A

Assignment Report on:

"TWITTER BASED ELECTION PREDICTION AND ANALYSIS"

Prepared by : **ADICHERLA VENKATA SAI**

Roll. No. : **P19CO016**

Class : **M. Tech. –I (Computer Engineering) 2st Semester**

Year : **2019-20**

Department of Computer Engineering

Sardar Vallabhbhai National Institute of Technology,

Surat -395007 (Gujarat), India

Sardar Vallabhbhai National Institute of Technology,

Surat -395007 (Gujarat), India

CERTIFICATE

This is to certify that the Assignment report entitled **TWITTER BASED ELECTION PREDICTION AND ANALYSIS** is prepared and presented by **Mr. ADICHERLA VENKATA SAI** bearing Roll No. : **P19CO016**, 1st Year of M**. Tech. (Computer Engineering)** and his work is satisfactory.

GUIDE **JURY(s)** **HOD**

Contents Page.no.

List of Figures

ABSTRACT

Elections are conducted to have the public opinion, and they choose the candidate by voting them. Many methods are used to predict the results of elections like agencies and media conduct pre polls survey and expect views to predict the election result. In this twitter data is used to predict outcome of the election, by collecting the data and analysing it and sentiment analysis of twitter data for the predictions of election outcome. Method used is Lexicon based approach with machine learning to find emotions in tweets and predict the sentiment score.

1. Introduction

In democracy, election has a major role in selecting the candidates as leaders. It's called the instrument of democracy where the voters communicate with representatives. It's been very interesting to predict the outcome of the elections due their important role in politics.

The vital component in an election is that election polls/survey. An opinion poll has existed since the early 19th century based on [1]. And currently, there are many scientifically proven statistical models to forecast an election, as shown in [2]. But sometimes, even in the developed countries, the polls failed to accurately predict the election outcomes. [3] listed several failed polls result such as in the 1992 British General Elections, the 1998 Quebec Elections, the 2002 and 2007 French presidential elections, the 2004 European elections in Portugal, the 2006 Italian General Elections, and the 2008 Primary Elections in the States.

Lately, it is observed that traditional polls may fail to make an accurate prediction. The scientific community has turned its interest in analysing web data, such as blog posts or social networks' users' activity as an alternative way to predict election outcomes, hopefully more accurate. Furthermore, traditional polls are too costly, while online information is easy to obtain and freely available. This is an interesting research area that combines politics and social media which both concern today's society. It is interesting to employ technology to solve modern-day challenges.

Trying to resolve the accuracy and high cost problem, we study the possibility of using data from social media as the data source to predict the outcome of an election. Social media has become the most popular communication tool on the internet. Hundreds of millions of messages are being posted every day in the popular social media sites such as Twitter4 and Facebook5. [4] Stated in their paper that social media websites become valuable sources for opinion mining because people post everything, from the details of their daily life, such as the products and services they use, to opinions about current issues such as their political and religious views. The social media providers enable the users to express their feelings or opinions as much as possible to increase the interaction between the users and their sites. This means that the trend on the internet is shifting from the quality and lengthy blog posts to much more numerous short posts that are posted by a lot of people. This trait is very valuable as now we can collect different kind of people's opinions or sentiments from the social web.

One of the social media that allows researchers to use their data is Twitter. Twitter is a micro blogging web service that was launched in 2006. Now, it has more than 200 million visitors on a monthly basis and 500 million messages daily. The user of twitter can post a message (tweet) up to 140 characters. The message is then displayed at his/her

personal page (timeline). Originally, tweets were intended to post status updates of the user, but these days, tweets can be about every imaginable topic. Based on the research in [5], rather than posting about the user's current status, conversation and endorsement of content are more popular. The advantages of using tweets as a data source are as follows; first, the number of tweets is very huge and they are available to the public. Second, tweets contain the opinion of people including their political view.

1.1 Objective

The objectives are:

- To implement an algorithm for automatic classification of text into positive and negative.
- Sentiment analysis to determine the attitude of the mass is positive or negative or neutral towards the subject of interest.
- Graphical representation of the sentiment.(Pie Chart/Bar Diagram /Scatter Plot).

1.2 Problem Definition

A major benefit of social media is that people talk good and bad about particular things (like brands/ items/ trending issues/ personality). The bigger your company gets difficulties it becomes to keep a handle on how everyone feels about your brand. For large companies with thousands of daily mentions on social media, news sites and blogs. It's extremely difficult to do this manually.

To combat this problem, sentiment analysis is necessary, here we use people sentiment about particular things and analyze them.

1.3 Proposed System

The main goal and challenge of the system is analyzing the data (twitter data)to see the impact of twitter on particular state election system. The proposed system is analyzing data which is based on the mechanism that analyses User Tweets using Hash tags and Keywords. Collets the twitter data using Hash tags of popular personalities/parties who might be participating in the election. General Public orientation toward these parties can be studied using the tweets of the people having posted on the Twitter. Tweets are generally launched by academics, journalists and politicians, for its potential political value. Many politicians make use of this micro blogging site to express themselves.

These tweets can be categorized on various policies such as geo location analysis to analyze the peoples view for that particular area which might help parties to design their winning strategy. The proposed system mainly focuses on collection of tweets to make volume analysis to and out the popular days of election. A trend analysis to and a popular or trending party/candidate and a sentiment analysis to actually bifurcate the positive and negative tweets for the party/candidate so that making trend analysis on this tweets can help user to actually make a clear opinion about any party/candidate.

This will be done by collecting data from the twitter and we then deal with loading the data for further analysis like Volume analysis. Trend analysis, Sentiment analysis.

