TWITTER SENTIMENT ANALYSIS USING R

Submitted as a part of DATA SCIENCE AND BIG DATA ANALYTICS
Course Requirement

By

Rahil Khan
RA1611003010567

Under the supervision of

Dr.B.Baranidharan
Associate Professor
Department of Computer Science and Engineering
SRM Institute of Science and Technology
Kattankulathur, Chennai

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
FACULTY OF ENGINEERING AND TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DECLARATION

I Rahil Khan (RA1611003010567) studying III year B.Tech in Computer Science and Engineering at SRM Institute of Science and Technology, Kattankulathur, Chennai, hereby declare that this Mini project is an original work of mine and I have not verbatim copied / duplicated any material from sources like internet or from print media, excepting some vital company information / statistics and data that is provided by the Technical organisations itself.

Signature of the Student

Date:

Place:

ACKNOWLEDGEMENT

This project would not have been possible without the kind support and help of many individuals. I would like to extend my sincere thanks to all of them.

I am highly indebted to Dr.B.Baranidharan for his guidance and constant supervision as well as for providing necessary information regarding the project & also for his support in completing the project.

I would also like to express my gratitude towards my parents for their everlasting support and my fellow peers of SRMIST for their kind co-operation and encouragement, allowing me to complete this project. My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

ABSTRACT

As times have progressed, the usage of social media has exponentially increased. Public and private opinions about a wide variety of subjects are expressed and spread continually via numerous social media platforms. Twitter is one of such platforms that has gained a lot of popularity. Twitter offers organizations and individual users a fast and effective way to advertise and communicate their ideas and thoughts without much hassle. Thus, analyzing customers' perspectives toward day to day events is crucial to success in the market place. Developing a program for sentiment analysis is an approach to be used to computationally measure people's perceptions.

This project applies sentiment analysis to a dataset containing thousands of tweets relating to a given string that is searched, all using R libraries. Searched strings could include hashtags, usernames, specific words etc. Using the processed output, we are able to determine the sentiments of people regarding any trending topic.

Tweets are imported using R and the data is cleaned by removing emoticons and URLs. Lexical Analysis is used to predict the sentiment of tweets and subsequently express the opinion graphically through ggplots, histogram, pie chart and tables.

TABLE OF CONTENTS

LIST OF TABLES

sentiment.R* × users ×

Show Attributes

Name	Type	Value
users	list [10000]	List of length 10000
[[1]]	S4 [1] (twitteR::status)	
.->created	double (S3: POSIXct, POSIXt)	2019-04-05 13:18:26
.->favoriteCount	double [1]	3448
.->favorited	logical	FALSE
.->id	character [1]	'1114155334510219264'
.->isRetweet	logical	FALSE
.->latitude	character [0]	
.->longitude	character [0]	
.->replyToSID	character [0]	
.->replyToSN	character [0]	
.->replyToUID	character [0]	
.->retweetCount	double [1]	2840
.->retweeted	logical	FALSE
.->screenName	character [1]	'TarekFatah'
.->statusSource	character [1]	'<a href="http://twitter.com/download/iphone" rel="nofollow">Twitter for iPhone< ...
.->text	character [1]	'Muslim Indians returning from a #RahulGandhi rally in #Wayanad, Kerala, wave gr ...
.->truncated	logical	TRUE
.->urls	list [1 x 5] (S3: data.frame)	A data.frame with 1 rows and 5 columns
.refClassDef	S4 (methods::refClassRepresenta	S4 object of class refClassRepresentation
.self	S4 [1] (twitteR::status)	
created	double (S3: POSIXct, POSIXt)	2019-04-05 13:18:26
favoriteCount	double [1]	3448
favorited	logical	FALSE
field	function (methods::refMethodDe	function(name, value) { ... }
getCreated	function (methods::refMethodDe	function() { ... }

