Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 173 (2020) 325–334

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on *Smart Sustainable Intelligent Computing and Applications* under ICITETM2020

Sentimental Analysis of Twitter Data with respect to General Elections in India

Ankita Sharma[a], Udayan Ghose[b] *

[a,b] *USICT, Guru Gobind Singh Indraprastha University, New Delhi, India*

Abstract

It is known that social media is one of the largest sources of unstructured data. Analyzing that data and harvesting meaning out of that is a tedious job. Recently opinion mining has become an emerging topic due to the vast amount of opinioned data available on the various social networking sites. Microblogging has appeared relatively recently, and twitter is the most popular microblogging sites used by the people. It is one of the biggest free, open data-source. In the world today, twitter often witnesses a lot of opinions. Opinion mining and sentiment analysis help researchers to gain insight into public emotions. In this paper, Twitter is used as a source of opinioned data. Twitter APIs are used for the collection of tweets. In this paper, R is used for the acquisition, pre-processing, analyzing the tweets, then sentiment analysis is performed based on the different approaches. In this paper, Tweets were collected from the period of Jan 2019 to March 2019. Using that tweets, sentiment analysis was performed to gain the opinion polarity of the folks concerning general elections held in India. Two candidates were considered for this study: Candidate-1 and Candidate-2. It was concluded that Candidate-1 is more liked and is famous as compared to Candidate-2. The result obtained in our paper was in full compliance with the actual election results obtained in May 2019.

Keywords: Twitter mining; Text Mining; Sentiment Analysis; Social media

1. Introduction

It is known that just a small percentage of the whole data which is produced daily all over the world is structured

* Corresponding author. Tel.: +011-25302719
E-mail address: udayan@ipu.ac.in

10.1016/j.procs.2020.06.038

in nature, rest of about 79% of the data is unstructured in nature meaning thereby they don't have any predefined archetype. Social media is one the largest source of unstructured data.

Now days, people have become so habitual to social media, that they cannot imagine a day without it. People express their happiness, anger, sadness and every other emotion of the various social networking sites. They not only use it as a communication medium, also they keep on posting their day to day experiences on it.

Social media, in other words have become a pool of sentiments. The unstructured data is dirty in nature. Therefore, working with it is a tedious job. Recently text mining and sentiment analysis have gained huge attention because of the availability of masses of data from number of social networking sites.

In the older days, when one wanted to know about what people feel and think about for instance a particular product/item they used to do surveys. Surveys were expensive and time-consuming activity of gaining insight of public emotions. Now days opinion mining is being done on the large amount of data. One can obtain large amount of opinioned data from these social networking sites. Twitter is one among the most popular micro blogging site, it has huge amount of opinioned data [1].

Text mining or text analytics is the process of harvesting the meaning out of large amount of data. Sentiment analysis and opinion mining can be used interchangeably. It is the process of extracting sentiments, emotions, thinking, opinions out of data.

Sentiment Analysis is widely practiced to get insight about the feelings and attitude of the people on the internet; it is done to measure the performance of various brands, products etc. For instance: Macy's is a leading retail chain/store. It uses Sentiment analysis as one of the ways to improve customer experience and business growth.

1.1. Contributions

Our contributions in this paper is as follows:

- In this paper, an efficient approach is proposed to analyze the sentiments of the twitter data.
- We perform the preprocessing and analysis on the collected corpus, followed by sentiment analysis.
- The experimental evaluations were conducted on the set of real tweets to prove that our presented approach is worthy and can be used in other similar applications for opinion mining.

1.2. Organization of the paper

The rest of the paper is organized as follows. In Section 2, A Walk Through is given about the social media, twitter mining and sentiment analysis. In Section 3, discussion on prior work on sentiment analysis is done. In Section 4, the approach and design methodology used is described. Implementations and Results are in Section 5. Finally, in Section 6, we have concluded our work.

2. The Virtual World: Social Media

In the world today, people are connected to each other via internet. Three decades ago, expressing opinions and making one's voice reachable on the fly to the world was a dream. Now with the advent of the social media, it has become a reality. In world today one cannot imagine their life without the social media; it is a growing phenomenon due to advances in information technology.

Virtual world has arisen not only as a communication media, but people also consider it as a necessary platform for expressing their views, opinions, sentiments on the various topics. People here not only connect by their blood relations or if acquainted with each other, but also people connect on social media based on the similar interests, hobbies, specialization etc. There are millions of active users of social media and users of social media are expected to increase up to 3 billion by 2021 [5][8].

US, China, India has the largest social media users. It is known that only about 21% of the whole data produced is structured in nature, rest of the data is unstructured data. Most of the unstructured data is the textual data. It is mainly produced by the various social networking sites. This data, if processed properly can be an information source and can be of great importance to us. Social media is used by both professional and non-professional users.

