BIA658C Social Network Analysis

FINAL PROJECT REPORT

Neha Mansinghka (CWID: 10420458)

2016

# EXECUTIVE SUMMARY

Demonetization refers to the recent move by Indian prime minister of ceasing high value currency notes (Rs. 500 and Rs. 1000) as legal tender from 9th November onwards. This report describes the sentiment analysis performed on a sample of tweets related to this topic. The sample consists of data collected from 22nd Nov – 23rd Nov of 8000 tweets and from 1st Dec – 4th Dec of around 800 tweets. Both samples were collected using “#demonetization” as the search term. A simple word matching algorithm has been used to perform the sentiment analysis. Tweets have been classified in three sentiment categories – positive, negative and neutral using this algorithm. The analysis shows that neutral tweets are the dominant category in both samples. Moreover, both the positive and negative tweets are comprised of majority of tweets with score 1 and -1 respectively. These are considered slightly positive or negative. Tweets with scores 3 or 4 or 5 or conversely -3 or -4 or -5 which would represent strongly positive or strongly negative tweets only form about 10% of the entire data set (combination of two samples). The neutral tweets are more informative and promotional rather than expressing a neutral sentiment. Although, this seems a counterintuitive result for such a significant and sudden change that has resulted in nationwide chaos, possible explanations for this are the time period lapsed between the implementation of the change and the data collection and the limited percentage of the population that are active social media users. Finally, there are limitations related to the samples and the algorithm used and the results of the analysis should be viewed in context of these limitations.

# INTRODUCTION

Recently, the prime minister of India implemented a demonetization policy in which [₹500](https://en.wikipedia.org/wiki/Indian_500-rupee_note) and [₹1,000](https://en.wikipedia.org/wiki/Indian_1000-rupee_note) bank notes were discontinued as a legally valid currency. This move created quite a lot of conversation on Twitter due to its substantial impact on the entire nation. This report attempts to perform a sentiment analysis on a sample of the tweets related to this topic. It first provides some background on the demonetization move. Then a description of the methodology used to perform sentiment analysis is provided including information on software tools and packages used. Following that the results of the analysis are presented. It then presents a conclusion and a short discussion of it. Finally, it addresses the limitations of the analysis and further work that can be undertaken to improve the accuracy of such an analysis.

# BACKGROUND

On November 8, 2016, the prime minister of India, Narendra Modi, announced the discontinuation of high value currency notes (Rs. 500 and Rs. 1000) as legal tender from 9th November onwards in a sudden and unprecedented move to curb corruption and the untaxed ‘black money’ (Team, 2016). The announcement was made in an unscheduled live telecast at 8pm on 8th November and the policy was effective from mid-night, within a few hours of the announcement (*Demonetisation of Rs. 500 and Rs. 1000 notes: RBI explains*, 2016). The move caused a huge chaos in the country in which estimates say around 90% of the transactions are still performed in cash and the bills discontinued are estimated to be around 86% of the currency in circulation (Team, 2016).

# SENTIMENT ANALYSIS

## METHODOLOGY & TOOLS

A sentiment analysis has been performed on a sample of tweets related to the demonetization move using RStudio. The sample tweets are from 22nd - 23rd November and from 1st – 4th December and collected using “#demonetization” as the search term. The sample data for 22nd - 23rd November has been obtained from the Kaggle website (https://www.kaggle.com/arathee2/demonetization-in-india-twitter-data) and consists of 8000 tweets. The sample data for 1st – 4th December has been collected using the ‘twitteR’ package of R and consists of around 800 tweets. The script used to collect the second set of tweets is called “twitter\_search.R”. The initial analysis results for sample data from 22nd- 23rd November generated curiosity regarding any change in sentiment as more time had passed since this massive change was implemented. Therefore, another sample was collected from 1st – 4th December. It would have made for a better comparative analysis if another equal sized sample was used but due to Twitter API restriction of 180 tweets in one call, only a sample of around 800 tweets could be collected. The visualizations have been created using both RStudio and Excel.

A simple word matching methodology has been used to identify the sentiments expressed in the tweets towards the demonetization move (Breen, 2011). Two lexicons, one of positive words and another of negative words, have been used as the base (Sanchez, 2012). Each tweet is then matched against each of these lexicons and receives a positive score for the number of positive matches and a negative score for the number of negative matches. The negative score is then subtracted from the positive score to get a final score for the tweet. If the final score is positive, then the tweet has more positive matches than negative matches and is classified as a positive sentiment. If the final score is negative, then the number of negative matches is higher than the number of positive matches and the tweet is classified as a negative sentiment. If there are no matches to either of the two lexicons, then the score is 0. If the number of negative matches is equal to the number of positive matches, then the final score is 0 too. Tweets with score 0 are classified as neutral tweets. The script used to perform the sentiment analysis is called “demonetisation\_sentiment\_analysis\_1.R”

For example, a tweet such as *“Now my bank sends me* ***propaganda*** *emails on why #demonetization is* ***good****. I am* ***scared****, very* ***scared****!”* receives a score of -2 using this method. The words in red bold are negative matches to the negative words lexicon and the words in green bold are positive matches.