Table 1: 'Users' containing JSON data parsed from the Twitter API

sentiment.R* × tweet.df ×

Filter

	text	favorited	favoriteCount	replyToSN	created	truncated	replyToSID	id	replyToUID	statusSource	screenName	retweetCount	isRetweet	retweeted
1	Muslim Indians returning from a #Rahul...	FALSE	3448	NA	2019-04-05 13:...	TRUE	NA	1114155334510...	NA	<a href="http://...	TarekFatah	2840	FALSE	FALSE
2	Modi is not only worst or liar PM, but he...	FALSE	3578	NA	2019-04-06 17:...	FALSE	NA	1114575580907...	NA	<a href="http://...	waglenikhil	1054	FALSE	FALSE
3	Excellent that groups of writers, artists, s...	FALSE	2089	NA	2019-04-05 10:...	TRUE	NA	1114120254643...	NA	<a href="http://...	chitraSD	766	FALSE	FALSE
4	RT @ashish30sharma: Such nice words t...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114678880936...	NA	<a href="http://...	shilpa29patel	22	TRUE	FALSE
5	RT @waglenikhil: Modi is not only worst ...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114678863232...	NA	<a href="http://...	GOURAB90	1054	TRUE	FALSE
6	RT @ashish30sharma: Have you watched...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114678799797...	NA	<a href="http://...	shilpa29patel	27	TRUE	FALSE
7	Have trust in Narendra Modi. Either he ...	FALSE	1	NA	2019-04-06 23:...	TRUE	NA	1114678605718...	NA	<a href="http://...	aanunim	1	FALSE	FALSE
8	RT @yashaveeryadav: baby Kids saying ...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114678193213...	NA	<a href="http://...	yash_wanti	1	TRUE	FALSE
9	RT @skarthikdotin: More strong messeg...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114677940036...	NA	<a href="http://...	Sathish_AIADMK	3	TRUE	FALSE
10	RT @ErosNow: This story will truly inspir...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114677619960...	NA	<a href="http://...	aju_wadkar	628	TRUE	FALSE
11	RT @waglenikhil: Modi is not only worst ...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114677334290...	NA	<a href="http://...	ilovemyindia21	1054	TRUE	FALSE
12	RT @IndiraTandon1: @ashish30sharma ...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114677198798...	NA	<a href="http://...	Sun1ashish	3	TRUE	FALSE
13	Ye #ABM #ABM kya hai? It is a Billion dol...	FALSE	0	NA	2019-04-06 23:...	TRUE	NA	1114677132532...	NA	<a href="http://...	amir_banday	0	FALSE	FALSE
14	RT @waglenikhil: Modi is not only worst ...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114677053209...	NA	<a href="http://...	SaiyedKhilafat	1054	TRUE	FALSE
15	I had recently been to Bali... the whole cit...	FALSE	0	NA	2019-04-06 23:...	TRUE	NA	1114676841720...	NA	<a href="http://...	BLKiran	0	FALSE	FALSE
16	RT @waglenikhil: Modi is not only worst ...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114676675189...	NA	<a href="http://...	ashualam10	1054	TRUE	FALSE
17	RT @ashish30sharma: Have you watched...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114676557644...	NA	<a href="http://...	IndiraTandon1	27	TRUE	FALSE
18	RT @waglenikhil: Modi is not only worst ...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114676449619...	NA	<a href="http://...	faizabul212	1054	TRUE	FALSE
19	RT @ErosNow: This story will truly inspir...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114676314458...	NA	<a href="http://...	bohravishal07	628	TRUE	FALSE
20	RT @Sanjeev40330782: @anupam_mitra...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114676165229...	NA	<a href="http://...	anupam_mitra	2	TRUE	FALSE
21	@RajeshK39804978 @jyotsnavarma9 @r...	FALSE	4	RajeshK39804978	2019-04-06 23:...	TRUE	1114576391725...	1114676104080...	9111895550086...	<a href="http://...	Amaresh458148...	2	FALSE	FALSE
22	RT @waglenikhil: Modi is not only worst ...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114676005048...	NA	<a href="http://...	rahulj_09	1054	TRUE	FALSE
23	RT @waglenikhil: Modi is not only worst ...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114675884500...	NA	<a href="http://...	maulinshah9	1054	TRUE	FALSE
24	RT @chitraSD: Excellent that groups of w...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114675515002...	NA	<a href="http://...	threadspanner	766	TRUE	FALSE
25	#SushmaSwaraj #PMOIndia #narendram...	FALSE	0	NA	2019-04-06 23:...	TRUE	NA	1114674832015...	NA	<a href="http://...	Crypto_Shower	0	FALSE	FALSE
26	RT @sonamketweet: #IndiaWantsModiA...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114674824603...	NA	<a href="http://...	Krishan92651204	9	TRUE	FALSE
27	RT @ashish30sharma: Such nice words t...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114674756899...	NA	<a href="http://...	kanomthom1	22	TRUE	FALSE
28	RT @ashish30sharma: Have you watched...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114673834429...	NA	<a href="http://...	kanomthom1	27	TRUE	FALSE
29	RT @waglenikhil: Modi is not only worst ...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114673307180...	NA	<a href="http://...	rkmarar9	1054	TRUE	FALSE
30	RT @RuthlessIndia: This part of modi int...	FALSE	0	NA	2019-04-06 23:...	FALSE	NA	1114673241132...	NA	<a href="http://...	romantic2809	6	TRUE	FALSE

Table 2: 'tweet.df' is a large data frame converted from the 'Users' table

sentiment.R* result

Filter

	score	text
1	0	Muslim Indians returning from a #RahulGandhi rally in #Waya...
2	-4	Modi is not only worst or liar PM, but he indulges and encour...
3	1	Excellent that groups of writers, artists, scientists & acad...
4	1	RT @ashish30sharma: Such nice words to start the day with..th...
5	-3	RT @waglenikhil: Modi is not only worst or liar PM, but he ind...
6	0	RT @ashish30sharma: Have you watched it yet, simple things ...
7	3	Have trust in Narendra Modi. Either he will find way or make ...
8	0	RT @yashaveeryadav: baby Kids saying @narendramodi doba...
9	2	RT @skarthikdotin: More strong messege of #Modi is winning...
10	1	RT @ErosNow: This story will truly inspire you! Check out the t...
11	-3	RT @waglenikhil: Modi is not only worst or liar PM, but he ind...
12	1	RT @IndiraTandon1: @ashish30sharma @ErosNow @Ridhima...
13	-1	Ye #ABM #ABM kya hai? It is a Billion dollar fake news factor...
14	-3	RT @waglenikhil: Modi is not only worst or liar PM, but he ind...
15	0	I had recently been to Bali... the whole city is "strawfree".... wh...
16	-3	RT @waglenikhil: Modi is not only worst or liar PM, but he ind...
17	0	RT @ashish30sharma: Have you watched it yet, simple things ...
18	-3	RT @waglenikhil: Modi is not only worst or liar PM, but he ind...
19	1	RT @ErosNow: This story will truly inspire you! Check out the t...
20	0	RT @Sanjeev40330782: @anupam_mitra @ImranKhanPTI Wo...
21	0	@RajeshK39804978 @jyotsnavarma9 @rajeshk234178 @Dhar...
22	-3	RT @waglenikhil: Modi is not only worst or liar PM, but he ind...
23	-3	RT @waglenikhil: Modi is not only worst or liar PM, but he ind...
24	1	RT @chitraSD: Excellent that groups of writers, artists, scientis...
25	0	#SushmaSwaraj #PMOIndia #narendramodi #RahulGandhi #P...

Table 3: 'result' contains tweets and their respective scores (positive, negative, neutral)

LIST OF FIGURES / GRAPHS

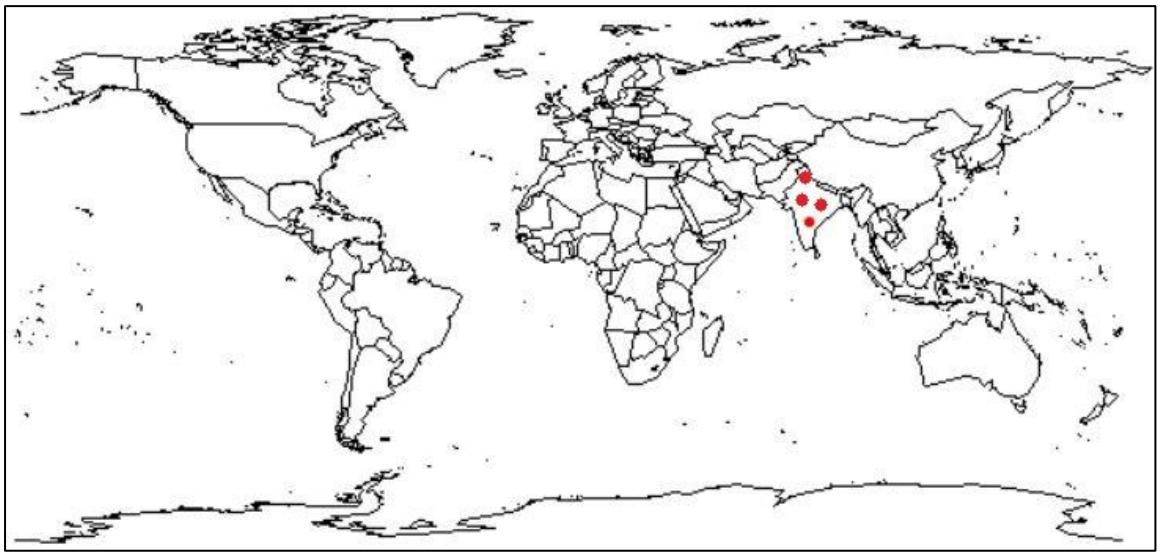


Fig.1: Tweet geolocation plotted on the world map. In this case all tweets are from India.

