

SENTIMENT ANALYSIS ON DEMONETIZATION BY GOVT OF INDIA

Sachin Kakkookal

Student, PGDM-Business Analytics

REVA Academy for Corporate Excellence, REVA University

Bangalore, India

Sachink.ba02@reva.edu.in

Nidhil C. H

Student, PGDM-Business Analytics

REVA Academy for Corporate Excellence, REVA University

Bangalore, India

Nidhil.ba02@reva.edu.in

ABSTRACT

Any system is subject to change, modifications and amendments and same go with government policies as well. However, all such changes will have a downside where the common man must face most of the repercussions. This paper analyses one such policy which has been trending in the social media since November 2016.

Since the announcement of Demonetization, the Indian economy has been fluctuating in terms of inflation and GDP rate. This has affected several small-scale businesses and individuals drastically.

This paper is aimed at reviewing the general implications of demonetization on people. This Research is based on Sentiment Analysis or opinion mapping using Naïve Bayes classification algorithm.

In recent years, microblogging websites have evolved to become a source of varied kind of information. One such tool is Twitter. In this paper, the data for analysis is gathered from Twitter, Facebook and other public forums and then sentiment analysis is applied using Natural Language Processing API's. Sentiment analysis is a method of classifying sentiments from a given text. This helps us to understand how an entity has influenced the minds of the general population. Sentiment analysis is also important to know what the common public think about demonetization. It is also a computational study of opinions, sentiment and emotions expressed in the text.

Keywords: *Demonetization, Sentiment Analysis, Twitter, Opinion mining, Natural Language Processing. Naïve Bayes algorithm*

1.0 INTRODUCTION

Demonetization as a Government policy brought changes to the Indian economy and was initiated to boost the financial growth for a better future. India is one of the fastest growing economies in the world with an average rate of seven percent GDP over several years in spite of massive corruption and other serious issues such as terrorism. The Central Government of India declared a currency ban of 500 INR and 1000 INR on November 8th, 2016. But the twist was that the declaration was made on the evening of November 8th, 2016 without any prior notice which took the entire nation by surprise. This demonetization procedure resulted in a crisis soon after the declaration of the government could not generate the currency on such a large scale that could reach out to all the walks of people in the society. Moreover, the process of demonetisation involved recalibration of ATM's all over the country since the size of the newly introduced currency was different from what was already in circulation. Initially, the economic experts, statisticians and global leaders supported the move by the Government. On the other hand, it was heavily criticized by the public and opposition party leaders, ultimately triggering organized protests against the approach of the government in several parts of India. Demonetization was implemented with an intention to fight corruption, terrorism financing and inflation. However, it was often criticized as a foolish act with serious repercussions on the Indian economy due to its poor implementation.

The aim of this study is to perform opinion mining on Social media to examine the impact of this demonetization from public opinions which in turn, could be used by concerned authorities to smoothen such processes in future.

2.0 LITERATURE REVIEW

The general purpose of Text Mining techniques is to process unstructured data and extract meaningful insights from it and convert the information into more structured data which can be used for various statistical and machine learning algorithms (Text mining by statsoft.com). the data extracted can be summarised into words contained in the text. The text analytics became more popular in the late 1990's, though it has been existing from mid-1980's done mostly manually (Roy 2017). At a basic level, words from various documents are indexed and counted to form a table of documents and words (Hill et al., 2006). However, once the table of unique words has been derived which is also called as tokens, all possible data mining techniques can be applied to create a cluster of words, and this can also identify important themes that best predict other variables of interest. Over the time text mining techniques have improved significantly with the help of many research and cutting-edge complex techniques (Weiss et al., 2010).

Once a data matrix is built with important themes derived from the input documents, various popular analytic methodologies such as clustering, decision trees can be used for further processing of the data (Hill et al., 2006).

A few of the text mining tools are considered as "black box" methods which can extract "deep meaning" from the documents with little human intervention. These text mining tools rely on standard algorithms for deriving meaningful insights from the text and can summarize a large amount of data automatically, retaining vital meaning from those documents. While there are several algorithmic approaches to do this, these technologies are very much still in its early phase, and summarizing such large data still remain difficult (Hill et al., 2006).

