



SENTIMENT ANALYSIS ON GST TWEETS

Ekta
Student-Department of Computer Science
GGSIPU
Dwarka, Delhi, India

Dr. Kavita Pabreja
Associate Professor-Department of Computer Science
Maharaja Surajmal Institute, GGSIPU
Dwarka, Delhi, India

Abstract: Social networking websites are considered as major sources of opinions and emotions of the public on social issues. Twitter a microblogging website where users share millions of views on a wide range of topics on a daily basis which can be used for Sentiment analysis in order to identify user's opinions, likes, dislikes, feedbacks, etc. Analysis of people's sentiment in a text and classification in terms of positive, negative or neutral is a challenging task. In this paper, Twitter has been used as a forum to analyze and graphically represent people's opinion of India towards recently introduced Goods and Services Tax by Indian Government on 1st July 2017. The tweets originating in India on 30th June and 1st July have been extracted then classified as positive, negative and neutral sentiments.

Keywords: Twitter, Sentiment analysis, Goods and Services Tax, Word Cloud, R language.

I. INTRODUCTION

The word tax originates from the Latin word 'taxare' which means 'to estimate' [1]. The Constitution One Hundred and First Amendment Act, 2016, introduced a national Goods and Services Tax in India from 1 April 2017 [2]. GST is an indirect tax that subsumes almost all the indirect taxes of central government and states governments into a unified tax [3]. GST was first introduced by France in 1954 [4]. Shri Atal Bihari Vajpayee brought this system to India in 2000 but no one paid attention to it and due to some reasons, it was not passed [5]. Goods and Services Tax (GST) is considered as biggest tax reform since 1947 [6]. It is an indirect tax levied in India from July 1, 2017 [7] through the implementation of One Hundred and First Amendment of the Constitution of India by Indian Government [8].

The GST was launched on the midnight of June 30, 2017 inside Parliament's Central Hall [9] by the honorable Prime Minister of India Mr. Narendra Modi. Members of Congress have listed for boycotting the GST and several other opposition parties too were staying away [10]. The GST replaced existing multiple cascading taxes levied by the central and state governments on the sale of goods and services. GST has two components - Central GST levied by the Centre and State GST levied by the states [11]. Goods and services are taxed at the rate of 0%, 5%, 12%, 18% and 28% [12]. There is a special rate of 0.25% on rough precious and semi-precious stones and 3% on gold [7]. GST was initially proposed to replace a slew of indirect taxes with a unified tax and was therefore set to dramatically reshape the country's 2 trillion dollar economy.

With the explosive growth of social media in the last decade, it is becoming an integral part of life. Nowadays billions of people all around the globe are expressing their views on almost anything in a discussion. This enables us to communicate with each other anytime without geographical

boundaries. Web 2.0 is the term that refers to the second generation [13] of World Wide Web where it changed web pages from static to dynamic. Social media is an aspect of Web 2.0. Twitter is a most popular social networking website where users communicate in 140-character messages called tweets. On November 7, 2017, the limit was doubled to 280 characters for all languages except Japanese, Korean and Chinese [14]. A Twitter user can follow any other user, and the user being followed need not follow back. Being a follower on Twitter means that the user receives all the tweets from those the user follows. Twitter's audience varies from regular users to celebrities, politicians, and even country presidents. Therefore, it is possible to understand opinions of users from different social and interests groups on a variety of topics. According to statistical website number of Twitter users in India are 30.4 million in the year 2018 and by the year 2019 number of users will be 34.4 million [15].

This paper describes various steps required to perform sentiment analysis on Goods and Services Tax using the tweets and conducts a fair judgment about this scheme launched by Government of India. We gathered Twitter data using the streaming API to extract tweets related to GST. This paper is organized as follows; first, we present the Literature Review of Research papers in the second section. Then we have provided a brief overview of techniques applied and approach used for sentiment analysis in section third and fourth respectively. Section five describes the different steps performed to generate the result. And finally, section six gives the conclusion of the whole paper.

II. LITERATURE REVIEW

A. Khurana, A. Sharma 2016 [1] mentioned that GST will provide relief to producers and consumers by providing wide and comprehensive coverage of input tax credit set-off, service tax set off and including several taxes. Efficient formulation of GST will lead to resource and revenue gain for both Central and State government. There will be a positive impact of GST on various sectors and industry. A. Yadav, 2017 [4] discussed that there will be a positive impact of GST on sectors like Infrastructure, textile, IT, Agriculture, Food Industry,

