Twitter Sentiment Analysis of Demonetization on Citizens of INDIA using R

Shiv Kumar Goel[1] , Sanchita Patil[2]

Deputy HOD, Dept. of MCA, VES Institute of Technology, Chembur, Mumbai, Maharashtra, India[1]

P.G. Student, Dept. of MCA, VES Institute of Technology, Chembur, Mumbai, Maharashtra, India[2]

ABSTRACT: Twitter a free microblogging tool allow people to express their point of view on a wide range of topics ranging from marketing to customer service which can be used for sentiment analysis in order to identify customer likes, opinions, dislikes, feedback, etc. The use of natural language processing, text analysis and computational linguistics which involves identifying people's opinion in a given data is known as sentiment analysis also known as opinion mining. Social media plays a significant role in sentiment analysis which can be used for better decision-making approach. This paper introduces a lexicon-based approach to classify the tweets on #Demonetization in terms of positive, negative and neutral polarity. A shiny dashboard web application is been created to visualize the data. The aim of the paper is to discover what the people of India felt post demonetization and thus to present the results of analysis in the shiny dashboard.

KEYWORDS: Sentiment Analysis, Twitter, R Programming, Shinydashboard, Shiny.

I. INTRODUCTION

Twitter is a popular social networking service where users interact and post messages ("tweets") restricted to 140 characters. Opinions about various topics are been expressed by these tweets. On 8 November 2016, a massive step of demonetization of ₹500 and ₹1000 banknotes was been taken by the Government of India, terminating the usage of all ₹500 and ₹1000 banknotes of the Mahatma Gandhi Series as a form of legal tender in India since 9 November 2016.Prime minister of India Narendra Modi declared circulation of all ₹500 and ₹1000 banknotes of the Mahatma Gandhi Series as invalid and announced the issuance of new ₹500 and ₹2000 banknotes of the Mahatma Gandhi New Series in exchange for the old banknotes. The announcement was made in an unscheduled live televised address to the nation at 20:15 Indian Standard Time (IST) the same day.

For sentiment analysis to be conducted, data is been collected from Twitter in two phases from April 2, 2017 to April 7, 2017 and April 8, 2017 to April 14, 2017. To study the trend change from one week to another of how people felt splitting of data collection was taken place. This paper describes various steps required to perform sentiment analysis on Demonetization using the tweets and conducting a fair judgment about this scheme launched by Government of India.

II. BACKGROUND OF DEMONETIZATION POLICY

Unexpected demonetization is not a new to India. Since 1946 and 1978, this is the third demonetization. During that period, the circulation of the higher denomination banknotes was limited and a majority of the higher denomination banknotes been grasped with the banks. According to Reserve Bank of India (RBI) records 2016, Indian rupee banknotes worth 16,664 billion are being circulated among the public. Of this, 86% (14.180 billion) are in Rs 500 and Rs 1000 banknotes [5]. So, in order to restrain the black money holders, the government emphasized demonetization of Rs 500 and Rs 1000. Exchange of old banknotes with the banks was provided to the people from November 10, 2016, which was a relief. Another announcement was people could also deposit the old banknotes in their respective bank accounts. Usage of old bank notes for necessary services such as the purchase of petrol, diesel, air tickets and rail tickets was granted by the government. A chaos occurred on November 10, 2016, when huge crowd saturated every single bank in the country. The major criticism was been noticed by the government as the banks lacked the new

