

# **A Project Report**

on

## **Twitter based Election Prediction and Analysis**

Submitted in partial fulfilment of requirements for the award of degree of

### **Master Of Technology**

**IN**

### **Computer Science and Engineering**

**BY**

**DASARI SHIVAKUMAR**

160118742009

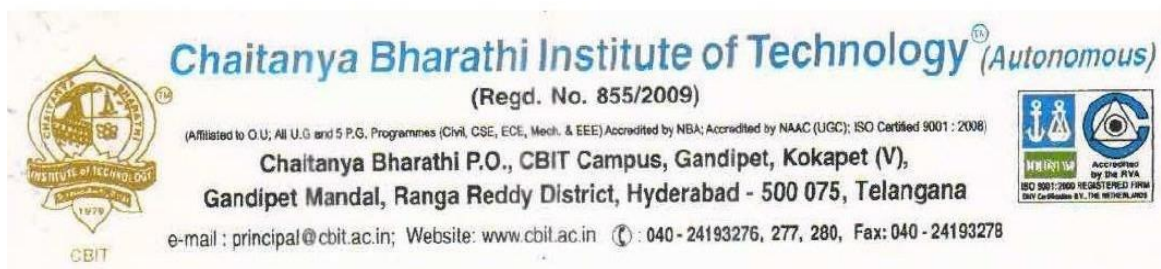
Under the guidance of

**Mr. V Madhusudan Rao**

Associate Professor



Department of Computer Science and  
Engineering Chaitanya Bharathi Institute of  
Technology (A) Hyderabad – 500075



## CERTIFICATE

This is to certify that the project work entitled **Twitter based Election Prediction and Analysis** submitted by **Dasari Shivakumar (160118742009)** in partial fulfilment of requirements for the award of degree of Master of Technology in Computer Science and Engineering as specialization is a record of the bonafide work carried out under the supervision of **Mr. V Madhusudan Rao** and this has not been submitted to any other university or institute for award of degree or diploma.

Project Guide

**Mr. V Madhusudan Rao**

Head of the Dept

**Dr. M Swamy Das**

## **DECLARATION**

I here by declare that the research work entitled **Twitter based Election Prediction and Analysis** is original and bonafide work carried out by me as a part of fulfilment for Master of Technology in Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Gandipet, Hyderabad, under the guidance of **Mr. V Madhusudan Rao , Associate Professor**, Department of CSE, CBIT.

No part of the project work is copied from books/journals/internet and wherever the partition is taken, the same has been duly referred in the text. The results are based on the project work done entirely by me and not copied from any other source.

Dasari Shivakumar

160118742009

## **ACKNOWLEDGEMENT**

The satisfaction that accompanies the successful completion of any task would be incomplete without introducing the people who made it possible and whose constant guidance and encouragement crowns all efforts with success. They have been a guiding light and source of inspiration towards the completion of the project.

I would like to express our sincere gratitude and indebtedness to guide Mr. V Madusudan Rao who has supported me throughout our project with patience and knowledge.

I am also thankful to Head of the department Dr. M. Swamy Das for providing excellent infrastructure and a conducive atmosphere for completing this project successfully.

I am also extremely thankful to our Project Coordinator Mr. V Madhusudan Rao, Associate Professor, Dept. of CSE, for his valuable suggestions and interest throughout the course of this project.

I convey my heartfelt thanks to the lab staff for allowing me to use the required equipment whenever needed.

Finally, I would like to take this opportunity to thank our families for their support through the work. I sincerely acknowledge and thank all those who gave directly or indirectly their support in completion of this work.

Dasari Shivakumar

160118742009

## **ABSTRACT**

Election is conducted to view the public opinion, where group of people choose the candidate by using votes, many methods are used to predict result. Many agencies and media companies conduct pre poll survey and expert views to predict result of election. We use twitter data to predict outcome of election by collecting twitter data and analyze it to predict the outcome of the election by analyzing sentiment of twitter data about the candidates. We used lexicon based approach with machine learning to find emotions in tweets and predict sentiment score.

## List of Figures

Sl.No.	Figure Name	Page No.
1	Data collecting Methods	9
2	Proposed System	13
3	Data Flow Diagram	14
4	Use case Diagram for Tweezer	16
5	Sequence Diagram	18
6	Activity Diagram	19
7	System Flow Diagram	19
8	Flow Chart	20
9	Retrieve the Tweets for Narendramodi	22
10	Retrieve the Tweets for Rahulgandhi	23
11	Term Document Matrix for Sentiment Words	23
12	Calculate the presence of eight different emotions on each Tweet.	24
13	India election Sentiment and Emotion analysis.	24

# TABLE OF CONTENTS

Certificate	i
Declaration	ii
Acknowledgments	iii
Abstract	iv
List of Figures	v
1. INTRODUCTION	1
1.1. Objective	2
1.2. Problem Definition	2
1.3. Proposed System	2
1.4. Organization of Report	3
2. LITERATURE SURVEY	4
3. METHODOLOGY	5
3.1. Data Collection	5
3.2. Pre Processing	6
3.3. Sentiment analysis	7
4. ARCHITECTURAL DESIGN	8
4.1 System Requirements	8
5. PROPOSED ALGORITHM	9
6. DIAGRAMMATIC REPRESENTATION (UML DIAGRAMS)	11
6.1 Technologies Used	17
7. RESULTS	22
8. CONCLUSION	26
9. References	27

# 1. Introduction

An election is a most important part in the democracy. It's the most instrument of democracy wherever the voters communicate with the representatives. Due to their important role in politics, there always has been a big interest in predicting an election outcome. It is the main instrument of democracy where the citizens communicate with the representatives. One vital component in an election is that the 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 analyzing 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. Hundreds of millions of messages are being posted every day in the popular social media sites such as Twitter<sup>4</sup> and Facebook<sup>5</sup>. [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).



## 1.1 Objective

The objectives of this project 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, negative or neutral towards the subject of interest
- Graphical representation of the sentiment in form of Pie-Chart, Bar Diagram and Scatter Plot.

## 1.2 Problem Definition

A major benefit of social media is that we can see the good and bad things people say about the particular brand or personality. The bigger your company gets difficult it becomes to keep a handle on how everyone feels about your brand. For large companies with thousand of daily mentions on social media, news sites and blogs, It's extremely difficult to do this manually.

To combat this problem, sentimental analysis software are necessary. This software's can we use to evaluate the people sentiment about particular brand or personality.

## 1.3 Proposed System

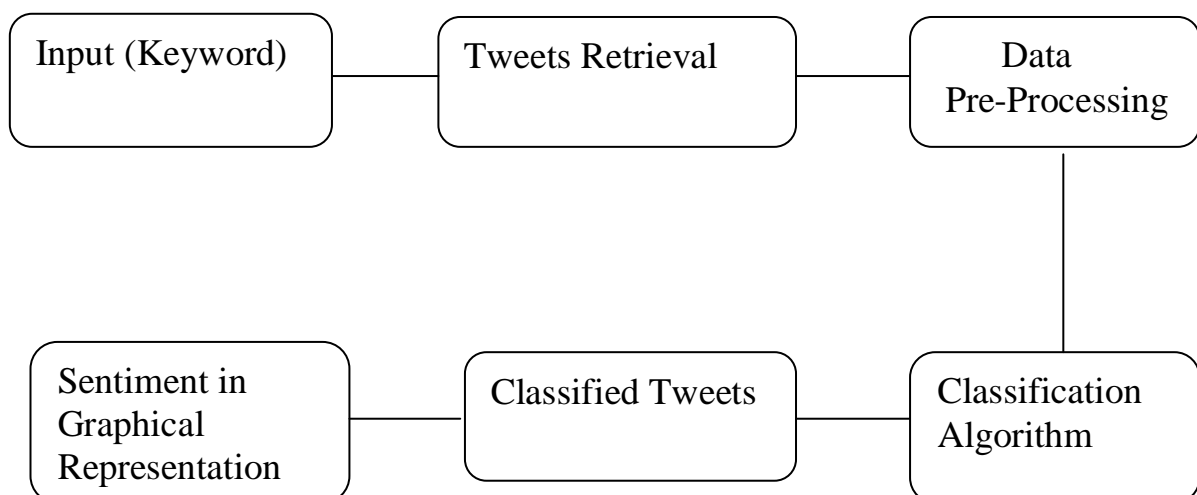


Fig 2: Proposed System.