Transport, Real estate industry. GST will improve tax collection and boost India's economic development. M. Kour, K. Chaudhary, S. Singh, B. Kaur, 2016 [5] stated that through GST Indian goods would be taxed at the same rate. GST play a dynamic role in the growth and development of our country. K. Pabreja, 2017 [7] concluded that sentiment analysis of emotions of citizens of India towards introduced Goods and Services Tax shows people's acceptance for it but with too much of anticipation feeling. R. Vasanthgopal, 2011 [17] stated that switching from an indirect tax system to GST will be a positive impact on Indian economy. More than 140 countries in the world have been introduced GST and a new preferred form of an indirect tax system in Asia Pacific region also. S. Gupta, Sarita, M. K. Singh, Komal, H. Kumawat, 2017 [18] explained that tax rate of GST must be implemented in such a way which will be beneficial to both the people and the Government. Through GST, the revenue which the government earns from indirect taxes will reduce. R. Sharma, 2017 [19] explained that implementation of GST will lead to more employment opportunities and flourish GDP by 1-1.5%. People will be able to do business at low cost and making domestic products more competitive in the local and international market by implementing GST. It will emerge as a world-class tax system in India. S. Poddar, E. Ahmad, 2009 [20] discussed that benefits from GST are critically dependent on a neutral and rational design of the GST. It will make simpler and more transparent tax system in India with an increase in output and productivity of economy in India. S. Shaik, S. A. Sameera, Sk. C. Feroz, 2015 [21] explained that GST will lead to commercial benefits and economic development in Indian framework. Through GST there will be a collective gain for the industry, trade, agriculture, common consumers as well as for Central Government and the State Government.

As none of these studies emphasize on the response of Indian public towards GST, in this paper, we have tried to observe the reactions of Indians towards GST.

III. TECHNIQUE APPLIED

We have used the R language which is an extremely flexible statistics programming language. It was developed in 1995 which was designed by Ross Ihaka and Robert Gentleman [16]. It consists of its own inbuilt statistical algorithms, which makes it easy for users to perform statistical computing. Third-party packages are staggering and continue to grow. R provides a list of a wide collection of tools for data analysis [8]. Sentiment analysis of 5,000 tweets has been done to understand the sentiments of public towards the introduction of Goods and Services Tax.

IV. SENTIMENT ANALYSIS APPROACH

Tweets related to GST have been extracted from the twitter using twitterR package. The input text has been broken down into words called tokens. Every new token encountered is then matched to the lexicon in the dictionary. A polarity score has been assigned to each of the tokens. In order to determine the sentiment behind the text the aggregated sum of the score has been calculated.

V. METHODOLOGY

Now let's have a look at different phases involved in the approach. The different steps from the collection of relevant data to finally analyzing it are as follows.

A. Data collection

An easy way to extract tweets containing the word "GST" from a user account or public tweets, the twitterR package is been introduced. From June 30, 2017, to July 1, 2017, the total number of tweets collected was 5000 to carry out sentiment analysis using R programming language. By creating a Twitter API, tweets were extracted and stored in a csv file named gst.csv.

B. Data pre-processing

The first step is to convert all the tweets that have been extracted to the lowercase using tolower function because R is case-sensitive [3]. However, Twitter doesn't distinguish between cases while a search is being carried.

Once the tweets have been converted into lowercase, the next step is to remove punctuations, Stopwords, numbers, URL's. These are not an essential element to be considered in the case of sentiment analysis, hence removed. These steps have been shown in Figure: 1.

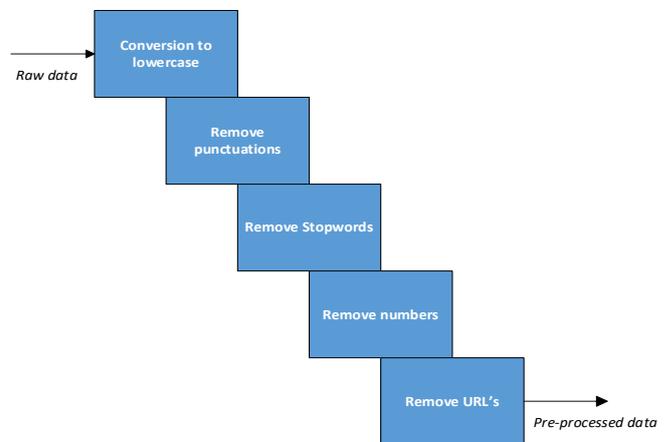


Figure 1. Steps to clean raw data

C. Data probing

This method consists of two dictionaries – one of the positively tagged words and other of negatively tagged words. Search is carried on each individual word of a tweet within those dictionaries then polarity score is assigned depending upon the location of the word.

1. Scoring: If a token is present in the dictionary of positive words, +1 score is assigned and if present in the dictionary of negative words, -1 score is assigned, else 0 score is assigned.
2. Aggregation: The total sum of the scores allocated to each word of a tweet is calculated and based on the final polarity value, tweets can be classified as positive, negative or neutral.