Tons of reviews or opinions produced and shared by users daily all over the world. If this data is refined and processed properly, then it can be used in various sectors for improving the various aspects of life.

2.1. Twitter Mining

Twitter mining is the hot topic these days, as it provides the important information which is used and applied in various fields. It is one of the major research areas. By using the various public APIs various tweets can be collected and analysed for research purpose. Through authenticated requests twitter APIs are established.

OAuth that is open standard authorization, which is required to get access of the protected data of the twitter. Users of the twitter come from every grounds of life. Its users vary from the celebrities to the normal users, company owners and representatives, politicians, also the country's presidents and prime ministers [14]. Therefore, it is possible to collect tweets of the persons from the different social and interest groups. Researchers find this information very useful to know about people's sentiments and opinions in defined number of words, political parties tend to get to know about whether people are in their support or not.

People use casual approach in writing tweets; therefore, the tweets has the noisy data, spelling mistakes, grammatical and punctuation errors. Therefore, analysing tweets is a tedious job. Through twitter we get a vast amount of data, but the data obtained is unstructured in nature. If that data is analysed, then more value can be added to it.

Text mining is the process of analysing the day to day produced data which is written in our natural language or spoken language. To accomplish this text mining make use of the natural language processing [16][17]. Text mining/analytics uses natural language processing to accomplish this. NLP is a component of the text mining which basically helps the machine to read the text and to derive high quality information from the text.

NLP uses the principles of the computer science and artificial intelligence to deal with the human language, also study the data and derive useful information from it. In general terms the text analytics is a process and NLP are the method to carry it out.

Opinion mining can be performed on the twitter data to get insights of the sentiments of the public. R is a statistical analysis tool and a programming language. In R software, "twitteR" package is used for extracting the tweets from the twitter R language will be used for twitter mining in this paper [9].

2.2. Sentiment Analysis of the Tweets

In this, an attempt to gain an insight of the textual tweet posted on the twitter about the candidates for the prime minister's position of the two most popular parties prior to the election is made that is Candidate-1 and Candidate-2. It is known that abundance of the information can be gathered from the tweets, if proper processing is done then the information obtained can be used for decision making, management and political campaigns.

Opinion mining can be performed on the twitter data to get insights of the sentiments of the public. This information is very useful for various organizations as well as the different political parties. It is known that feelings, emotions, sentiments are of great importance in human life. Mining such opinions is called sentiment analysis [3].

It is an emerging field which identifies the polarity, relevance and objectivity of the text. With the help of this, text can be classified into categories like positive, negative and neural. In this; reviews, blogs, microblogs written by the humans are examined, to get insight about the feelings, attitudes of the people on the internet for monitoring the performance of the political parties, governments, products, movies etc.

Assigning labels of sentiments to a text, which tells about the polarity of the text is called Sentiment Analysis (SA). It can be classified into three levels: document level, sentence level or aspect level. By document level it is meant that the opinions are classified for the entire document. For this to happen, document should consist of the single topic. There is no provision for the document containing multiple topics. In sentence level, all the sentences are examined to find their polarity, it could be either positive or negative, mixed opinions may or may not be considered for a sentence. Aspect level can also be known as entity level, in this all the aspects of the sentences are considered, meaning thereby in what context sentence talks about. Entity level approach is a detailed approach but it quite complicated to implement [13][24].

3. Related Work

Twitter Microblogs have been mined for getting insights of the customer tweets regarding two stores in US. Methodology used is, data is accessed using the "TwitteR" package to get tweets from the twitter API. Cleaning is the data is performed to remove stop words, spaces, URLs, punctuation marks etc. Several mining tasks are performed such as finding frequent terms, association rules and finally sentiment analysis is done using lexicon based approach. There is list of positive and negative words. Scoring function is there which is assigned to the tweets [1].

The point in study was to analyze the tweets of six hockey teams using opinion lexicon and LDA scoring scheme. It is observed that the type of the lexicon selected affects the results of the Sentiment Analysis. When lexicon is small with lesser relevant words, the spectrum obtained is more positive. On the other hand, if the lexicon is large enough with more relevant words then the spectrum obtained consist of both positivity and negativity [7].

In this, researchers performed opinion mining on the tweets collected from the Twitter API on the topic Modi ji's Digital India Campaign. Opinions are classified as positive, negative and neural. Dictionary based approach and machine learning approach is used for performing sentiment analysis. It is concluded that 50% of the opinions are positive, 20% are negative and left are neutral [6].

