Sentimental Analysis on the Union Budget, India-2020

PROF.KALAM NARREN
HOD of Business Analytics,
ISBR Business School
Bengaluru, India

PROF.V.VINAY KRISHNA
Department of Finance and Business Analytics, ISBR.
Bengaluru, India

Dr.CHANDRA MOHAN
Department of Finance and Business Analytics, ISBR.
Bengaluru, India

Dr.SAI
Department of Marketing and Business Analytics, ISBR.
Bengaluru, India

Abstract-

In this work, the sentiment analysis of annual budget of India for the Financial year 2020-21 is performed. Text mining is used to extract text data from twitter using hashtags related to annual budget and to compute the word affiliation of full-size phrases and their correlation in computed with the related phrases. To understand the outcome of the research we visualize the results using word cloud for the word frequency. R programming is used to facilitate the analysis and research. The resulting sentiment score is computed and analyzed. This analysis is of substantial significance keeping in mind the sentiment of the citizens towards the particulars of the budget and how it is going to add to the growth of the country.

Keywords - Sentiment Analysis, Union Budget, Text mining, Twitter, New Tax Policy.

Highlights – This paper brings out the sentiments of citizens regarding the Union Budget for India

Introduction

Sentiments are a representation of one's emotions, thoughts, and perceptions of individuals which defines the base upon which the constructs framing the mindset of inundated methods of understanding and comprehension are established and nurtured. In the world of today, every individual who has a perception and understanding of their own seeks forward to enhance it by professing it to others so that by the contribution of the particular subject or matter the subject itself can flourish upon. As a result of a world which is ruled upon by sentiments and perceptions, it is imperative to be able to analyze and draw conclusions of the collective emotions and perceptions as everything in the world which is given reverence upon by the people is a result of the collective belief of a group.

To resolve the major issues of the previous Union Budget the current one aimed to enhance the financial outlook and better the governance, The Union Budget 2020 was mainly focused upon the central ideas of "*Aspirational India, Economic development and A Caring society*".

These ideas were to be implemented by adequate governance and a corruption-free financial sector. The budget atoned the perspectives of Tax-payers by ensuring a positive impact portraying a continued focus of the government upon Healthcare, also highlighting a reasonable solution which was pressed from the consolidation of fiscal perspectives to acknowledge the importance of it together.

Methodology

Here for disclosure upon the methodological practices, one of the most prevalent social media platforms, Twitter was chosen, which enables an individual to broadcast his or her opinions towards prevalent topics to be broadcasted all across the world [Alec et.al.,2009] and [Pak and Paroubek, 2010]. The various tweets were downloaded by using the hash tags

"#UnionBudget" and/or "#UNIONBUDGET2020" over a period of one month coinciding with the budget announcement i.e. from 2nd February 2020 to 29th February 2020, this period has been chosen to get the most accurate and genuine responses of all those individuals who have held high hopes patiently awaiting the forthcoming budget.This was cleaned to remove duplicate tweets and retweets.

A. Data Cleaning

The Main purpose of this study is to comprehend the sentimental reactions of people towards the Union Budget proposal of 2020, and the data has been captured through python and was exported to excel. The Data has been cleaned by removing repetitions which included the retweets, opinions of the same individuals recounted and remaining relevant tweets were retained. Further, the neutral words were removed so that they don't pose an obstacle in the classification of sentiments. The data collected was cleaned by removing punctuations, stop-words in English language, words lesser than three alphabets were removed, numbers and special characters were removed [Hu and Liu, 2004-].

B. Word Frequency

After cleaning the text data, the next step is to count the occurrence of each word. Using the function TermDocumentMatrix() from the text mining package, we can build a Document Matrix – a table containing the frequency of words. The ten most repeated words have been tabulated in the results, which has also been visualized as a word cloud to give a general sense of the frequency of all the words used. Here the size of the word represents the number of times it has been repeated relative to the others.

C. Sentiment Analysis

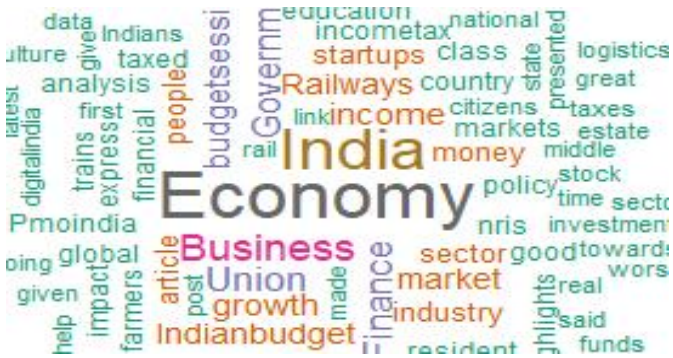
The sentiments can be positive, negative or neutral. Emotion classification as defined by "NRC Emotion Lexicon" is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy and disgust) and two sentiments (negative and positive).

Emotion classification here is performed with the NRC Word-Emotion Association Lexicon (EmoLex). It is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two sentiments (negative and positive). The annotations of which were manually done through crowdsourcing."

The function get_nrc_sentiments, returns a data frame with each row representing a sentence from the original file. The data frame has ten columns (eight columns for each of the emotions and two columns for the sentiments- positive and negative) [Liu et.al., 2005].

Results and Interpretation

The world cloud is a figure made up of those words which are most frequently tweeted and most relevant for our analysis which pertains to economy, education, growth taxation government policies and business the specific domains that the Union Budget governs, this enables us to determine those words which are relatively more relevant than others in order to comprehend the Sentiments and Emotions which are expressed in response to the Union budget.



