

# Review Paper on Sentiment Analysis of The Demonetization of Economy

Jyoti Yadav

Department of computer Science and Engineering  
DPGITM  
Gurgaon, India  
jkyadav649@gmail.com

Sonal Arora

Department of computer Science and Engineering  
DPGITM  
Gurgaon, India  
Sonalarora431@gmail.com

**Abstract**—Sentiment analysis is one of the major tasks of NLP (Natural Language Processing). Sentiment analysis has increased the interest and attention of many researchers in recent years. Today online reviews of sentiment analysis on Demonetization has become a hot research field. Demonetization was a unique event for Indians which was organized by PM of India. On 8th nov,2016 PM of India declared a decision of Demonetization under which notes of 500 & 1000 were banned to tackle the problem of corruption, terrorist funding and black money. Sentiment analysis on Demonetization mainly focus on framework of lexicon construction, feature extraction, polarity determination, maximum entropy and SVM. This paper explains the survey on the latest development in sentiment analysis. The methods used in current research are especially emphasized to analysis the sentiments of Indians on Demonetization. Finally, in this paper we pointed out some possible future directions of research.

**Keywords**-Demonetization, currency, black money, corruption, cashless, digital economy, SVM, Machine learning

\*\*\*\*\*

## I. INTRODUCTION

Demonetization is the process of banning or ending the currency in the form of coins and notes as a legal tender or decision of a country. After demonetization the new currency was introduced instead of old currency for daily needs. India had also faced this problem before 1.6 months. So, we have decided to analysis the sentiments of Indian public on Demonetization. Sentiment analysis is the study of attitude, thought, or judgment prompted by feelings of the people. Sentiment analysis is also known as opinion mining. Internet is a resource for getting sentiment information of public by their Facebook posts, message, twitter and microblogging.

Today for analysis and data collection there are many APIs (Application Programming Interfaces). Recently Twitter has three different versions of APIs:

- Rest APIs
- Search APIs
- Streaming APIs

Rest APIs:

Rest APIs helps us in gathering status data and user information.

Search APIs:

Search APIs helps us in finding queries.

Streaming APIs:

Streaming APIs helps us to collect Twitter content in real time.

The online data have several flaws that create difficulties for the process

of sentiment analysis. The first flaw in which people can freely post their own

content, thoughts and the quality of their opinions cannot be guaranteed. For example, online spammers post spam on

forums instead of sharing topic-related opinions. Some spam is fake or that have meaningless opinions.

The second flaw in which online data didn't have truthful or ground truth information.

A ground truth indicates that the opinion is positive, negative, or neutral.

Corpus is one of the ground truth dataset that is public available.

The corpus contains millions machine-tagged Twitter messages and each message is tagged on the base of the emoticons as positive, negative and neutral discovered inside the message. Each review or received message are inspected before it can be posted and each review must have a rating that can be used as the ground truth. The rating is based on a star-scaled system, where the 5 stars denotes the highest rating and the 1 star is for lowest rating.

Star Rating	General Meaning
	I hate it.
	I don't like it.
	It's okay.
	I like it.
	I love it.

Figure 1. Star rating on reviews.

In this paper we tackle the problem of sentiment analysis known as sentiment polarity categorization.

The flowchart depicts our proposed process for sentiment categorization.

In phase 2 A mathematical approach is proposed for Computation sentiment score. A feature vector generation method is used for sentiment polarity categorization.

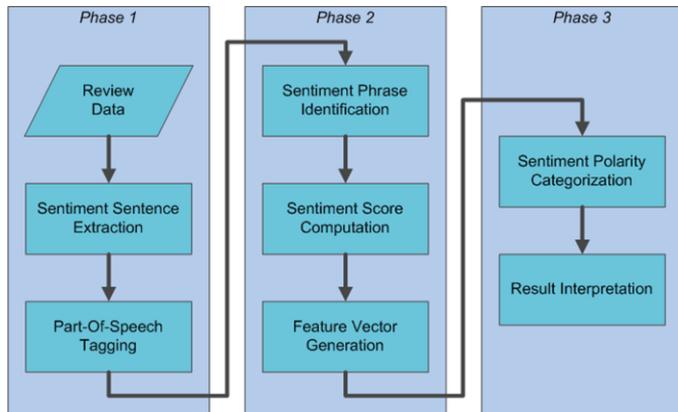


Figure 2 Sentiment Polarity Categorization Process.