Authors in this, collected 3000 tweets in Arabic and Saudi language on a particular topic. Sentiment analysis was performed using Supervised and Unsupervised approach also the mélange of both that that is Hybrid approach. It was concluded that hybrid approach performs better than the previous two approaches [2].

Researchers have applied k-means clustering on the gathered corpus, with the aim to find clusters in data with already given number of clusters. The tweets which were considered for the clustering contained geographical information. Tweets were classified into geographic clusters and visualized using ggplot package [4].

Authors proposed an ontology-based Aspect level method to analyze the opinions from the tweets about the diabetics. In this, tweets were collected and corpus is formed out of that. Each tweet is manually labeled as positive, negative or neural. The outstanding results was obtained from N-gram method [13].

In this researcher have proposed a model for opinion mining that considers typographic, infographic visual content along with the textual content of the tweets. The proposed model will serve as a tool for visual listening as well as a tool for social media monitoring and analytics. The obtained results improve the generic sentiment analysis task [28].

The point in study was to include sentiment diffusion information, while performing sentiment analysis on the twitter messages to achieve better performance. SentiDiff, an iterative algorithm is used to perform the opinion mining on the twitter data; which is a relationship between the textual information and sentiment diffusion. This was concluded that this approach performs well in comparison to prior state-of-the-art- textual information-based sentiment analysis algorithms [29].

4. Approach and Design Methodology

The aim of the paper is grounded in analyzing the sentiments of the public, regarding the two most popular candidates of the two most powerful parties, for the prime minister's position in India (2019) that is Candidate-1 and Candidate-2. The data set used is primary. Twitter's API is being used to collect the tweets of the two respective candidates. Two Corpus are made out of the tweets collected from twitter using R.

The steps taken are as follows. The first step was the understanding of the data, meaning thereby reviewing of data, this includes collection of data, describing and exploring the data and the most important step is to verify the quality of data. Since acquisition of the data is done through online source, hence it contains noise, so therefore, it cannot be put directly for analysis purpose, therefore its pre-processing is required. Pre-processing and cleaning of the tweets was performed using R [20][25].

Data preparation was done with the help of tm package. Cleaning involves conversion of Twitter feeds into lowercase, removal of the English stop words, URLs, numbers, punctuation marks and alphanumeric characters etc. After performing all the steps, cleaned data was read as a data frame in R [12]. Later again it was converted into corpus and finally was outputted as a CSV file.

It is also important to get insight about the most frequently occurring words in tweets; through them one can get to know about the structure and the type of tweets, word cloud can be used to depict this. Word cloud is made with the help of "wordcloud" package. The larger the word inside the word cloud, the frequent that word is in corpus.

Second step is the modelling of the data, this step includes choosing the appropriate technique for the experiment and access them. Named entities are extracted from the tweets because with the help of this one can access any situation giving in depth information about names of person, locations and organizations in tweets enabling others to go quickly through tweets and discarding the tweets which are irrelevant and unnecessary.

It is very frequent that one finds completely off-topic tweets. To go through and every tweet is not feasible, therefore extracting name entities is useful. It is done about the polarity of the sentiment, sentiment strength and sentiment subjectivity etc. In this paper, the named entity extraction is done using AYLIEN, which is a text analysis extension of Rapid Miner.

It is not only important to get the named entities from the tweets but getting insights of the emotions, fears and sentiments of the people are equally important. At the end sentiment analysis is done through R, Rapid-Miner AYLIEN [15]. Through them one can get insight about polarity of sentiment, sentiment strength and sentiment subjectivity etc. By Analyzing the sentiments in tweets, humanitarian organizations get insight about public's opinions, issues etc. In the last step, all the final results are outputted.

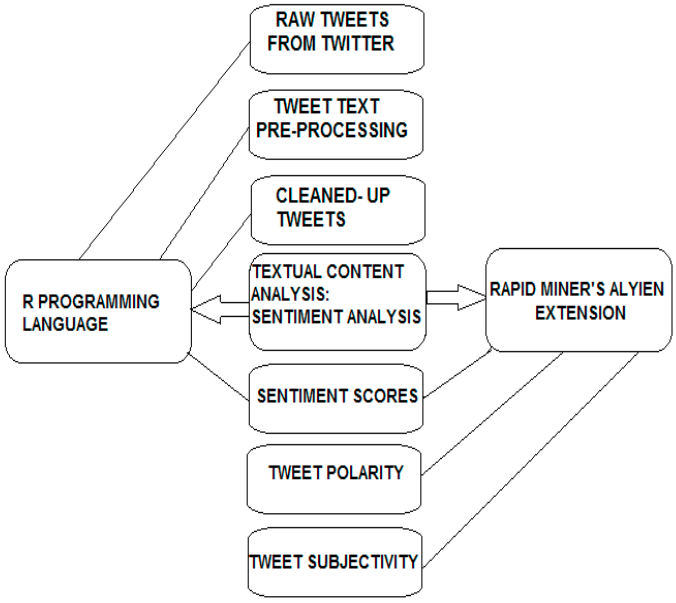

Fig. 1. Broad Methodology.