(Number of positive matches) – (Number of negative matches) = 1 – 3 = -2

Another example, a tweet such as *“#Demonetization #IndiaFightsCorruption* ***Immense******courage*** *needed to take such a* ***momentous*** *decision”* receives a score of 3.

(Number of positive matches) – (Number of negative matches) = 3 – 0 = 3

Once the tweets were classified using this method, the following results were obtained from the analysis.

## ANALYSIS RESULTS

The below chart shows the score distribution of all tweets combined. This includes 8000 tweets from 22nd - 23rd November 2016 and 790 tweets from 1st – 4th December 2016.

Figure 1

It can be seen from the above chart that neutral tweets have the highest frequency followed by slightly positive tweets having the second highest frequency and slightly negative tweets having the third highest frequency. The below charts show the trend of positive, negative and neutral tweets over the two periods for which data has been collected.

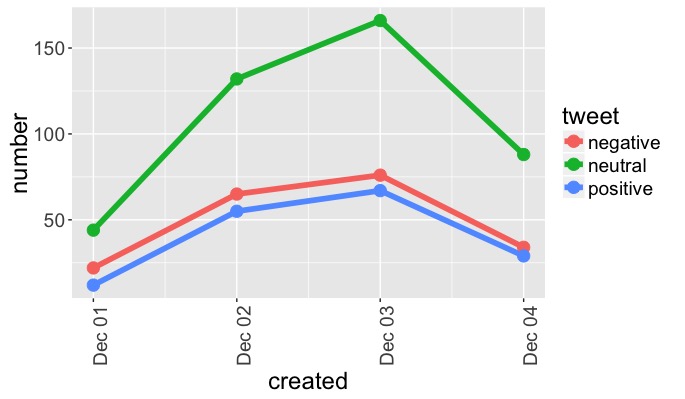
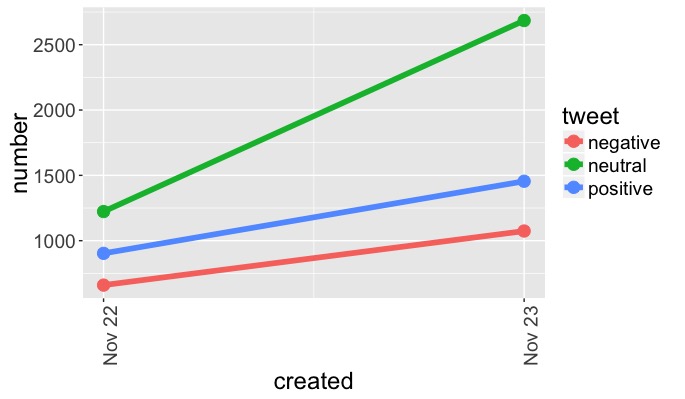


Figure Figure

Both charts show that the positive and negative tweets increase and decrease by almost the same amount while the increase and decrease in neutral tweets is much higher than the other two categories. For example, from Nov 22nd to Nov 23rd the count for neutral tweets increases by 1500 tweets but the count for positive and negative tweets only increases by 600 and 400 tweets respectively. Similarly, from Dec 1st to Dec 2nd the count for neutral tweets increases by about 125 tweets but the count for positive and negative tweets only increases by 40 tweets each. The increase in neutral tweets is almost 3 times as much as the increase in other two types of tweets. The following charts show the percentage wise split of the neutral, positive and negative tweets for the two periods:

Figure 4 Figure 5

The neutral tweets constitute half of all the tweets collected and analyzed for both periods. The percentage of positive tweets is higher than the negative ones in the earlier period but this situation reverses in the latter period from 1st Dec – 4th Dec.

From this analysis, it appears that no strongly positive or strongly negative sentiments are dominant on Twitter related to the demonetization topic. The neutral sentiment is the most dominant one. Therefore, the tweets data was examined more closely to study the neutral tweets and understand the general nature and content of these tweets.