In phase 3 Two sentiment polarity categorization experiments based on sentence level and review level of the sentiments are respectively performed. Performance of three classification modals of sentimental analysis are evaluated and compared based on their experimental results.

## II. METHODS OF SENTIMENTS ANALYSIS

There are many methods which are used for the sentiment analysis. The most common way of analysis is the graphical representation. Machine learning and Proposed methodology are used for sentiment analysis.

### A. Graphical Representation

Graphical representation is a simple method to present something. The analysis of sentiments towards Demonetization follow 5 steps in graphical presentation.



#### 1) Data collection

At the Social networking sites like Facebook, Twitter and Instagram the data available in different fonts, different vocabularies, different context of writing, short form and slung. The manual analysis is impossible from such type of data. So, first of all some special process is done for the same collection of data for analysing.

#### 2) Text preparation

Text preparation is the process of filtering the data before analysing and it identifying and eliminating non-textual content is included.

#### 3) Sentiment Detection

In the sentiment detection the sentiments are detected. The sentiment analysis at the sentiment detection is done at different levels.

#### 4) Sentiment Classification

Sentiments are classified into three groups; positive, negative and neutral. In sentiment analysis methodology each subjective sentence are classified into groups; positive, negative, like, dislike, good and bad.

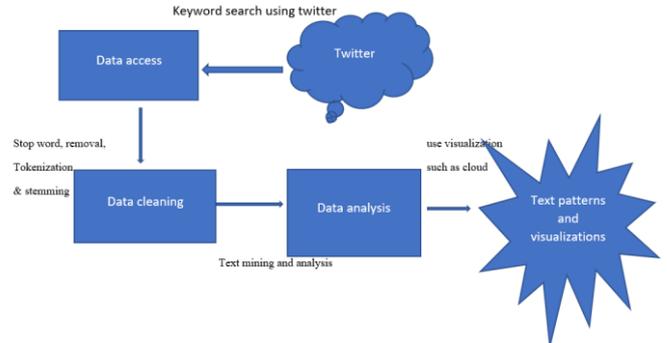
#### 5) Presentation of Output

The main task in sentiment analysis is to convert unstructured text into structured and meaningful information. After the completion of analysis of sentiments, the output or the result are displayed on the graph as pie chart, bar chart and line graph.

### B. Proposed Methodology

The Proposed methodology is used for data mining and sentiment analysis. In the Proposed Methodology various phases are involved for the sentiment analysis.

- Twitter Authentication
- Access Twitter Data Set
- Text pre-processing
- N-Grams
- Sentiment Classification Score



#### 1) Twitter Authentication

In the Proposed Methodology before the mining of data the data from the twitter which is using APIs, have to authentication from twitter for using the application created on the twitter. Once the application is created, we can access the Twitter Authentication server.

```
consumer key<-',,xxxxxx'  

consumer secret<-'xxxxxxxx'  

access token<-'xxxxxxxx'  

access secret<-'xxxxxxxx'  

setup_twitter_oauth(consumer_key,consumer_secret,access_token,access_secret)
```

#### 2) Access Twitter Data Set

After the authenticated of APIs with Twitter Authentication service, a token is generated and this made available API for every transaction with the Twitter server. Using these token,

tweets are mined using hashtags. We use search Twitter() function to access the data, and data set stored in .txt files.

### 3) *Text pre-processing*

Text pre-processing use the clear text and remove stop words.

### 4) *Clean text*

The clean text is used cleaned the unnecessary data from twitter data set, these clean texts are HTML Tags, White spaces, Numbers, Special symbols.

### 5) *Remove stop words:*

Stop words in the text pre-processing are the bag of words based on the dictionary, are the unnecessary words which are removed from the twitter data set, then the resultant data set contains only required information.

