Sentiment Analysis in Measuring Impact of Demonetization based on Social Media Interaction

[1]Manisha Shinde-Pawar, [2]Pallavi Jamsandekar
[1]Assistant Professor, [2] Professor
[1,2]Bharati Vidyapeeth Deemed to be University, Pune
[1]Institute of Management, Kolhapur, [2]Institute of Management & Rural Development Administration, Sangli

Abstract: As usage of the social media, internet and E commerce, digital marketing has become gold mine of data for business analytics. Social media like facebook and twitter have given increased and easy access, open platform to informal communications. Every digital moment of life is being captured and analysed through using business models and analytics. Online, real time data and its analysis need application of Sentiment analysis thought to such unstructured data.

The Researcher reveals the application of sentiment analysis to unstructured social media data. By applying syntactic context positive and negative classification of demonetization decision related tweets can predict overall sentiments and common concerns of people expressing them on social media. Similarly any administrative or government decision influencing to societal life can be studied for analysis of overall positive or negative attitude towards decision.

***Keywords* – Social Media, Unstructured data, syntactic context, Sentiment Analysis, Attitude.**

I. INTRODUCTION:

Sentiment analysis has become important, as major organizations and governments are trying to predict and judge overall sentiments of common people influencing any particular decision made and its impact on society. Social media is immense, easy source of ample data but due to its unstructured form, analysis of such data is challenging task which has given rise to domain of sentiment analysis research. So Business analysts, researchers and academicians are being attracted to sentiment analysis.

The proposed study simply focus social media text mining to come up with overall sentiments of common people are positive or negative. The researcher aims to analyze pre processing techniques, tokenization and term frequency calculations to form result classes. And also to represent these classes using visual techniques to identify and judge overall level of positive or negative impact of demonetization on common people through their tweet discussions.

II. RESEARCH METHODOLOGY:

Researcher has extracted data for twitter social media tweets dataset from online data source available in .csv dataset format [4]. And then preprocessing will be applied to get refined set of terms or words, which will to be used for sentiment analysis to track positive terms which supports to demonetization decision or frequency of negative terms will represent negative impact on society. Total 500 tweet records (sample size) are considered to study positive or negative impact of demonetization.

II.1 Sample Size and Scope:

- **Sample Size – 500 tweet records**
- **Source-**https://www.kaggle.com/arathee2/demonetization-in-india-twitter-data
- **Scope**- Only Twitter Social Media data is considered for analysis.

II.2 Tools and standards:

Researcher has applied the experiment using R- programming tool.

- RStudio Version- 1.1.383
- R -3.4.3
- tidytext text ming package[8]

II.3 Task Plan/Schedule:

A tidytext package provides functionality to tokenize data and to handle data with more effective and easier ways by considering text format being a table with one-token per row. With tidytext package one can break text into individual tokens and can apply necessary analysis techniques and visualization for the same can be created using ggplot2 package.

- Apply Pre-Processing:
- Removal of stopwords, punctuation marks, whitespaces, numbers and noisy data
- Detect sentiment term in refined dataset terms.
- Determine term words, frequency and probability of occurrence
  i. e. Quantify Polarity

A common method of text mining is applied here where frequency of identified terms and probability is calculated. And Comparative analysis for the terms (words) with predefined set of terms can track sentiment in two classes.

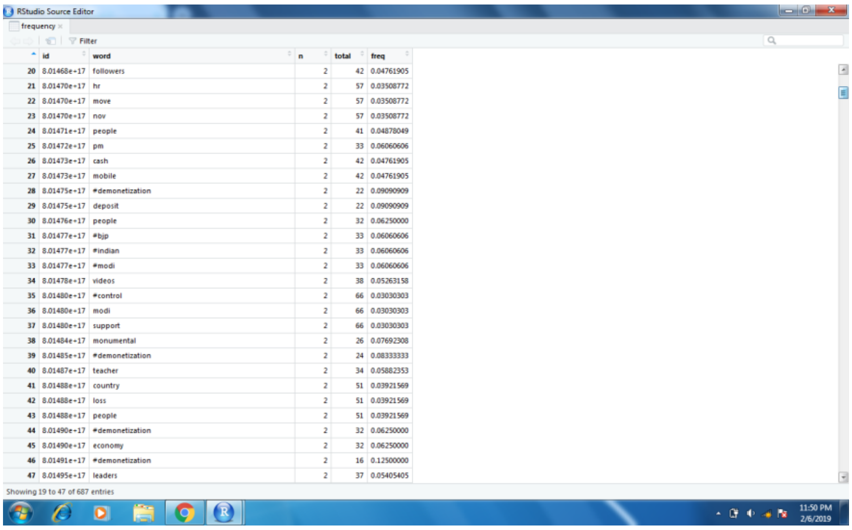
- Identify positive and negative terms.
- Determine positive and negative classes for entire data with positive attitude and negative attitude.
- Analyze the classes to get overall sentiment on tweets of demonetization.
- Create Visualization – plot graphs using gplot2 for generated dataset.
- Apply measures of interestingness as accuracy, recall, precision and F score.