Source: AuthorsFigure 1. Word Cloud

The Figure 1 shows the particular phrases relevant to our study which are gathered based on frequency of relevant tweets which helps us to narrow down our search and help us locate the main phrases allowing us to efficiently perform this analysis and analyze the relevant sentiments. This specific word cloud is generated from cleaned data wherein the size and the thickness of the alphabets proclaims its relevancy and its frequency, which means the words with the highest frequency and relevancy are the largest. The Table 1 shown below depicts the relevant phrases and the frequency of their expression. In below table we can conclude to state the word which encompasses all of the above with the highest frequency is the word "economy " with a maximum usage in 427 Tweets within the observed time frame, which enables us to circumnavigate around these specific terms used to garner the relevant expressions for our findings.

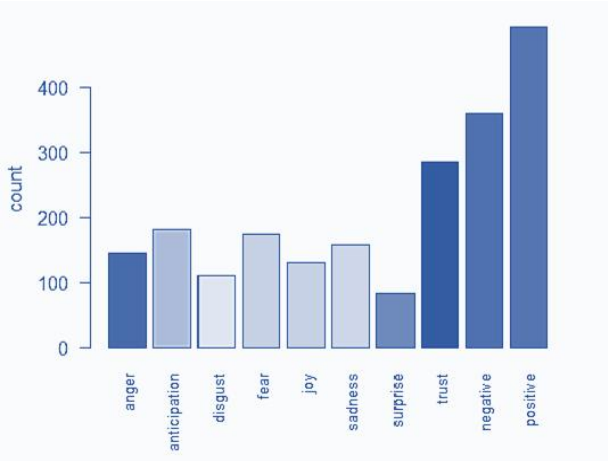
Table 1. Top 10 most frequently used words.

Ranking	Word	frequency
1	Economy	427
2	India	360
3	Business	165
4	Union	145
5	Finance	135

6	Government	126
7	Budget session	107
8	income	104
9	growth	99
10	Railways	92

Source : Authors

Upon analysis of the sentiments exhibited by the numerous tweets, the results have been



Source : Authors Figure 2.Results of analyzing Emotions and Sentiments of tweets.

summarized in the figure 2. EmoLex, segregates the emotions portrayed into 8 different categories which include- anger, anticipation, disgust, fear, joy, sadness, surprise and trust [Goel et.al., 2016]. From the graph it can be observed that trust with the Government policies was the predominant emotion that was portrayed with the tweets, while anticipation was the next predominant emotion, which reflects the populous wanting to learn of the ways the budget impacts them. The other emotions are equally significant and points to the general sense of a mixed feeling towards the budget with anger, fear and sadness being a predominant non-positive emotion being expressed which reflects the sense of non-conformance of the budget by the general public [Naiknawareand Kawathekar, 2016].

The sentiments expressed by the tweets are summarized as 'Positive' and 'Negative' [Guha et.al. 2015], wherein a positive score is taken up when the overall sentiment expressed in the individualtweet is positive, that is there are a greater number of positive sentiment words in the sentence. Similarly, for the negative sentiment score.

In the result it is seen that the number of tweets with positive sentiment is greater than the number of tweets with a negative sentiment. With this it can be concluded that though there is general sense of acceptance towards the policies enlisted in the budget, the proportion of tweets with a negative score too is significant which should be paid attention to in the design of the future budgets.

Conclusion

The tool of sentiment analysis of tweets provides the government with an important means to measure the reaction of the common people towards the various initiatives that it takes up. This can be used as a feedback tool to ensure that the steps being taken are in line with the expectation of the general public which should be the case in a democratic setup.

Here, the major sentiment of the people towards the budget has been positive while the negatives too are significant and thus the government should work towards bettering its offering making adequate corrections before the passing of the budget in the parliament at the end of the budget session or in the future budget proposals.

Future scope of work

This tool highlights the potency of using those platforms which empower the people to voice out their opinions as means of easily accessible information source to understand human sentiments, emotion and perceptions [Hutto and Gilbert, 2014]. The effective usage of such a tool shall help to keep in check the future policymakers, governing & regulating agencies by

ensuring the decisions which they implement for the greater good of the people are actually transpiring to do as they were intended.

Thereby placing the tool to measure the behavioral responses owing to the sentimental characteristics of human beings applicable to any spectrum, in the hands of those who are accountable and responsible for the greater good of others.

References

[1]. Alec Go, Richa Bhayani and Lei Huang. 2009. Twitter Sentiment Classification using Distant Supervision. Stanford University, USA.

[2]. Alexander Pak and Patrick Paroubek.2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Proceedings of the International Conference on Language Resources and Evaluation, LREC.

[3]. Minqing Hu and Bing Liu. 2004. Mining and Summarizing Customer Reviews. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004).pp 22-25. Seattle, USA.

[4]. Bing Liu, Minqing Hu and Junsheng Cheng. 2005. Opinion Observer: Analyzing and Comparing Opinions on the Web.Proceedings of the 14th International World Wide Web conference (WWW-2005). Chiba, Japan.

[5]. Ankur Goel, Jyothi Gautam and Sitesh Kumar. 2016. Real time sentiment analysis of tweets using Naive Bayes. 2nd International Conference on Next Generation Computing Technologies (NGCT).

[6]. Bharat Naiknaware and Seema S. Kawathekar. 2018. Peoples Opinion on Indian Budget Using Sentiment Analysis. International Journal for Research in Engineering Application & Management

[7]. Satarupa Guha, Aditya Joshi and Vasudeva Varma. 2015. Sentibase: Sentiment Analysis in Twitter on a Budget. SEM 4th Joint Conference on Lexical and Computational Semantics. Denver, Colorado, USA.

[8]. Hutto, C.J. and Gilbert, E.E. 2014. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI.