COMMUNICATING FIVE-YEAR BUDGETS FOR THE INDIAN ECONOMY: COMPARATIVE TEXT AND SENTIMENT ANALYSIS

Neelam Kaushal
Department of Business Administration, National Institute of Technology,
Kurukshetra, Haryana 136119, India

Suman Ghalawat*
Department of Business Management, CCSHAU, Hisar 125004, India
***Corresponding Author**

Apul Saroha
Department of Business Administration, National Institute of Technology, Kurukshetra, Haryana
136119, India

ABSTRACT

The content on social media is full of useful information that helps in communicating people's preferences and opinions. The various examples in this context are that people frequently express their opinions about films and other social issues using Twitter, Facebook, etc. In this work, Sentiment Analysis of the Annual Budget for five financial years, namely, *2017–2018, 2018–2019, 2019–2020, 2020–2021*, and *2021–2022* was initiated with the help of Twitter. Firstly, the researcher applied Text Mining to extract the budget's text data documents and computed correlation to know the association of influential words. Then, in analysis section plotted the occurrence of the words and the accompanying word cloud. The analysis was performed employing R software. Finally, the sentiment score for each item was calculated and assessed. This research is crucial because conducting a comparative text and Sentiment Analysis of five-year budgets for the Indian economy would communicate the previously prevailing positive and negative forecasts and thinking, which will aid future policymakers in planning future budgets.

Keywords: Social-media, Annual Budget, Twitter, sentiment analysis, Text Mining

INTRODUCTION

Indian economy is said to be the developing economy of the world. It has the sixth-largest nominal Gross Domestic Product (GDP) and the third-largest Purchasing Power Parity (PPP) in the entire globe. India was the fastest-expanding economy in the world from *2014 to 2018*. If we check India's long-term perspective, we can see a positive curve due to the young generation population and low dependency ratio, investment rates, increasing integrity in the global market, and health savings. Due to demonetization in 2017, the Indian economy was severely impacted. India's GDP mainly depends on government spending, exports to foreign countries, and investments done by the government. Since January 1, 1995, India has been a part of the World Trade Organization. However, there is enormous inequality in India's income distribution among different classes. Indian economy integrated with the global economy, and the main trading partners of India are the United States (US), the United Arab Emirates (UAE), Saudi Arabia, Germany, Hong Kong, South Korea, Indonesia, Malaysia, Switzerland, and China. India also practices free trade policies with the Association of Southeast Asian Nations (ASEAN) countries, South Asian Free Trade Area (SAFTA), South Korea, Japan, etc. The Indian economy is organised into three sectors: primary (agricultural), secondary (manufacturing and industry), and tertiary. According to research published by the Reserve Bank of India in 2021, the tertiary sector provides more than half of GDP. India is fourth in the world in terms of foreign exchange reserves, at $585 billion.

BUDGET

The budget can also be referred to as the Union Budget or Annual Financial Statement. The ruling government present the Union Budget on February 1 or every year. The main reason behind presenting the budget in February is that the government gets proper time to materialize it before starting the new

financial year in April. The presiding Finance Minister presents it in the Parliament. The current study includes the budgets presented by the earlier finance ministers during the mentioned periods: Arun Jaitley presented 2017–2018 and 2018–2019 budgets and Nirmala Sitharaman presented 2019–2020, 2020–2021, and 2021–2022 budgets.

The Union Budget contains all the details of allocation of budget in various sectors by the government. Union Budget always consisted of two categories, namely, Revenue Budget and Capital Budget. In all budgets, the finance minister discusses the factor that will focus on the upcoming financial year. The budget also includes the alterations related to taxation policies that the government will impose in the future.

BUDGET ANALYSIS

Budget Analysis is the analysis of different year budget's expenditures and receipts. The Budget Analysis is analyzing the different year budgets and is based on the expenditure distributed for different sectors, schemes, sources of revenues, and other factors. The Budget Analysis includes both differentiation and similarities between different year budgets and helps in understanding the issues related to budget expenditure and revenues.