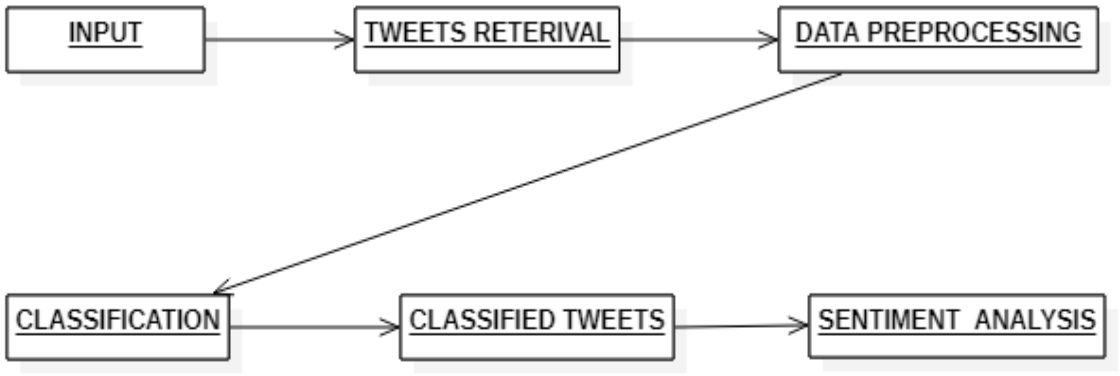


Fig. 1.3 Proposed System

2. Related Work

The literature survey is most important step in the software development process. Different strategies such as profile details, user behavior, Twitter specific feature (reply/re-tweet), user graph and sentiment from tweet content can be used for inferring political leaning. For example, In [6], the authors used tweet containing parties' name in several political events to assign a political/ideological leaning of the user who posted the tweets. Similar to the previous method, [7] used the tweets and re tweets of a user regarding a political party to infer the political leaning. [8] Assigned a score to every congress member which a Twitter user is following, then a political preference is assigned based on that score. In [9], the authors compared several features such as user's bio and avatar, posting behavior, linguistic content, follower, reply and re tweet. They found out that the combination between user profile and linguistic outperform other feature. They then applied to classify the ethnicity of the user and whether the user is a Starbucks fan, but their result showed that information from user bio is more accurate for classifying Starbucks fan, and user's avatar for classifying user's ethnic.

The second approach is by using selected data just days or weeks prior to the election. The prediction could be derived by comparing the number of tweets mentioning each candidate or by comparing the number of tweets that has positive sentiments towards each candidate. The earliest research stated that the number of tweets mentioning a party reflects the election result was shown in [10] where they found out that the prediction result from Twitter were only better than other. While [11] is the first research in which argued that sentiment detection approach from Twitter can replace the expensive and time intensive polling?

Researchers have tried to compare these two methods, for example, [12] that tried to predict congress and senate election in several states of the US. They showed that though the method is the same, the prediction error can vary greatly. The research also showed that lexicon based sentiment analysis improves the prediction result, but the improvements also vary in different states. Same result was shown in [13] where they predict the result of Irish general election using both methods and [14] which predicts the Italian primary election. All of the research showed that sentiment detection does reduce the error of the prediction result. Because of that, several researchers focused on improving the sentiment analysis, such as[14] and [15] who used more sophisticated sentiment analysis than lexicon based in the US presidential election, France legislative election, and Italy primary election.

Other than using sentiment analysis, the prediction result from Twitter can be improved by using user normalization. This is based on the fact that in an election, one person only has one vote. [16] Implemented this method and showed that the prediction result of 2011 Dutch senate election was improved. [17] Takes further step by adding census

correction on the user normalization. [18] Also implemented this method in several South American countries. He collected more than 400 million of tweets, and got a very good result (low difference with the election result) predicting Venezuela presidential election. But when applying in Ecuador and Paraguay presidential election that has much less dataset, the error of the prediction increases significantly.

Other methods proposed by researchers are by:

(1) Utilizing interaction information between potential voter and the candidates.
(2) Creating trend line from the changes in follower of the candidates.

[19] Used interaction information such as the number of interaction, the frequency of interaction, the number of positive and negative terms in the interactions in the Canadian legislative election. The candidates were grouped into four parties, and based on their result; they argued that that the generated content and the behaviour of users during the campaign contain useful knowledge that can be used for predicting the user's preference. [20] Tried to utilize the size of candidates' network (follower in Twitter and friend in Facebook), but the result showed that it was not a good predictor of election results. One interesting result from their research is that despite the huge size of social media, it has small effect on the election results. Therefore, it only makes a difference in a closely contested election.

However, there are several researchers arguing that research in this area is still premature and requires a lot of development before it can give satisfying prediction result. [21] Argued that prediction model using Twitter only able to predict the result from the top candidates/parties and slight variable changes in the model did impact the prediction result. In [22], the authors listed several drawback of the research in this topic such as, most predictions are actually a post-hoc analysis, no commonly accepted way exists for "counting votes", the sentiment analysis methods are not reliable, no data cleansing step, demography and self selection bias has not been addressed. In [23], in addition to previously stated drawbacks, gave several suggestions such as the importance of geographical and demographical bias, the noise in the social media, the reproducibility of proposed methods, and MAE should be use rather than only winner prediction.

3. Methodology

3.1 Data Collection

The data collection step is the initial phase, where data is collected from twitter. There are two methods on how to connect and collect tweets from Twitter. The first method is by searching tweets matching to keywords. The second method is by collecting all the tweets provided by Twitter through streaming API, or all the tweets in a specific language, or all the tweets in a specific location then put all of them into the database.

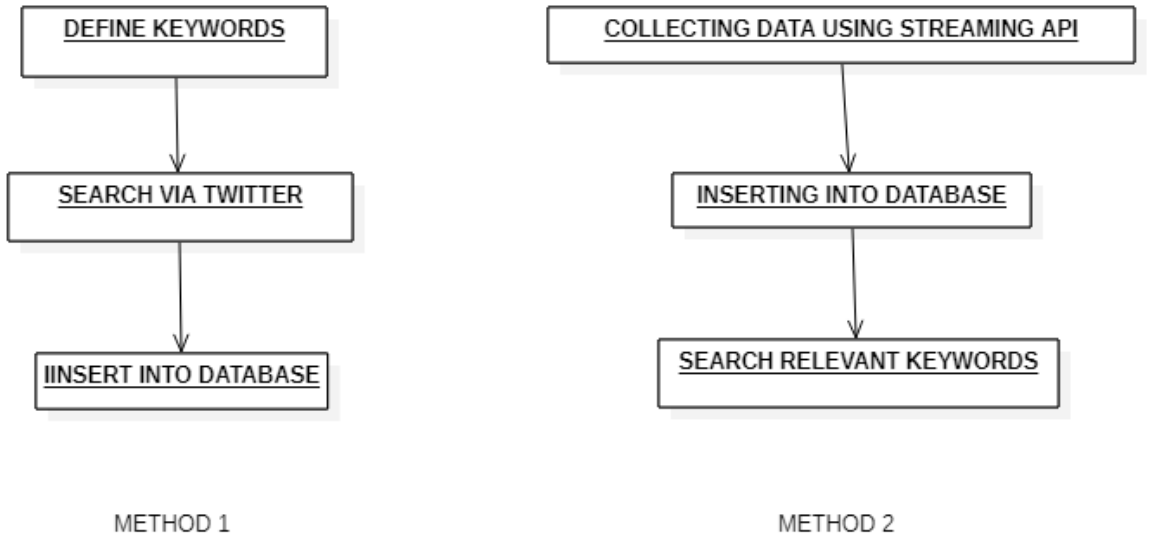


Fig. 3.1 Data Collection methods

Both methods have their own advantages and disadvantages. The first method requires only small storage as the data is relatively small. The downside is that researcher cannot get data from other keywords (if her needs to), from an earlier time. Twitter allows the search API only for 7 days backwards. This data collection method is suitable if the focus of the research is on the feature extraction or the prediction method. With the second method, researcher can apply many set of keywords to get the best result.

3.2 Pre processing

Many current methods for text sentiment analysis contain various pre-processing steps of text. One of the most important goal of pre-processing is to enhance the quality of the data by removing noise. Another point is the reduction of the feature space size.

a) Lower Case Conversion:

Because of the many ways people can write the same things down, character data can be difficult to process. String matching is another important criterion of feature selection. For accurate string matching we are converting our complete text into lower case.

b) Removing Punctuations and Removing Numbers:

All punctuations, numbers are also need to remove from reviews to make data clean and neat. Unnecessary commas, question marks, other special symbols get removed in this case. Here we are not removing dot (.) symbol from our review s because they split our text into sentences.

c) Stemming:

It is the method of conflating the variant styles of a word into a standard illustration, the stem. For example, the words: “presentation”, “presented”, “presenting” could all be reduced to a common representation “present”. The is a widely used procedure in text processing for information retrieval based on the assumptions that words presentation and presented. Stemming in this case is helpful in correct words matching and counting case.

d) Striping White Spaces:

In this pre-processing step all the text data is cleaned off. All the unnecessary white spaces, tables, newline character get removed from the text.

4. Sentiment Analysis

4.1 Machine Learning Approach:

There are two approaches of machine learning, supervised and unsupervised. in our research we used supervised machine learning approach.

In supervised machine learning approach there is finite set of classes for classification. Training dataset is also available. Most research papers do not use the neutral class, which makes the classification problem considerably easier, but it is possible to use the neutral class. Given the training data, the system classifies the document by using one of the common classification algorithms such as Support Vector Machine, Naïve Bayes etc. We used naive bays for classification of tweets. We classified tweets into polarity and emotion also using naive bays classifier.

Naive Bayes is a machine learning algorithm for classification problems. It is based on Bayes' probability theorem. It is primarily used for text classification that involves high dimensional knowledge sets. A few examples are spam filtration, sentimental analysis, and classifying news articles.

It is not only known for its simplicity, but also for its effectiveness. It is fast to build models and make predictions with Naive Bayes algorithm.

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right) * P(A)}{P(B)}$$

Where,

$P\left(\frac{A}{B}\right)$: $Conditional\ Probability\ of\ occurence\ of\ event\ given\ the\ event\ B\ is\ true$

$P(A)\ and\ P(B)$: $Probabilities\ of\ the\ occurance\ of\ event\ A\ and\ B\ respectively$

$P\left(\frac{B}{A}\right)$: $Probability\ of\ the\ occurance\ of\ the\ event\ B\ givent\ the\ event\ A\ is\ true$

4.2 Lexicon Based Approach:

There three main approaches to compile sentiment words. They are

- MANNUAL APPROACH
- DICTIONARY APPROACH
- CORPUS-BASED APPROACH

In this research we used dictionary based approach, used some variables for classification, that are sadness, tentativeness, anxiety, work, anger, certainty, achievement, positive words, negative words, positive hash tag and negative hash tag. And collected various related words to the variable to classify them.