currency to meet the daily requirement of people. Still, in order to defeat the black money holders, the government insisted people bear initial problem being faced.
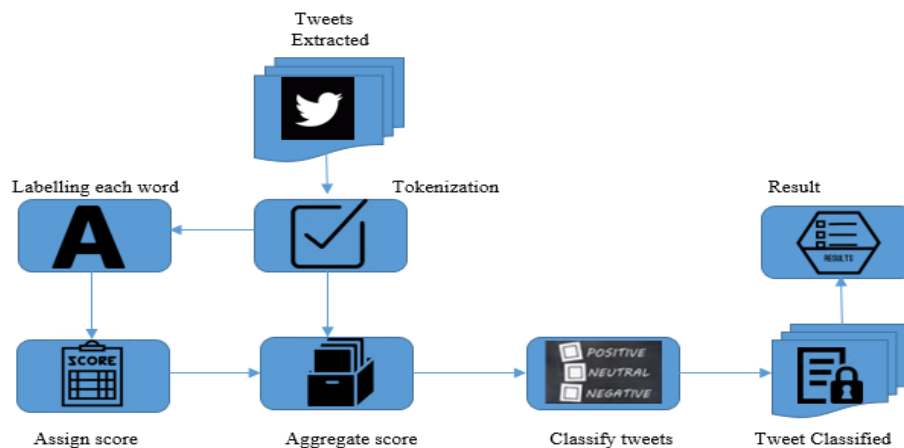
III. APPROACH

Fig. 1 Overview of general lexicon based approach.

Tweets related of #Demonetization are been extracted from the twitter using twitterR package.A sequence of strings been broken down into pieces like keywords, words, symbols, phrases and other elements called tokens is the job of tokenization.Often punctuations are been removed in this process. A polarity score is then been assigned to each of the element. In order to determine the sentiment behind the text the aggregated sum of the score is been calculated.Depending on the calculated score the text is been classified as positive, negative and neutral.

IV. METHODOLOGY

Now let's have a glance on different stages and procedures in the retrieval of insightful results and finally analyzing it. The Different phases involved in the approach are as follows

A. *Planning and Data Collection*

An easy way to extract tweets containing a hashtag say #Demonetization from a user account or public tweets, the twitteR package is been introduced. The searchTwitter function can be used to retrieve tweets been tweeted for #Demonetization. Hashtags are basically used to categorize your tweets making it easy to search.

An app needs to be created on dev.twitter.com, before loading twitteR library and using its functions. The Search API could produce only 1500 tweets at a time which is one of the constraints imposed by Twitter.

From April 2, 2017 to April 8, 2017, total number of tweets collected were 4500 and From April 9, 2017 to April 15, 2017, a total of 3790 tweets related to #Demonetization were gathered, to carry out sentiment analysis using R programming language. The total number of tweets extracted from April 2, 2017 to April 15, 2017 were merged and stored in a csv file named stack.csv.

Create Twitter Account

Login
(https://dev.twitter.com/)

Click on My Apps
Click on Create new App
Fill required Details
Create App

Click on "Keys and Access Token" tab
Click "Create my access token" button

Copy following four keys in your application
Consumer Key, Consumer Secret
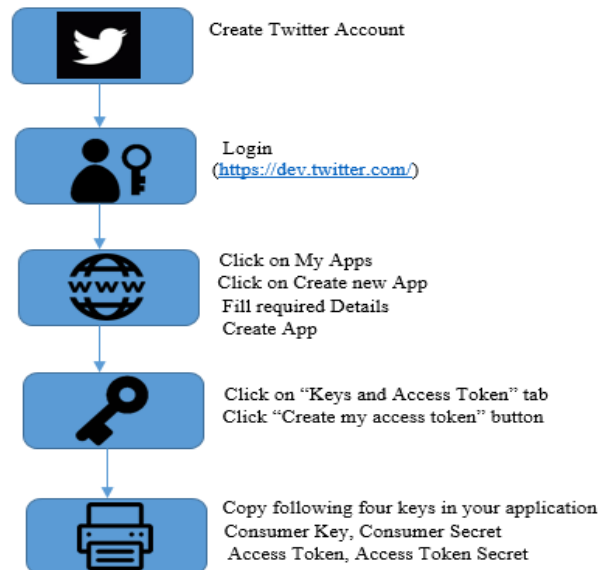Access Token, Access Token Secret

Fig. 2 Steps to create Twitter App to collect data by calling Twitter API's.

