Impact of Financial Resolution and Deposit Insurance on Indian Economy

[1]I.M.V.Krishna, [2]G.Tanuja, [3]M.Tejaswini, [4]Dr. B.Srinivasa Rao
[1]Assistant Professor, Dept. of Computer Science & Engineering, P.V.P. Siddhartha Institute of Technology , Kanuru,Vijayawada. India
[2]System Engineer, Infosys, Hyderabad
[3]Software Engineer, Allsec Technologies, Hyderabad
[4]Professor, Dept. of Information Technology, Lakki Reddy Bal Reddy Collgee of Engineering Mylavaram Vijayawada (Rural)
([1]imvkrishna@gmail.com,[4]buragasrinivasarao@gmail.com)

Abstract-The Financial Resolution and Deposit Insurance (FRDI) Bill, which was proposed in Parliament in the year 2017, have given rise to many concerns (worry) over protection for bank deposits. "Does this bill have any impact on depositors, Banking factor and Economy?" This question is elevated (raised) to many people. But the solution regarding this was not yet materialized. To do this task, sentiment analysis is used in this paper. Sentiment Analysis is used to extract and analyze subjective information from the web - mostly social media. Opinions will extract from the twitter and pre-processed tweets are given for classifiers. The main aim of this project is to analyze and conclude whether it gives a positive or negative impact on Indian Economy.

I. INTRODUCTION

"What does FRDI bill mean?" "Does the bill is really necessary?" "Does depositors need to worry about this bill?" "What is the impact of this paper in society?" These are all the questions that are seen now-a-days; this paper could clear this ambiguity in society. When any bank faces insolvency problem then that particular bank is merged into some other bank even before bill is passed. So, no depositor would face any loss in their deposits, but this FRDI Bill will make changes in this process. Because of the resultant bill if any bank faces insolvency problem, then that particular bank will not be merged into any other and the deposits in bank are converted into shares of bank which means the deposited money cannot be refunded to depositors this decision will be taken by resolution corporation which is being assigned by this FRDI Bill. The resolution corporation analyses financial firms, checks the risk of failure and takes correct action and resolve them in the case of failure.

Sentiment analysis, an application of natural language processing, used for the extraction of emotion analysis this is done by classifying the polarity of text in terms of positive, negative, neutral. Based on this mood extraction decision making of human can be known. The main aim of sentiment analysis is to minimize the gap between human and computer. Thus, it is collection of human intelligence and electronic intelligence for mining the text and classifying user sentiments, likes, dislikes and wishes. The contents generated by users can be seen in many forms such as web logs, reviews, news and discussion forums. Web 2.0 & 3.0 has provided a platform to share the feelings and views about the products and services. Social network revolution plays a crucial role in gathering information containing public opinion. Sentiment analysis plays a very important role to know public opinion about the product, marketing campaigns, political issues, social events and company strategies.

In this paper the impact of FRDI would be told using sentiment analysis. To achieve this, task python language is being used in this paper. The main reason for the usage of python is due to its features, python contains many advanced features such as simple, easy to learn, free and open source, high-level language, portable, interpreted, object oriented, extensible, embedded, extensive libraries. Not only because of these features but also it is exciting and powerful language. Writing programs in python is both fun and easy.

II. LITERATURE SURVEY

[1] used Twitter, which is a popular micro-blogging service where users create status messages (called "tweets"). Till then no research has been done on micro-blogging sites like Twitter and also tells about the pre-processing steps needed for analysis in order to achieve high accuracy. It uses multiple classifiers such as Naive Bayes, Maximum Entropy (MaxEnt), and Support Vector Machines (SVM) which leads to ambiguity to find optimum accuracy. [2] investigated on the structure of micro-blog postings, types of expressions, and sentiment fluctuations and made analysis on the tweets using multi nominal bayes classifier. [3] in order to gather a large dataset of current public opinions on a particular topic, they decided to use Twitter and expanded on simply positive or negative sentiment towards a candidate by classifying tweets by emotions like happy, sad, fear, laughter, and angry,..etc.

For that they used Support Vector Machines, Nearest Neighbours, and Naive Bayes classifiers.