Fig. 2. Word Cloud (a) Candidate -1 and (b) Candidate -2 respectively.

5. Implementation and Results

The “SentimentAnalysis” package is used for calculating the polarity of the text sentiment at the sentence level. The “sentimentr" package is used for the purpose of opinion mining or sentiment analysis on textual content in R. It used various existing dictionaries for this. It is a commonly used dictionary-based approach, people can create their own dictionaries also. From the literature review it is concluded that unsupervised techniques are better than supervised techniques [5][11][12].

5.1 Lexicon Based Approach for Sentiment Analysis

- The words, phrases, idioms, adjectives that have a semantic meaning is known as lexicon, this approach relies on the lexicons for classifying the data.
- In other words, the lexicon is composed of set of positive, negative or neutral. It requires a scoring function to score each of the sentences according to negative, positive words, for determining the sentiment polarity.
- There are numerous ways of making the lexicons, one such method is based on dictionary approach. It begins with the basic sentimental words and is made large by adding similar words and antonyms to it. Once it is made, it can be used to perform matching in the words in text document and the lexicon words.
- Here words are treated as numbers. Positive number implies the positive words (PW) and Negative number implies the negative words (NW). Polarity is assigned as +1, -1, 0 for positive, negative and neutral text respectively.
- Hiu and Liu developed an opinion lexicon for general purpose, which will classify English words into positive and negative words [26][27]. There are about one thousand nine hundred sixty-seven positive words and four thousand seven hundred eighty-three negative words. Final Sentiment score of the tweets is calculated by subtracting the positive tweets from the negative tweets.

$$Score\ of\ Sentiment\ = \sum PW\ -\ \sum NW \tag{1}$$

$$Positive\ Score\ of\ Sentimen: \sum PW > \sum NW \tag{2}$$

$$Negative\ Score\ of\ Sentiment: \sum PW < \sum NW \tag{3}$$

$$Zero\ Score\ of\ Sentiment: \sum PW = \sum NW \tag{4}$$

Zero score states that there are no opinion words in the text, whereas positive score implies positive polarity of the text and negative score implies negative polarity of the text. This approach produces good results for opinion classification.

It was an earliest approach to the classification of the sentiments and has good accuracy on the single phase [11]. That is why, this approach is used here. Figures below show a glimpse of the sentiment scores for both the candidates.

```
> modiscore$score
  [1]  1 -1  2  0  0  0  0 -2  2 -1  0  0  1  0  0  0 -1 -1  1  0  0 -2  0  1  2 -1 -2  0  0  0  0  0 -1 -1  0 -4  0  1
 [39]  0  1  0  1 -1  1  0  0  0  0  1 -1  0 -3  1  0  0  1  0  0  0 -2  0  0  0 -1  0  0 -1  1  0  0  0  0 -1  0  0  1
 [77]  0  0 -1  0  0 -1  0  0 -1  0  0  1  0  0  0 -2  0  2 -2 -1  0 -1  0  0  1 -1 -1 -1  0  0  1  1  0  1  0 -1  0 -1
[115] -1  1  1 -1 -1  2  0  0  0 -1 -1  0 -2  0 -2  0 -2  0  1  0  2  0  2  1 -2  1 -1  0  1  1  0  1  2  0  0  0  0  1
[153] -1  0  0  0 -1 -1  1  0 -1  1  0 -1  1  1 -2  2  0 -1 -2 -1  0  2  0  1  0  0  0  0 -1  1 -2 -2 -2  0  3  0  0  1
[191]  0  0 -1  0  2  0 -1  1 -1  0  0 -2  1  0  3  0  0  0  0 -1  0 -2  0  0  0  1 -2  0  2 -1  2  0  2  1  0 -2 -4  3
[229] -2  2  0 -1 -1  0  0  0 -1 -1 -1  1  2  0  0  0  0 -2  1  0  0  0  0  0  1  0  0 -1  0  1 -1 -2  0  1  1  0 -1
[267]  2  0 -2  0  1 -1  0  0  0 -1 -1 -2  0  0 -1  1 -1 -1  1  0 -1 -2 -1 -1  0  1 -3  0  0 -1  1  1  0  0  0  0  0  1
[305]  2 -1  0  0  0  0  0  1  2 -1  0  1  0  0  0  0 -1  0  1  1  0  0 -1  1 -1  1  0 -1  0  0  1  0  0  1  0  1  0
[343]  1  0 -1  1 -1  0 -1  0 -1  1  0 -1 -1  0  0  0  0  0  0  1  1  0  0  1  0  1  0  0  0  0  0  1  0 -1  1  0  1 -1  0
[381]  0 -1  1 -1  0  1 -2  0  1  0  0  0 -1 -1 -2 -1 -1  0  0  0  0 -1 -1  0  0  1  0  0 -2  1  0 -3  0 -1  0  0  0  0
[419] -2 -1  0  0  0 -1 -1  1 -1  1  0  0  0 -1  0 -1 -2  0  0 -2 -1 -1 -1  1  0  2  0  0  0 -1  0  0  2  0 -2 -1  0  0
[457]  1  1  1  2  0  1  0  0  0  0 -1  1 -2  1 -1  1  1  0  0  0  1  2  0  0  0  0 -1  0  0  0  0 -2  0  0  0  0 -1  0
[495]  0  0  0  2  1  1 -3 -1 -1  0  0  0  0  2  0  0 -1  1  0 -1  0  1 -2  2  0  1  0  1 -1 -1  0  0  0  0  1  1  0 -1
[533]  1  0 -1  0 -1  0  0  1  1  1  0  0 -1  1 -1 -1  0  0  0  1  0  1 -1  0 -1 -1  0 -1 -1 -1  1 -1  1  0  0  0  0 -1
[571]  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 -2  0  0 -1  0  0 -1  1  0  0  0 -1 -1  1  0 -1 -1  1  0  0 -1  2  0
[609]  0  1 -1  2  0  0 -2  1  0  1  1  0  0 -1 -1  1  0 -2  0 -1  2  0  0  3  0  0 -1  0  0  2  1 -1  0  0  0  0  0  0
[647] -1  0  0 -1  0  1 -1  0  1 -1  0  0  0  0 -1  0 -2  0  0 -1  0  1  0 -1  0  0  0 -1  0  0  0 -1  1  0  0 -1  0
```