B. Data Pre-Processing

- First step is to convert all the tweets been extracted into lower case using tolower function. Lower-case is suitable if the hashtag is of one word or a shorter version. Twitter however doesn't distinguish between cases while a search is been carried.

  For example: Search for #Demonetization and #demonetization while return the same result.

- Once the tweets are been converted into lowercase, the next step is to remove punctuations using removePunctuation function in the tm package. Using multiple punctuations and spaces will result in your hashtag making no sense.

- Stopwords are a set of commonly used words not only in English but in any other language. In order to concentrate on important words instead of very commonly used in a given language, stopwords are been removed. For tweets, terms like "#", "RT", "@username" can be likely regarded as stop words.

  For example: If we search for "how to develop a web application" the search engine will show up with multiple pages containing the terms "how", "to", "a" , instead of main keywords "develop" , "web", "applications".

- Just for simplicity, tweets consisting of numbers are been removed.URL's are not an essential element to be considered in the case of sentiment analysis, hence removed. Also, a long URL consumes a huge portion of allocated 140 characters.
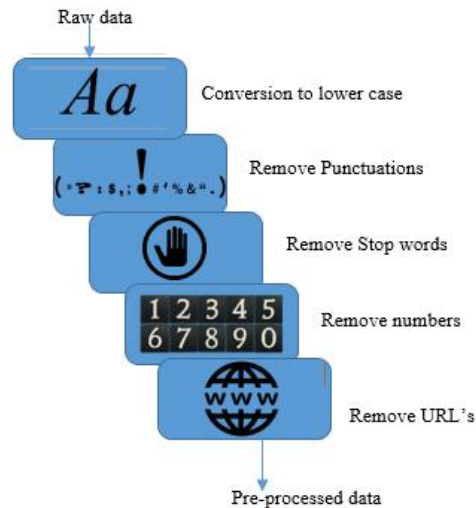
Fig. 3   Steps to prepare/clean data for sentiment analysis.

C.   Data Probing

The baseline approach also known as "Bag of Words Approach" is the simplest and most widely used lexicon based approach. This method consists of two dictionaries – one of the positively tagged words and other of negatively tagged words. After tokenization, search is carried for each individual word of the tweet within those dictionaries, and a polarity score is been assigned depending upon the location of the word.

Consider a tweet: "Demonetization is good in the long run and will bring such a positive change for Indian economy". At the end of data pre-processing, the resultant text for analysis is – "demonetization is good in the long run and will bring such a positive change for Indian economy".

According to the technique explained above, the words - "good","positive" are allocated a sentiment score of +1 since they are present in the dictionary of positive word. The total polarity score of +2 is obtained on aggregation, indicating the sentiment behind the tweet as positive.

```
1  Score = Number of positive words - Number of negative words
2  If Score > 0, means that the tweet has 'positive sentiment'
3  If Score < 0, means that the tweet has 'negative sentiment'
4  If Score = 0, means that the tweet has 'neutral sentiment'
```

Fig. 4   Formula used to calculate sentiment score.

- Scoring: If an individual token is been found in the dictionary of positive words, it is assigned a polarity score of +1 and if present in the dictionary of negative words, a score of -1 is been assigned, else a score of 0 is assigned.

- Aggregation: The total sum of the scores allocated to an individual word in the text is calculated and based on the final polarity value, tweets can be categorized as positive, neutral or negative.

D.   Shinydashboard Application

Shiny/Shinydashboard is a package that enables you to easily create flexible, attractive, interactive dashboards with R. Shinydashboard package can be easily installed from CRAN (Comprehensive R Archival Network). Shinydashboard is a standard HTML document, hence can be deployed on any web server. For additional interactivity, Shiny components can be added and then deployed on your own Shiny Server or shinyapps.io. The basic structure of Shinydashboard application is separated in a user-interface script and server script file. The layout and appearance of the application is been controlled by user-interface (ui), defined in a source script named ui.R. Instructions that the computer requires

building an application is contained in the server.R script. A dashboard build with shinydashboard consists of three main elements

- The dashboard header
- The dashboard sidebar
- The dashboard body

The Shiny and Shinydashboard package are designed primarily to run applications locally and are free and open source.