[4] performed analysis upon the collected tweets, by using the TextBlob Python library. TextBlob also allows the use of a built-in classifier to classify text depending on a set of training data. [5] performed sentiment Analysis, which is a natural language processing(NLP) task that is used mine opinion information from various text forms like reviews, news, and blogs and classify them by their polarity as positive, negative or neutral. [6] Described about the people's primary emotions, i.e., love, joy, surprise, anger, sadness and fear, which can be sub-divided into many secondary and tertiary emotions. Each emotion has different intensities. The strengths of opinions are related to the intensities of certain emotions, e.g., joy, anger, and fear by using Rule-Based and Machine Learning based techniques. [7] used Naive Bayes classification algorithm to determine the helpfulness of the review because it was the most efficient and stable basic algorithm in accordance with the study. Based on our experience, this algorithm has advantages like being able to handle data with many features, handling missing data conditions, and can be used for small data as training data.

8] choosed Twitter because 1) tweets are small in length, thus less ambigious; 2) unbiased; 3) are easily accessible via API; 4) from various socio-cultural domains and shows in detail about mining sentiments from tweets by pre-processing techniques. Text classification using machine learning algorithms is a well studied field [9] and [10] carried out research on the performance using various machine learning techniques (i.e., Naive Bayes, maximum entropy, and support vector machines) in the specific domain of movie reviews. It achieved best accuracy on Naive Bayes classifier. [11] it shows that Naive Bayes algorithm improves accuracy of classification of tweets, by providing positivity, negativity and objectivity score of words present in tweets. For actual implementation of this system, python with NLTK and python-twitter APIs are used. [12] present general background information on Twitter and micro- blogging. There are number of popular press articles on leverage micro-blogging press articles on leverage micro-blogging applications for different purposes. According to the work done by [13] describes about sentiment analysis is a current research area in text mining and stem of natural language processing or machine learning methods. It is the important sources of decision making and can be extracted, identified, evaluated from the online sentiments reviews. It tells about how to connect on Twitter and search for the tweets that contain a particular keyword and then evaluate the polarity of the tweets as positive or negative. For the purpose of collecting keywords, Twitter API is used and the extracted raw data are pre-processed using Natural Language Toolkit techniques. To obtain the best features, Naïve Bayes classifier is used. It also evaluates the sentimental polarity.

III. METHODOLOGY

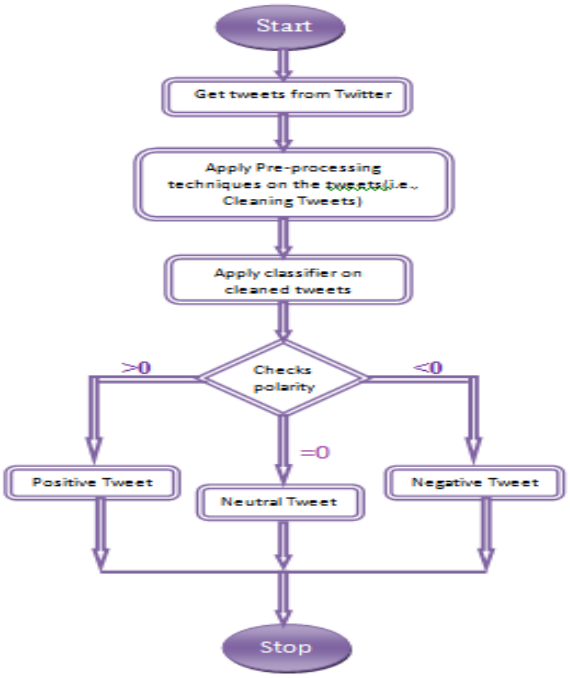


Fig.1: tells about the step wise procedure involved in this proposed work

IV. IMPLEMENTATION

A. Data Collection:

The main aim is to analyse sentiment on the extracted tweets based on FRDI. The first and foremost step is to excerpt tweets from the twitter API. In the process of getting tweets user must have access on twitter database. To access Twitter database, user requires consumer key, consumer secret key, access token key and access token secret key. To acquire these keys, user must sign up with their Twitter account and authorize in order to establish a connection with the Twitter API. Then, an application is to be created in developer level. Once the application is created the required four keys are ready for usage. Using these four credential keys user needs to setup the Twitter's API connection. In this paper, implementation was done using python programming language. An extractor object is created, by using that user gets list of tweets based on a certain search key word. The list of tweets that are extracted from the twitter database are imported into a CSV (Comma Separated Values) file.

