Opinion Mining and Analysis on Demonetization in India

Ashish Jitendra Roy[1], Rahil Rajesh Parikh[2]

[1]Shree L. R. Tiwari College of Engineering (Mumbai University), Kanakia Park, Mira Road East, Thane 401107, Maharashtra, India

[2]Dwarkadas J. Sanghvi College of Engineering, No. U-15, J.V.P.D. Scheme, Bhaktivedanta Swami Road, Vile Parle West, Mumbai, Maharashtra, India

Abstract: *Demonetization was the burning topic when implemented on the night of 8th November, 2016. At one stroke, 86% of the currency in circulation was demonetized, causing confusion, chaos and endless misery to the common Indian. We have used opinion mining method to deal with unstructured tweets on Twitter to evaluate the emotions classification and polarity towards public opinion on demonetization. A significant percentage of people also showed emotions of disgust, sadness and anger in their tweets. Furthermore, the polarity classification revealed. This paper aimed at reviewing the general implication of demonetization on people using Hadoop Ecosystem.*

Keywords: Demonetization, Twitter, Opinion Mining, Sentiment Analysis, #narendramodi, #currencyban

1. Introduction

On 8 November 2016, the Government of India announced the demonetization of all ₹500 (US$7.40) and ₹1,000 (US$15) banknotes of the Mahatma Gandhi Series. The government claimed that the action would curtail the shadow economy and crack down on the use of illicit and counterfeit cash to fund illegal activity and terrorism. The sudden nature of the announcement and the prolonged cash shortages in the weeks that followed created significant disruption throughout the economy, threatening economic output. The move was heavily criticized as poorly planned and unfair, and was met with protests, litigation, and strikes.

2. Literature Review

Go, A., Bhayani, R., & Huang, L [2009]Asthey discussed the more established method for getting information and also performing the sentiment analysis on those data. Here they used different techniques for crawling the data from the twitter where they extracted the data from the Twitter web pages by using some code that may be written either in JAVA, Python etc. For those they are downloaded the libraries that are provided by the twitter API by using this they crawled the data that they want particularly.

Tang, H., Tan, S., Cheng, X& [2013]After got the twitters raw information they separated by utilizing some old procedures and furthermore they discovered the positive, negative and moderate words from the rundown of gathered words in All these words should be collected by the user to filter out or do some sentiment analysis on the filtered data.

Komal S, Snehal K, Sneha K, Pooja D[2016]There are many ways to fetch data from twitter they got twitter data from twitter API ,then data is tokenized through the TFIDF and Porter Stemming Algorithm and found the root of word that compared to affine dictionary and compared the weight and the final result calculated with the k means algorithm.

Monika Sharma [2017] There are three approaches of sentiment analysis machine based, lexical and hybrid In this paper they used machine learning approach .they gathered the tweet of 11th November 2017 to 12th November 2017 and based on tweet they did sentiment by using R technology and for displaying the result they used python application.

Some Drawbacks of existing system:
- It takes a lot of time for performing analysis in large amount of data
- It may result in system failure when a large amount of data is passed to the system
- Major functions and operations for sentiment analysis like stemming and NLP processing takes a huge amount of time.

3. Impact of Demonetization

- Effect on parallel economy: The removal of these 500 and 1000 notes and replacement of the same with new 500 and 2000 Rupee Notes is expected to - remove black money from the economy.
- Effect on Money Supply: With the older 500 and 1000 Rupees notes being scrapped, until the new 500 and 2000 Rupees notes get widely circulated in the market.
- Effect on Demand: The overall demand is expected to be affected to an extent.
- Effect on Prices: Price level is expected to be lowered due to moderation from demand side.
- Effect on various economic entities With cash transaction lowering in the short run, until the new notes are spread widely into circulation
- Effect on GDP: The GDP formation could be impacted by this measure, with reduction in the consumption demand.
- Effect on Online Transactions and alternative modes of payment: With cash transactions facing a reduction, alternative forms of payment will see a surge in demand.

4. Opinion Mining based on Twitter

Over Twitter offer associations' a quick and viable approach to screen the publics' sentiments towards their image, business, and executives. Sentiment analysis over Twitter offer associations' a quick and compelling approach to screen the publics' emotions towards their image, business, and executives

4.1 HDFS

Hadoop is an open source framework for processing and storing large datasets over a cluster. It is used in handling large and complex data which may be structured, unstructured or semi-structured that does not fit into tables. Twitter information falls into the classification of "semi structured" information which can be best put away and examined utilizing Hadoop and it's hidden document framework.