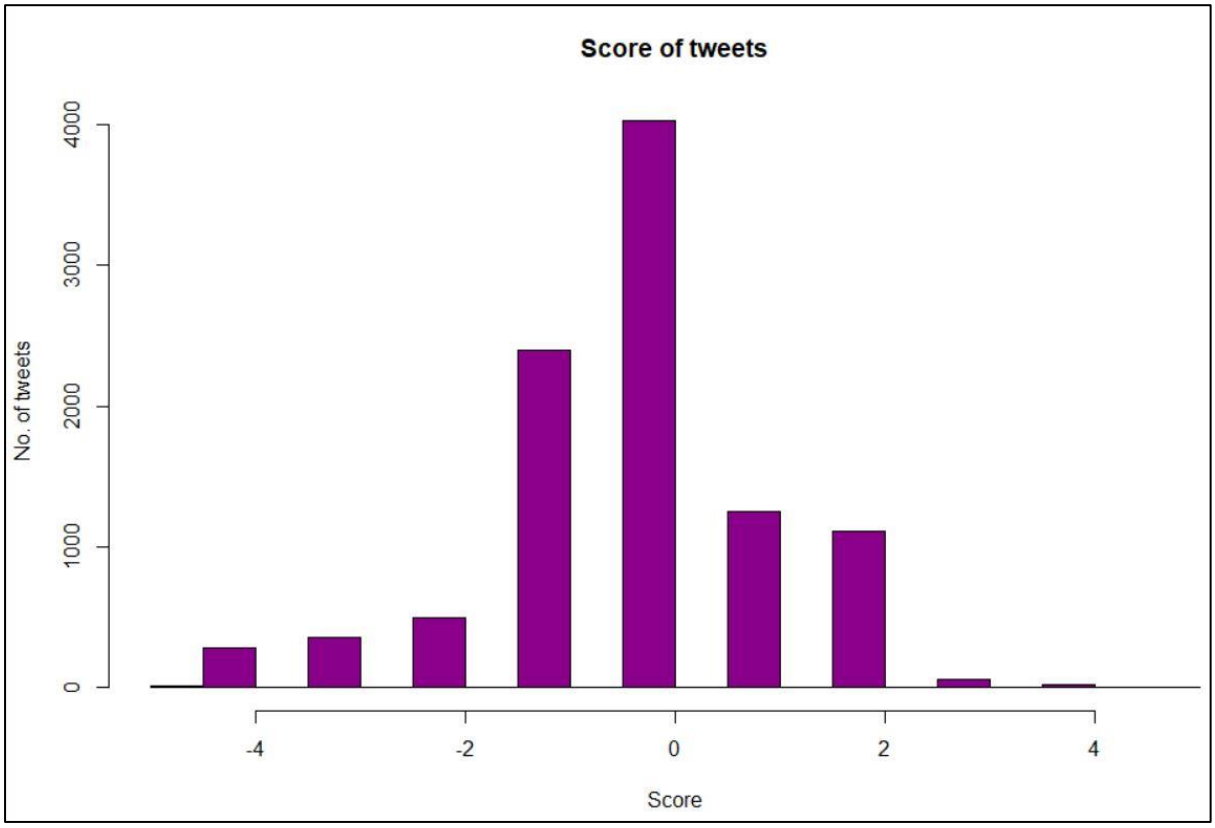


Fig 2: Histogram plotting the score of tweets and their frequencies.