The main goal and challenge of the system is analyzing twitter data to see the impact of twitter on Indian or particular state election. The proposed algorithm consists of three main steps. Our proposed system is analyzing system which is based on the mechanism that analyses User Tweets using Hashtags and Keywords. The proposed system collects tweets using Hash Tags which are nothing but the popular personalities/Parties. General public orientation toward these parties can be studied using the tweets the people have posted on the Twitter. Twitter is generally lauded by academicians, journalists and Politicians; for its potential political value. Many politicians make use of this micro blogging site to express themselves in the limit of 140 characters. These tweets can be categorized on various policies such as geolocation analysis to analyses 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 this party/candidate to act accordingly to improve their reputation at the same time it might help user to actually make a clear opinion about any party/candidate. This will be conducted in 3 phases. To brief about it the phase one is connecting with tweeter and downloading the tweets. The second phase deals with loading these tweets on HDFS for further analysis and the third phase is the actual analysis and they are volume analysis, Trend Analysis and Sentiment analysis.

## **1.4 Organization of Report**

This project is mainly divided into 6 modules as follows:

- Literature Survey discusses about the literature survey of this project which includes an insight into the core part of our project along with the technologies used.
- The System Architecture part deals with the design of our proposed system. The Implementation part deals with the implementation of our system which discusses about the algorithms used in building our system.
- The Result section displays our results and discussions through a series of screenshots. The final part talks about the conclusions and the future scope of our project.

## 2. Literature Survey

The literature survey is the most important step in the software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength.

Wong, F. M. proposed that Quantifying Political Leaning from Tweets and Retweets. ICWSM. 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 learning.[6]

Boutet, A. K. proposed that 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. Boutet used the tweets and retweets of a user regarding a political party to infer the political leaning.[2]

Golbeck, J. Computing political preference among twitter followers. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Golbeck Assigned a score to every congress member which a Twitter user is following, then a political preference is assigned based on that score.[4]

pennacchiotti, M. &. (2011). Democrats, republicans and starbucks aficionados: user classification in twitter. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining , 430-438. the authors compared several features such as user's bio and avatar, posting behavior, linguistic content, follower, reply and retweet. They found out that the combination between user profile and linguistic outperform other feature. [5]

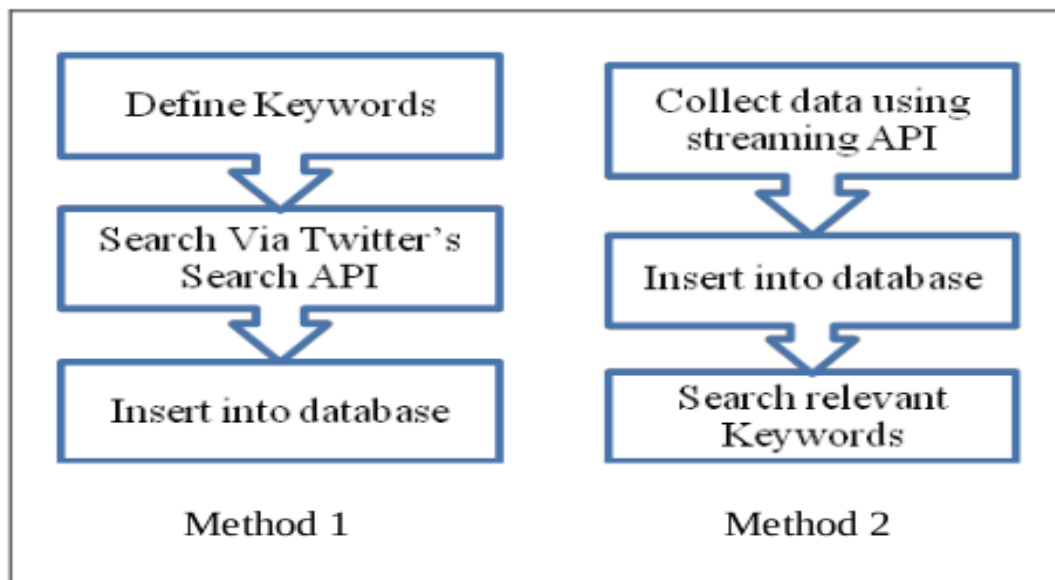
Tumasjan, A. S. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. ICWSM, 10, 178-185 The earliest research stated that the number of tweets mentioning a party reflects the election result .where they found out that the prediction result from Twitter were only better than other. [7]

O'Connor, B. B. (2010). From tweets to polls: Linking text sentiment to public opinion time series. ICWSM, 11, 122-129 is the first research in which argued that sentiment detection approach from Twitter can replace the expensive and time intensive polling? [8]

### 3.Methodology

#### 3.1 Data Collection

The data collection step is the initial phase in the research, 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 the 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-1: Data Collection Methods**

Both methods have their own advantages and disadvantages. For example, the first method requires only small storage as the data are relatively small. The downside is that researcher cannot get data from other keywords (if he 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 goals 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 reviews because are splitting our text into sentences.

### c) Stemming:

Stemming is that 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”. This is a widely used procedure in text processing for information retrieval (IR) based on the assumption that posing a query with the term presenting implies an interest in documents containing the words presentation and presented. Stemming in our case helpful in correct words matching and counting case.

### d) Striping White Spaces:

In this preprocessing step all text data is cleansed off. All unnecessary white spaces, tabs, newline character get removed from the text.

### 3.3 Sentiment analysis

#### A) 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.

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

$P(A|B)$ : Probability (conditional probability) of occurrence of event A given the event B is true.

$P(A)$  and  $P(B)$ : Probabilities of the occurrence of event A and B respectively.

$P(B|A)$ : Probability of the occurrence of event B given the event A is true.

#### B) Lexicon Based Approach:

There are three main approaches to compile sentiment words. Three main approaches are: manual approach, dictionary-based approach, and corpus-based approach. In my project I have used dictionary based approach. I have used eleven different variables for classification, that variables are sadness, tentativeness, anxiety, work, anger, certainty, achievement, positive words, negative words, positive hashtag and negative hashtag. I have collected various word related to that eleven variable and classified them.

## 4 . Architectural design

The Architectural design of a system emphasizes the design of the system architecture that describes the structure, behavior and more views of that system and analysis.

### Logical design

The logical design of a system pertains to an abstract representation of the data flows, inputs and outputs of the system. This is often conducted via modelling, using an over abstract .

### Physical design

The physical design relates to the actual input and output processes of the system. This is explained in terms of how data is input into a system, how it is verified/authenticated, how it is processed, and how it is displayed. In physical design, the following requirements about the system are decided.

1. Input requirement,
2. Output requirements,
3. Storage requirements,
4. Processing requirements,
5. System configuration,
6. Control and backup or recovery.

## 4.1 System Requirements

**Operating System :** Windows 8 or Windows 10, Mac OS X 10.8.3+

**Processor** : 2GHz Dual Core or above

**RAM** : 4GB or above

**Disk Space** : 350MB

## 5. Proposed Algorithm

The system uses Naïve Bayes Approach for text categorization. For the categorization of the text, Naïve Bayes classifiers assume 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, we propose an approach involving 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 find whether the sentiment word reflects the feelings of the user.

