SENTIMENT ANALYSIS OF SOCIAL MEDIA CONTENTS

HIMANSHU RAWAT

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 will be analyzing the sentiments that are reflected from the post of people on social media networks.

➢ For our study, we have chosen “Twitter” as our social media platform. And the coding language that we used to perform the analysis is R.

TWITTER

•An online social networking service that enables users to send and read short 140-character messages called “tweets” (Wikipedia).

• Over 320 million monthly active users (as of 2016).

• Creating over 500 million tweets per day

Techniques and Tools

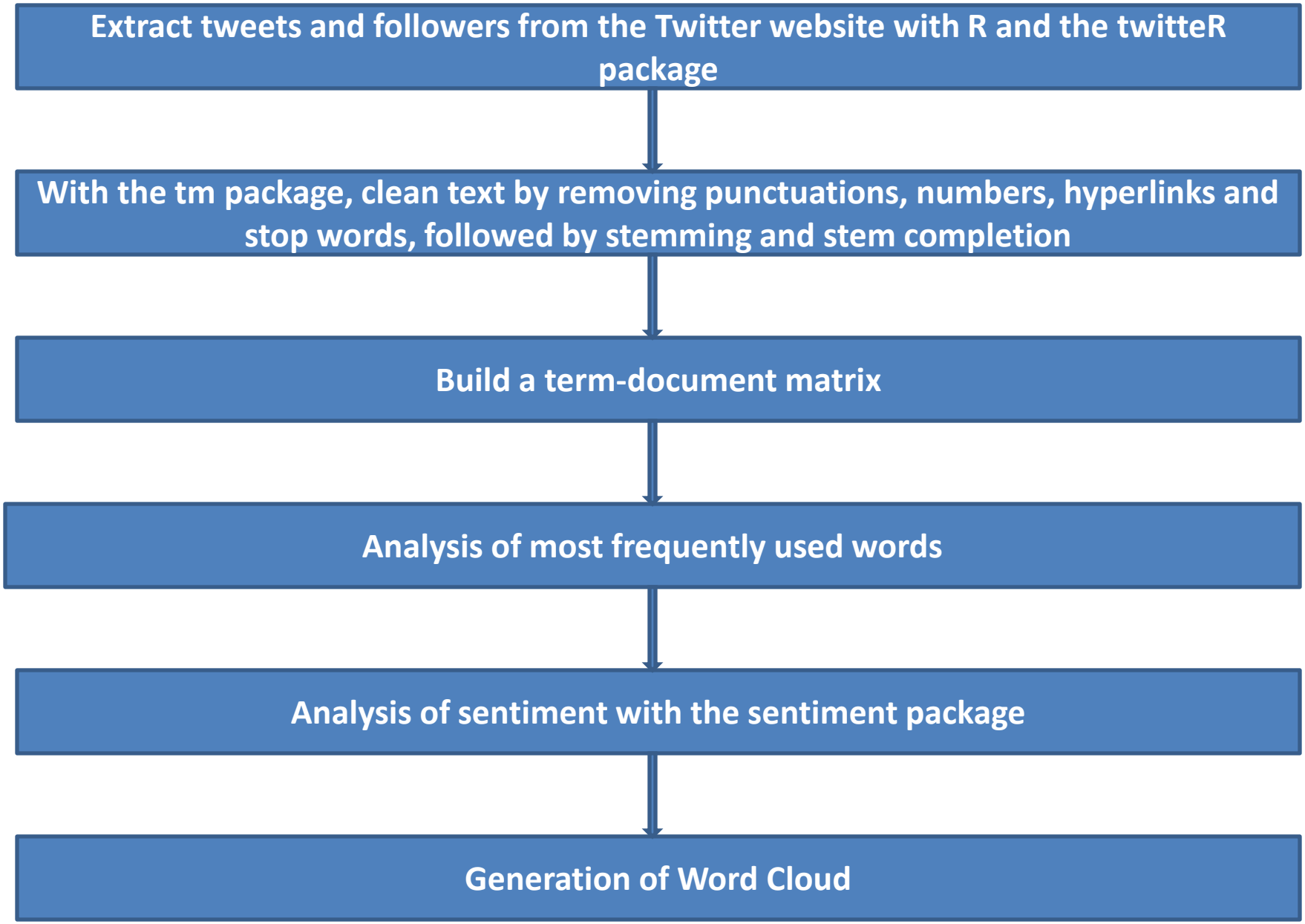
Techniques

- Text mining
- Topic modelling
- Sentiment analysis
- Social network analysis

Tools

- Twitter API
- R and its packages:
  - twitteR
  - GGPLOT2
  - ROAuth
  - sentiment
  - Word Cloud