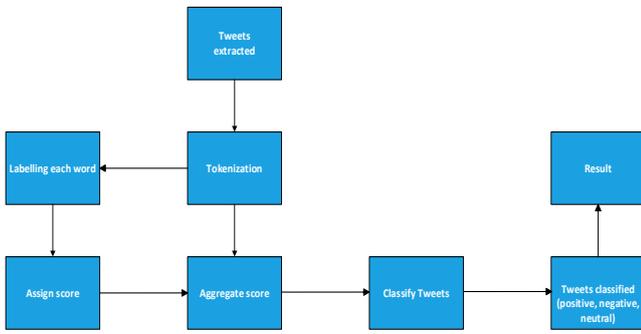


Figure 2. Overview of Sentiment analysis approach

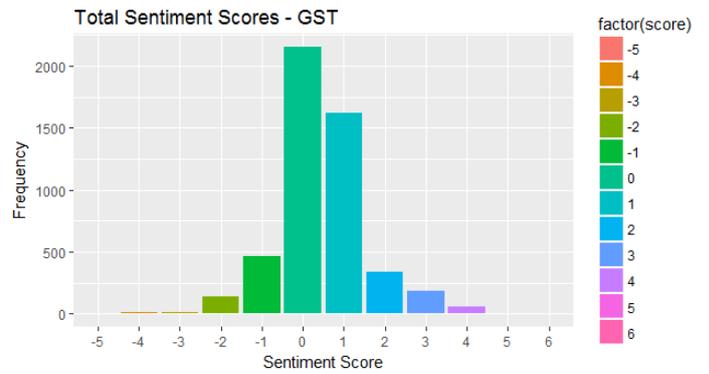


Figure 4. Total Sentiment Scores

D. Results

All tweets mentioning “GST” have been analyzed from the following perspective:-

- Analyzing user’s sentiment score
- Word cloud generation

Sentiment analysis of the set of tweets has been done to understand the opinions of the public for GST. Opinions are positive, negative or neutral. From the bar graph in Figure: 3 there are 2203 Twitter users have who commented in a positive way, 638 Twitter users have commented in a negative way and the rest 2159 are neutral, as mentioned in Table: 1.

Table 1. Total tweets of each polarity

Polarity	Number of tweets
Positive	2203
Negative	638
Neutral	2159

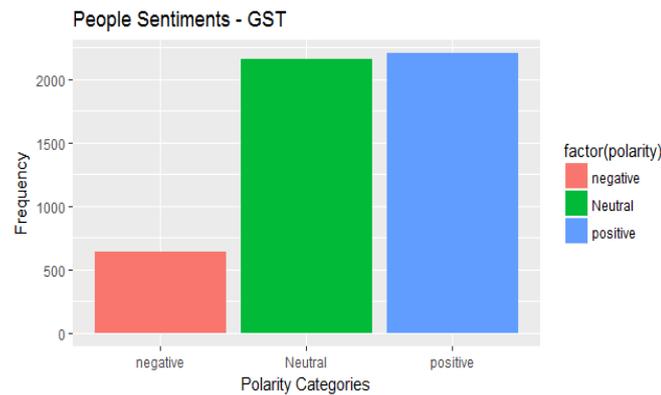


Figure 3. Sentiment analysis of tweets mentioning “GST”

The bar graph below in Figure: 4 depicts Twitter user’s sentiment score, the positive score plus (+) symbol denotes that users are quite happy whereas the negative score denoted by the minus (-) symbol indicates unhappiness with GST and zero represents that users are neutral. Table: 2 shows the total number of tweets of a particular score.

Table 2. Total tweets of each score

Score	Number of Tweets
-6	0
-5	1
-4	13
-3	16
-2	141
-1	467
0	2159
1	1618
2	340
3	183
4	60
5	1
6	1

From the below pie chart in Figure: 5 it is very clear that there are 44% of Twitter users of India in favor of GST, 13% against and rest 42% neutral (may or may not be in favor of GST).

The Word Cloud has been generated corresponding to tweets mentioning “GST” as shown in Figure: 6. The word “gst” is contained by every tweet on which analysis has been performed. As the size of a word depends on the frequency of its usage, so the size of word “gst” is big in comparison to all other words in the Word Cloud.

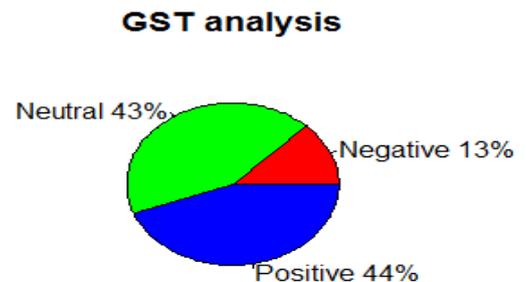


Figure 5. Pie chart for tweets mentioning “GST”