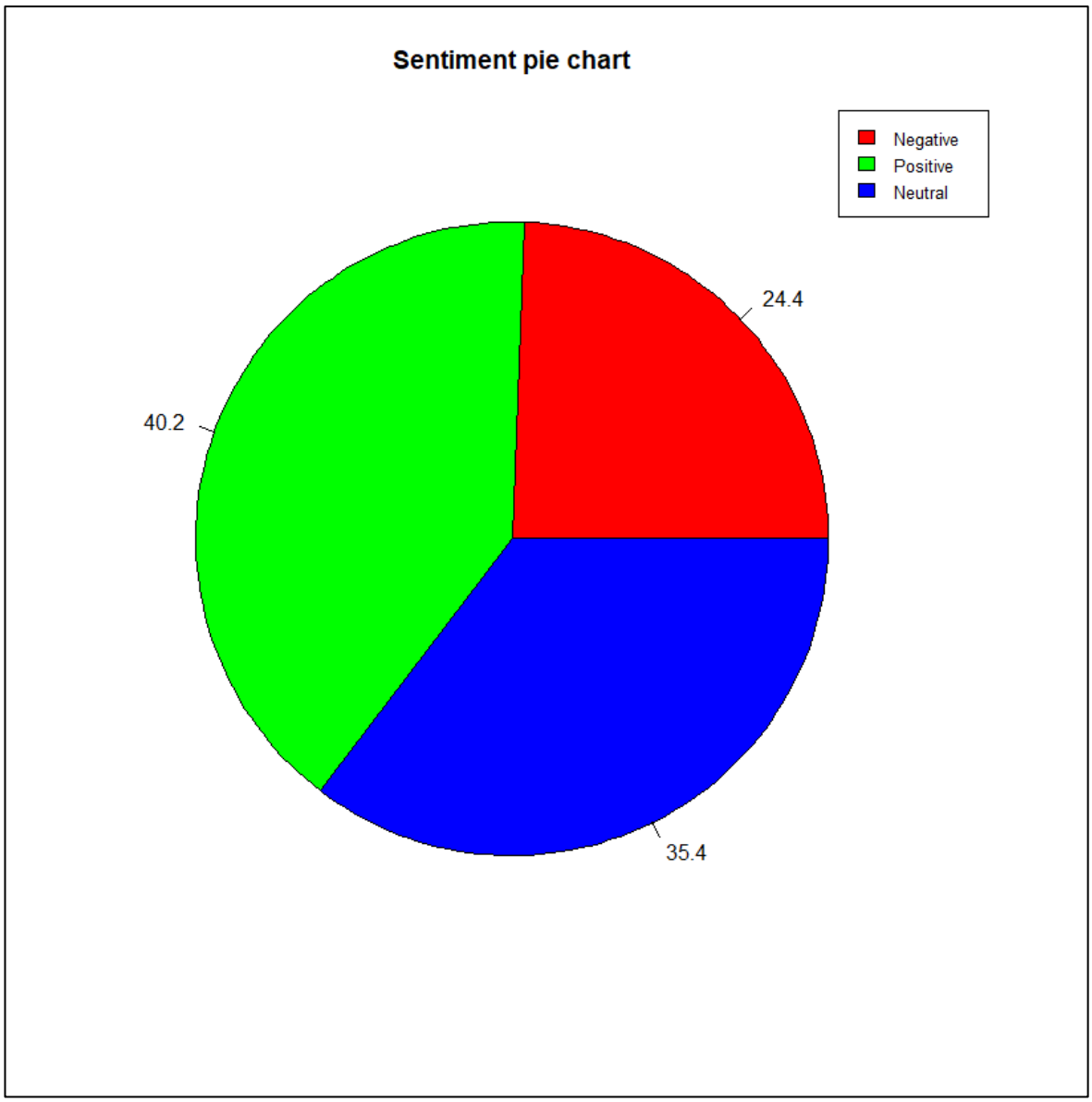


Fig 3: A pie chart with the percentage of sentiments.

INTRODUCTION

In the past one decade, there has been an exponential surge in the online activity of people across the globe. The volume of posts that are made on the web every second runs into millions. To add to this, the rise of social media platforms has led to flooding to content on the internet.

Social media is not just a platform where people talk to each other, but it has become very vast and serves many more purposes. It has become a medium where people

- Express their interests.
- Share their views.
- Share their displeasures.
- Compliment companies for good and poor services.

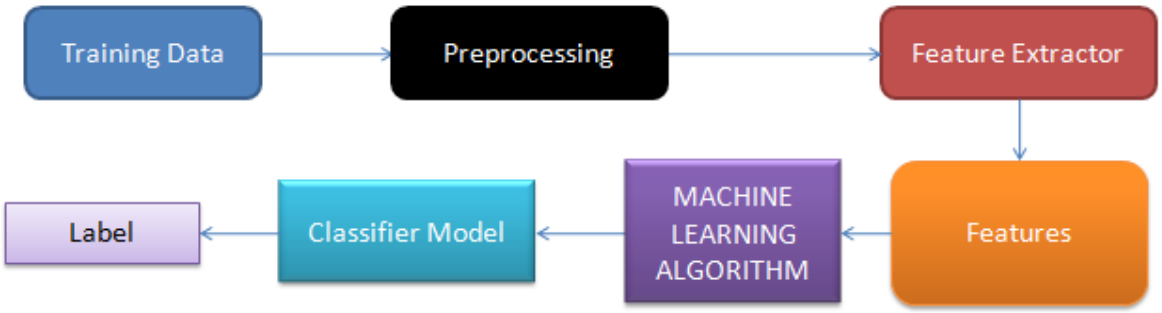
In this project, we are going to learn how we can analyze what people are posting on social networks (Twitter) to come up a great application which helps companies to understand about their customers.