5. Proposed Algorithm

The system uses Naïve bays approach for text categorization. For the categorization of the text, Naïve bays classifier assumes that the effect of a variable value on a given class is independent of the values of other variables. This assumption is called as conditional independence. In this paper, the proposed approach involve in both dictionary corpus based techniques which finds the semantic orientation of the sentiments in the tweets. Emoticons, neutralization, negation handling and capitalization is also considered as they are the huge part of the modern internet language. To uncover the sentiments, we will first extract the opinion words from the tweets and then we find out their orientation that is to determine whether the sentiment word reflects the feelings of the user.

The following steps will brief the process of the proposed system:

a) Retrieval of tweets:

As Twitter is the most exaggerated part of the social networking sites it consists of various blogs which are related to various topics worldwide. Instead of taking whole blogs, we'll rather search on a particular topic and extract all the tweets related to that topic.

b) Pre-processing of extracted data:

After retrieval of tweets, sentiment analysis is applied to raw tweets but in most of the cases results in very poor performance. Therefore, pre techniques are necessary for obtaining better results.

The process involves the following steps:

i. Filtering: It is nothing but the cleaning of raw data. In this step , URL links (E.g. https://twitter.com), special words in twitter(E.g. “RT” which mean Re-tweets), usernames in twitter are removed and emotions are replaced with special strings.

ii. Tokenization: It is segmentation of sentences. Here we’ll tokenize or segment the text with the help of splitting text by spaces and punctuation marks to form a container of words.

c) Sentiment scoring module:

The basic feature of this model is Polarity of the words. A dictionary which contains a list of English words and score which ranges from 1 to 3. The scoring module is used to determine the sentiment of the textual data.

d) Output sentiment:

To show how the social media like twitter can be used to make prediction of future outcome such as election. Specifically by using R, to extract the sentiment or view of people who are likely to vote in the general election or have an influence on those who will vote, and sentiment analysis to classify their sentiment.

6. Implementation

R Programming language is(language and software environment) for statistical analysis, graphics representation and reporting. R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, it is currently developed by the R Development Core Team.

Code:

Install the respective packages:

```
#install.packages("twitteR")
#install.packages("RCurl")
#install.packages("ROAuth")
```

Load the packages:

```
library(twitteR)
library(RCurl)
library(ROAuth)
```

API keys and Tokens for Retrieving tweets from Twitter :

```
api_key<-"TCZ9ExykX4BzItaKseWueIpSs"
api_secret<-"1WPMZ85Cko6DtkmkUFzKpsWP0ZbmK16bmzeeYgVvYJHJeP512y"
access_token<-"3258828348-QL7t9Ej40qR9sSNaiJeJhYC8KRmJmmx3YvbFAgH"
access_token_secret<-"Yx1clNeau00XpY0h6auwN2wBY3CCtaqgkgxDRQeV7qlii"
```

Setup Twitter Oath connection to API keys :

```
setup_twitter_oauth(api_key,api_secret,access_token,access_token_secret)
```

Search Twitter :

```
mytweet<-searchTwitter('$narendramodi',n=1500,lang='en')
mytweet

mytweet1<-searchTwitter('$rahulgandhi',n=1500,lang='en')
mytweet1

n.tweet <- length(mytweet)
n.tweet1 <- length(mytweet1)
head(n.tweet)
head(n.tweet1)
```

Convert tweets to Data Frames:

```
mytweet.df<-twListToDF(mytweet)
mytweet1.df<-twListToDF(mytweet1)
head(mytweet.df$text)
head(mytweet1.df$text)
```

Retrieved Tweets are stored in .csv file as election:

```
write.csv(mytweet.df,file='/Users/DELL/Desktop/election.csv')
write.csv(mytweet1.df,file='/Users/DELL/Desktop/election1.csv')
#head(election)
#head(election1)
```

Read File :

```
election<-read.csv(file.choose(),header=T)
str(election)
election1<-read.csv(file.choose(),header=T)
str(election1)

#look over data
head(election)
head(election1)
```

Build Corpus:

```
#install.packages("tm")

#loading data into corpus
library(tm)
corpus <- iconv(election$text,to = "utf-8")
corpus <- Corpus(VectorSource(corpus))
inspect(corpus[1:1000])

library(tm)
corpus1<- iconv(election1$text, to = "utf-8")
corpus1 <- Corpus(VectorSource(corpus1))
inspect(corpus1[1:1000])
```

Clean Text :

Lower text Conversation:

```
corpus <- tm_map(corpus, tolower)
```

```
inspect(corpus[1:1000])
corpus1 <- tm_map(corpus1, tolower)
inspect(corpus1[1:1000])
```

Removing Panctuations:

```
corpus <- tm_map(corpus, removePunctuation)
inspect(corpus[1:1000])
corpus1 <- tm_map(corpus1, removePunctuation)
inspect(corpus1[1:1000])
```

Removing Numbers:

```
corpus <- tm_map(corpus, removeNumbers)
inspect(corpus[1:1000])
corpus1 <- tm_map(corpus1, removeNumbers)
inspect(corpus1[1:1000])
```

Remove Common English words:

```
cleanset <- tm_map(corpus, removeWords, stopwords('english'))
inspect(cleanset[1:1000])
cleanset1 <- tm_map(corpus1, removeWords, stopwords('english'))
inspect(cleanset1[1:1000])
```

Remove URLs:

```
removeURL <- function(x) gsub('http[[:alnum:]]*', '', x)
cleanset <- tm_map(cleanset, content_transformer(removeURL))
inspect(cleanset[1:1000])
removeURL <- function(x) gsub('http[[:alnum:]]*', '', x)
cleanset1 <- tm_map(cleanset1, content_transformer(removeURL))
inspect(cleanset1[1:1000])
```