PROCESS FLOW



Functions & Generated Results

(Case Study on Moto E 5 Plus Launch)

Launched: 10 July-18

Retrieve Tweets

Filter

	text	favorited	favoriteCount	replyToSN	created	truncated	replyToSID
1	RT @amitbhawani: Motorola #MotoE5Plus Unboxing - S...	FALSE	0	NA	2018-07-10 15:07:30	FALSE	NA
2	Here's the Unboxing of the #MotoE5Plus from @Motoro...	FALSE	1	NA	2018-07-10 15:07:22	TRUE	NA
3	RT @amitbhawani: Motorola #MotoE5Plus Unboxing - S...	FALSE	0	NA	2018-07-10 15:06:57	FALSE	NA
4	Here's the Unboxing of the #MotoE5Plus from @Motoro...	FALSE	0	NA	2018-07-10 15:06:17	TRUE	NA
5	Here's the Unboxing of the #MotoE5Plus from @Motoro...	FALSE	0	NA	2018-07-10 15:05:33	TRUE	NA
6	RT @amitbhawani: Motorola #MotoE5Plus Unboxing - S...	FALSE	0	NA	2018-07-10 15:04:40	FALSE	NA
7	RT @deepak_rsingh: Do not miss what our tech expert ...	FALSE	0	NA	2018-07-10 15:04:36	FALSE	NA
8	RT @KartikAaryanHQ: Starboy Kartik Aaryan ready to lau...	FALSE	0	NA	2018-07-10 15:04:31	FALSE	NA
9	RT @amitbhawani: Motorola #MotoE5Plus Unboxing - S...	FALSE	0	NA	2018-07-10 15:04:30	FALSE	NA

Showing 1 to 10 of 10,000 entries

Following functions have been used:

- **twitteR**-for connecting with twitter rest api
- **ROAuth**- for authorization with twitter

*Tweets were and extracted and then converted to a data-frame as shown here.

*No of Tweets extracted =10,000

Text Cleaning

Show Attributes

Name	Type	Value
myCorpus	list [10000] (S3: SimpleCorpus, C	List of length 10000
1	list [2] (S3: PlainTextDocument, T	List of length 2
content	character [1]	'rt amitbhawani motorola motoeplus unboxing smartphone mah battery video buy pho ...
meta	list [7] (S3: TextDocumentMeta)	List of length 7
2	list [2] (S3: PlainTextDocument, T	List of length 2
content	character [1]	'heres unboxing motoeplus motorolaindia phoneradarblog amitbhawani '
meta	list [7] (S3: TextDocumentMeta)	List of length 7
3	list [2] (S3: PlainTextDocument, T	List of length 2
content	character [1]	'rt amitbhawani motorola motoeplus unboxing smartphone mah battery video buy pho ...
meta	list [7] (S3: TextDocumentMeta)	List of length 7

myCorpus

“tm” function is used to pre-process(cleaning, removing hashtags ,hyperlinks etc)

The following operations were performed in a step wsie manner

Tweets were converted to lowercase

Removal of URL’s

Removal of number punctuations

Removal of Stop Words

Removal of white spaces.

Stemming

```r
# Stem & Stemming
myCorpus <- tm_map(myCorpus, stemDocument) # stem words
writeLines(strwrap(myCorpus[[190]]$content, 60))
## r refer card data mine now provid link packag cran packag
## mapreduc hadoop ad
stemCompletion2 <- function(x, dictionary) {
  x <- unlist(strsplit(as.character(x), " "))
  x <- x[x != ""]
  x <- stemCompletion(x, dictionary=dictionary)
  x <- paste(x, sep="", collapse=" ")
  PlainTextDocument(stripWhitespace(x))
}
myCorpus <- lapply(myCorpus, stemCompletion2, dictionary=myCorpusCopy)
myCorpus <- Corpus(VectorSource(myCorpus))
writeLines(strwrap(myCorpus[[190]]$content, 60))
```

This is also a part of “tm” function and was used for stemming so that root words were consolidated and arbitrary words removed.

