 Sentiment Analysis of Apple Tweets

Prof Shital Sobale   
Computer Engineering  
Vishwakarma Institute of TechnologyPune,India  
shital.sobale@vit.edu

Rahul Nagpure  
Computer Engineering  
Vishwakarma Institute of TechnologyPune,India  
[rahul.nagpure21@vit.edu](mailto:rahul.nagpure21@vit.edu)

Prajwal Kadam  
Computer Engineering  
Vishwakarma Institute of TechnologyPune,India  
[prajwal.kadam21@vit.edu](mailto:prajwal.kadam21@vit.edu)

Rushikesh Kale   
Computer Engineering  
Vishwakarma Institute of TechnologyPune,India  
[rushikesh.kale21@vit.edu](mailto:rushikesh.kale21@vit.edu)

Suyash Gaikwad  
Computer Engineering  
Vishwakarma Institute of TechnologyPune,India  
[suyash.gaikwad21@vit.edu](mailto:suyash.gaikwad21@vit.edu)

Deepak Gavit  
Computer Engineering  
Vishwakarma Institute of TechnologyPune,India  
[Deepak.gavit21@vit.edu](mailto:Deepak.gavit21@vit.edu)

***Abstract*— Twitter is an online news and social networking service that enables users to send and read short 140-character messages called "tweets". Registered users can read and post tweets, but those who are unregistered can only read them. Hence Twitter is a public platform with a mine of public opinion of people all over the world and of all age categories. As of October 2016, Twitter has more than 315 million monthly active users. Twitter Sentiment Analysis is the process of determining the emotional tone behind a series of words, used to gain an understanding of the attitudes, opinions and emotions expressed within an online mention.**

***Keywords:twitter,tweets,apple,earnings,report,sentiments,users***

Introduction

The applications of sentiment analysis are broad and powerful. Shifts in sentiment on social media have been shown to correlate with shifts in the stock market. For example, the Obama administration used sentiment analysis to gauge public opinion to policy announcements and campaign messages ahead of 2012 presidential election. The ability to quickly understand consumer attitudes and react accordingly is something that Expedia Canada took advantage of when they noticed that there was a steady increase in negative feedback to the music used in one of their television adverts.

With the rapid growth of the internet, millions of people are sharing their views and opinions on a variety of topics on micro blogging sites. On these websites user makes real time short and frequent posts about everything. These posts also include Sentiments which refers to emotions, feelings, attitude or opinion. Sentiment analysis is basically study of emotions and opinions from text. The basic idea is to analyze the results that are based on customer feedback or opinions. It is helpful for consumers who want to find out the sentiment of products before earning report publishing, or companies that want to monitor the public sentiment of their brands. Twitter sentiment analysis is tricky as compared to broad sentiment analysis because it contains slang words, misspellings and repeated characters.

**Objectives :**

* Understand the concepts of sentiment analysis.
* Understand the process of how to extract useful information from huge datasets.
* Perform text/data pre-processing activities.
* Analyse and visualize the sentiment analysis of Apple Company before and after quarterly earnings report.

**IMPLEMENTATION:**

**Code :**

#read dataset

apple = read.csv(file.choose(),header=T)

str(apple)

View(apple)

#build corpus

library(tm)

corpus <- iconv(apple$text,to='UTF-8', sub = "byte")

corpus <- Corpus(VectorSource(corpus))

inspect(corpus[1:5]) #inspect 1st 5 tweets

# Clean text

corpus <- tm\_map(corpus, tolower)

inspect(corpus[1:5])

#remove punctuation marks

corpus <- tm\_map(corpus, removePunctuation)

inspect(corpus[1:5])

#remove numbers

corpus <- tm\_map(corpus, removeNumbers)

inspect(corpus[1:5])

#remove common english words that don't add enough value

cleanset <- tm\_map(corpus, removeWords, stopwords('english'))

inspect(cleanset[1:5])

#remove URL's

removeURL <- function(x) gsub('http[[:alnum:]]\*', '', x)

cleanset <- tm\_map(cleanset, content\_transformer(removeURL))

inspect(cleanset[1:5])

#remove extra space left after cleaning

cleanset <- tm\_map(cleanset, removeWords, c('aapl', 'apple'))

cleanset <- tm\_map(cleanset, gsub,

pattern = 'stocks',

replacement = 'stock')

cleanset <- tm\_map(cleanset, stripWhitespace)

inspect(cleanset[1:5])

#term document matrix : essentially structurize the data

tdm <- TermDocumentMatrix(cleanset)

tdm

tdm <- as.matrix(tdm)#convert it into matrix

tdm[1:10, 1:20]

# Bar plot

w <- rowSums(tdm)

w <- subset(w, w>=25)#when the word frequency is more than 25 only those words are included in w

barplot(w,

las = 2,

col = rainbow(50))

# Word cloud

library(wordcloud)

w <- sort(rowSums(tdm), decreasing = TRUE)

set.seed(222)

wordcloud(words = names(w),

freq = w,

max.words = 150,

random.order = F,

min.freq = 5,

colors = brewer.pal(8, 'Dark2'),

scale = c(9, 0.3),

rot.per = 0.7)

library(wordcloud2)

w <- data.frame(names(w), w)

colnames(w) <- c('word', 'freq')

wordcloud2(w,

size = 0.9,

shape = 'star',

rotateRatio = 0.5,

minSize = 1)

#letterCloud(w,

# word = "Apple",

# size=1)

# Sentiment analysis

library(syuzhet)

library(lubridate)

library(ggplot2)

library(scales)

library(reshape2)

library(dplyr)

# Read file

apple <- read.csv(file.choose(), header = T)

tweets <- iconv(apple$text, to='UTF-8', sub = "byte")

# Obtain sentiment scores

s <- get\_nrc\_sentiment(tweets)

head(s)

tweets[4]

get\_nrc\_sentiment('delay')#testing the sentiment function for accuracy

# Bar plot

barplot(colSums(s),

las = 2,

col = rainbow(10),

ylab = 'Count',

main = 'Sentiment Scores for Apple Tweets Before Earnings Report')

#After earnings report

apple <- read.csv(file.choose(), header = T)

tweets <- iconv(apple$text, to='UTF-8', sub = "byte")

# Obtain sentiment scores

s <- get\_nrc\_sentiment(tweets)

head(s)

tweets[4]

get\_nrc\_sentiment('delay')#testing the sentiment function for accuracy

# Bar plot

barplot(colSums(s),

las = 2,

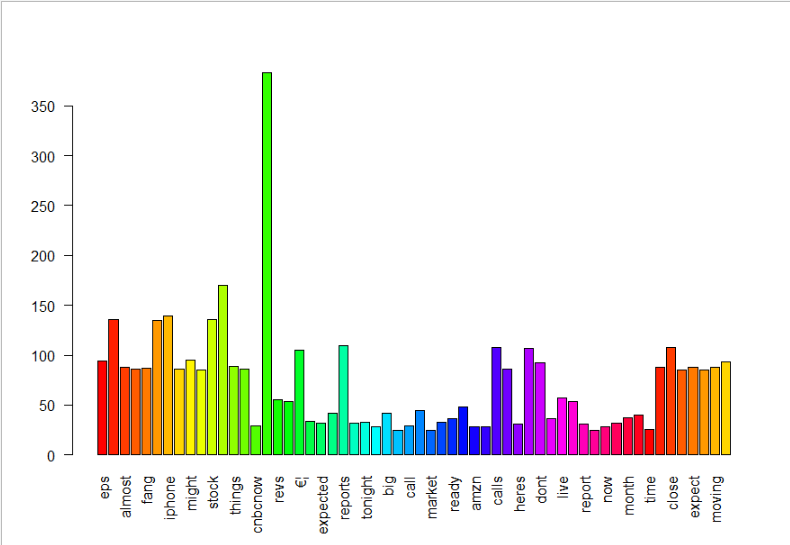
col = rainbow(10),

ylab = 'Count',

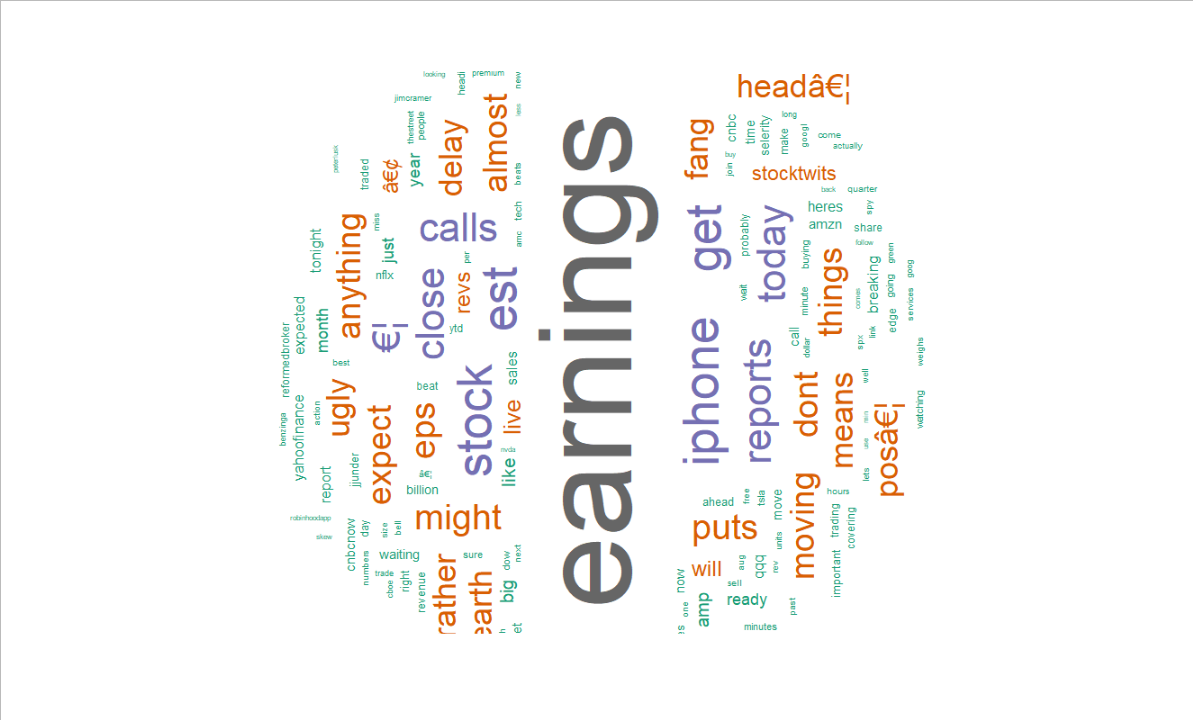
main = 'Sentiment Scores for Apple Tweets After Earnings Report')