Text mining is also used to search documents. Here a large number of documents are scanned based on keywords or key phrases. This is predominantly used by popular internet search engines such as Google, Bing to have efficient access to Web pages with certain content.

There have been prior studies in the field of analysing social media with text mining approaches. Bae et al., (2012) demonstrated how the sentiment analysis can be used to measure a valid popularity indicator. The authors developed a positive to negative measure to distinguish themes based on their key findings. First, the differentiation was made between the positive and negative audiences of popular users. Second, they found that the sentiments expressed by these users influenced the sentiment of their respective audience. Lastly, the positive-negative measure was used to find their influence on Granger causality analysis. Further, the time series based change in positive - negative sentiments of a particular audience was conducted related to the real-time sentiment landscape of popular users.

Paltoglou et al., (2016) analysed if sentiment analysis technique can be used for detecting significant events that occur in the world. They explored whether sudden changes in the positive or negative mentality that keywords are typically associated with can be exploited for this purpose. He demonstrated that the number of tweets related to event detection is more important factors than the number of days used to extract token frequency. Focus on event detection concluded that all approaches are dependent on the level of coverage such events receive in social media.

Also, Hridoy and colleagues (2015) discussed a methodology which allows utilization and interpretation of Twitter data to determine public opinions. The main focus of this research was to analyze the tweets about iPhone 6 to understand the opinions of people based on feature-specific popularity analysis and male-female specific analysis. The tweets were analysed found to be mixed but there was general consistency in the customer's reviews and comments.

3.0 DATA COLLECTION

The data was collected randomly from several sources such as Twitter, Facebook and some of the public forums. Data collection process started with the identification of the hashtags that were used for the tweets related to this act. A corpus manager called sketch engine is used to label the data. Sketch Engine is generally used to enable people studying language behaviour to search for large text collections based on the complexity of the queries. Sketch Engine gained its popularity after one of its key features known as word sketches was introduced.

Hashtags used

#Demonetization, #Currencyban, #Modifightscorruption, #Demonetizationresponse , #Noteban

4.0 DATA PREPARATION

The collected data needed to be cleaned before sentiment analysis. A corpus was created using sketch Engine before they were labelled as positive and negative. Natural language processing tools were used to tokenize the collected data by words and gave them parts of speech tagging. The next step was to filter about 5000 most popular adjectives from labelled corpus data Then a quick function was built to find top 5,000 words in labelled data, marking their presence as either positive or negative. A balanced dataset was used with two classes of sentiment polarity – positive and negative.

5.0 ANALYTICAL APPROACH

Model preparation is the process of assembling different techniques which will work as a single unit with an aim to produce an intended result. In this project, A combination of Naïve Bayes classifier and other Scikit Learn classification methods were used to classify the text as positive or negative. Naïve Bayes is a popular algorithm used in text classification. Training sets were created from labelled data and the classifier algorithms were made to learn from them. This is a good example of supervised machine learning because the machine is trained to differentiate between positive and negative before a different set of data is fed to it to see what it thinks about the category of new data.

Few more variations of Naïve Bayes algorithm were used to enhance the accuracy. Scikit learn has got some of the variations such as Multinomial Naïve Bayes, Bernoulli Naïve Bayes, Logistic Regression and were trained on the same data fed to original Naïve Bayes algorithm. Then a voting system was constructed involving all trained classifiers where each classifier has to cast their vote on the classification for each text. An additional parameter called “Confidence” is also used which is nothing but a tally of votes against the winning vote. This means if 4 out of 5 classifiers agree that a given text is positive, then the confidence value will be 80 percent.

Finally, the functions were wrapped to create a sentiment analysis module which returns the most voted “classification” value and it's corresponding confidence indicator when a text is passed. This module was used to evaluate the sentiment polarity of a new set of data collected randomly from social media and the results were plotted using ‘matplotlib’.

6.0 EVALUATION AND RESULTS

The Python scripts and models are uploaded to GitHub. Please check the link [here](#). Most of the algorithms that were trained showed an accuracy above 70% (Table 1)

Classifier	Accuracy
Original Naïve Bayes	75.45
Multinomial Naïve Bayes	71.81
Bernoulli Naïve Bayes	71.81
Logistic Regression	70.9

Table 1 Model scores

The sentiment analysis performed on random tweets and data from social media shows positive on the demonetisation step by India government. Though the sentiment distribution shows negative in some areas, it appears positive most of the time (Figure 1).