In Budget Analysis, Text Mining is performed to extract the main words of the budget speech, and then, Sentiment Analysis is done on the budget speeches to find the emotions of the government behind the budget, with the cutting-edge technology existing to extract text data and examine it. It has grown to be prolific to get a few meaningful facts from textual content statistics. It is challenging to search suitable patterns and trends to essence crucial information from massive data. However, vast data like textual information is formless or unshaped and contravenes easy efforts to form sense. Social media proved very effective in gathering information and ongoing trends. It is an irrational task to analyse the unstructured and mixed textual data. Data mining approaches extract important information from unorganized data. Further, it can help examine the information with the help of pictorial representations and tables.

Text Mining is an instrument and easy technique to scrutinize and penetrate

extensive databases that are not responsible for establishing a qualitative research approach. There are two ways to perform Text Mining. These are as follows:

- **Text refining**: It transforms the document into a specific form based on the choices of the user and its desired needs. Thus, the user can remove desired words and extract the crucial information according to his needs.
- The intermediate form is described as a graphical representation or structural representation of the data desired by the user. For example, a user can extract the data from a document and convert it into graphs, bar charts, tabular representation, etc.

Text Mining is an essential and valuable tool for researchers and can help them analyze the data very quickly with the help of tables, graphs, etc. Text Mining also cleans the unwanted data from any document and can filter it in a simple and easily understandable form. The R software is both a speech and domain aligned towards analytical computing and pictorial formation. It is made accessible under the GNU (GNU's Not Unix) General Public License due to decisive group involvement, several updates, known as packages, developed with time, and substantial documentation. Due to this adjustability and functionality, R software has endured well-liked data and Text Mining implementations consistently across many domains. It contains powerful Text Mining instruments. It is such an advanced software that anything can be done with the help of R software very quickly. The updating of new packages on the site of R studio helped in many ways. We can do any calculation in a fraction of seconds and a large amount of data filtration with R software's help.

In this study, R studio was used as a data mining tool to extract pertinent data from budget speeches and analyze them by making word clouds and doing Sentiment Analysis with the help of R software. The years in which budget speeches were chosen are between 2017 and 2021.

The study compares the budget speeches as the same government is ruling in all five budget sessions. Sentiment Analysis is done in

this study to communicate the emotions of the government while making the budget. The act of recognising and characterizing a piece of text as favourable or negative in order to appreciate the users' feelings is known as sentiment analysis. This search would lay hold of the reader through fundamental sentiment classifiers such as building, comparing clouds, word clouds and categorizing words based on emotions.

As a result, the remaining sections of this study are stated as follows: literature review, methodology, and analysis, which includes word clouds, Text Mining, and Sentiment Analysis. The methodology part of the paper will go over all of the software and tools that were used in the current investigation. In addition, there is a discussion of the findings of current research objects and ideas for future study in the latter sections of the report.

RELATED WORKS
The topic of using social media to envisage the future is very common in recent years. Furthermore, substantial work has been done on Sentiment Analysis (SA) in recent years. Jain (2013) concluded the study on tweets about movies to forecast several facets of the movie popularity. The primary outcomes offered were that whether a movie would be efficacious at the box office. Guha, Joshi, and Varma's (2015) paper specifics the portrayal of the system submitted by team Sentibase for SemEval 2015. The purpose of this work was to put together a complete sentiment analysis for Twitter in a day that achieves enviable performance without going through multiple modelling procedures, as the paper's title suggests. Shakeel et al. (2016) study executed a SA of the Annual Budget executed for 2016–2017.