### 6) *N-grams:*

N-grams plays an important role in text and sentiment analysis. These are uni-grams( $n=1$ ),

bi-grams( $n=2$ ), tri-grams ( $n=3$ ),...N-grams. The general equation of N-grams is of the next word in a sequence would be;

$P(w_n | w_{1n-1}) \approx P(w_n | w_{n-N+1n-1})$ , where word sequence  $w_1, w_2, \dots, w_{n-1}$  is represented as  $w_{1n-1}$ .

$$P(w_n | w_{1n-1}) \approx C(w_{n-1}w_n)/C(w_{n-1}).$$

### 7) *Sentiment classification score*

The sentiment classification on the cleaned data and n-gram features added to the basic twitter data sets.

Here two types of polarity methods of sentiment classification are used in this application,

1. Polarity score for sentence level.
2. Polarity score for paragraph level.

Polarity score for sentence level: - The function polarity score is used to sentiment classification. This polarity score function is applicable for finding the sentiments according to the polarity scores at each sentence level.

## III. MACHINE LEARNING APPROACH

Machine learning approach is a type of artificial intelligence (AI). It provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the development of computer programs that can change when exposed to new data. The machine learning is similar to data mining. However, instead of extracting data for human comprehension -- as is the case in data mining applications -- machine learning uses that data to detect patterns in data and adjust program actions accordingly. Machine learning algorithms are often categorized as being supervised or unsupervised. Supervised algorithms can apply what has been learned in the past to new data.

Unsupervised algorithms can draw inferences from datasets.

IN Machine Learning we use many techniques like Maximum Entropy, SVM, KNN, Linear Regression and Naïve Bays for the analysis of sentiments for the Demonetization.

## IV. FUTURE WORK

This survey is on the Sentiment Analysis on Demonetization. Sentiment Analysis is also known as opinion mining. In this survey we Analysis the review on twitter and after that recognized the how many

positive, negative and neutral text are used. This demonetization

will help us to eradicate the 5% of the black economy which is put in cash. Also, we move towards the digital economy. After this evaluation and receiving user feedback, it is clear that the several improvements to the system are needed. First to speech to text transcription module should be adapted to a smoother conversational flow. Second "Happy" and "Sad" expression should be redesigned to denote happiness actively and sadness. The Sentiment Analysis need to incorporate negation handling and emphasis handling in order to improve its classification accuracy.

## V. CONCLUSION

In this work, a new method is proposed to get the sentiment analysis on the demonetization from the twitter data sets, tweets are samples of the society opinions, with this sample tweets we can get the sample positive sentiments and negative sentiments of the demonetization decision of the Indian government, in this process data cleaning, bigrams, polarity, and sentiment scores and graphical methods are used for this Twitter sentiment analysis. In this survey we find that machine learning is best technique as compared to the Proposed Approach and other technique like Lexicon. The accuracy in Machine Learning is best. So, we used Machine learning for the analysis of sentiments. Hence, we can only speculate future macroeconomic effects of demonetization.

## REFERENCES

- [1] What is Demonetization. Retrieved from: <http://www.investopedia.com/terms/d/demonetization.asp>.
- [2] Jose, 16 November 2016, What is the impact of demonetization. Retrieved from:
- [3] Anand Pandey, Sentiment Analysis on Demonetization. Retrieved from: <https://bigishere.wordpress.com/2016/12/12/sentiment-analysis-on-demonetization-pig-use-case/>.
- [4] Priya Ananthraj, Sentiment Analysis of tweets. Retrieved from: <https://www.kaggle.com/priyaananthram/d/arathie2/demonetization-in-india-twitter-data/sentiment-analysis-of-tweets>.
- [5] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with

- 
- coherent OFDR,” in Proc. ECOC’00, 2000, paper 11.3.4, p. 109.
- [6] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital-to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997.
- [7] (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [8] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: <http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/>
- [9] FLEXChip Signal Processor (MC68175/D), Motorola, 1996.
- [10] “PDCA12-70 data sheet,” Opto Speed SA, Mezzovico, Switzerland.
- [11] A. Karnik, “Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP,” M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
- [12] J. Padhye, V. Firoiu, and D. Towsley, “A stochastic model of TCP Reno congestion avoidance and control,” Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
- [13] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.