In consideration of the upcoming elections, we shall perform a **sentiment analysis on our current president Narendra Modi** using the hashtag **#Modi** and **#Narendra Modi**.

SYSTEM REQUIREMENTS

- R Studio with required libraries
- Twitter Authentication to access API

FEATURES AND IMPLEMENTATION



Extraction of Tweets

(i) Create twitter application for consumer and access key generation.

(ii) twitteR - Provides an interface to the Twitter web API

(iii) ROAuth - R Interface for OAuth

(iv) Create twitter authenticated credential object. It is done using consumer key, consumer secret, access token, access secret.

(v) During authentication, we are redirected to a URL automatically where we click on Authorize app as shown in the image below and enter the unique 7-digit number to get linked to the account from which feeds are being taken.

Searching Twitter like a Pro

Here's a complete list of Twitter search operators that help perform more accurate searches on Twitter:

1. **from:BarackObama**
   All tweets sent by a particular Twitter user
2. **filter:verified cool OR amazing**
   Only show tweets from verified Twitter accounts (with the blue tick)
3. **gangnam style filter:replies**
   Only show tweets that are replies. You can use exclude:replies to remove @reply tweets from search results.
4. **gangnam style filter:retweets**
   Only show tweets that are retweets. You can use exclude:retweets to remove RTs from search results.

5. **to:BarackObama -filter:links**
   Tweets sent to @BarackObama but not containing any links
6. **elections list:TIME/time-staff**
   Search for tweets from users who belong to a particular Twitter list
7. **youtube.com min_faves:100**
   Tweets containing YouTube videos that are favorited by at least 100 users
8. **earthquake min_retweets:10**
   Tweets that have been retweeted at least 10 times

Cleaning Tweets

The tweets are cleaned in R by removing:

(i) Extra punctuation
(ii) Stop words (Most commonly used words in a language like the, is, at, which, and, on)
(iii) Redundant Blank spaces
(iv) Emoticons
(v) URLS
(vi) Hashtags
(vii) Handle

Loading Word Database

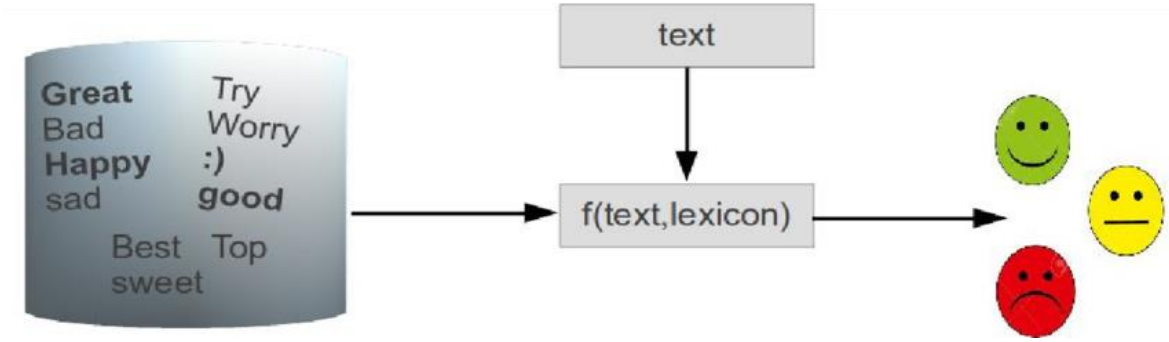
A database, created by Hui Lui containing positive and negative words, is loaded into R. This is used for Lexical Analysis, where the words in the tweets are compared with the words in the database and the sentiment is predicted.

AFINN is a list of English words rated for valence with an integer between minus five (negative) and plus five (positive). The words have been manually labeled by Finn Årup Nielsen in 2009-2011. The file is tab-separated. The version used is: AFINN-111: Newest version with 2477 words and phrases.

ALGORITHMS USED

Lexical Analysis:

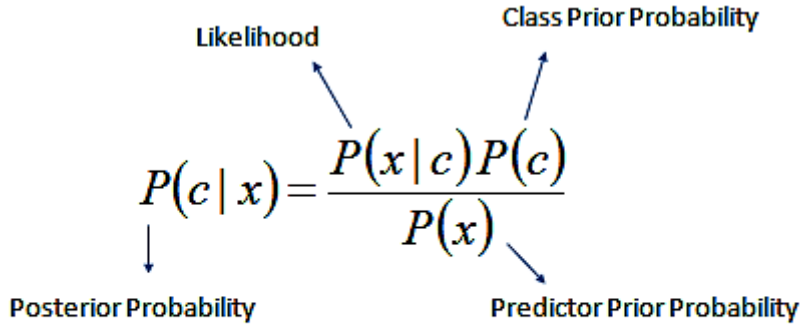
By comparing uni-grams to the pre-loaded word database, the tweet is assigned sentiment score - positive, negative or neutral and overall score is calculated.