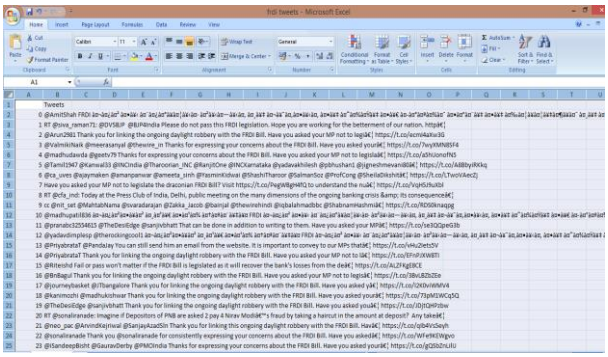


Fig.2: shows the collected tweets related to FRDI in CSV format

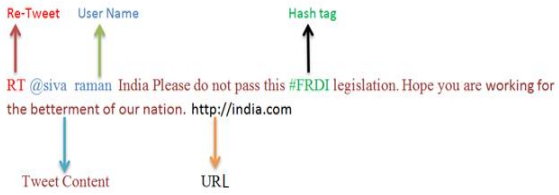


Fig.3: is a detail struture of a tweet

B. Data Pre-processing

Pre-processing must be performed to improve classification effectiveness. In this paper, we use a regular expression to pre-process. This regular expression returns a cleaned tweet. Cleaning or pre processing tweets include removing hashtags, removing punctuation marks, removal of URLs and convert tweet into lowercase and replace the usernames.

For example: Consider a tweet,

Raw tweet: RT @siva_raman71: India Please do noooot pass this #FRDI legislation. Hope you are working for the betterment of our nation. httpâ€¦

Removal of hashtags:

Hashtags are generally used in many of the tweets to highlight specific word which resembles their expressions or emotions in their tweets. During analysis we ignore the hashtags but the word associated with it is considered for classification. After removing the hashtag in the above example the resultant tweet is as below

RT @siva_raman71: India Please do noooot pass this FRDI legislation. Hope you are working for the betterment of our nation. httpâ€¦

Punctuation Marks Removal:

Punctuation Marks don't affect the tweet or text. So, we remove punctuation marks, in order to improve the classification effectiveness. On punctuation marks removal for the above example the tweet is as follows

RT @siva_raman71 India Please do noooot pass this FRDI legislation Hope you are working for the betterment of our nation httpâ€¦

Removal of URLs:

To share additional information URL links are used. This URLs may consists of video links, NEWS, YouTube links, News movie releases. This is one of the pre-processing techniques that helps to classify the tweets quickly. By removing the URL for the above example it is transformed as below.

RT @siva_raman71 India Please do noooot pass this FRDI legislation Hope you are working for the betterment of our nation

Convert into lowercase:

The tweet is a sentence which represents different expressions. So, the tweet may contain either uppercase or lower case. In order to ease the polarity process the tweet should be represented in specific case either upper or lower. Generally we convert the upper case letters to lower case letters. We convert the total tweet into lowercase letters then it appears as

rt @siva_raman71 india please do noooot pass this frdi legislation hope you are working for the betterment of our nation

Replacement of @username:

Generally username contains @. For each and every tweet first parameter is username. As it does not affect the performance of the classifier '@user name' is replaced with white space. By replacing the username with a white space for the above stated example it will be as follows

rt india please do noooot pass this frdi legislation hope you are working for the betterment of our nation

Stop words:

The words which consist of duplicated alphabets or symbols to carry additional effect or to extend the expression are called stopwords. Some of examples of stopwords are happpppppy, coooooool, gooooooooooood, saaaaaaaaad. For the above stated example there is one stop word (noooot) it will be reformed as

rt india please do not pass this frdi legislation hope you are working for the betterment of our nation

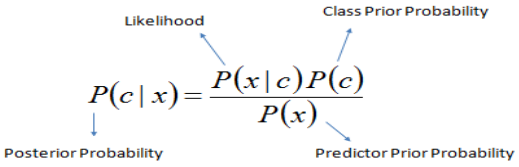
Tokenization:

A tweet is embedded with links, username, punctuation marks, numbers etc... A tweet is indeed in the form of a sentence. To classify a tweet, every word need to be checked for its polarity rate i.e., either positive or negative or neutral tweet. The technique of splitting downs the sentence or tweet into individual words is called tokenization. This ideology of tokenization simplifies the process of polarity checking. For the above tweet on tokenization, it splits into tokens as below 'rt' 'india' 'please' 'do' 'not' 'pass' 'this' 'frdi' 'legislation' 'hope' 'you' 'are' 'working' 'for' 'the' 'betterment' 'of' 'our' 'nation'

C. Classification:

In this paper we are analyzing the performance of the classifier by using naive bayes algorithm. A naive bayes

classifier is used mainly for large amount of data sets. Now let us consider a small example for better understanding of naive bayes algorithm. A fruit may be considered to be a mango if it is green, and about four inches in diameter, and may be of round/oval/heart shaped, or kidney-shaped. These features depend upon each other/upon the existence of many other features. All these properties will independently contribute to the probability to voice that the given fruit is mango. The algorithm works using training data set and testing data set.