Remove White Space:

```
cleanset <- tm_map(cleanset, stripWhitespace)
inspect(cleanset[1:1000])
cleanset1 <- tm_map(cleanset1, stripWhitespace)
inspect(cleanset1[1:1000])
```

Convert Unstructured data to stuructured data using Term Document Matrix:

```
tdm <- TermDocumentMatrix(cleanset)
```

```
tdm
tdm1 <- TermDocumentMatrix(cleanset1)
tdm1
```

Build Matrix for data:

```
tdm <- as.matrix(tdm)
tdm[1:10, 1:20]
tdm1 <- as.matrix(tdm1)
tdm1[1:10, 1:20]
```

Calculating the words how many times repeated and Build barplot:

```
w <- rowSums(tdm)
w1 <- rowSums(tdm1)
barplot(w,w1, las = 2, col = rainbow(50))
```

Sentiment analysis:

```
#install the respective packages into Rstudio(uncomment is not installed)
#install.packages("syuzhet")
#install.packages("lubridate")
#install.packages("ggplot2")
#install.packages("scales")
#install.packages("reshape2")
#install.packages("dplyr")

library(syuzhet)
library(lubridate)
library(ggplot2)
library(scales)
library(reshape2)
library(dplyr)
```

Reading file:

```
election <- read.csv(file.choose(), header = T)
mytweet <- iconv(election$text, to = 'utf-8')
election1 <- read.csv(file.choose(), header = T)
mytweet1 <- iconv(election1$text, to = 'utf-8')
```

Obtain sentiment scores:

```
s <- get_nrc_sentiment(mytweet)
s1 <- get_nrc_sentiment(mytweet1)
head(s)
```

```
head(s1)
```

Barplot of Sentiment Analysis:

```
c1<-colSums(s)
c2<-colSums(s1)

sampledata<-data.frame(c1,c2)
barplot(height <- rbind(c1, c2),
        las = 2,
        col = c("blue","red"),
        ylab = 'Count',
        beside = TRUE,
        main = 'Sentiment Scores for Narendra Modi Vs Rahul Gandhi')
```

7. Results and Discussions

The data (India Rajya Sabha election) collected through twitter API, after performed pre-processing on it and classified the polarity of Positive and Negative. The sentiment analysis positive tweets were more in Narendra Modi data than that of Rahul Gandhi. As prediction we can predict that Narendra Modi will win elections from twitter data we collected. The emotional analysis for the variables such as joy. Surprise, anger, disgust, fear, and sadness.