The following steps will brief the process of the proposed system which are

1. Retrieval of tweets
2. Pre-processing of extracted data
3. Sentiment scoring module
4. Output sentiment

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

**2. Pre-processing of extracted data:** After retrieval of tweets, Sentiment analysis tool 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: Filtering is nothing but the cleaning of raw data. In this step, URL links (E.g. <http://twitter.com>), special words in twitter (e.g. “RT” which means Retweets), usernames in twitter are removed and emoticons are replaced with special strings.

ii) Tokenization: Tokenization is nothing but Segmentation of sentences. In this step, we will tokenize or segment text with the help of splitting text by spaces and punctuation marks to form a container of word.



**3. 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.

**4. Output Sentiment:** able 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 views 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. Diagrammatic Representation

### 6.i Data Flow Diagram

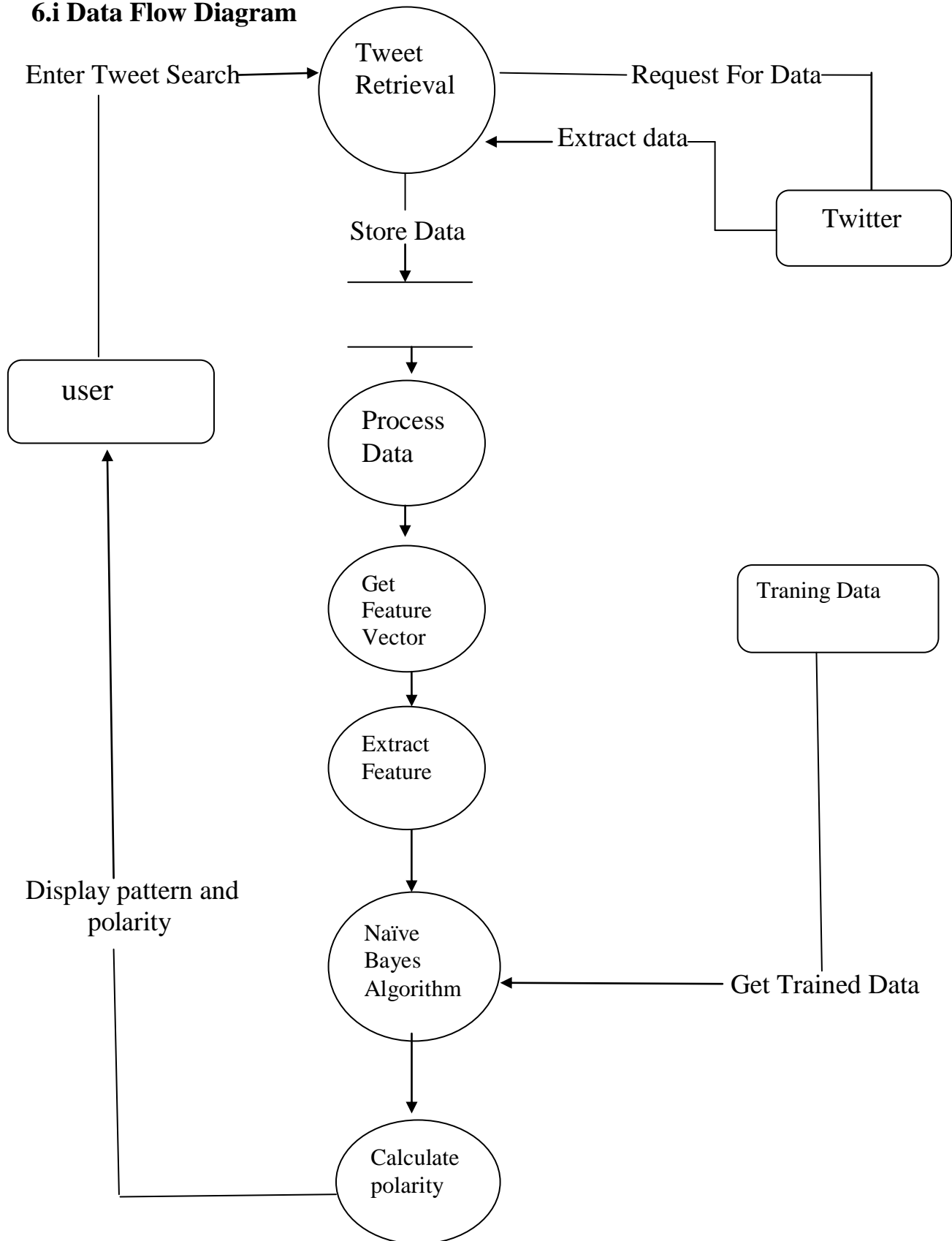


Fig3: Data Flow Diagram

## 6.ii Use Case Diagram

### Twitter

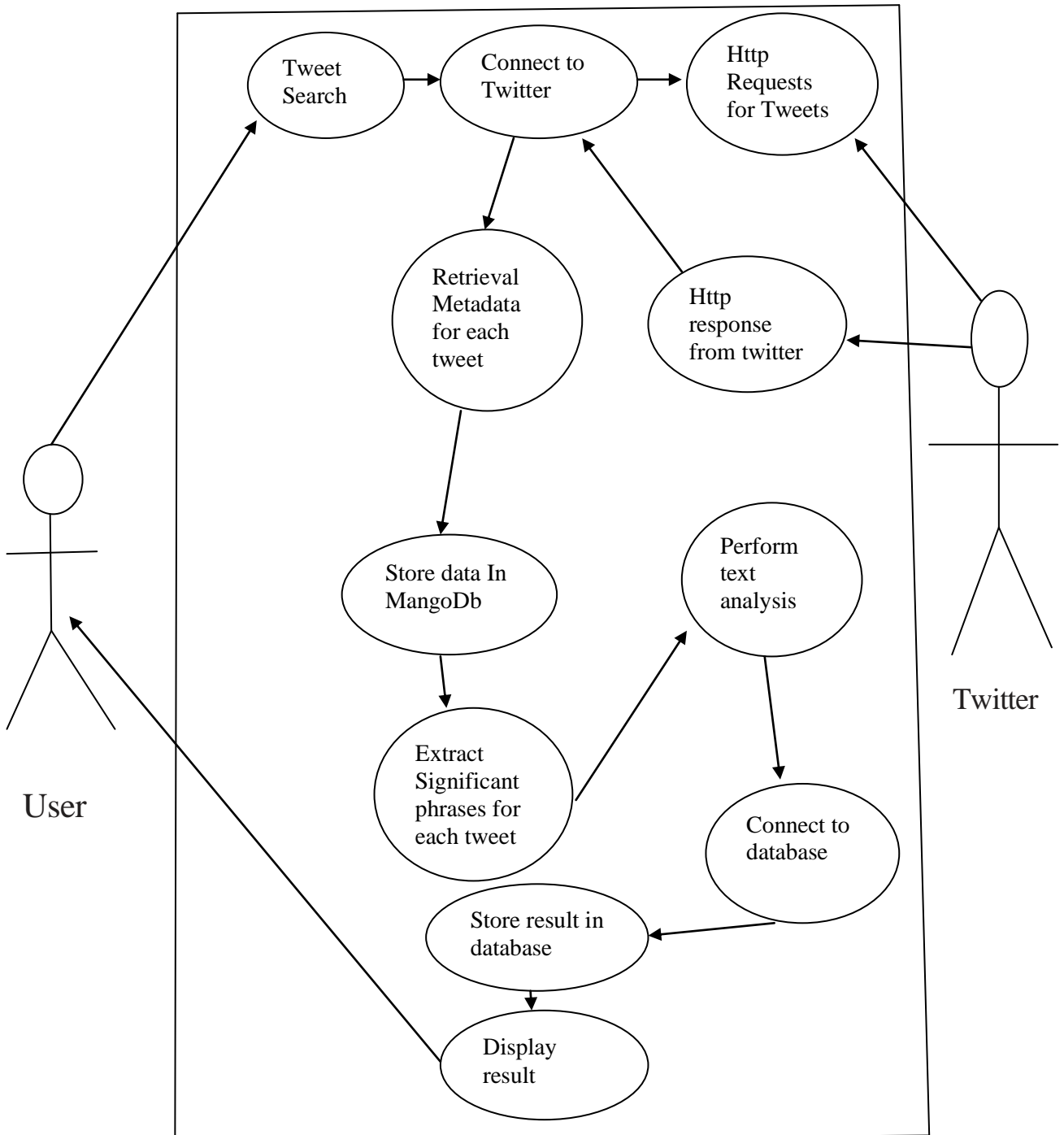