Measures of interestingness considers positive, negative with more precision for comparative analysis in form of true positive, false positive, true negative and false negative classes. In future work researcher aims to identify these classes as input parameter to pass to get precision, accuracy, recall and F Score. With the same intention researcher focuses to track classified outcome for sentiment attitude in the proposed study.

III. RESULTS AND DISCUSSIONS:

As shown in figure no.1, quantification of term frequency to probability helps to identify proposition and size of occurrence of term in entire tweets set.

Figure No.1

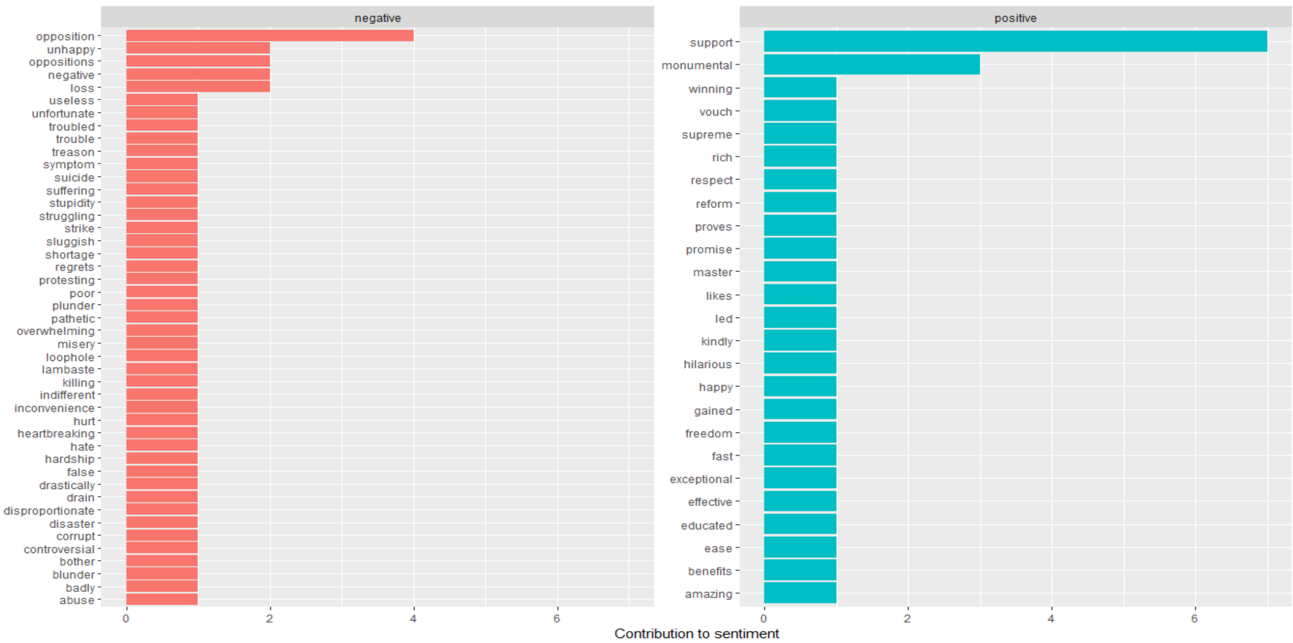


	id	word	n	total	freq
20	8.01468e+17	followers	2	42	0.04761905
21	8.01470e+17	hr	2	57	0.03508772
22	8.01470e+17	move	2	57	0.03508772
23	8.01470e+17	nov	2	57	0.03508772
24	8.01471e+17	people	2	41	0.04878049
25	8.01472e+17	pm	2	33	0.06060606
26	8.01473e+17	cash	2	42	0.04761905
27	8.01473e+17	mobile	2	42	0.04761905
28	8.01475e+17	#demonetization	2	22	0.09090909
29	8.01475e+17	deposit	2	22	0.09090909
30	8.01476e+17	people	2	32	0.06250000
31	8.01477e+17	#bjp	2	33	0.06060606
32	8.01477e+17	#indian	2	33	0.06060606
33	8.01477e+17	#modi	2	33	0.06060606
34	8.01478e+17	videos	2	38	0.05263158
35	8.01480e+17	#control	2	66	0.03030303
36	8.01480e+17	modi	2	66	0.03030303
37	8.01480e+17	support	2	66	0.03030303
38	8.01484e+17	monumental	2	26	0.07692308
39	8.01485e+17	#demonetization	2	24	0.08333333
40	8.01487e+17	teacher	2	34	0.05882353
41	8.01488e+17	country	2	51	0.03921569
42	8.01488e+17	loss	2	51	0.03921569
43	8.01488e+17	people	2	51	0.03921569
44	8.01490e+17	#demonetization	2	32	0.06250000
45	8.01490e+17	economy	2	32	0.06250000
46	8.01491e+17	#demonetization	2	16	0.12500000
47	8.01495e+17	leaders	2	37	0.05405405

Showing 19 to 47 of 687 entries