Naive Bayes Classification:

Naive Bayes classifier is the simplest and the fastest classifier. Many researchers claim to have gotten best results using this classifier.

For a given tweet, if we need to find the label for it, we find the probabilities of all the labels, given that feature and then select the label with maximum probability.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

PACKAGES USED

twitteR: Provides an interface to the Twitter web API

stringr: For string operations in R.

ROAuth: Provides an interface to the OAuth 1.0 specification allowing users to authenticate via OAuth to the server of their choice.

RCurl: Provides functions to allow one to compose general HTTP requests and provides convenient functions to fetch URIs, get & post forms, etc. and process the results returned by the Web server.

ggplot2: An implementation of the grammar of graphics in R. It combines the advantages of both base and lattice graphics: conditioning and shared axes are handled automatically, and you can still build up a plot step by step from multiple data sources.

reshape: Flexibly restructure and aggregate data using just two functions: melt and cast

tm: A framework for text mining applications within R.

RJSONIO: This is a package that allows conversion to and from data in Javascript object notation (JSON) format. This allows R objects to be inserted into Javascript code and allows R programmers to read and convert JSON content to R objects

plyr: Tools for Splitting, Applying and Combining Data

SOURCE CODE

#List of Packages used

```
library(RColorBrewer)
library(tm)
library(twitteR)
library(ROAuth)
library(plyr)
library(stringr)
library(base64enc)
library(SnowballC)
library(ggplot2)
library(maps)
```

#Setting up a connection with the Twitter API

```
consumerKey <- "KUkOmiLu4wfYYnlTQzUshfUPf"
consumerSecret <- "wVdLDplrlKWgFdfWT8IFcPM04F5ZDujV9jPFaxHhDuzpYSUp0q"
accessToken <- "901448980038537217-Naot7eB1bk8Ue47vTUEW6A2cVAp5Ym4"
accessTokenSecret <- "xSMHnNGnoXvk78VlEHnJEWtSwqMgw6RJWGdKB0rfwOlYz"
requestURL<- "https://api.twitter.com/oauth/request_token"
accessURL<- "https://api.twitter.com/oauth/access_token"
authURL<- "https://api.twitter.com/oauth/authorize"
setup_twitter_oauth(consumerKey,consumerSecret,accessToken,accessTokenSecret)
```

#Searching Twitter and converting to data Frame

```
users<- searchTwitteR("#modi OR #narendramodi", resultType="mixed", n=10000,
lang="en", since='2019-04-01', until='2019-04-07')
#Converting into Dataframe
tweet.df = do.call("rbind",lapply(users,as.data.frame))
```

#Plotting data on map

```
map('world')
points(tweet.df$longitude,tweet.df$latitude, pch=20, cex=1, col="red")
```

#Viewing the data

```
View(tweet.df)
```

#Reading sentiment analysis data from Txt document

```
pos.words = scan('./positive-words.txt', what='character', comment.char=';')
neg.words = scan('./negative-words.txt', what='character', comment.char=';')
```

#Appending some more words to actual words

```
pos.words = c(pos.words, 'new','nice' ,'good', 'horizon')
neg.words = c(neg.words, 'wtf', 'behind','feels', 'ugly', 'back','worse' , 'shitty', 'bad',
```

```
'no','freaking','sucks','horrible')

#Converting the 'users' data into dataFrame
test <-ldply(users,function(t)t$toDataFrame())

#Calcuating sentiment analysis result
result <- score.sentiment(test$text,pos.words,neg.words)

#Summarizing data
summary(result$score)

#Plotting a histogram
hist(result$score,col="yellow", main="Score of tweets",ylab="No. of tweets", xlab="Score")

#Counting no. of Tweets
count(result$score)

#plotting a percentage pie chart.
sentinum <- c(sum(result$score<0),sum(result$score>0),sum(result$score==0))
piepercent<- round(100*sentinum/sum(sentinum), 1)
pie(sentinum, labels = piepercent, main = "Sentiment pie chart",col =
rainbow(length(sentinum)))
legend("topright", c("Negative","Positive","Neutral"), cex = 0.8, fill =
rainbow(length(sentinum)))
#score_Sentiment function: used to remove all unwanted data
score.sentiment = function(sentences, pos.words, neg.words, .progress='none'){
   require(plyr)
   require(stringr)
   scores = laply(sentences, function(sentence, pos.words, neg.words) {
   sentence = gsub('[[:punct:]]', '', sentence)
   sentence = gsub('[[:cntrl:]]', '', sentence)
   sentence = gsub('\\d+', '', sentence)
   sentence = tolower(sentence)
   word.list = str_split(sentence, '\\s+')
   words = unlist(word.list)
   pos.matches = match(words, pos.words)
   neg.matches = match(words, neg.words)
   pos.matches = !is.na(pos.matches)
   neg.matches = !is.na(neg.matches)
   score <- sum(pos.matches) - sum(neg.matches)
   return(score)
   }, pos.words, neg.words, .progress=.progress )
   scores.df = data.frame(score=scores, text=sentences)
   return(scores.df) }
```