It is noted that majority of the tweets that have been scored 0 and hence considered neutral are either more of an informative tweet (sharing news) or a promotional tweet rather than expressing a neutral sentiment towards the topic in question or the negative or positive sentiment in the tweet has not been picked up by the algorithm due to limitations discussed in the next section. Some examples of tweets that are informative rather than expressive and are scored neutral are:

*“Round table discussion on #Demonetization with @theicai Vice President @NileshVik”*

*“Stay tuned for a few anecdotal tweets about #demonetization from speaking with a small cross section of people in Delhi. 1/n”*

Examples of tweets that are promotional and are scored neutral are:

*“Now link multiple #bank #accounts with same virtual address in #UPI on @mypoolin https://t.co/OPTHZLphSF #Demonetization”*

*“The latest The Anglo-India Central Daily! https://t.co/3yDj8iOHlJ #demonetization #ausvnz”*

*“Did you vote on #Demonetization on Modi survey app?”*

# CONCLUSION AND DISCUSSION

The analysis presents some interesting insights into the social media activity related to the demonetization move on Twitter. It would be logical to expect more of the strongly positive or strongly negative tweets in the aftermath of such a sudden move resulting in nationwide chaos. The analysis shows quite different results though. The neutral tweets make up about 50% of the entire dataset while the strongly positive and strongly negative tweets (scores with absolute value of 3, 4 or 5) only make up about 10% of the entire dataset. One reason for this could be the time period of the data collection. The first set of data is from two weeks after the change was implemented and the second set is from three weeks after it was implemented. It would aid the analysis if twitter data from few days immediately after the change was implemented could be accessed. These tweets could be analyzed to see if neutral tweets have been the dominant activity throughout or only after the immediate aftermath of the change. Furthermore, the results seem less counterintuitive when one considers that out of a population of 1.27 billion (PTI, 2015) in India, only 12% are active social media users (Velayanikal, 2016) and the percentage of Twitter users is likely to be lower than this number too. Therefore, the Twitter data provides a very small and not well represented sample of the population.

# LIMITATIONS AND FUTURE WORK

There are certain limitations of this analysis that are pointed out so the reader can view the results in the context of these limitations. This would aid in a more informed decision making based on such an analysis. The limitations related to the representativeness of the sample, the time period and unequal sample sizes have already been discussed in previous sections.

The word matching methodology used here also has certain limitations. It considers each word in the tweet in isolation. This results in inaccurate identification of sentiment in certain scenarios. For example, words “not good” appearing in a sentence would be analyzed separately and the tweet would receive a score of 1 based on “good” having a match to the positive word lexicon. However, here the sentiment expressed is negative and only if the words are analyzed together that this would be identified. Machine learning techniques can be utilized to build deep learning models that analyze contiguous blocks of words together and provide a more accurate identification of the sentiment (Socher et al., 2013).

Emojis have been excluded from this analysis. However, emojis can be an equally important part of sentiment analysis as text words are. They enhance the sentiment expressed in the sentence. In some instances, they are the only means of identifying the sentiment. For example, a sentence like “I am going home” is a neutral sentence on its own. However, with a happy or sad smiley at the end it turns into a positive or negative sentiment respectively. Therefore, including them in such an analysis would result in a more complete identification of the sentiment.

# BIBLIOGRAPHY

Breen, J. (2011) *slides from my R tutorial on Twitter text mining #rstats*. Available at: <https://jeffreybreen.wordpress.com/2011/07/04/twitter-text-mining-r-slides/> (Accessed: 26th November 2016).

*Demonetisation of Rs. 500 and Rs. 1000 notes: RBI explains* (2016). Available at: <http://www.thehindu.com/news/national/Demonetisation-of-Rs.-500-and-Rs.-1000-notes-RBI-explains/article16440296.ece> (Accessed: 9th December 2016).

PTI (2015) *India’s population as of 5pm today: 127,42,39,769 and growing*: The Indian Express. Available at: <http://indianexpress.com/article/india/india-others/indias-population-as-of-today-1274239769-and-growing/> (Accessed: 2nd December 2016).

Sanchez, G. (2012) *Mining twitter with R*. Available at: <https://sites.google.com/site/miningtwitter/questions/sentiment/analysis> (Accessed: 26th November 2016).

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y. and Potts, C. (2013) *Deeply Moving: Deep Learning for Sentiment Analysis*. Available at: <http://nlp.stanford.edu/sentiment/> (Accessed: 2nd December 2016).

Team, T. (2016) *Demonetization Will Impact Amazon's Growth In India*. Great Speculations. Available at: <http://www.forbes.com/sites/greatspeculations/2016/11/29/demonetization-will-impact-amazons-growth-in-india/#59e075151de4> (Accessed: 4th December 2016).

Velayanikal, M. (2016) *The latest numbers on web, mobile, and social media in India*. Available at: <https://www.techinasia.com/india-web-mobile-data-series-2016> (Accessed: 1st December 2016).