E. Sentiment Analysis

In order to identify the sentiments from the text, sentiment analysis also known as opinion mining is been performed. Consider tweets collected from April 02 to 08, 2017 as dataset1 and from April 09 to 15, 2017 as dataset2. Fig. 6 reveals that out of 4,500 tweets from dataset1, 1,187 twitter users have commented in a positive way, 1,532 tweeter user have commented in a negative way and the rest 1781 are neutral. According to Fig.7 in total 3,790 tweets were extracted, from which 924 tweets were positive tweets, while 962 tweets were negative tweets and the rest 1,904 were neutral.

	Dataset1					Dataset2		
	tweet	created	number			tweet	created	number
1	negative	02-04-2017	103		1	negative	09-04-2017	183
2	negative	03-04-2017	390		2	negative	10-04-2017	195
3	negative	04-04-2017	134		3	negative	11-04-2017	132
4	negative	05-04-2017	478		4	negative	12-04-2017	145
5	negative	06-04-2017	169		5	negative	13-04-2017	73
6	negative	07-04-2017	207		6	negative	14-04-2017	118
7	negative	08-04-2017	51		7	negative	15-04-2017	116
8	neutral	02-04-2017	121		8	neutral	09-04-2017	214
9	neutral	03-04-2017	329		9	neutral	10-04-2017	317
10	neutral	04-04-2017	126		10	neutral	11-04-2017	177
11	neutral	05-04-2017	441		11	neutral	12-04-2017	261
12	neutral	06-04-2017	99		12	neutral	13-04-2017	315
13	neutral	07-04-2017	590		13	neutral	14-04-2017	242
14	neutral	08-04-2017	75		14	neutral	15-04-2017	378
15	positive	02-04-2017	157		15	positive	09-04-2017	120
16	positive	03-04-2017	100		16	positive	10-04-2017	119
17	positive	04-04-2017	40		17	positive	11-04-2017	43
18	positive	05-04-2017	293		18	positive	12-04-2017	275
19	positive	06-04-2017	181		19	positive	13-04-2017	117
20	positive	07-04-2017	298		20	positive	14-04-2017	124
21	positive	08-04-2017	118		21	positive	15-04-2017	126

Fig. 5   The above table depicts the sentiment score of the classified tweets as per day.

1)   Creating wordcloud to analyse the tweets.

Once all the unwanted features are been removed during data pre-processing, a word cloud can be prepared out of the final pre-processed data. Wordcloud is an image composed of words used in a particular set of data or subject, where the size of each word indicates its frequency or importance. R consists of the package named wordcloud for the same and the function wordcloud can be used to prepare a wordcloud out of the pre-processed data. Wordcloud is basically created to understand the opinion of people and to analyze customer feedback which allows one to find out their likes and dislikes regarding a particulate subject.

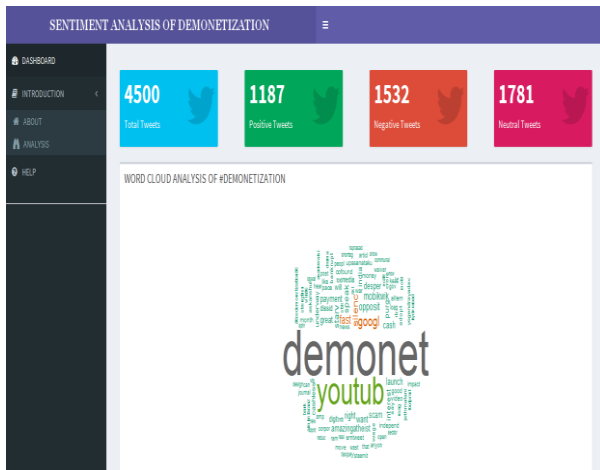# International Journal of Innovative Research in Computer and Communication Engineering