OBSERVATIONS

Due to Twitter API limitations, only 10,000 tweets on Narendra Modi were retrieved for data analysis. After parsing, cleaning and analysis using the 'score.Sentiment' function, we have the following results (out of 10,000).

Negative Sentiment – 2439
Positive Sentiment – 4023
Neutral Sentiment – 3538

```
> #Negative #Positive #Neutral
> sentinum
[1] 2439 4023 3538
```

Here's a summary of the result data:

```
> summary(result$score)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-5.0000 -1.0000  0.0000 -0.1887  0.0000  5.0000
```

We may dive in further by also looking at the no. of tweets that have a specific score (-5 being most negative, 5 being most positive).

As we can observe, most of the tweets are either neutral or nearly neutral, containing only a few elements of positivity/negativity.

```
> #Count No of Tweets
> count(result$score)
    x freq
1  -5    8
2  -4  181
3  -3  360
4  -2  492
5  -1 1398
6   0 3538
7   1 2248
8   2 1109
9   3  645
10  4   19
11  5    2
```

Most negative tweet sample:

970	-5	#gen "It worries some people that we killed #terrorists. When...

Most positive tweet sample:

5087	5	Yet, another award. Congratulations sir @narendramodi Ji. All ...

Neutral tweet sample:

8	0	RT @yashaveeryadav: baby Kids saying @narendramodi doba...

LIMITATIONS

1. The Twitter Search API can get tweets only up to a maximum of 7 days prior.
2. Cannot get 100% efficiency in analysing sentiment of tweets.
3. Cannot detect sarcasm.
4. Limitation on rate of tweets obtained from API.
5. Unable to analyse views of people using different languages.
6. Data redundancy as many copies of retweets appear.

SCOPE FOR IMPROVEMENT

1. **Investigating Support Vector Machines**: Several text mining papers have discussed using Support Vector Machines (SVMs). The next step would be to test our approach on SVMs. However, a paper by Go, Bhayani and Huang states that SVMs do not increase accuracy.
2. **Building a classifier for Hindi tweets**: There are many users on Twitter that use primarily Hindi language. The approach discussed here can be used to create a Hindi language sentiment classifier.
3. **Improving Results using Semantics Analysis**: Understanding the role of the nouns being talked about can help us better classify a given tweet.
4. Add Telugu words to dataset.
5. Find no of mentions of n particular organizations.
6. Parallelizing code.

CONCLUSION

After robust data analysis, we can conclude that the majority of tweets mentioning our current Prime Minister are positive (40.2%), while neutral tweets (35.4%) outnumber negative ones (24.4%). However, we cannot say that the majority of people are positive, as the sum of both neutral and negative tweets (59.8%) outweighs total positive tweets (40.2%).

Accuracy is currently indeterminable, but can surely be improved by using a larger positive/negative word database and by extending our lexical analysis to different regional languages. The data redundancy caused by multiple retweets entering the same database has to be handled in the future for precise results.

This analysis was done for the first week of April 2019, amidst great political hubbub. It is for certain that if a similar analysis was to be done at a later date, the results would vary.

Sentimental analysis proves yet again to be a useful tool for gaining insight on the perception of people on various brands, personas, events etc.

BIBLIOGRAPHY

1. https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
2. https://www.quora.com/How-can-I-read-Twitter-data-directly-in-R
3. https://www.r-bloggers.com/emoticons-decoder-for-social-media-sentimentanalysis-in-r/
4. https://eight2late.wordpress.com/2015/11/06/a-gentle-introduction-to-naive-bayes-classification-using-r/
5. https://www.labnol.org/internet/twitter-search-tricks/13693/
6. https://github.com/ayushoriginal/Sentiment-Analysis-Twitter