Fig. 3. Sentiment Scores for Candidate-1.

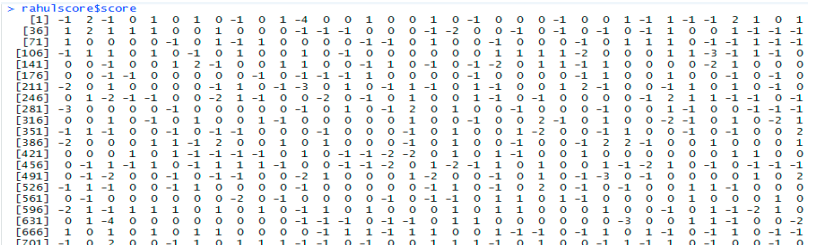

Fig. 4. Sentiment Scores for Candidate-2.

5.2 NRC Dictionary Based Approach:

"Tidytext" package was put to use into this work. In the previous approach the tweets were classified as positive, negative and neutral, to add multi-sentiment values NRC sentiment dictionary was used. It classifies the tweets based on eight emotions along with simple positive and negative [27].

Positive sentiment is depicted by joy, surprise, trust and anticipation. On the other hand, negative sentiments were depicted using words such as sadness, anger, fear and disgust. The sentiment score using NRC dictionary for both the candidates are depicted in figure 8 and figure 9.

The lexicons provided by this package do not necessarily contain every English word, as most of them are neutral in polarity. It is based on the uni-grams only and hence cannot catch sarcasm etc. Another advantage is, since data is sentence sized, the positive and negative sentiment scores aren't averaged to zero [22][23]. Therefore, it gives good results. Rapid Miner is a GUI based tool and was used to obtain opinion polarity and subjectivity from tweets [21]. AYLIEN was used to obtain sentiment polarity, subjectivity as well as sentiment scores. There are several extensions available, other than the built-in operator in this software.

Text Processing and AYLIEN Text Analytics Extension was downloaded for this paper. Sentiment Analysis was performed on the tweet dataset for both candidates that is candidate 1 and candidate 2 using AYLIEN extension. AYLIEN is the Text Analysis extension of the Rapid Miner, which is made up of different operators. With the use of AYLIEN one can analyse the input data and make sense of the input data supplied to it. This extension includes Sentiment Analysis, Entity Extraction, Language Detection, Hashtag Suggestion and Related Phrases operators [15][20][21]. Sentiment Analysis operator of the AYLIEN was used for performing opinion mining on tweets related to both the candidates.