With syntactic reference of tidytext analysis, it is identified that each processed frequent term is postive or negative, as shown in Figure no. 2. The figure shows that, the graph specific negative or postive term and relative occurrence level(frequency) and data matrix for each term is generated as shown in figure no.4.

Figure no. 2



The resultant terms are identified for each positive and negative occurrence exact sentiment measure to either yes or no in form of 1 or -1 as shown in figure No. 3.

Figure No.3

	word	index	negative	positive	sentiment
1	abuse	0	1	0	-1
2	amazing	0	0	1	1
3	badly	0	1	0	-1
4	benefits	0	0	1	1
5	blunder	0	1	0	-1
6	bother	0	1	0	-1
7	controversial	1	1	0	-1
8	corrupt	0	1	0	-1
9	disaster	0	1	0	-1
10	disproportionate	0	1	0	-1
11	drain	0	1	0	-1
12	drastically	0	1	0	-1
13	ease	0	0	1	1
14	educated	0	0	1	1
15	effective	0	0	1	1
16	exceptional	0	0	1	1
17	false	0	1	0	-1
18	fast	0	0	1	1
19	freedom	0	0	1	1
20	gained	0	0	1	1
21	happy	0	0	1	1
22	hardship	0	1	0	-1
23	hate	0	1	0	-1
24	heartbreaking	0	1	0	-1
25	hilarious	0	0	1	1
26	hurt	0	1	0	-1
27	inconvenience	0	1	0	-1
28	indifferent	0	1	0	-1

Showing 1 to 28 of 71 entries

The Figure No. 3 shows that, the exact attitude for particular term is determined with binary representation for existence of positivity or negativity of sentiment. Further for the same term total sentiment is represented in -1 representing negative attitude and +1 representing positive attitude for the term used.

Figure No.4

	word	sentiment	n
1	support	positive	7
2	opposition	negative	4
3	monumental	positive	3
4	loss	negative	2
5	negative	negative	2
6	oppositions	negative	2
7	unhappy	negative	2
8	abuse	negative	1
9	amazing	positive	1
10	badly	negative	1
11	benefits	positive	1
12	blunder	negative	1
13	bother	negative	1
14	controversial	negative	1
15	corrupt	negative	1
16	disaster	negative	1
17	disproportionate	negative	1
18	drain	negative	1
19	drastically	negative	1
20	ease	positive	1
21	educated	positive	1
22	effective	positive	1
23	exceptional	positive	1
24	false	negative	1
25	fast	positive	1
26	freedom	positive	1
27	gained	positive	1
28	happy	positive	1