Fig. 6 Shinydashboard application with the overall sentiment score andwordcloud of dataset1.
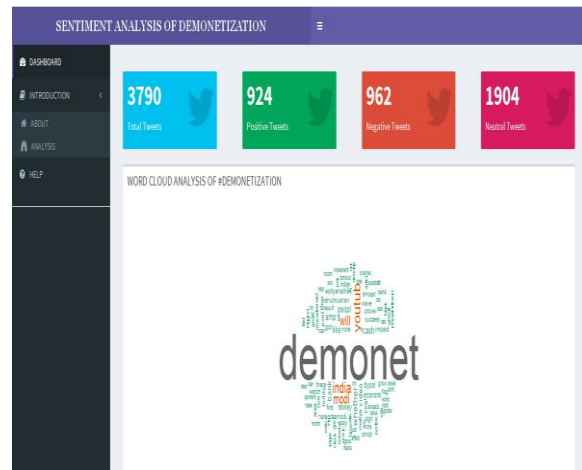

Fig. 7 Shinydashboard application with the overall sentiment score and Wordcloud of dataset2.

2) To create histogram and pie-chart.

Histogram helps to make the comparative study of the sentiments on day to day basis. Fig. 5 consists of the classified tweets along with the sentiment score as per day. The visual representation of the Fig. 5 is created using histogram. It consists of x and y axis with various bars of different size. The graphical representation of the distribution of tweets for dataset1 and dataset2 across positive, negative and neutral polarity as per day is shown in Fig. 8 and Fig. 9 respectively.
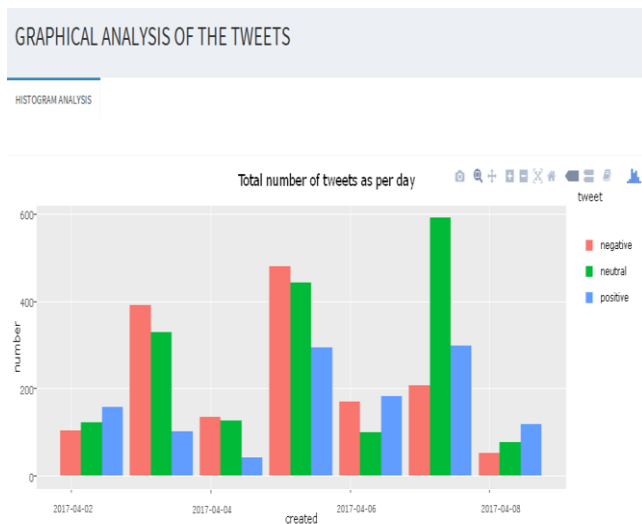

Fig. 8 The histogram above showcases the distribution of tweets for Dataset1across positive, negative and neutral polarity as per day.
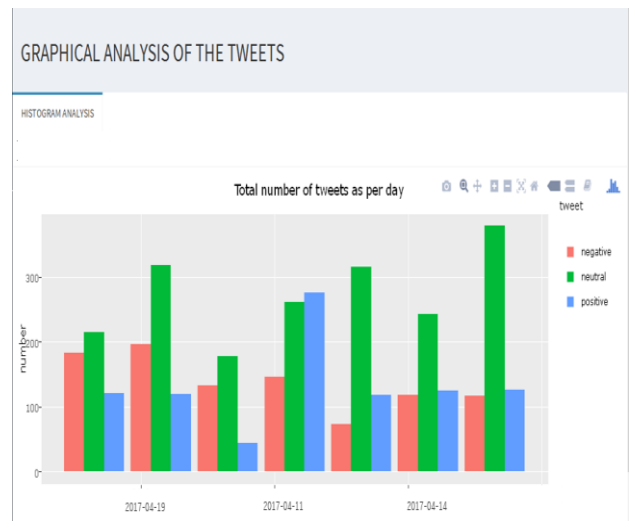

Fig. 9 The histogram above showcases the distribution of tweets for Dataset2 across positive, negative and neutral polarity as per day.

