SENTIMENT ANALYSIS ON GST IN SOCIAL MEDIAUSING R

Dr.SP.Chokkalingam[1], N.DuraiMurugan[2]
[1]Professor, Department of Information and Technology
[1]Saveetha School of Engineering, Saveetha University, Thandlam, Chennai, Tamilnadu, INDIA.
[2]Research Scholor, Saveetha School of Engineering, Saveetha University
[2]Assistant Professor, Department of Computer Science and Engineering, Rajalakshmi Engineering College, Tamilnadu, INDIA.
E-Mail: cho_mas@yahoo.com

ABSTRACT

The rapid development in social media has given the method for exploring and tracing the all the sentiments in all sort of issues in India. In recent past months there has been a change in the Goods and Services Tax of financial standards in India which leads to major talks in all over India through social media and led to a big debates between every individuals' emotions. People are sharing expectations, perspectives and Sentiments with large amount of text or image or video data. To discover bits of knowledge from this Data produced and settle on certain choice we represent microblogging web application that gathers twitter information which shows various measurable structures. In this paper we perform analysis of twitter information to investigate the common mans' sentiment or emotionsusingsensitivity in R. Our essential approach was to concentrate on data as positive, negative or unbiased from twitter datasets using streaming API to extract tweets related to GST.

Keywords:*GST,Sentiment Analysis, Word cloud, TwitteR,*

1. INTRODUCTION

Social Mediaand Analytics

The social media has taken rapid increase in the number of users of internet over the last decade. It is alsoplays a major rolein digital marketing and continues to grow. It is now one of the substantial partsof our life and it's been served both as a benefit as well as a misery for the society. People from across the globe take to the social media to express their thoughts regarding a trending topic. We witness various discussions regarding important topics by the people of the world across different social media platforms. One can easily express his or her thoughts by simply tweeting about it or posting about it on Facebook[10]. Other social media users can express their regard or disregard on the very same topic by retweeting it or by

commenting on it respectively. This way, we get a clear idea about the number of users in favour of the topic and the number of users against the matter in discussion.

Recent years have witnessed the explosive growth of the social media content on the internet; people now express their views on almost anything in discussion [7,8,9]. There are many micro-blogging websites like Twitter, Facebook, etc. Twitter has become a very popular communication tool among Internet users and it is one of the most open and simplest platforms to share their opinions on different topics. The outgrowth of social media in expression of thoughts has resulted in the availability of huge volumes of data from masses has startedaground for steering research in the area that combinesboth politics and social media.

Twitter is one of the most popular social networking services used by today's generation. It enables users to send and read short 140-character messages called "tweets" [5]. When someone posts a new tweet on some topic, other like-minded users also join in the talk and this formulates a Twitter's network of like-minded users related to some topic which can provide anpointer of poll of assuranceassociated to the topic of conversation.

In India usage of socialnetworking isgrowing movementsnowadays in participating the sites like Facebook,Twitter, whatsapp, Instagram, ect are become a public domain is portrayed clearlyto express their opinions regarding major hot topic. The active user of social media users in India were more than 18 million in 2015 and by in late 2016 there was massive increase in number of active internet user tomore than 100%.Consider a scenario where in we are required to determine whether the discussion about a trending global topic is on a positive side or a negative side or unbiased side. Suppose we consider many tweets on the topic of discussion, it becomes a difficult task to go through each tweet [11,15].

Data on social media can be broadly classified into textual data such as tweets and comments, network data such as Facebook friendship network or the twitter follow-following network, and actions such as likes, shares, views, and retweets. Get-together this data, pre-processing it, and received at consistent and resounding insights on the data become an integral part of social media analytics.

Goods and Services Tax (GST):

GST is an indirect tax validall over India which replaced multiple fallingof taxes imposed by the central and state governments. The GST is directed by a GST Council and its Chairman is the Finance Minister of India. GSTfollows[0, 5, 12, 18, and 28] %rates as tax

where 50% of these taxesis to state government and 50% taxes is to central government. India's biggest tax modificationwas passed onmidnight (Jun 30/Jul 1,017) by the Prime Minister of India which marked by a historic session of both the houses of parliament convened at the Central Hall of the Parliament.