Showing 1 to 28 of 70 entries

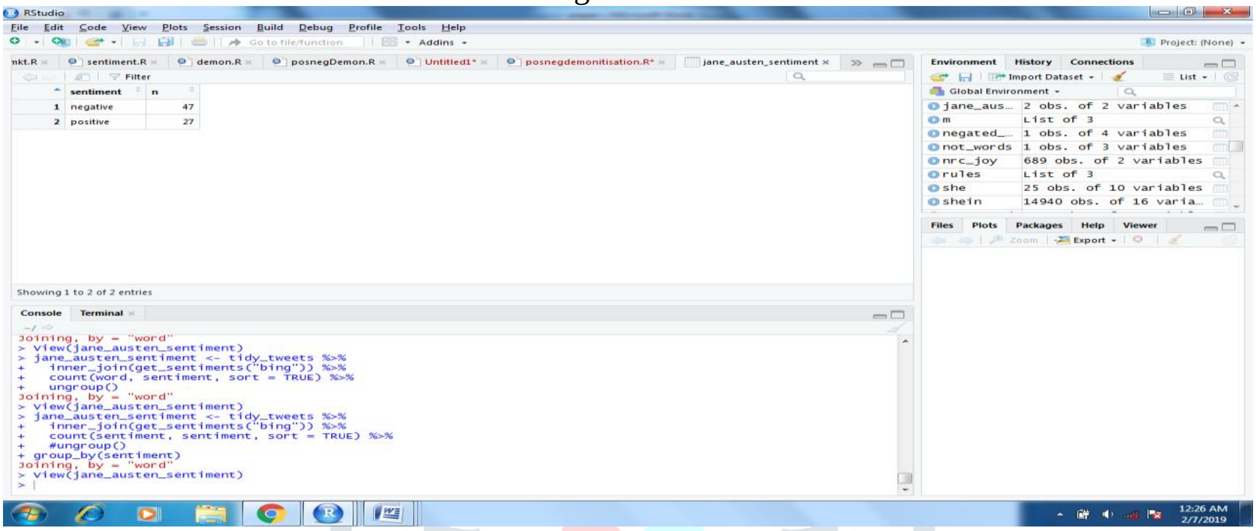
To analyse overall sentiments on twitter dataset, the researcher considered 500 tweets and after preprocessing filtered and identified total terms expressing sentiments which are aggregated to form two classes as overall Positive and Negative Sentiment. As shown in figure No. 4, and Table No. 1

Table No.1

Sr. No.	Sentiment	n
1	Positive	27
2	Negative	47
	Total	74

Where n= number of frequency for total positive terms(words) and total negative terms(words)

Figure No. 5



The figure no. 5 shows that, 47 total words are negative and 27 total no. of words are positive. So, 63.52% sample data from 500 twitter records representing negative opinion/attitude towards demonetization decision. And rest 36.48 only is representing positive attitude for demonetization decision.

That means overall attitude predicts that the common people have the opinion that the demonetization have negative impact in their lifestyle.

IV. CONCLUSION AND FUTURE RESEARCH:

The above study considers and predicts only based on syntactic context with reference to tidytext package. In future research, the researcher has planned to focus on other dimensions semantic in form of literary and psychological parameters and further to classify with multivalve logic or in the form of certainty factor to define ranges of not having exact meaning.

Further the researcher would like to apply classification of positive and negative words and their probability for quantifying into the scale of sentiments as shown in table no.2.

Table no.2

Sr. No.	Sentiment Class	Probability Range
1	True Positive	0.5 to 1
2	False Positive	0 to 0.5
3	True Negative	0.5 to 1
4	False Negative	0 to 0.5

So based on resultant data researcher can apply measures of interestingness as accuracy, recall, precision and F score to analyze and predict specific overall sentiment.

V. REFERENCES:

1. Saif Mohammad, Xiaodan Zhu, and Joel Martin!, " Semantic Role Labelling in Emotions in Tweets.", **National Research Council Canada,** Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 32–41, Baltimore, Maryland, USA. June 27, 2014
2. Sara Rosenthal, PreslavNakov, Svetlana Kiritchenko, Saif M Mohammad, Alan Ritter, VeselinStoyanoy, "SemEval-2015 Task 10: Sentiment Analysis in Twitter", International Workshop on SemanticEvaluation (SemEval 2015), pages 451–463,Denver, Colorado, June 2015.
3. Ashish Shukla, Rahul Misra, " Sentiment Classification and Analysis Using Modified K-Means and Naïve Bayes Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, Vol 5, Issue 8, Aug. 2015
4. Data Source : https://www.kaggle.com/arathee2/demonetization-in-india-twitter-data
5. https://www.tidytextmining.com/sentiment.html
6. BEN, "On Social Sentiment And Sentiment Analysis", DECEMBER 16, 2013 , http://brnrd.me/social-sentiment-sentiment-analysis/
7. Mojdeh Jalali-Heravi, Osmar R. Zaïane, "A Study on Interestingness Measures for Associative Classifiers", University of Alberta, Canada, 2010Steven Finlay, "Predictive Analytics, Data Mining and Big Data", published in 2014 by Palgrave MacMillan.
8. Julia Silge, David Robinson, "Text Mining with R, a tidy approach", O'Reilly publications active on 10th feb 2019 https://www.tidytextmining.com/.