4.2 The core objective of the project is:-

- Tweets Retrieval: A large amount of data is collected Hadoop online streaming tool using Apache Flume.
- Storage: This data is stored in a certain format (HDFS: Hadoop Distributed File system) so as to form a key-value pair which is needed to feed to mapper in map-reduce programming approach. The data is stored in Hadoop Distributed File System.
- Data Processing: Data collected over a period of time is processed by using java and distributed processing software framework developed by Apache Hadoop and using map-reduce programming model and Apache hive framework.
- Data Analysis: The output obtained from reducer phase is analyzed.
- Sentiment Representation: Representation of sentiments classified data in the form of pie charts and map and the table.
- Final output we will get the output of tweets in the sorted form of classified tweets that is Positive, Negative and Neutral tweets.
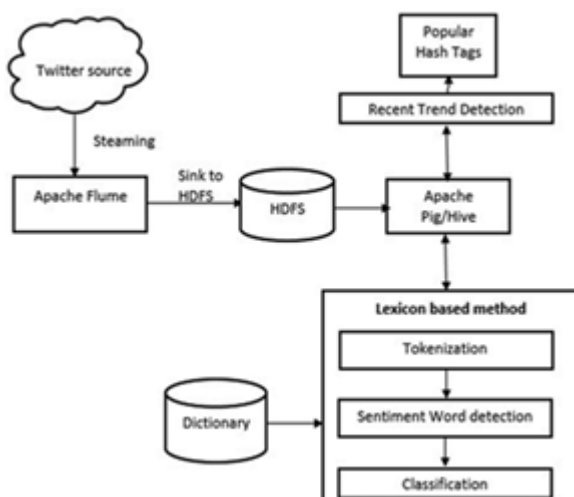
4.3 Proposed System

Figure 1: Proposed System

4.4 Getting Data Using Flume

From the Twitter, we are going to take the sentiment on demonetization of 500 & 1000 rupee banknotes. After successfully creating an application in the Twitter developer site we will get the credential along with the secret keys and the token values by which we will access the Twitter and we will get the tweets that what we want exactly here we will get everything in JSON format and this will be stored in the HDFS on the path that we are going to give location where to save all the data that comes from the Twitter

4.5 Querying using Hive Query Language (HQL)

After running the Apache Flume by setting the above configuration, the Twitter data will automatically save into HDFS. All the data will be in the form of JSON format that data is in unstructured data. now we will create a table that can be filter our unstructured data into structured data that is nothing but how we are going to read the data that is in the form of JSON format for that we are using the custom serialize and deserialize for JSON so that our hive can read the JSON data and can create a table in our defined format.

4.6 AFFIN Dictionary

The AFINN is a dictionary which consists of 2500 words which are rated from +5 to −5 depending on their meaning. The Twitter words are analyzed using AFINN dictionary, and rate the Twitter words as per their meaning from +5 to −5. Initially load the dictionary into pig by using the below statement:

dictionary = load'/AFFIN.txt' using PigStorage('/t') As (word:char array, rating:int);

4.7 Naive Bayes Classification

Naive Bayes classifier is used to predict the particular class of given words. It is used because of its easiness in both during training and classifying steps. To train the input set using naïve Bayes classifier, preprocessed data is supplied as input data. After training phase, test data is given as input to classify sentiment words. As shown in Eq. 1, the Bayes theorem shows good performance as follows.
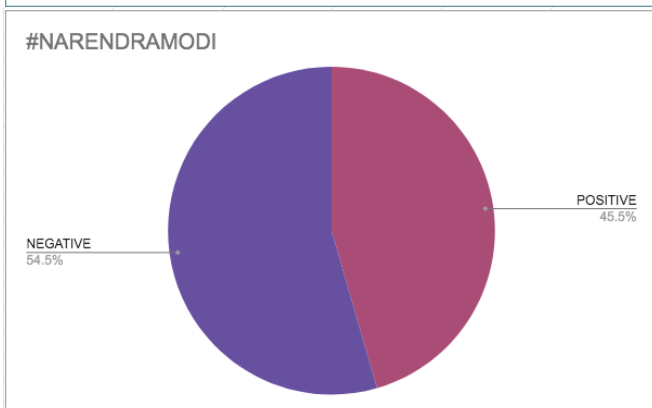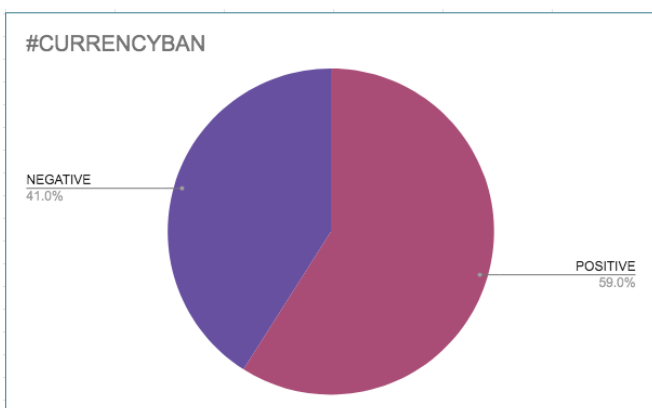
$$P(C|m) = P(C)\prod_{i=1}^{n}P(f_i|C)$$ [2]

