

# **Sentiment Analysis on Demonetization in India**

Varshini Sriram, Jeyassri B, Shivani Chander, DataByte, NIT Trichy

## **Goal**

To do Sentiment Analysis on Demonetization in India based on opinions expressed via Twitter

## **Introduction**

Sentiment analysis (also called as opinion mining) is the task of identifying whether the opinion expressed in a text is positive or negative in general about a particular topic or context.

Prime Minister of India, Narendra Modi announced the demonetization in an unscheduled live televised address at 20:00 Indian Standard Time (IST) on 8 November. In the announcement, Modi declared that use of all ₹500 and ₹1000 banknotes of the Mahatma Gandhi Series would be invalid past midnight, and announced the issuance of new ₹500 and ₹2000 banknotes of the Mahatma Gandhi New Series in exchange for the old banknotes.

This was received by the masses with mixed feelings. This project aimed at structuring and analyzing those feelings, especially the ones that were expressed on Twitter.

Twitter is an online micro-blogging and social-networking platform which allows users to write short status updates of maximum length 140 characters. It is a rapidly expanding service with over 200 million registered users generating nearly 250 million tweets per day. Due to this large amount of usage we hope to achieve a reflection of public sentiment by analyzing the sentiments expressed in the tweets.

## **Data Collection and Analysis**

The dataset (containing 8000 tweets) was obtained from Kaggle and R was used for the analysis.

## **Analysis**

The sentiment (score/mood) of each tweet was analyzed using 'bag of words' method. In this method, the sentences/text is broken down into words which are treated as a unit/token. The text field of each tweet is pulled out, cleaned and preprocessed. Lexicon based Sentiment Analysis was used for calculating the sentiment score of the tweets. Lexicon is a list of words tagged positive and negative. Each tweet was parsed and scanned for the presence of words matching the words from the list of positive and negative words from Lexicon. Based on the count of positive and negative words in the text, scores of +1 and -1 were allotted to each positive and negative word respectively. The score of a tweet was the net score of all its words. This process is repeated for all 8000 tweets. Based on the score, the tweet was identified as positive ( $>0$ ), negative ( $<0$ ) or neutral ( $=0$ ).

Example: "Demonetization is brilliant. But people are suffering a bit."

Score = 0 (+brilliant, -suffering), Sentiment = Neutral

## Results

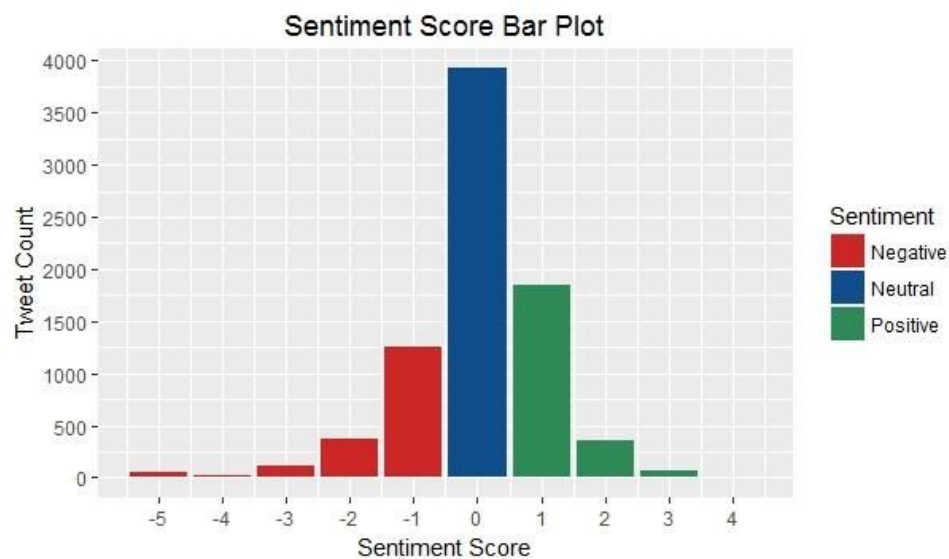
### Sentiment Score

The distribution of scores obtained was as follows:

Score	-5	-4	-3	-2	-1	0	1	2	3	4
Tweets	47	26	112	371	1246	3922	1848	357	63	8

Mean score: 0.015

The mean score is slightly positive inferring that people are not having negative sentiment towards demonetization announcement.

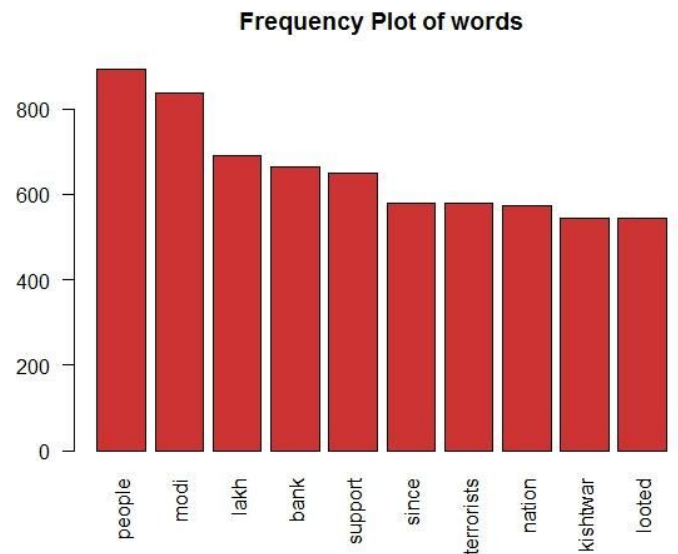


Sentiment	Tweets
Negative	1802
Neutral	3922
Positive	2276

Clearly, the sentiment of most of the tweets are neutral. Also, more number of positive tweets exist than negative ones which shows that the sentiments of people regarding demonetization are positive.