Fig4 : Use Case Diagram

### 6.iii Sequence Diagram

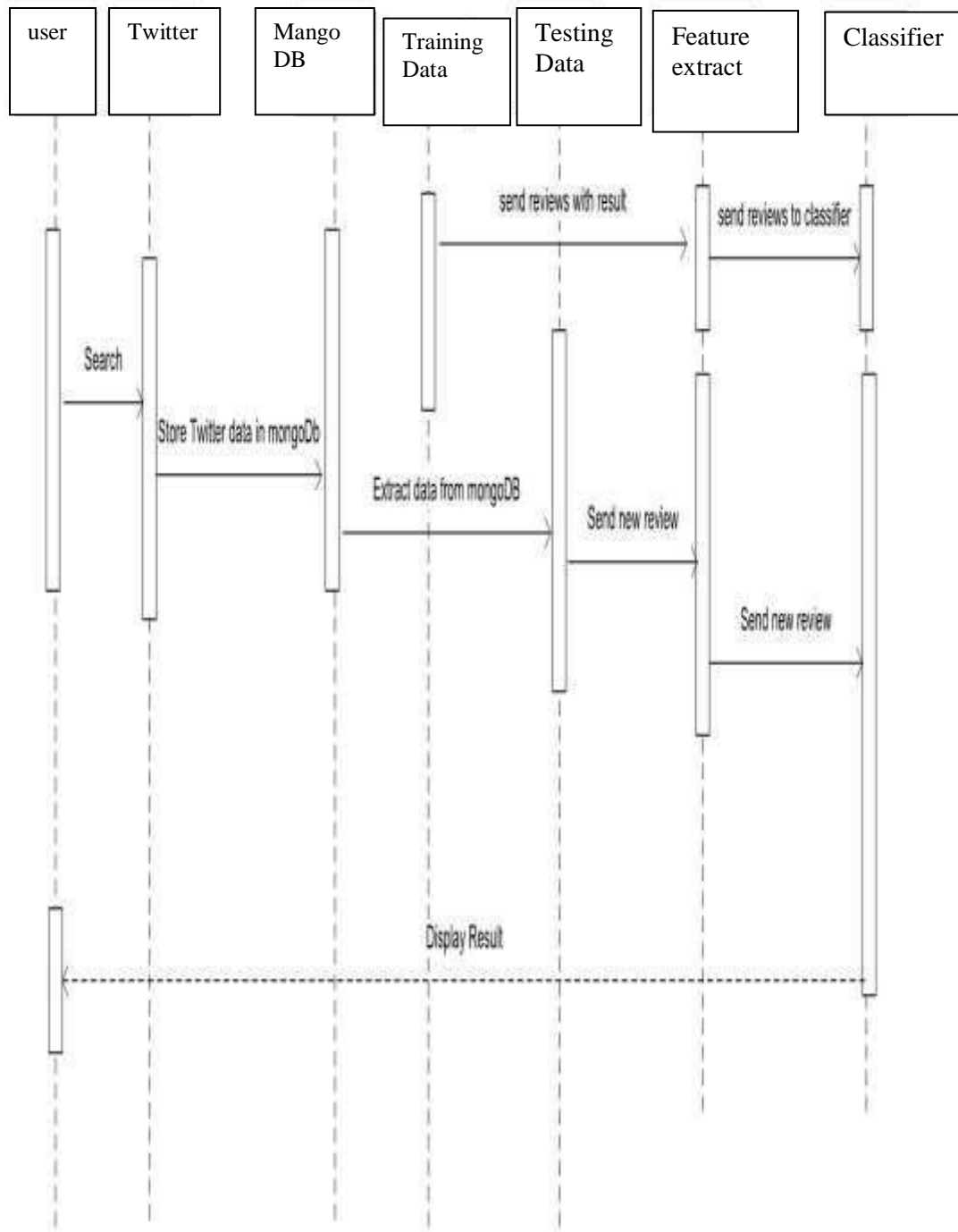


Fig 5 :Sequence Diagram

## 6.iv Activity Diagram

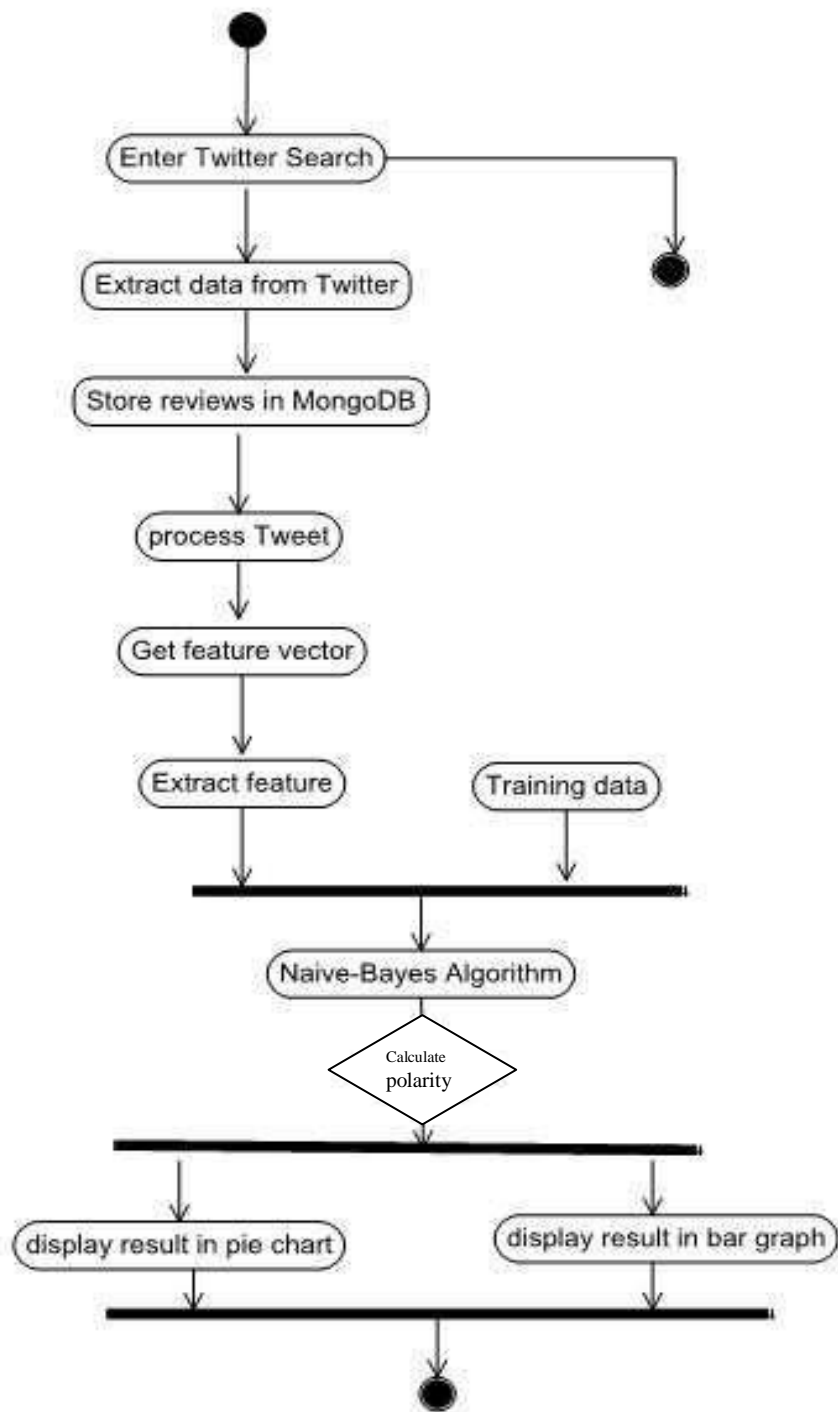


Fig: 6 :Activity Diagram

## 6.v System Flow Diagram

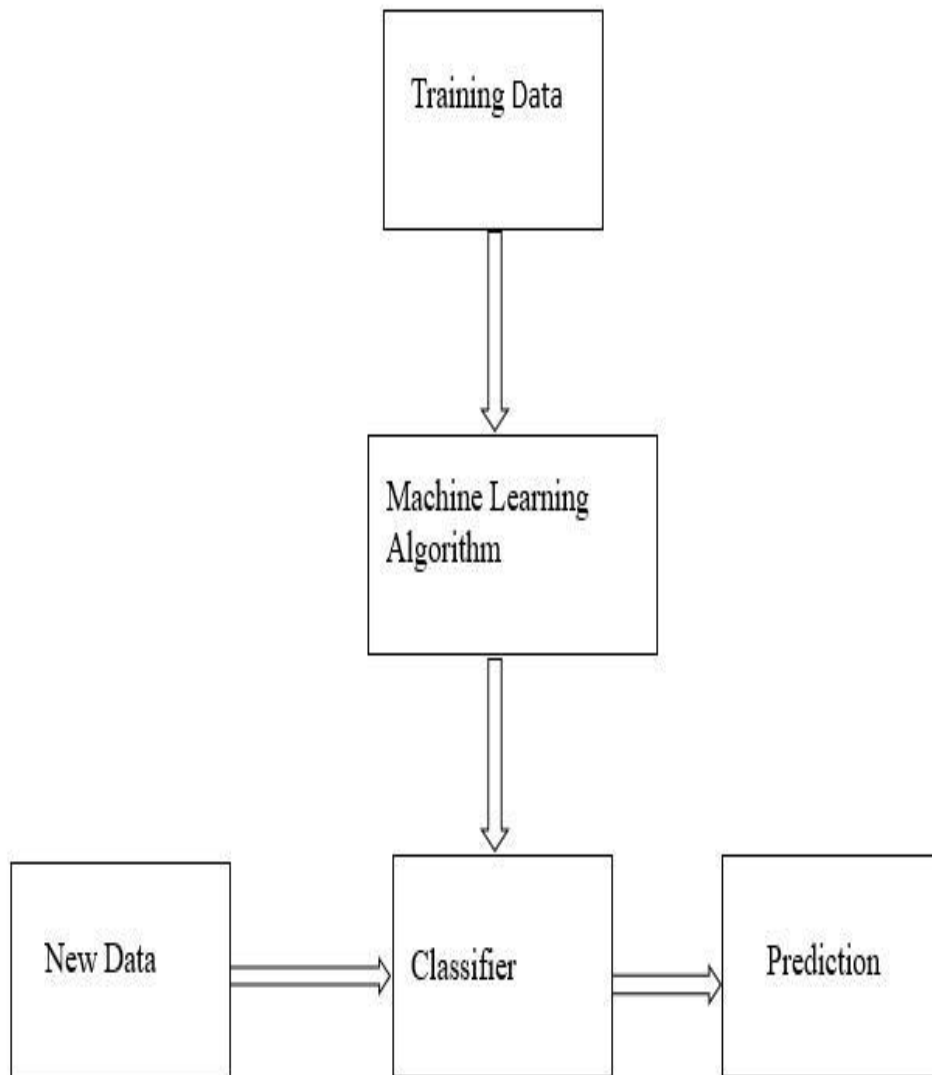


Fig 7: System Flow Diagram

## 6.vi Flow Chart

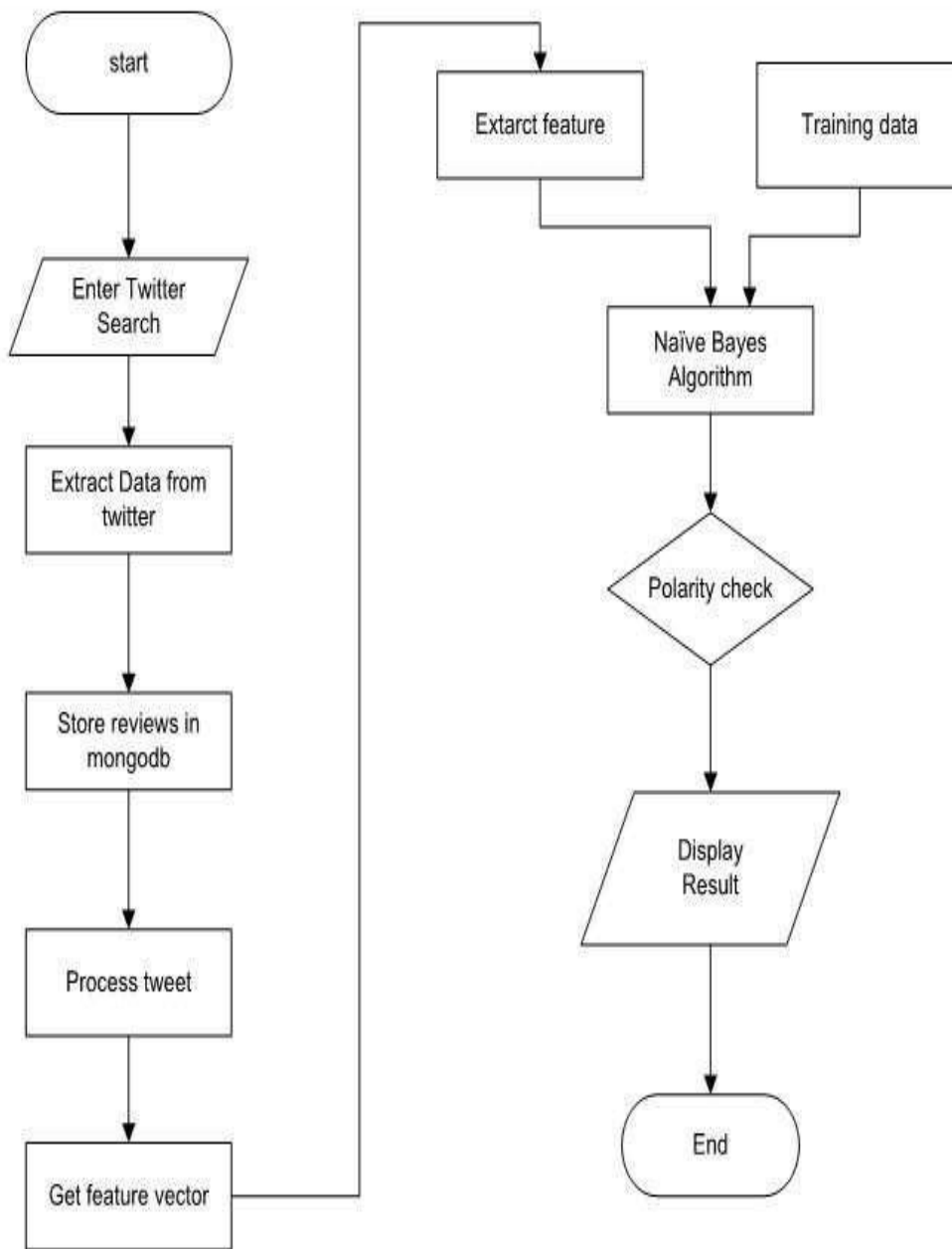


Fig 8 : Flow chart

## 6.1 Technologies used

### R Programming Language

R is a programming 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, and is currently developed by the R Development Core Team.