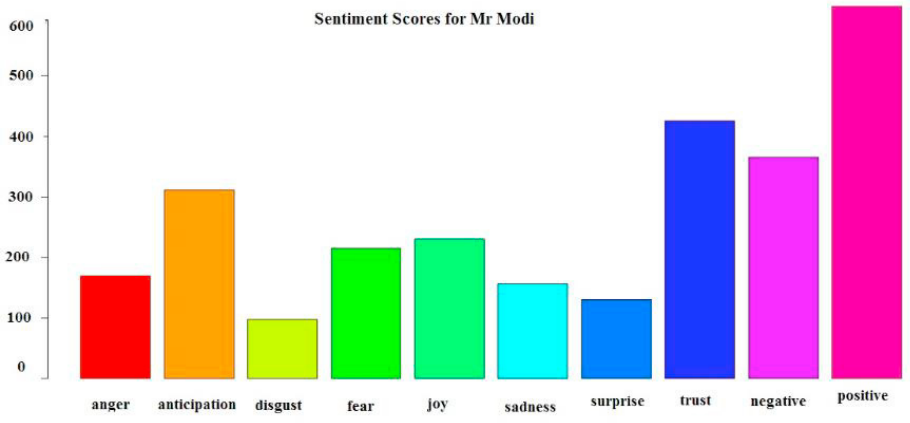

Fig. 5. Sentiments for Candidate-1 using NRC dictionary approach.

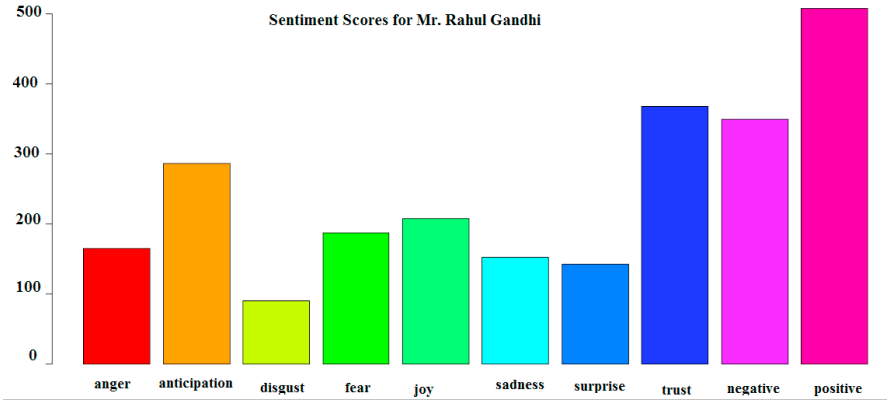

Fig. 6. Sentiments for Candidate-2 using NRC dictionary approach.

This operator performs two tasks:

- This operator classifies the tweets according to polarity as predicted into three classes- Positive, Neural and Negative.
- Also, it classifies the tweets according to subjectivity as expressed into two classes- Objective and Subjective.

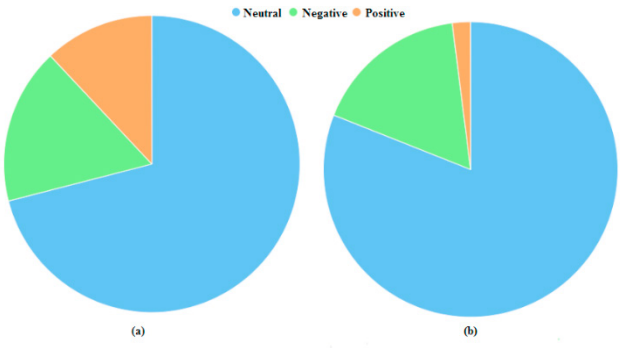

Fig. 7. Polarity Pie Chart for (a) Candidate-1 (b) Candidate-2.

From the pie chart above it was concluded that people have more positive sentiments for Candidate-1 as compared to Candidate-2. Although most of the sentiments for both the candidates are neutral.

While getting insights into the polarity of the tweets only involves observing the strength of the sentiment words used, understanding the objectivity and content of the tweet.

Objective Tweet refers to the factual information, it is not a personal opinion, while the Subjective Tweets refers to one's personal opinions, beliefs and feelings. It is not a fact. Subjectivity Classification refers to classifying a tweet text as Opinioned or Non-Opinioned.

While Polarity Classification refers to classifying a tweet text as expressing positive, negative or neutral sentiments. In addition to the Subjectivity and Polarity Classification AYLIEN also provides a confidence level measure. It was found that for the Candidate-1 there were more subjective tweets, whereas for the Candidate-2 there were more objective tweets. Confidence Level measure is a number between 0 and 1, which denotes the confidence with which classification task is performed.