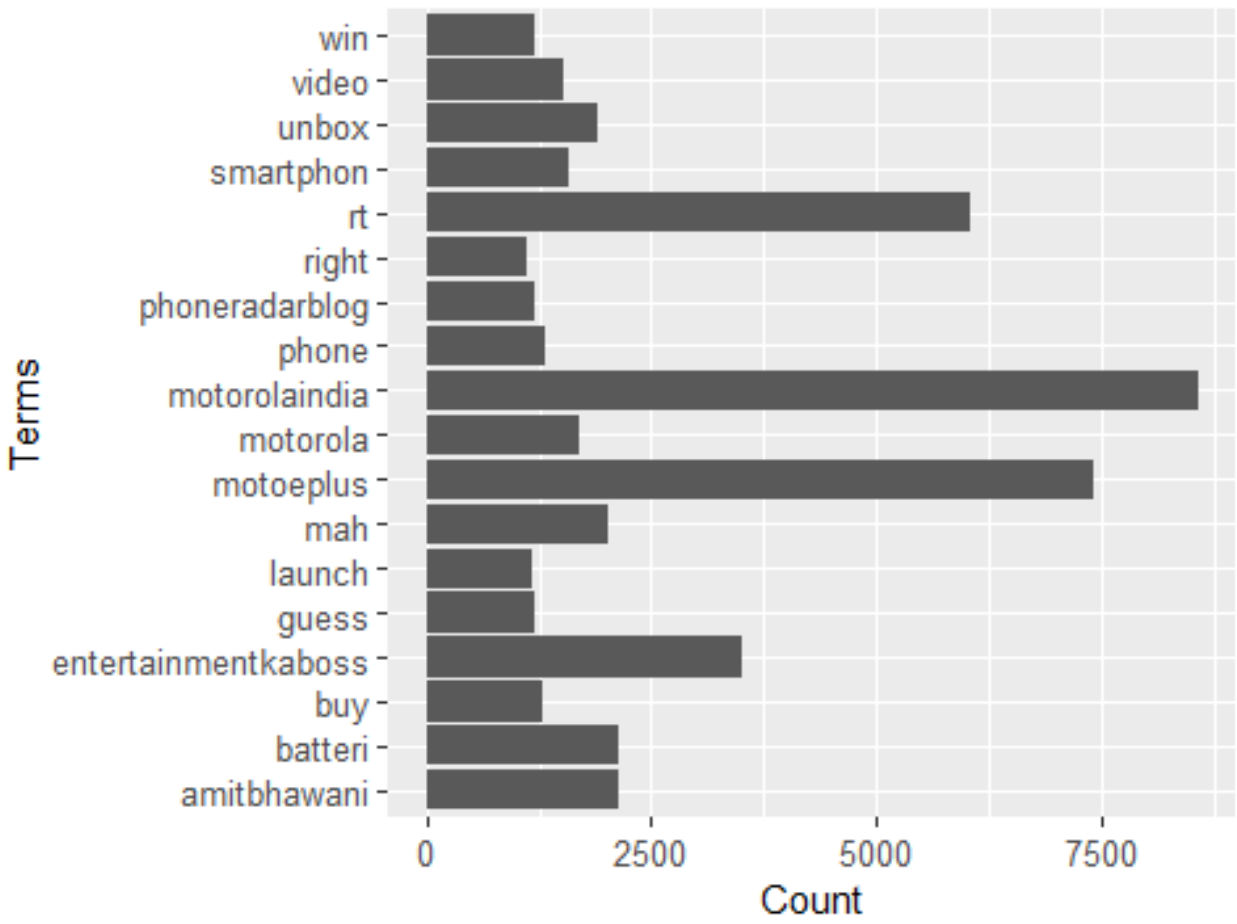
Term Doc Matrix

```r
##Build Term Document Matrix

tdm <- TermDocumentMatrix(myCorpus,
                          control = list(wordLengths = c(1, Inf)))
tdm

idx <- which(dimnames(tdm)$Terms %in% c("r", "data", "mining"))
as.matrix(tdm[idx, 21:30])
```