### Implementation:

#### Install Packages:

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

#### API keys and Tokens for Retrieving tweets from Twitter :

```
api_key<-"TCZ9ExyX4BzItaKseWueIpSs"
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('$narendramodi',n=50,lang='en')
mytweet
mytweet1<-searchTwitter('$rahulgandhi',n=50,lang='en')
mytweet1
```

#### Convert tweets to Data Frames:

```
mytweetdf<-twListToDF(mytweet)
mytweet1df<-twListToDF(mytweet1)
```



**Retrieved Tweets are stored in .csv file as election.**

```
write.csv(mytweetdf,file='/Users/dshiv/Desktop/election.csv')
head(election)
write.csv(mytweet1df,file='/Users/dshiv/Desktop/election1.csv')
head(election1)
```

**Read File :**

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

**Build Corpus:**

```
library(tm)
corpus <- iconv(election$text, to = "utf-8")
corpus1 <- iconv(election1$text, to = "utf-8")
corpus <- Corpus(VectorSource(corpus))
corpus1 <- Corpus(VectorSource(corpus1))
inspect(corpus[1:50])
inspect(corpus1[1:50])
```

**Clean Text :****Lower text Conversation**

```
corpus <- tm_map(corpus, tolower)
inspect(corpus[1:50])
corpus1 <- tm_map(corpus1, tolower)
inspect(corpus1[1:50])
```

**Removing Punctuations:**

```
corpus <- tm_map(corpus, removePunctuation)
inspect(corpus[1:50])
corpus1 <- tm_map(corpus1, removePunctuation)
inspect(corpus1[1:50])
```

**Removing Numbers:**

```
corpus <- tm_map(corpus, removeNumbers)
inspect(corpus[1:50])
corpus1 <- tm_map(corpus1, removeNumbers)
inspect(corpus1[1:50])
```

**Remove Common English words:**

```
cleanset <- tm_map(corpus, removeWords, stopwords('english'))
inspect(cleanset[1:50])
cleanset1 <- tm_map(corpus1, removeWords, stopwords('english'))
inspect(cleanset1[1:50])
```

**Remove URLs:**

```
removeURL <- function(x) gsub('http[[:alnum:]]*', '', x)
cleanset <- tm_map(cleanset, content_transformer(removeURL))
inspect(cleanset[1:50])
removeURL <- function(x) gsub('http[[:alnum:]]*', '', x)
cleanset1 <- tm_map(cleanset1, content_transformer(removeURL))
inspect(cleanset1[1:50])
```

**Remove White Space:**

```
cleanset <- tm_map(cleanset, stripWhitespace)
inspect(cleanset[1:50])
cleanset1 <- tm_map(cleanset1, stripWhitespace)
inspect(cleanset1[1:50])
```

**Convert Unstructured data to structured 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:**

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

**Obtain Sentiment Score:**

```
s <- get_nrc_sentiment(mytweet)
s1 <- get_nrc_sentiment(mytweet)
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("red","blue"),
```

```
  ylab = 'Count',
```

```
  beside = TRUE,
```

```
  main = 'Sentiment Scores for Narendra Modi Vs Rahul Gandhi')
```

## 7. Results and Discussions

I have collected data through twitter API, after that performed preprocessing on that data. For collected data for India Rajya Sabha election we classified polarity of Positive and Negative.

I observed that in sentiment analysis positive tweets were more in Narendra modi data than that of Rahul gandhi and vice versa. As prediction we can predict that Narendra modi will win elections from twitter data we collected.

After take six different variables for emotion analysis. six different variables such as joy, surprise, anger, disgust, fear and sadness.