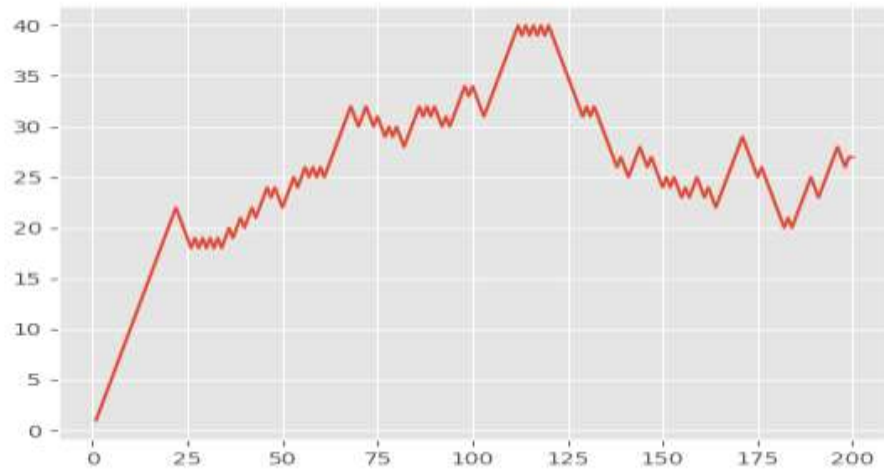


Figure 1 Sentiment Distribution

Some of the most informative features are listed below with their relative presence in positive and negative categories (Table 2).

Most informative Features	%Negative	%Positive
Due	5.3	1
More	1	3
worth	1	2.5
massive	2	1
informal	2	1
poor	3	1
fake	2	1
large	1	2.5

Table 2 Most Informative Features

7.0 CONCLUSIONS AND FUTURE WORK

In this paper, we performed sentiment analysis on a sample collected from social media to sense the people's sentiment on Demonetization step by the BJP government. We employed four variations of naïve Bayes classification on our analysis for better confidence on the results. The classification algorithms were trained against a good amount of labelled data and the results gave strong evidence that the people of India have generally POSITIVE sentiment towards the Demonetization step by the BJP government. In the future work, even richer linguistic analysis can be performed and the data would be collected location wise to find out the sentiment distribution across every state in India.

REFERENCES

- Bae, Y., & Lee, H. (2012). Sentiment analysis of Twitter audiences: Measuring the positive or negative influence of popular Twitterers. *Journal of the Association for Information Science and Technology*, 63(12), 2521-2535.
- Hill, T., Lewicki, P., & Lewicki, P. (2006). *Statistics: methods and applications: a comprehensive reference for science, industry, and data mining*. StatSoft, Inc.
- Hodeghatta, U. R. (2013). Sentiment analysis of Hollywood movies on Twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*
- Hridoy, S. A. A., Ekram, M. T., Islam, M. S., Ahmed, F., & Rahman, R. M. (2015). Localized twitter opinion mining using sentiment analysis. *Decision Analytics*, 2(1), 8.
- Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50, 723-762.
- Paltoglou, G., & Thelwall, M. (2010, July). A study of information retrieval weighting schemes for sentiment analysis. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 1386-1395). Association for Computational Linguistics.
- Roy, K., Kohli, D., Kumar, R. K. S., Sahgal, R., & Yu, W. B. (2017). Sentiment Analysis of Twitter Data For Demonetization In India—A Text Mining Approach. *Issues in Information Systems*, 18(4).
- Weiss, S. M., Indurkha, N., Zhang, T., & Damerau, F. (2010). *Text mining: predictive methods for analyzing unstructured information*. Springer Science & Business Media.

WEBLIOGRAPHY

- What is demonetisation – Retrieved from <https://www.importantindia.com> (Last accessed on 10/11/2017)
- https://www.researchgate.net/publication/247935218_Sentiment_Analysis (Last accessed on 3/11/2017)
- <https://www.sketchengine.co.uk/> (Last accessed on 1/11/2017)