### Wordcloud and Frequency plot of words

A word cloud was prepared based on the frequency of occurrence of words. The most frequently used keywords are towards the center of the word cloud. Size is related to the word frequency. As the frequency decreases, the size also decreases. Term Frequency plot gives the most frequent terms in the tweets along with their frequencies as a bar plot.



## Word Association Plot

Correlation of the word 'modi' with other words

words	Correlation value
narendra	0.50
app	0.41
strategy	0.34
rmg	0.34
pygk	0.34
feedback	0.34
taking	0.33
opposing	0.23
decide	0.22
vote	0.21
totally	0.21
ratantata	0.21
proves	0.21
mounting	0.21
misery	0.21
lth	0.21
indifferent	0.21
hards	0.21
frthy	0.21
demonetizat	0.21
bypolls	0.21

## Dendrogram

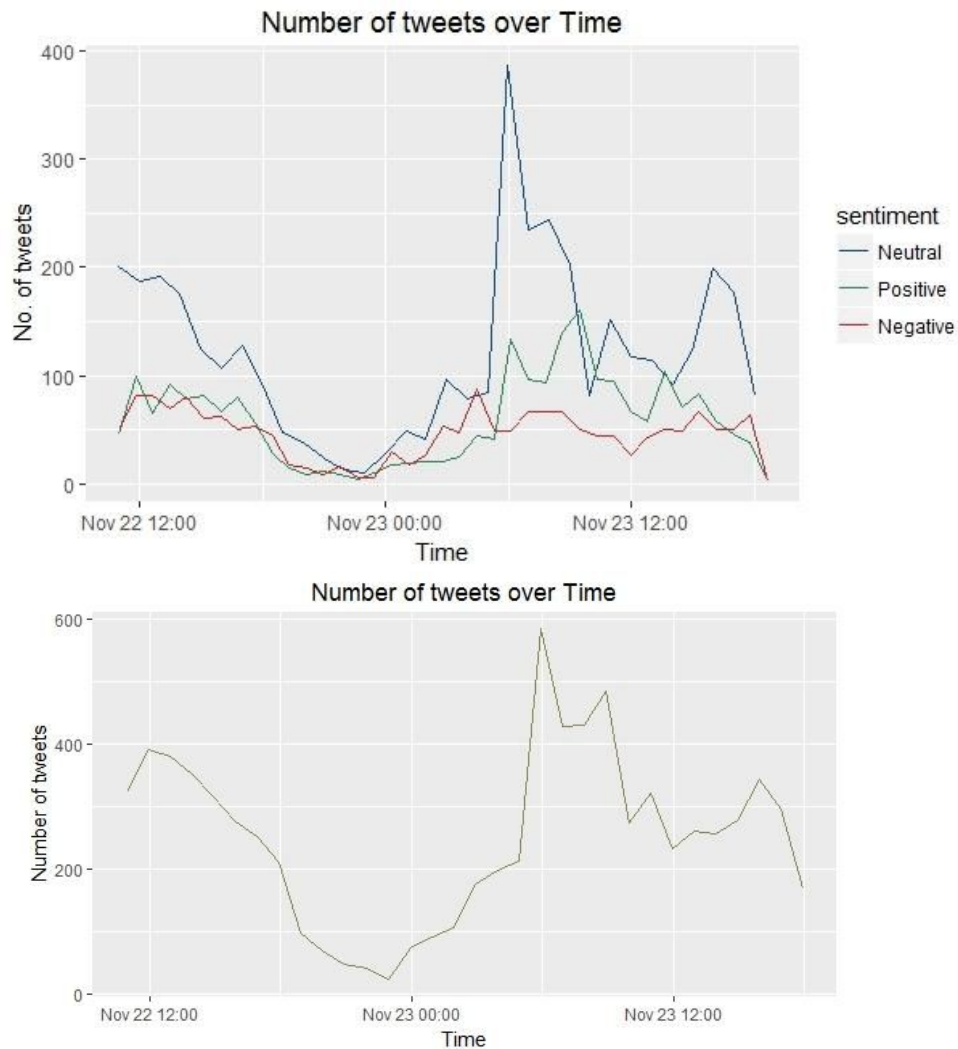
### Cluster Dendrogram

Height

distance  
hclust(\*, "complete")

### Variation of Tweets with time

The number of positive tweets are mostly greater than the number of negative tweets at any point of time. The number of neutral tweets are greater than both at any point of time.



## Inferences

A sentiment analysis was carried out on demonetization with twitter data to understand the people's reaction towards the demonetization step by PM Narendra Modi. Bag of words approach with lexicon based analysis was employed to calculate the sentiments. The results obtained gave strong evidence that the people of India support demonetization and are having a positive sentiment towards the Demonetization step. Social media gives a thumbs up to Demonetization!

## Limitations

The effect of negation words and sarcasm was not considered while calculating the sentiment score. Since the approach used was lexicon based, the list of positive and negative words can never be the complete list and hence most tweets were scored as neutral.