Pie-Chart provides us the exact percentage of people in or not in the favor of #Demonetization. From the analysis carried out in Fig. 10 and Fig. 11 it is clearly stated that people do not tend to be much in favor of #Demonetization. Fig. 10 depicts 26% of public in favour of #Demonetization, 34% against and rest 40% neutral (may or may not be in favour of #Demonetization).

Fig. 11 depicts 25% of public in favour of #Demonetization, 24% against and rest 50% neutral (may or may not be in favour of #Demonetization).
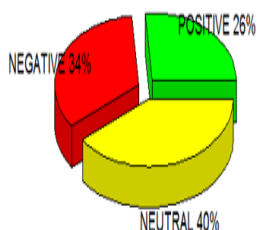

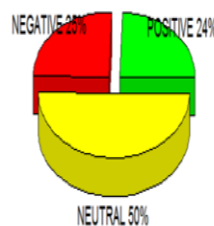Fig. 10 Pie-chart analysis for total tweets for dataset1

Fig. 11 Pie-chart analysis for total tweets for dataset2.

3) Analyzing user's Sentiment score.

The bar graph below in Fig. 12 and Fig. 13 illustrates twitter user's sentiment score, the positive score plus (+) denotes that users are quite happy whereas the negative score denoted by the minus (-) symbol indicates unhappiness with #Demonetization. Zero represents that users are neutral.
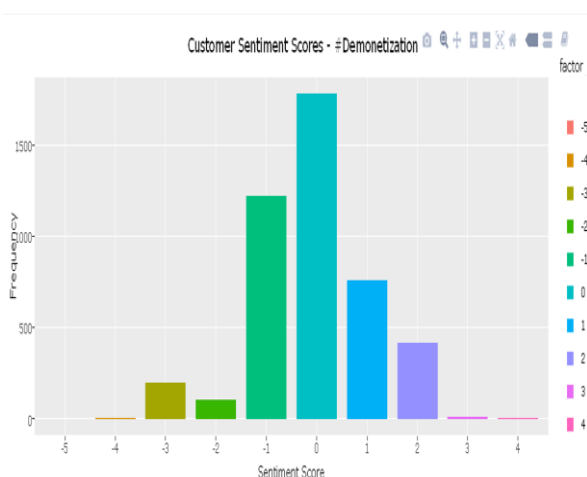

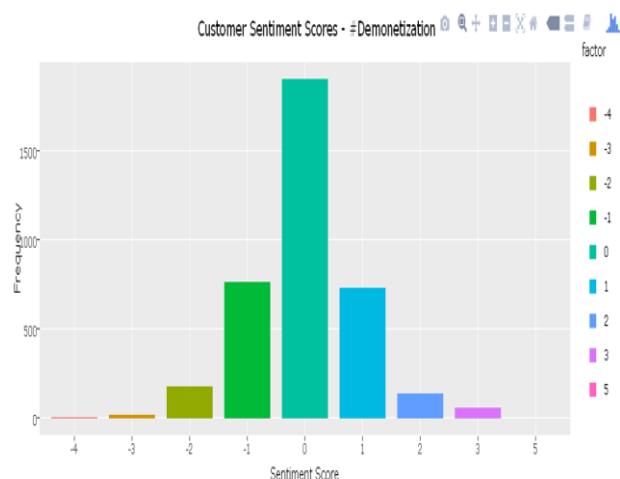Fig. 12 Total sentiment score of users from dataset1.

Fig. 13 Total sentiment score of users from dataset2.

F. Results and Discussion

The following are the finding from the above analysis:

- Dataset1 consisted of total number of 4500 tweets. Whereas, Dataset2 consisted of 3790 tweets in total.
- From the total 4500 tweets of dataset1, 1,187 twitter commented in a positive way, 1,532 tweeter user commented in a negative way and the rest 1,781 are neutral. Hence, the sentiments of people resulted to be negatively biased.
- From 3790 tweets of Dataset2, 924 tweets were positive tweets, while 962 tweets were negative tweets and the rest 1,904 were neutral resulting again into a negative edge.
- Pie-chart analysis results of dataset1 depicted 34% negative tweets, 26% positive tweets, leading to more negative sentiments of people post-demonetization.
- Pie-chart analysis results of dataset2 depicted 25% negative tweets, 24% positive tweets, leading to a minor difference between both the sentiments, still negative in nature.