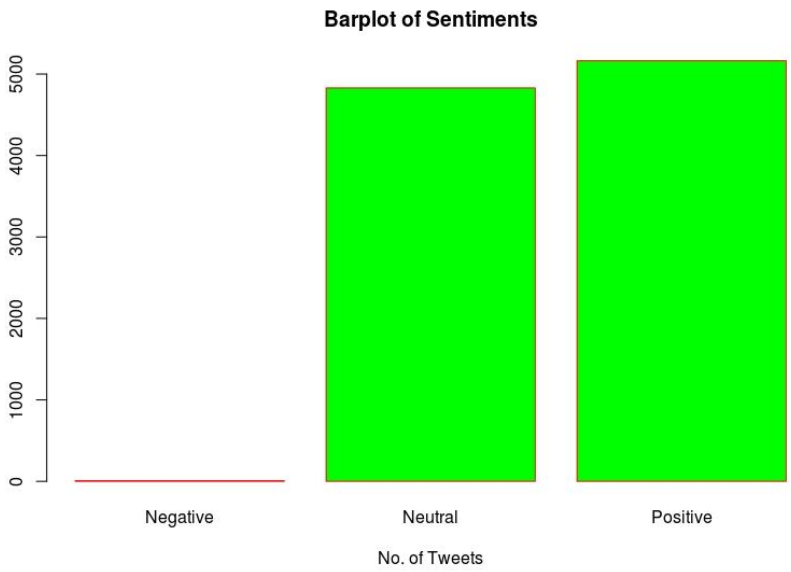
Term doc matrix was created to obtain the frequency of words in the collection of tweets. This matrix is then used to generate the words frequency plot.

The matrix is further employed to generate the sentimental plots.

Words frequency Plot



Sentiment Analysis- Maximum Entropy Classifier



The independent words extracted from each tweet are matched with predefined lexicon. The lexicon is generated containing positive, negative & neutral sentiments using Maximum Entropy Classifier Approach. The cumulative score of each word gives the tweet a value(-1,0,+1), on the basis of which the tweet is defined as negative, neutral or positive. The count of each individual tweet is then taken to determine whether the sentiment is negative, neutral or positive for a particular subject (MotoE5 plus in our study).

STATISTICS:

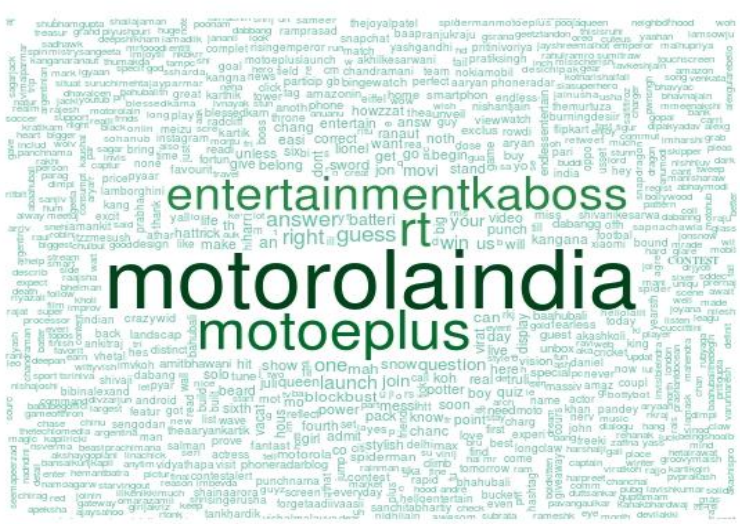
Total Tweets : 10,000 **Hashtag : Moto E5 Plus**

Positive :5163

Negative :7

Neutral :4830

Word Cloud



A Word cloud is generated basis the frequency of words appearing in the tweets.

The most frequently occurring words are displayed more prominently in the cloud of

All the words present in the tweets.

This is achieved by using “Word Cloud “ function .