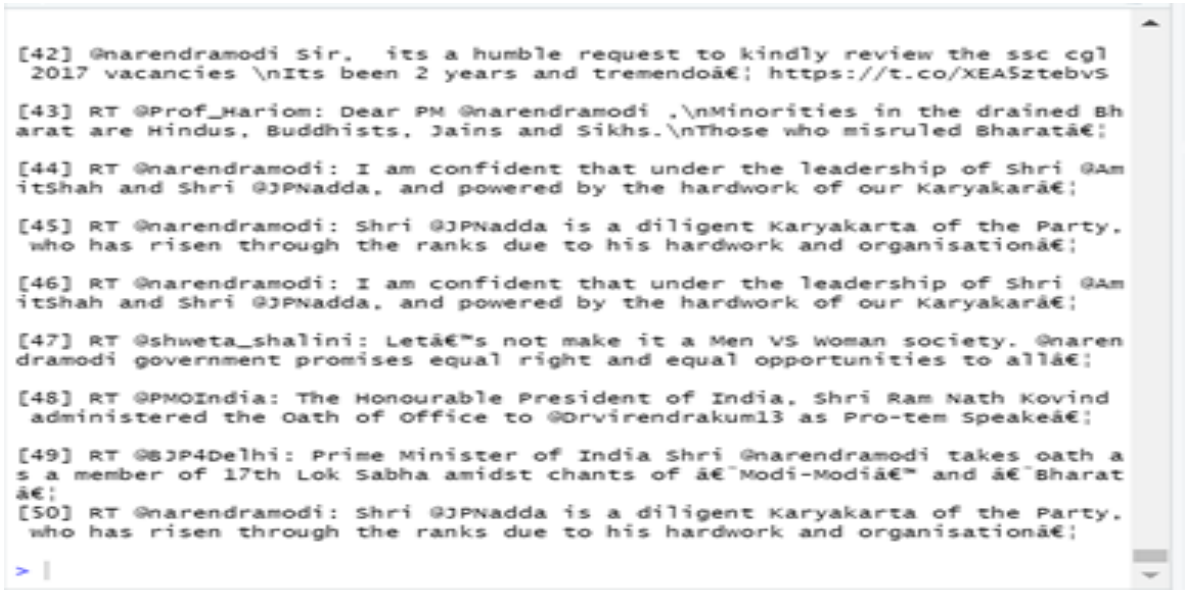


Fig. 7.1 Tweets for Narendra Modi

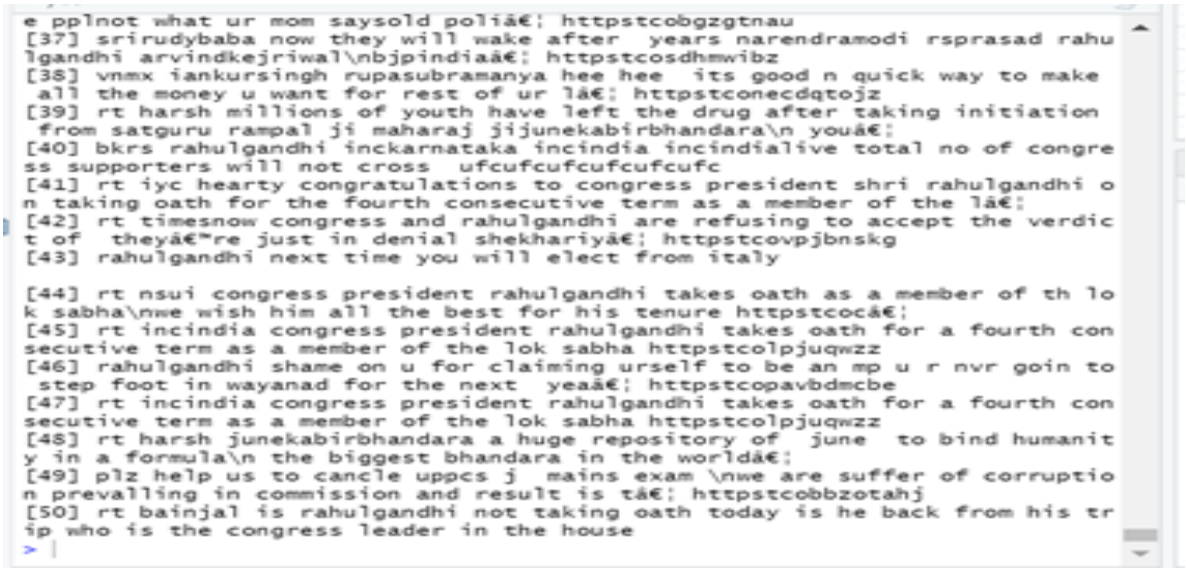
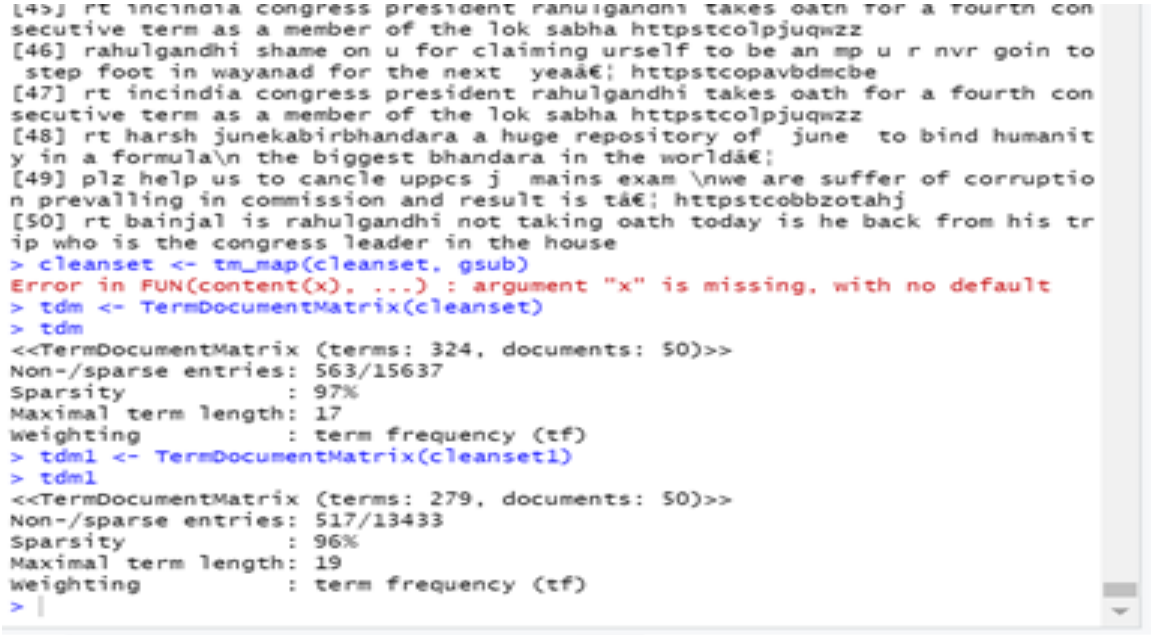
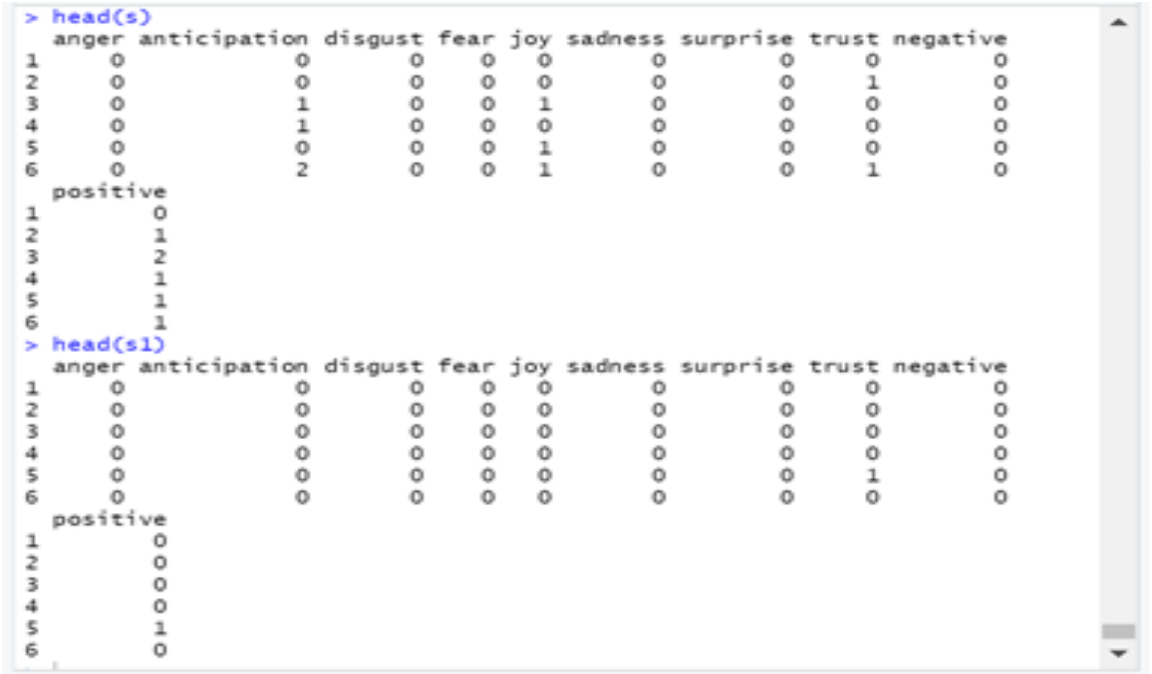