International Journal of Innovative Research in Computer and Communication Engineering

V. CONCLUSION

As seen, sentiment analysis enables one to understand public sentiments with respect to specific products/services by their Comments and feedback and hence can be used for better decision-making approach. As discussed, that out of the total 4,500 tweets collected during the first phase a total of 1,187 (26%) tweets were positive tweets, while 1,532 (34%) tweets were negative tweets. In the second phase, from 3,790 of total tweets, 924 (25%) tweets were positive, while 962(24%) tweets were negative.This shows a clear dip towards the negative side in both phases. Hence, based on the analysis carried out post demonetization it is observed that people were not in favor of #Demonetization.

REFERENCES

[1] Lei Zhang, Riddhiman Ghosh, Mohamed Dekhil, Meichun Hsu, Bing Liu, "Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis", http://www.hpl.hp.com/techreports/2011/HPL-2011-89.pdf, June 21, 2011.

[2] Amol S. Gaikwad , Anil S. Mokhade,"Twitter Sentiment Analysis Using Machine Learning and Ontology**",**International Journal of Innovative Research in Science, Engineering and Technology **(IJIRSET),** Volume 6,Special Issue 1, January 2017,pp. 173-180, ISSN (Online) : 2319 - 8753.

[3]Veera Raghava Reddy, "Sentiment Analysis using R language", 28 November,2016.[Online]Available: http://www.evoketechnologies.com/blog/sentiment-analysis-r-language/ [Accessed: 22-April-2017]

[4]PrabhsimranSingha, Ravinder Singh Sawhneyb, Karanjeet Singh Kahlona, "Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government", http://www.sciencedirect.com/science/article/pii/S2405959517300176 , Received 13 January 2017, Accepted 9 March 2017, Available online 22 March 2017.

[5]https://rbidocs.rbi.org.in/rdocs/AnnualReport/PDFs/0RBIAR2016CD93589EC2C4467793892C79FD05555D.PDF [Accessed: 15th April, 2017]

[6] DevangJhaver, AunshChaudhari, Lakshmi Kurup , "Twitter Sentiment Analysis on E-commerce Websites in India", International Journal of Computer Applications (0975 – 8887) Volume 127 – No.18, October 2015,pp. 14-18.

[7] Trivedi, Shrawan Kumar, Dey, Shubhamoy, Kumar, Anil, Panda, TapanKumar , "Handbook of Research on Advanced Data Mining Techniques and Applications for Business Intelligence." [Online] Available: https://books.google.co.in/books?id=yxwhDgAAQBAJ&pg=PA194&lpg=PA194&dq=why+tweete+need+to+be+converted+in+lower+case&source =bl&ots=QtNxrNv8m9&sig=N4Nl9UqNfRqrZzx7oMRt4RC7myk&hl=en&sa=X&ved=0ahUKEwjtoPKh0KvTAhVJv48KHXHFDEUQ6AEIJTAA #v=onepage&q=why%20tweete%20need%20to%20be%20converted%20in%20lower%20case&f=false[Accessed: 22- April- 2017]

[8] Sanchita Patil,"WhatsApp group data analysis using R",International Journal of Computer Applications (0975 – 8887) Volume 154 – No.4, November 2016, pp. 31-36.

BIOGRAPHY

Miss Sanchita Patil is currently a final year M.C.A student of Vivekanand Education Society's Institute of Technology (V.E.S.I.T), Mumbai.She had completed her BSc (I.T) graduation in the year 2014 from Vivekanand Education Society's of Arts, Commerce and Science ,Chembur, Mumbai, Maharashtra, India .She has an abiding interest in Data Mining, Database programming and Data analysis.