2. RELATED WORK

Social media has been explored to estimate the popularity of politicians, sentiments of general public towards some recently introduced policy maybe budget, tax reforms etc., to find out the sentiments of social media users. Social networking sites have also been used to compare people's political preferences expressed online with those observed by elections. Social media can be analysed on daily or hourly basis during an electoral campaign so as to get a detailed insight into emotions of voters[3,4].

It is possible to track in real-time trends and capture any sudden change by monitoring and analysing the conversation on social networking sites and get the public opinion well before declaration of results of polls. There are few studies that claim that analysing social media allows a reliable forecast of the final result. In a study by researchers, it has been stated that the number of times a candidate is mentioned in blog posts is a good predictor of electoral success and can achieve better predictions than election polls.On similar lines the authors, compared party pointed out on Twitter with the results of the 2009 German election and discussed that the relative number of tweets related to each party is a good predictor of its vote share and stands a improvedmethod to analyse tweets such that not just the total or mention of party name or candidate name is considered but the sentiment attached in tweets are also analysed as stated by authors [7,8].

3. DATA COLLECTION AND PRE-PROCESSING

Twitter data has been taken as aprogressiveenvironmentof collecting the public opinion and treat it as data elections. Which depicts their different emotions and their sentiments and they keeps on tweetingtheir opinion based on some statement or news by political parties. By usingtwitterAPI, we collected 1,000 tweets and observedthe GST proposalnoticefrom the day of declaration30thJun' 17 or 1stJul'17 to till date. The tweets included useful information related to GST as wellpunctuation marks,special characters, and emojis. The data collected has been cleaned so as to remove all the unwantedpunctuation

symbols, special characters. All tweets have been converted to lowercase and finally a word corpus has been generated.

4. TECHNIQUE APPLIED

We have used the R language which is used widely for data analysis and statistical computing. It is a powerful language developed in the year 1993. R has sufficientrequirements forimplementing and executing machine learning algorithms in a fast and simple manner. We have generated Word Cloud based on the set of 5,000 tweets concentrating on GST and Sentiment is also been analysedon these tweets which depicts the emotions of public in the direction ofGST introduction. We have used many packages viz. syuzhet, lubridate, ggplot2, scales, reshape2, dplyr to facilitate understanding of emotions and plotting the same.

5.RESULTS AND DISCUSSIONS

5.1. Word Cloud Generation

A word cloud is a method of mining textwhichconsents to focus themost frequently usedwords in a tweet.It is a photographicsign showing the most related or relevant words. We always wanted to produce this kind of images within R.we visualize tweets as a word cloud to bond outwhat people are tweeting about the GST (#gst).

We require three R packages for this, Extract tweets from Twitter, identify & create text to turninto a cloudthebest step is to identify & create text on which you want to create the word cloud is by creating a word corpus from the collectionof text.Corpus is just a way to store a collection of documents in R software readable format.The text mining packages like “tm” operate on a formatcalled corpus.

Data cleaning on the text is the most important step.It is done by converting all data into lowercase and removingpunctuation, links or replacing symbols like “/” or “@” witha blank space. We’ve transformed every word tolower case, so that ‘Tax’ and ‘tax’now count as the same word. We’ve removed all punctuation ‘gst’and ‘gst!’ will now be the same. We stripped out any extrawhitespace and we removed stop words andwe need to remove URL’s from text.

Stop words are the common uninterested words. If we look at the result of stop words for eg:”English” we can see what it isgetting removed.The information value of ‘stopwords’ is near zero due to the fact thatthey are so common in a language. Removing this kind of

words isuseful before further analysis. We can also remove our own stop word depending on what you are trying to achieve with your analysis to do the data cleaning step differently. get Transformations() is used to see what are all our data cleaning options.

Term Document Matrix() function would be building a table containing the wordsfrequencyFig.1. directly helps us identify the most frequently used words in the text.

Figure 1: Word cloud "GST"

5.2. Sentiment analysis

Sentiment Analysis is used to determine whether a part oftext is positive, negative or neutral called as opinionmining. A common practiceis to discover how people feel about aspecificissue. It is extensively applied to assessments and social media for avariety of applications.

We will classify the sentiment of a tweet based on the polarity of theindividual words. Each word will be given a score of +1 if classified aspositive, -1 if negative, and 0 if classified as neutral.This will be determined using positive and negative lexicon lists compiled in the #gst wordlist, 5000 words and phrasesrated from -5 [very negative] to +5 [very positive].