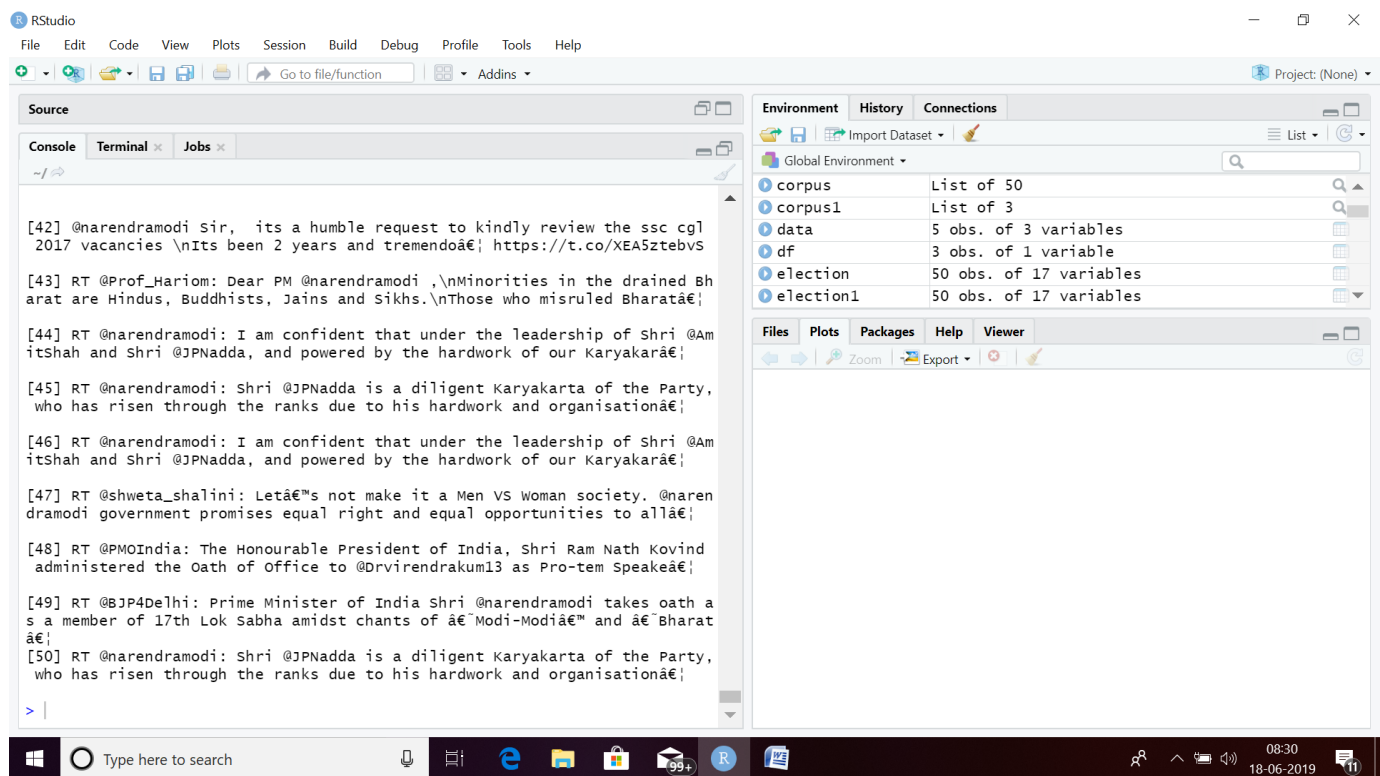


Fig 9: Retrieve the Tweets for Narendramodi

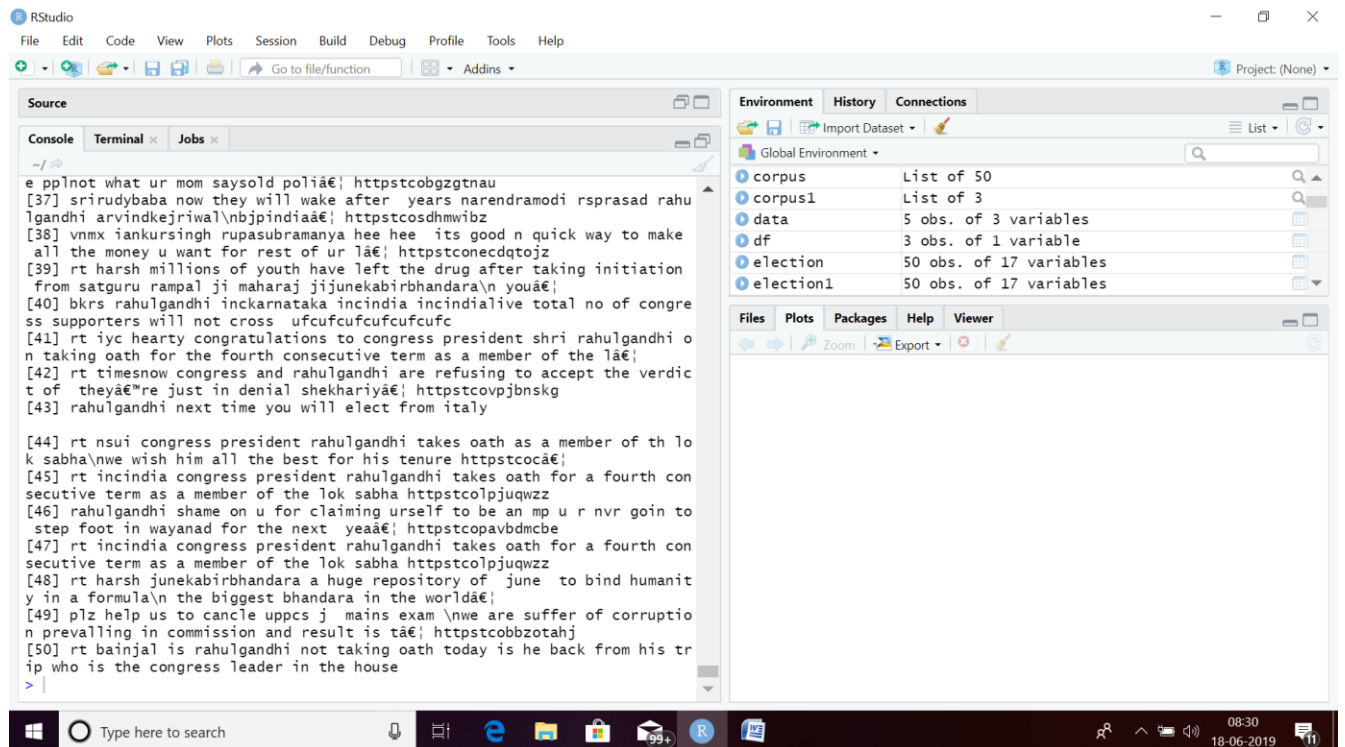


Fig10 : Retrieve the tweets for Rahulgandhi

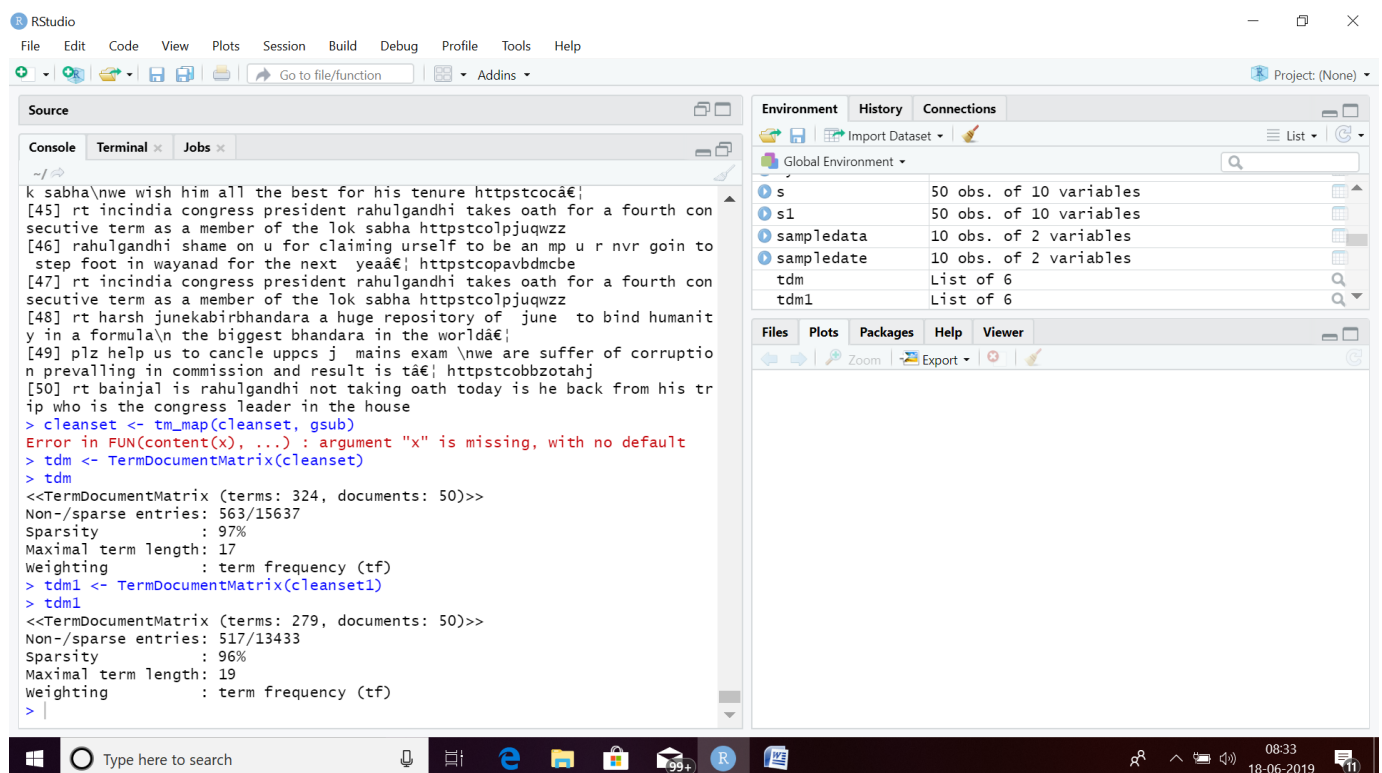


Fig11: Term Document Matrix of Sentiment Words

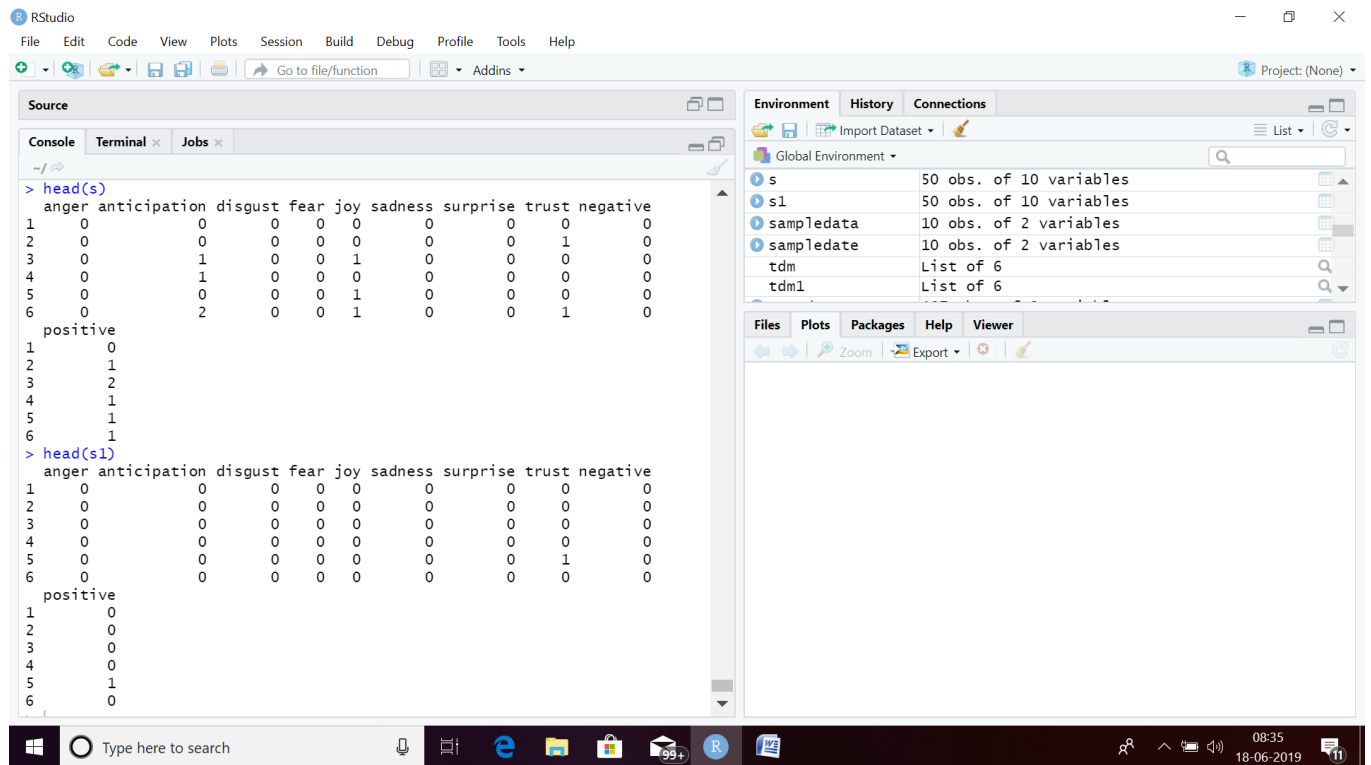


Fig12 : Calculate the presence of eight different emotions on each Tweet.

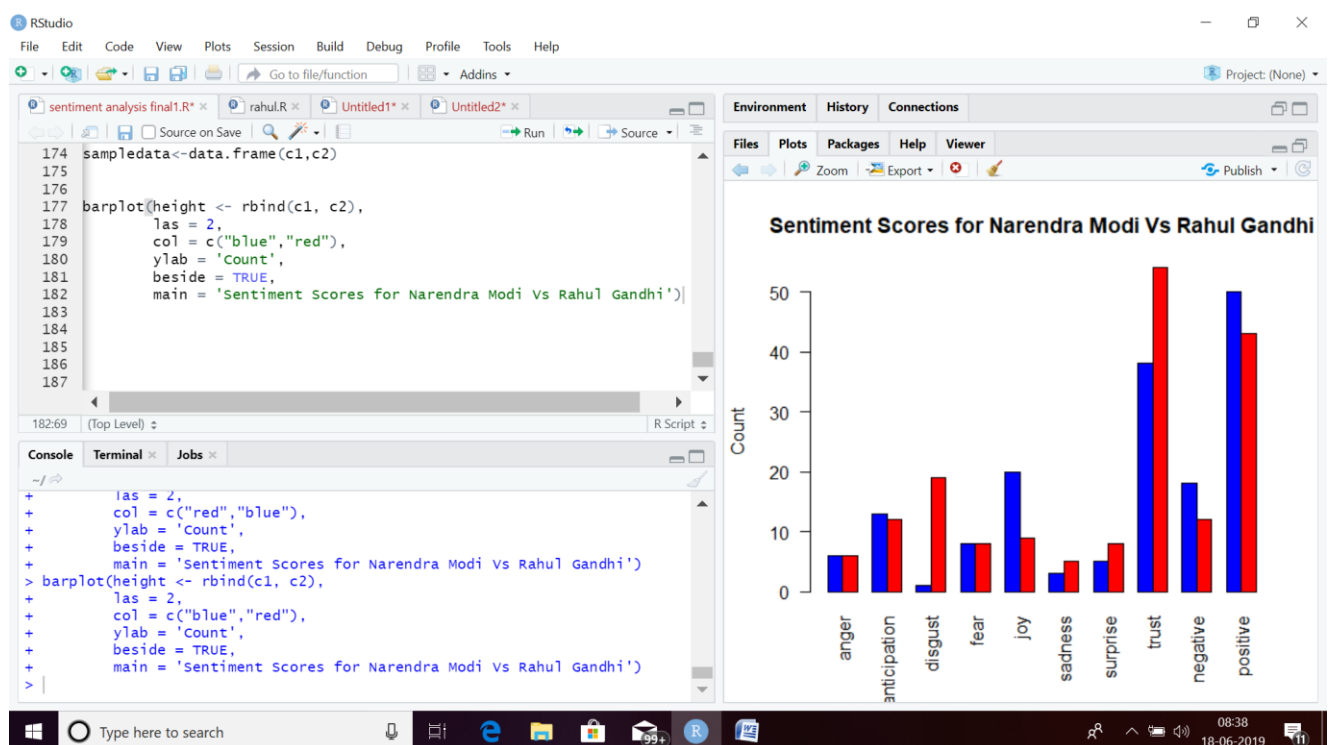


Fig13: India election Sentiment and Emotion analysis.

I observed that 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. Using naïve Bayes classifier to train our data, we observed that the output observations on tweeter data. This suggests that in election Narendra Modi will get benefit of positive tweets and Rahul Gandhi tweets has less tweets on all other emotions. It can be observed that large part of Narendra Modi tweets express Positive.



## **8. Conclusion and Future Work**

I am able 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 views 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.

Data Pre-Processing using more parameters to get the best sentiment. Updating Dictionary for new synonym and antonyms of already existing words. Web application can be converted to mobile applications. 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] Boutet, A. K. proposed that 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*. Boutet used the tweets and retweets of a user regarding a political party to infer the political leaning.
- [3] Lewis Beck, M. S. (2005). Election forecasting: principles and practice. *The British Journal of Politics & International Relations*, 7(2), 145-164.
- [4] Golbeck, J. Computing political preference among twitter followers. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Golbeck Assigned a score to every congress member which a Twitter user is following, then a political preference is assigned based on that score.
- [5] Pennacchiotti, M. &. (2011). Democrats, republicans and starbucks aficionados: user classification in twitter. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 430-438. the authors compared several features such as user's bio and avatar, posting behavior, linguistic content, follower, reply and retweet. They found out that the combination between user profile and linguistic outperform other feature.
- [6] Wong, F. M. (2013). Quantifying Political Leaning from Tweets and Retweets. *IC*
- [7] Tumasjan, A. S. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM*, 10, 178-185. The earliest research stated that the number of tweets mentioning a party reflects the election result. where they found out that the prediction result from Twitter were only better than other.
- [8] O'Connor, B. B. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11, 122-129. is the first research in which argued that sentiment detection approach from Twitter can replace the expensive and time intensive polling?