**Results:**

**Most frequently used words in the tweets :**

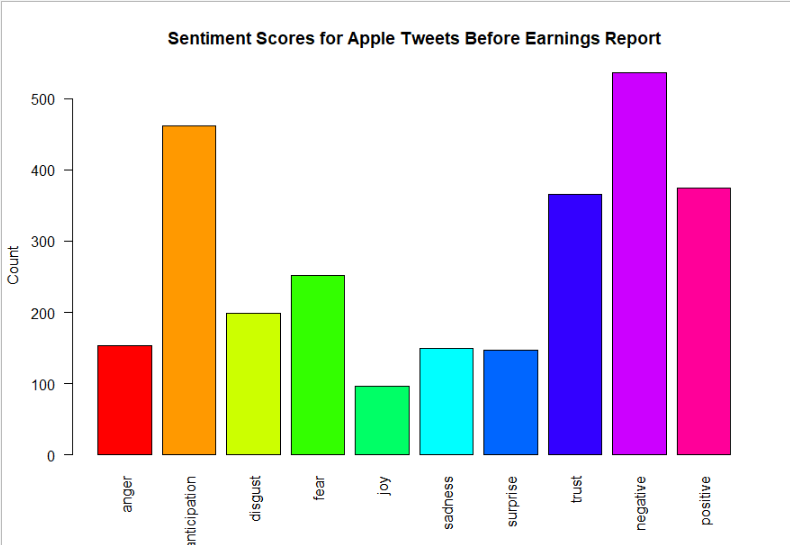


**Wordclouds for most frequently used words in the tweets:**

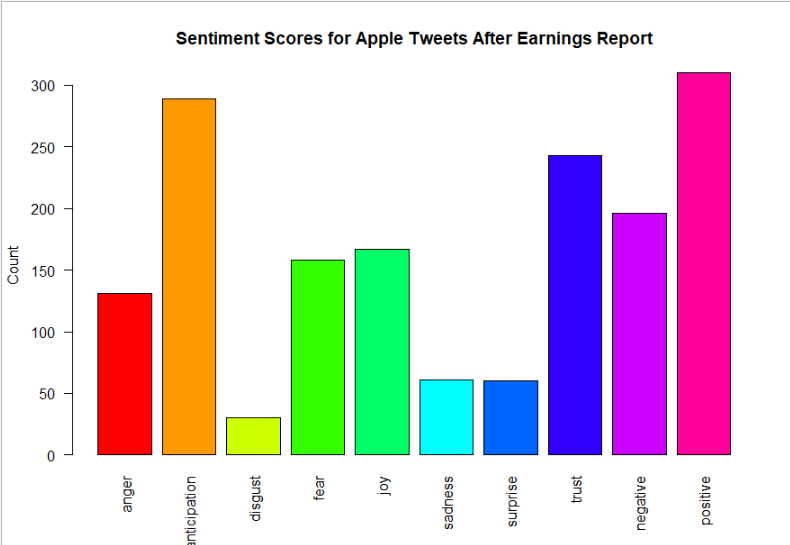




**Sentiment analysis plot before earnings report is published :**



**Sentiment analysis plot after earnings report is published:**



CONCLUSION

We can clearly derive and understand how the sentiment of common people has changed with respect to Apple Company before and after the earnings report. We also appreciated the most commonly used words in tweets. Also we learnt about different packages and cleaning the dataset especially in such cases where we are trying to extract crucial business data from subjective sentences.

References

[1] <http://www.rdatamining.com/docs/twitter-analysis-with-r>

[2]<https://jeffreybreen.wordpress.com/2011/07/04/twitter-text-mining-r->

[3] <https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html>

[4] <https://www.quora.com/How-can-I-read-Twitter-data-directly-in-R>

[5] <http://www.rdatamining.com/docs/twitter-analysis-with-r>

[6] <https://github.com/datumbox/twitter-sentiment-analysis>

[7] <https://www.r-bloggers.com/emoticons-decoder-for-social-media-sentimentanalysis-in-r/>

[8] <https://www.r-bloggers.com/twitter-sentiment-analysis-with-r/>