processing and Management, Vol.51, Issue 4, pp. 428-432, 2015.

[2] G.Vinodhini, R.M. Chandrasekaran, "A comparative performance evaluation of neural network based approach to sentiment classification of online reviews" Journal of King Saud University, Computer and Information Sciences, Vol. 28, Issue 1, pp. 2- 12, 2016

[3] B.Pang , L. Lee, "Opinion mining and sentiment analysis". Foundations and Trends in Information Retrieval Vol.2, Issue 1, pp.1-135, 2008.

[4] M.Bhuvanewari, V.Srividhya, "Enhancing The Sentiment Classification Accuracy of Twitter Data Using Preprocessing Technique" International Journal of Engineering Research, Vol. 4 , Issue 5, 2016

[5] BholaneSavitaDattu, Prof.Deipali V. Gore "A Survey on Sentiment Analysis on Twitter Data Using Different Techniques", International Journal of Computer Science and Information Technologies, Vol. 6, Issue 6, pp. 5358-5362, 2015.

[6] O'Connor B, Balasubramanyan R, Routledge BR, Smith NA. "From tweets to polls: Linking text sentiment to public opinion time series" Proceedings of the fourth International AAAI Conference on Weblogs and Social Media, Washington, pp. 122-129, 2010.

[7] Xin J, Gallagher A, Cao L. "The wisdom of social multimedia: Using flickr for prediction and forecast", Proceedings of the 18th ACM international conference on Multimedia, New York, pp.1235-1244, 2010.

[8] MJ.Jensen, L.Jorba E. Anduiza "Digitalmedia and political engagement worldwide: A comparative study", Cambridge University Press, New York, NY, pp1-15, 2012

[9] Tjong Kim Sang E, J.Bos, "Predicting the 2011 Dutch senate election results with twitter", Proceedings of SASN 2012, Avignon, France, pp.1-8, 2012.

[10] C.Nanda, M.Dua, "A Survey on Sentiment Analysis "International Journal of Scientific Research in Computer Science and Engineering, vol. 5, Issue 2, pp. 67-70, 2017.

[11] H.Yu. V.Hatzivassiloglou, "Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences" In the Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 129-136, 2003.

[12] Sunil B. Mane, Kruti Assar, PriyankaSawant, Monika Shinde. "Product Rating using Opinion Mining", International Journal of Computer Engineering in Research Trends, Vol.4, Issue 5, pp.161-168.

[13] S.R.Chheda, A.K.Singh, P.S.Singh, A.S.Bhole, "Automated Trading of Cryptocurrency Using Twitter Sentimental Analysis", International Journal of Computer Sciences and Engineering, Vol 6, Issue 5, pp.209-214.

Authors Profile

1. **Dr.V.Srividhya** has completed M.Sc., M.Phil., and Ph.D in Computer Science. She is working as Assistant Professor(SG) in the Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore. Her fields of research interest are data mining, text mining and Big Data. She has published papers in the international journals and presented research papers in international and national conferences.



2. **G. Raja Meenakshi** has completed B.Sc in Computer Science from Avinashilingam Institute for Home Science and Higher Education for Women. She has done her final year project in Data Mining. She is a Graduate Trainee in the Cognizant Technology Solutions.