Fig. 7.2 Tweets for Rahul Gandhi



```
[45] rt incindia congress president rahulgandhi takes oath for a fourth con
secutive term as a member of the lok sabha httpstcolpjuqwzz
[46] rahulgandhi shame on u for claiming urself to be an mp u r nvr goin to
 step foot in wayanad for the next  yeaâ€¦ httpstcopavbdmcbe
[47] rt incindia congress president rahulgandhi takes oath for a fourth con
secutive term as a member of the lok sabha httpstcolpjuqwzz
[48] rt harsh junekabirbhandara a huge repository of  june  to bind humanit
y in a formula\n the biggest bhandara in the worldâ€¦
[49] plz help us to cancle uppcs j  mains exam \nwe are suffer of corruptio
n prevalling in commission and result is tâ€¦ httpstcobbzotahj
[50] rt bainjal is rahulgandhi not taking oath today is he back from his tr
ip who is the congress leader in the house
> cleanset <- tm_map(cleanset, gsub)
Error in FUN(content(x), ...) : argument "x" is missing, with no default
> tdm <- TermDocumentMatrix(cleanset)
> tdm
<<TermDocumentMatrix (terms: 324, documents: 50)>>
Non-/sparse entries: 563/15637
Sparsity           : 97%
Maximal term length: 17
Weighting          : term frequency (tf)
> tdm1 <- TermDocumentMatrix(cleanset1)
> tdm1
<<TermDocumentMatrix (terms: 279, documents: 50)>>
Non-/sparse entries: 517/13433
Sparsity           : 96%
Maximal term length: 19
Weighting          : term frequency (tf)
>
```

Fig. 7.3 Term Document Matrix for Sentiment words



```
> head(s)
  anger anticipation disgust fear joy sadness surprise trust negative
1     0            0       0    0   0       0        0     0        0
2     0            0       0    0   0       0        0     1        0
3     0            1       0    0   1       0        0     0        0
4     0            1       0    0   0       0        0     0        0
5     0            0       0    0   1       0        0     0        0
6     0            2       0    0   1       0        0     1        0
  positive
1        0
2        1
3        2
4        1
5        1
6        1
> head(s1)
  anger anticipation disgust fear joy sadness surprise trust negative
1     0            0       0    0   0       0        0     0        0
2     0            0       0    0   0       0        0     0        0
3     0            0       0    0   0       0        0     0        0
4     0            0       0    0   0       0        0     0        0
5     0            0       0    0   0       0        0     1        0
6     0            0       0    0   0       0        0     0        0
  positive
1        0
2        0
3        0
4        0
5        1
6        0
```