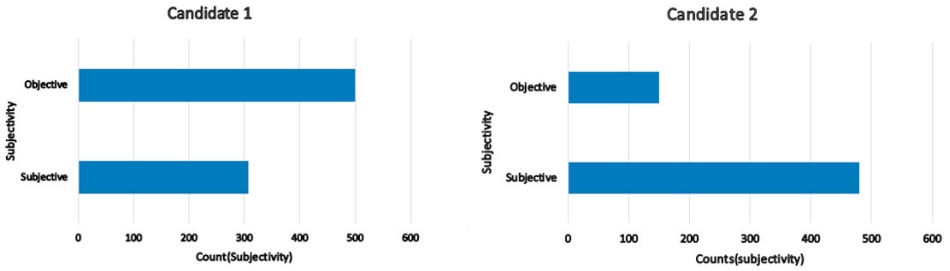

Fig. 8. Subjectivity Classification for (a) Candidate-1 (b) Candidate-2.

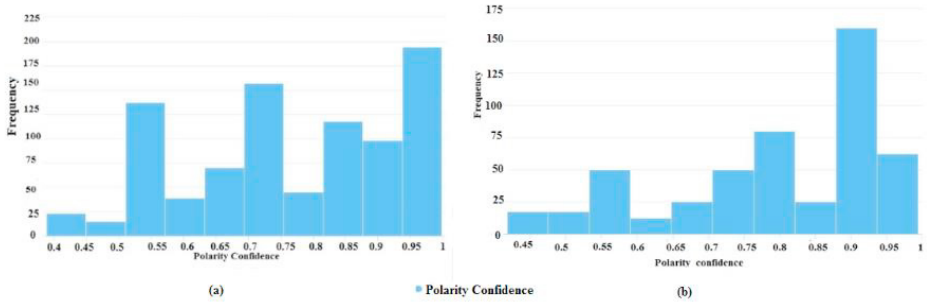

Fig. 9. Confidence Level Measure for (a) Candidate-1 and (b) Candiadate-2.

5. Conclusion and Future Direction

The beauty of this works brings together text mining and sentiment analysis which otherwise are discussed as separate topics, this work has amalgamated the two into one. The approach followed in this paper was sound and systematic. It included analysis of tweet text before performing sentiment analysis on them. The data set used in paper is primary and adequate in size. The tools selected in this paper was based on intensive literature review. It is known that the tweet has no particular format, the only limitation imposed is with the length of the text. As the tweets were collected from the online source, hence contained no sentiment labels. The absence of the sentiment labels and the noisy nature of the tweets added complexity to the paper.

The sentiment analysis was performed using different tools like R, Rapid Miner's ALYLIEN extension. Though the corpus of the tweets was collected in different point of time giving us a variety of views. Average of the views have been considered in this study. Almost same results were obtained from them with little difference, implying that the resulting output obtained is valid.

From the results and the analysis, it was concluded that Candidate-1 is more liked and is popular among the public, as compared to Candidate-2. Though the mixed feelings for both the candidates were almost same but people have more positive feelings towards candidate-1 than Candidate-2. This concluded that Candidate-1 was their first choice for prime minister's position in India.

The actual result of the election obtained in May 2019 was in full compliance with the results obtained during experiment from Jan 2019 to March 2019. This shows that the approach applied in this paper for performing sentiment analysis was apt. The scope of the paper is limited to the textual tweet data obtained from twitter. In future some additional data like pictures, sounds, online articles, multimedia etc. could be added to this data set. In this work, geo-spatial information like longitude, latitude is completely ignored. In future all these characteristics would be considered. Another future recommendation would be to use hash tags on the top of the tweet sentiments as features for classification of the text.

References

[1] Younis, Eman MG. (2015) "Sentiment analysis and text mining for social media microblogs using open source tools: an empirical study." *International Journal of Computer Applications*, 112(5) :0975 – 8887.

[2] Alhum oud, Sarah, Tarfa Albuhairi, and Wejdan Alohaideb. (2015) "Hybrid sentiment analyser for Arabic tweets using R." *7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K),* 1: 417-424.

[3] Pak, Alexander, and Patrick Paroubek (2010) "Twitter as a corpus for sentiment analysis and opinion mining." *Proceeding of LREC*, 10.

[4] Garg, Neha, and Rinkle Rani (2017) "Analysis and visualization of Twitter data using k-means clustering." *International Conference on Intelligent Computing and Control Systems (ICICCS), IEEE*, 670-675.

[5] Das, Sourav, and Anup Kumar Kolya (2017)"Sense GST: Text mining & sentiment analysis of GST tweets by Naive Bayes algorithm." *Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), IEEE*, 239-244.