$$P(c \mid X) = P(x_1 \mid c) \times P(x_2 \mid c) \times \cdots \times P(x_n \mid c) \times P(c)$$

Textblob package is used to implement this naive bayes algorithm. This package is facilitated with some features like sentiment analysis, classification, tokenization, spell correction, etc,. A textblob function with multiple parameters is created and passed to classify the polarity of a tweet. To analyse the sentiment of a tweet polarity is used, which is a measure to classify the tweets in terms of different dimensions such as positivity, negativity or neutrality. In most of the cases neutrality maybe ignored.

Positivity can be measured if a tweet probability is greater than zero i.e., >0.

Negativity can be measured if a tweet probability is less than zero i.e., <0.

Neutrality can be measured if a tweet probability is equal to zero i.e., =0.

On classification, the polarity of that particular tweet is returned. For the example stated above we obtain sentiment analysis as 'Neutral'. As it carries both positive expression and negative expression, it is considered as a Neutral tweet.

V. RESULTS

	Tweets	len	ID	Date	Source	Likes	RTs	SA
0	@AmitShah FRDI बिल को निर्दोष लोगों से बैंकों ...	138	970910118257684480	2018-03-06 06:33:05	TweetDeck	0	0	0
1	RT @siva_raman71: @DVSBJP @BJP4India Please d...	140	970577129053278208	2018-03-05 08:29:54	TweetDeck	0	1	0
2	@Arun2981 Thank you for linking the ongoing da...	140	970565191011221506	2018-03-05 07:42:28	TweetDeck	0	1	0
3	@ValmikiNaik @meerasanyal @thewire_in Thanks f...	139	968017978972282880	2018-02-26 07:00:45	TweetDeck	10	9	1
4	@madhudawda @geetv79 Thanks for expressing you...	140	968015992877072384	2018-02-26 06:52:52	TweetDeck	0	0	1
5	@Tamil1947 @Kanwal33 @INCIndia @Tharoorian_INC...	139	968015586994282496	2018-02-26 06:51:15	TweetDeck	0	0	0
6	@ca_uves @ajaymaken @amanpanwar @ameeta_sinh @...	134	968014654076174337	2018-02-26 06:47:33	TweetDeck	2	1	0
7	Have you asked your MP not to legislate the dr...	140	968012693084037121	2018-02-26 06:39:45	TweetDeck	0	0	0
8	RT @cfa_ind: Today at the Press Club of India,...	144	966913357209989120	2018-02-23 05:51:23	TweetDeck	0	1	1
9	cc @nit_set @MahtabNama @svaradarajan @Zakka_J...	128	966614280140279808	2018-02-22 10:02:57	TweetDeck	0	0	0

Fig.4: Displays the sentiment analysis of every tweet

The above table shows the tweet and its length and ID of the user who tweeted the tweet on particular Date by what means of Source. And for that tweet how many Likes, Re-tweets are there and Sentiment analysis for the tweet.

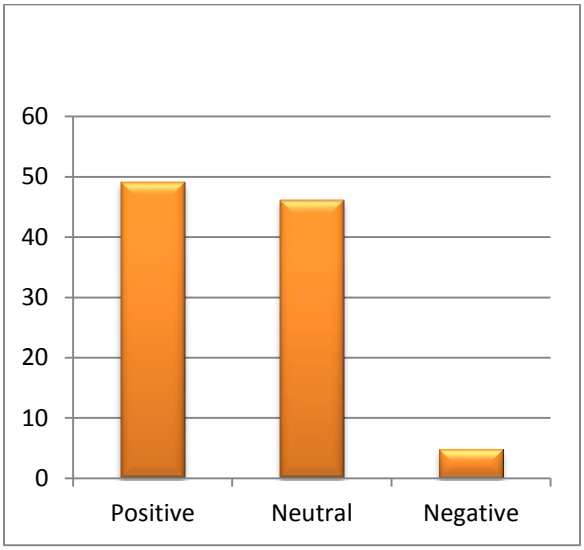


Figure 5 shows the percentage of positive, negative and neutral polarity on FRDI.

VI. CONCLUSION

In this paper, we have done analysis on FRDI (Financial Resolution and Deposit Insurance) by considering the opinion of the people from micro blogging site i.e., Twitter. Based on the analysis made, we conclude what is the impact of FRDI bill on Indian Economy. On analysis, we obtained the result as 49.09% as positive, 4.84% as negative, and 46.06% as neutral.