Words isdivided into four categories,

Very Negative (-5/-4),

Negative (-3/-2/-1),

Positive (1/2/3),

Very Positive (4/5/6),

Sentiment Function: Once we convert these tweets into some useful information. The main operationalvalue of sentiment analysis is to find thewords in the tweets thatrepresent

positive or negative sentiments.we use two packages plyr and string to operate strings. Put tweets into data frameand Apply sentiment Function to the tweets.

This means most of the people were having positive or happy about GST. From the plot we can see overall score liesbetween neutral and positive.So, in this series of three Articles we learned how to performSentimental Analysis in R by extracting data from Twitter.

Finally the sentiment analysis of the set of tweets has been done to understand the emotionsof public for GST. Emotions of anger, anticipation, disgust, fear, joy, sadness and surprise have been extracted using get_nrc_sentiment() function of "R" software and other preprocessing of tweets. From the sentiment analysis of "gst" tweets it is very clear that there is a feeling of anticipation for "GST". Here nearly equal joy and sadness emotion for "GST" in India.

The tweets show that the citizens of India have hardly any emotion of anger or fear asobserved from the tweets. The sentiment analysis has been shown in Fig. 2.

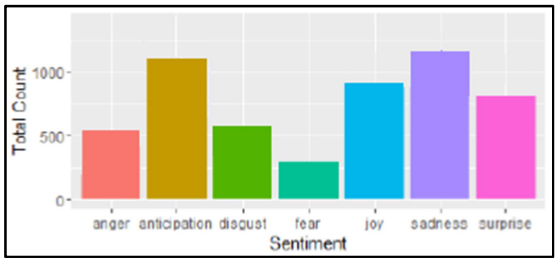

Figure 2: Sentiment analysis of tweets mentioning "GST"

6. CONCLUSIONS

The growingquantity of social media users has raisedthe interest about the opportunity to understand the relation between people's favourites and real political behaviour. This training focuses on the question that whether the data from social networking sites can be utilized to interpret the attitude of citizens of a nation towards various policies. We have observed that twitter is very commonly being used as a platform for discussion by citizens of India. We concluded that social media is a controlling and reliable source of public opinion as far as a nation like India is concerned. The discussions on twitter are corresponding to traditional debates and are skilledadequate to give animpartial idea of emotions of general public. We have done sentiment analysis of emotions of people which shows people's

acceptance for GST but with too much of sadness feeling. In future, we plan to convert this analysis in real time corresponding to tweets arriving on time-basedrule and geographically split and examine the tweets according to states.

REFERENCES

1. NehalMamgain, Ekta Mehta, Ankush Mittal and Gaurav Bhatt.2016 Sentiment Analysis of Top Colleges in India using Twitter Data.
2. KarthikGanesan, Akhilesh P Patil, SrinidhiHiriyannaiahRamaiah Institute of Technology, "Predictive Analysis of Tweets on Goods and Services Tax(GST) in India using Machine Learning", International Journal of Engineering Technology, Management and Applied Sciences, August 2017, Volume 5, Issue 8, ISSN 2349-4476
3. Dr.KavitaPabreja, "GST sentiment analysis using twitter data", International Journal of Applied Research 2017; 3(7): 660-662
4. The constitution (one hundred and first amendment) act, Amendment No. 101 of Retrieved from Internet, 2016, 8.
5. India's midnight 'tryst with destiny': GST rolled out. Dynamite News.com.(ANI).Retrieved from Internet, 2017.
6. Gloor PA, Krauss J, Nann S, Fischbach K, Schoder D. Web science 2.0: Identifying trends through semantic social network analysis CSE. International Conference on Computational Science and Engineering, Vancouver, BC, 2009; 4:215-222.
7. Barbera P. Birds of the same feather tweet together. Bayesian ideal point estimation using twitter data, 2012.
8. Conover M, Goncalves B, Ratkiewicz J, Flammini A, Menczer F. Predicting the political alignment of twitter users. Proceedings of 3rd IEEE Conference on Social Computing Social Com, 2011.
9. O'Connor B, Balasubramanyan R, Routledge BR, Smith NA. From tweets to polls: Linking text sentiment to public opinion time series. Proceedings of the International AAAI Conference on Weblogs and Social Media, Washington, DC, 2010.
10. Xin J, Gallagher A, Cao L. The wisdom of social multimedia: Using flickr for prediction and forecast. ACM Multimedia, International Conference, New York, NY, 2010.