[6] Mishra, Prerna, Ranjana Rajnish, and Pankaj Kumar. (2016) "Sentiment analysis of Twitter data: Case study on digital India." *International Conference on Information Technology (InCITe)-*The Next Generation IT Summit on the Theme-Internet of Things: Connect your Worlds. IEEE, 148 – 153.

[7] Fiaidhi, Jinan, et al. (2012) "Opinion mining over twitter space: Classifying tweets programmatically using the R approach." *Seventh International Conference on Digital Information Management (ICDIM). IEEE*, 313-319.

[8] Chen, Hsinchun, and David Zimbra. (2010) "AI and opinion mining." *IEEE Intelligent Systems*, 25(3): 74-80.

[9] Kohli, Shruti, and Himani Singal. (2014) "Data analysis with R." *Proceedings of the IEEE/ACM 7th International Conference on Utility and Cloud Computing.* IEEE Computer Society.

[10] Devika, R., and S. Revathy. (2018) "Survey on clustering techniques in Twitter data." *Second International Conference on Computing Methodologies and Communication (ICCMC). IEEE,* 1073-1077.

[11] Kamal, Rashid, et al. (2019) "Emotion Classification and Crowd Source Sensing; A Lexicon based Approach." *IEEE Access*, 27124-27134.

[12] Ravichandran, M., G. Kulanthaivel, and T. Chellatamilan. (2015) "Intelligent topical sentiment analysis for the classification of e-learners and their topics of interest." *The Scientific World Journal, 2015*, 617358.

[13] Salas-Zárate, María del Pilar, et al. (2017) "Sentiment analysis on tweets about diabetes: an aspect-level approach." *Computational and mathematical methods in medicine,* 2017: 5140631.

[14] Serrano-Guerrero, Jesus, et al. (2015) "Sentiment analysis: A review and comparative analysis of web services." *Information Sciences* 31(1): 18-38.

[15] Kumar, K. Sailaja, et al. (2016) "Identify the influential user in online social networks using R, Hadoop and Python." *International Conference on Circuits, Controls, Communications and Computing (I4C). IEEE*, 1 – 6.

[16] Geetha, R., Pasupuleti Rekha, and S. Karthika. (2018) "Twitter Opinion Mining and Boosting Using Sentiment Analysis." *International Conference on Computer, Communication, and Signal Processing (ICCCSP). IEEE,* 1 – 4.

[17] Mostafa, Mohamed M. (2013) "More than words: Social networks' text mining for consumer brand sentiments." *Expert Systems with Applications* 40(10): 4241- 4251.

[18] Colneriĉ, Niko, and Janez Demsar. (2018) "Emotion recognition on twitter: Comparative study and training a unison model." *IEEE Transactions on Affective Computing,* 1949 - 3045.

[19] Pavel, Andrei, et al. (2017) "Using Short URLs in Tweets to Improve Twitter Opinion Mining." *16th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE*, 965 – 970.

[20] Ohana, B., & Tierney, B. (2009). "Sentiment classification of reviews using SentiWordNet". *In 9th. it & t conference,* 13: 18-30.

[21] Tripathi, Pragya, Santosh Kr Vishwakarma, and Ajay Lala. 2015 "Sentiment analysis of English tweets using rapid miner." *International Conference on Computational Intelligence and Communication Networks (CICN). IEEE*, 668 – 672.

[22] Taj, Soonh, Baby Bakhtawer Shaikh, and Areej Fatemah Meghji. (2019) "Sentiment Analysis of News Articles: A Lexicon based Approach." *2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). IEEE*, 1 – 5.

[23] Joseph, George, and Vinu Varghese. (2019) "Analyzing Airbnb Customer Experience Feedback Using Text Mining." *Big Data and Innovation in Tourism, Travel, and Hospitality. Springer*, Singapore, 147-162.

[24] Anandarajan, Murugan, Chelsey Hill, and Thomas Nolan. (2019) "Sentiment Analysis of Movie Reviews Using R." *Practical Text Analytics. Springer,* Cham, 193-220.

[25] Joshi, Sunil. "Sentiment Analysis on WhatsApp Group Chat Using R. (2019) " *Data, Engineering and Applications. Springer,* Singapore, 47.

[26] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,* 168 -177.

[27] Plutchik, Robert. (2001) "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice." *American scientist 89* (4) :344-350.

[28] Kumar, Akshi, and Geetanjali Garg. "Sentiment analysis of multimodal twitter data." Multimedia Tools and Applications 78.17 (2019): 24103-24119.

[29] Wang, Lei, Jianwei Niu, and Shui Yu. "SentiDiff: Combining Textual Information and Sentiment Diffusion Patterns for Twitter Sentiment Analysis." IEEE Transactions on Knowledge and Data Engineering (2019).