VII. FUTURE SCOPE

We have done analysis on FRDI (Financial Resolution and Deposit Insurance) by considering the opinion of the people from micro blogging site i.e., Twitter. Based on the analysis made, we conclude what is the impact of FRDI bill on Indian Economy. Till now there was lot of work done in this sentiment analysis field, but there wasn't much work done that considered the location aspect of the tweets. As extension to this work, analysis can be made by filtering the tweets location wise and compare it with other popular approaches[14] in terms of accuracy and efficiency.

REFERENCES

[1]. Alec Go, Richa Bhayani, Lei Huang, "Twitter sentiment classification using distant supervision", Technical report Stanford Digital Library Technologies Project, 2009.

[2]. http://www.yuefly.com/Public/Files/2017-03-07/58beb0822faef.pdf

[3]. B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury, "Micro-blogging as online word of mouth branding". In CHI EA '09:

Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, pages 3859–3864, New York, NY, USA, 2009, ACM.
[4]. https://faculty.ist.psu.edu/jjansen/academic/pubs/jansen_Micro_blogging_as_Online_Word_of_Mouth_Branding.pdf
[5]. Delenn Chin, Anna Zappone, Jessica Zhao, "Analyzing Twitter Sentiment of the 2016 Presidential Candidates", News & Publications: Stanford University.
[6]. https://web.stanford.edu/~jesszhao/files/twitterSentiment.pdf
[7]. Abdalmounaam Rezgui, Daniel Fahey, Ian Smith, "AffinityFinder: A System for Deriving Hidden Affinity Relationships on Twitter Utilizing Sentiment Analysis", Internet of Things and Cloud Workshops (FiCloudW), IEEE International Conference, 2016.
[8]. http://ieeexplore.ieee.org/document/7592726/
[9]. M.P.Ashna, Ancy K Sunny, "A study on sentiment analysis in Malayalam language", Computing and Methodologies Communication, 2017 International Conference.777-783.
[10]. https://www.ijarcce.com/upload/2017/si/nCORETech-17/IJARCCE%20nCORETech%2014.pdf
[11]. Liu, Bing, and Lei Zhang. "A survey of opinion mining and sentiment analysis", Mining text data, Springer US, 2012, 415-463.
[12]. http://www.cs.unibo.it/~montesi/CBD/Articoli/SurveyOpinion Mining.pdf
[13]. Fetty Fitriyanti Lubis, Yusep Rosmansyah, Suhono Harso Supangkat, "Improving course review helpfulness prediction through sentiment analysis", ICT for Smart Society (ICTSS), 2017 International Conference.
[14]. http://ieeexplore.ieee.org/document/8288877/
[15]. Akshat Bakliwal, Piyush Arora, Senthil Madhappan Nikhil Kapre, Mukesh Singh and Vasudeva Varma, "Mining Sentiments From Tweets", e 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, pages 11–18,2012.
[16]. http://anthology.aclweb.org/W/W12/W12-3704.pdf
[17]. C. D. Manning and H. Schutze, "Foundations of statistical natural language processing", MIT Press, 1999.
[18]. http://ics.upjs.sk/~pero/web/documents/pillar/Manning_Schuetze_StatisticalNLP.pdf
[19]. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)", pages 79–86, 2002.
[20]. https://www.cs.cornell.edu/home/llee/papers/sentiment.pdf
[21]. Ankur Goel, Jyoti Gautam, Sitesh Kumar, "Real Time sentiment analysis of tweets using Naive Bayes". Next Generation Computing Technologies (NGCT), 2nd International Conference, 2016.
[22]. http://ieeexplore.ieee.org/document/7877424/
[23]. Milstein.S, Chowdhury.A, Hochmuth.G, Lorica.B and Magoulas.R, "Twitter and the micro-messaging revolution: Communication, connections, and immediacy — 140 characters at a time", O'Reilly Media, 2008.
[24]. http://weigend.com/files/teaching/haas/2009/readings/OReillyTwitterReport200811.pdf
[25]. M. Vadivukarassi, N. Puviarasan and P. Aruna, "Sentimental Analysis of Tweets Using Naive Bayes Algorithm", 2017.
[26]. https://www.idosi.org/wasj/wasj35(1)17/7.pdf
[27]. Dr.S.P.Setty, B.Srinivasa Rao, K.Nageswara Rao, "An approach for heart disease detection by enhancing training phase of neural network using hybrid algorithm" in Advance computing conference (IACC) , 2014.
[28]. http://ieeexplore.ieee.org/abstract/document/6779500/?reload=true