where C is the class of positive, negative, or objective sets, $m\psi$ is the Twitter message, and $f\psi$ is a feature. In our experiments, the features are POS tags, unigrams, or bigrams.
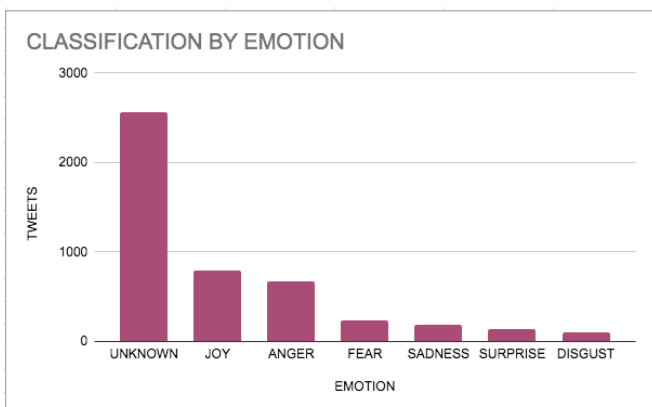
5. Analysis of Tweets

5.1 Positive v/s Negative Tweets

TWEETS vs SENTIMENT

5.2 Popular Hashtag

#CURRENCYBAN

#NARENDRAMODI
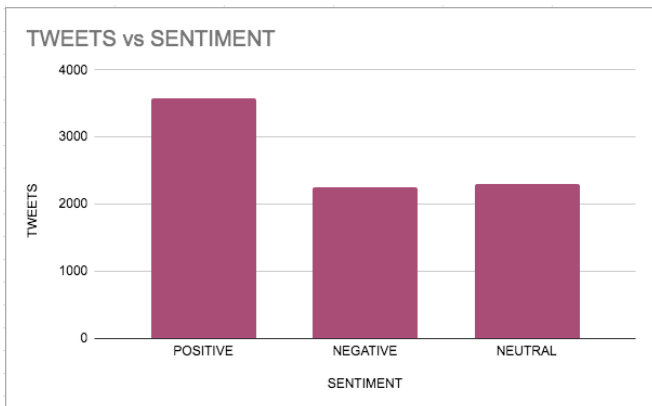
5.3 Emotion Classification

CLASSIFICATION BY EMOTION

6. Conclusion

Equalize the length of your columns on the last page. If you are using *Word*, proceed as follows:
Insert/Break/Continuous.

References

[1] Meera Sanyal, The Big Reverse: How Demonetization Knocked India Out, 10 November 2018.
[2] Singh, P., Sawhney, R. S., & Kahlon, K. S. (2017). Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government. ICT Express. doi:10.1016/j.icte.2017.03.001
[3] A Data Mining Approach for Sentiment Analysis of Indian Currency Demonetization, 22 August 2018. https://link.springer.com/chapter/10.1007/978-981-13-0514-6_2#enumeration
[4] Monika Sharma, Kamrah Institute of Information Technology, Gurgaon, MDU University, Twitter Sentiment Analysis on Demonetization an Initiative Government of India. IJRTER 2016

Author Profile

Ashish Jitendra Roy Computer Engineering Student at Shree L. R. Tiwari College of Engineering, Mumbai University.

Rahil Rajesh Parikh Production Engineering Student at Dwarkadas J. Sanghvi College of Engineering.