Fig. 7.4 Different emotions on each Tweet

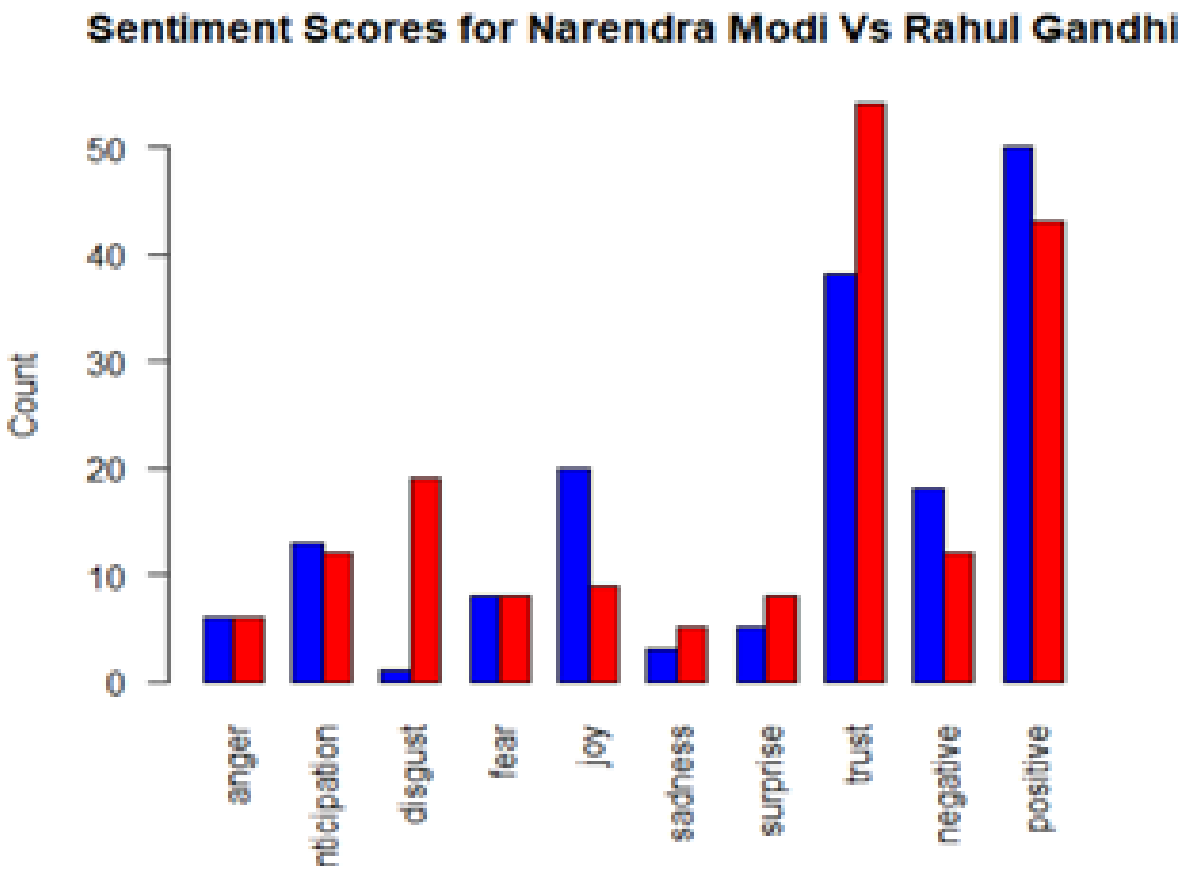


Fig. 7.5 Sentiment and Emotion analysis

Observation: Disgust and Joy were among most expressed emotions in our data. Rahul Gandhi had more tweets about trust and Narendra Modi has more tweets about Positive. And Narendhra Modi will get benefit of positive tweets and Rahul Gandhi has tweets are less on other emotions.

Note: The experimental results are based on limited no of tweets (50) in practise we could work on more.

8. Conclusion

In this paper, the social media like twitter can be used to make prediction of future outcome such as election. The sentiment or views of people who are likely to vote in the general election or have an influence on those who will vote, and sentiment analysis is done to classify their sentiment.

Data pre-processing is done to get more parameters to result in the best sentiment. Updating Dictionary for new synonym and antonyms of already existing words will help a lot better. Context sentimental analysis may be implemented in future for accuracy purposes.

9. References

[1] Hillygus, D. S. (2011). The evolution of election polling in the United States. Public opinion quarterly, 75(5),, 962- 981.

[2] Lewis Beck, M. S. (2005). Election forecasting: principles and practice. The British Journal of Politics & International Relations, 7(2), 145-164.

[3] Fumagalli, L. &. (2011). The total survey error paradigm and pre-election polls: The case of the 2006 Italian general elections. ISER Working Paper Series. 2011-29.

[4] Pak, A. &. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. LREC.

[5] Dann, S. (2010). Twitter content classification. First Monday, 15(12).

[6] Wong, F. M. (2013). Quantifying Political Leaning from Tweets and Retweets. ICWSM.

[7] Boutet, A. K. (2012). What's in your Tweets? I know who you supported in the UK 2010 general election. Proceedings of the International AAAI Conference on Weblogs and Social Media.

[8] Golbeck, J. &. (2011). Computing political preference among twitter followers. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

[9] Pennacchiotti, M. &. (2011). Democrats, republicans and starbucks afficionados: user classification in twitter. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining , 430-438.

[10] Tumasjan, A. S. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. ICWSM, 10, 178-185.

[11] O'Connor, B. B. (2010). From tweets to polls: Linking text sentiment to public opinion time series. ICWSM, 11, 122-129.

[12] Gayo-Avello, D. M. (2011). Limits of electoral predictions using twitter. ICWSM.

[13] Bermingham, A. &. (2011). On using Twitter to monitor political sentiment and predict election results.

[14] Ceron, A. C. (2014). Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters—Evidence From the United States and Italy. Social Science Computer Review.

[15] Ceron, A. C. (2013). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. New Media & Society, 16(2), 340-358.

[16] Sang, E. T. (2012). Predicting the 2011 dutch senate election results with twitter. the Workshop on Semantic Analysis in Social Media (pp. 53-60). Association for Computational Linguistics.

[17] Choy, M. C. (2012). US Presidential Election 2012 Prediction using Census Corrected Twitter Model. arXiv preprint arXiv:1211.0938.

[18] Gaurav, M. S. (2013). Leveraging candidate popularity on Twitter to predict election outcome.Proceedings of the 7th Workshop on Social Network Mining and Analysis. ACM., 7.

[19] Makazhanov, A. R. (2014). Predicting political preference of Twitter users. Social Network Analysis and Mining, 1-15.

[20] Cameron, M. P. (2013). Can Social Media Predict Election Results? Evidence from New Zealand. No. 13/08.

[21] Jungherr, A. J. (2012). Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, to, sander, pg, & welpe,"predicting elections with twitter: What 140 characters reveal about political sentiment". Social Science Computer Review, 30(2),, 229-234.

[22] Gayo-Avello, D. (2012). No, you cannot predict elections with twitter. Internet Computing, IEEE, 16(6), 91-94.

[23] Gayo-Avello, D. (2013). A meta-analysis of state-of-theart electoral prediction from Twitter data. Social Science Computer Review.