Syaifudin and Puspitasari (2017) in their study, concluded for the public positive or negative response to government policies. The review showed that people from 21 out of 30 states and union territories of India measured in the study displayed assist to the union government's decision. Thus, user feedback from social media can also be used by governments in public policymaking. Proksch, 19) addressed the challenge with a sentiment-based approach titled multilingual and presented that such a measure could efficiently capture different forms of conflict in

the parliamentary. They also demonstrated that an automated digital transformation produces effective consequences and significantly enables comparative work on legislation. Kaur, Kaur, Singh, and Ranjan's (2020) study conduct a real-time SA of the public response in the direction of the declaration of the Indian Union Budget 2020.

General public vents their thoughts concerning common problems on social media platforms. During the budget announcement hours, the general public's rapid reaction in the form of tweets was retrieved. The dataset was evaluated using a Sentiment Analysis-based model. For the dataset used to determine the sentiment for individual tweets, reliability and polarity values were collected. The public response to the Indian government's demonetization programme, which outlawed INR 500 and 1,000 currency notes with effective from November 8, 2016, was explored by monitoring tweets, according to Singh, Sawhney, and Kahlon (2018). Kaurav, et al. (2021) and (2020) Three topics emerged from the research: learners, language, and discipline. The proportion of persons were hopeful about the administration of the NEP (National Education Policy) in the higher education sector, as per the Sentiment Analysis. Higher Educational Institutes (HEIs), teachers, and researchers could all benefit from this research. Sharma and Gupta's (2021) the report used qualitative research to examine the influence of Covid-19 across a wide range of industries. To accomplish so, extracted tweets were analysed with NVivo, and the important metrics, such as themes, sentiment, word map, and word cloud, were observed. As a result, the report identified different themes under the heading "Atmanirbhar Bharat," with the top five relating to defence and products. According to the Sentiment Analysis, the most of the tweets come from Asia and are encouraging.

Based on an assembly of appropriate Sentiment Analysis pieces of literature, and the application for the same in various research areas like HEIs, movies, and social media, we design to recognize the affiliation between the five-year Budget Analysis of the Indian economy and Sentiment Analysis. Another fundamental goal of the research is to complete the successes that researchers have already accomplished, with the belief that

there is still much more to be added to provide a framework for future research. So, in this perspective, we'd like to look at the significant shifts in public opinion and emotions for the previous year's Indian budget.

METHODOLOGY

The data for the study is retrieved from the official websites of the Indian budget for the years *2017–2018, 2018–2019, 2019–2020, 2020–2021*, and *2021–2022*. The budget speeches were used for the study and retrieved from the official websites of the Indian budget for the years *2017–2018, 2018–2019, 2019–2020, 2020–2021*, and *2021–2022*. For the stabilized study of Union Budget, the speeches were taken for five consecutive years. Hence, the present study was conducted to compare the budgets and communicate the budget's primary focus. The word or term frequency and Sentiment Analysis were executed based on downloaded budget speeches. The R software has been used to plot the word cloud, comparative cloud, and Sentiment Analysis. The R software collects thousands of integrated or unsegregated packages comprising analytical or statistical programming languages and several other functions. These packages help evaluate vast data, statistical analysis of data, visualized outputs in charts, clouds, tables, etc. Figure 1 has summarized various stages for discussion of the current study.

1. To analyze the main focus of five-year consecutive budgets of the Indian economy.
2. To compute commonality in various five-year budgets implemented by the same government.
3. To compare the last five-year budgets to find the similarities and differences between them.
4. To know the sentiments of government for the last five-year budgets of the Indian economy.

TEXT MINING

Various technologies were studied to simplify and comprehend the information needed to gain business insight into the rapid growth of big social data such as twitter, the web and blogs (Park and Yu, 2015). These techniques are essential for handling unstructured data and finding useful information from the data. It is pertinent to group frequent words through Text Mining and group them according to the topic to understand and collect the information with the help of data mining, small and valuable information extracted from extensive data in text files. Text Mining encompasses the course of constructing input text and removing unnecessary characters from the data. Nouns and adjectives are also removed from the text. Thus, the main imperative words are extracted
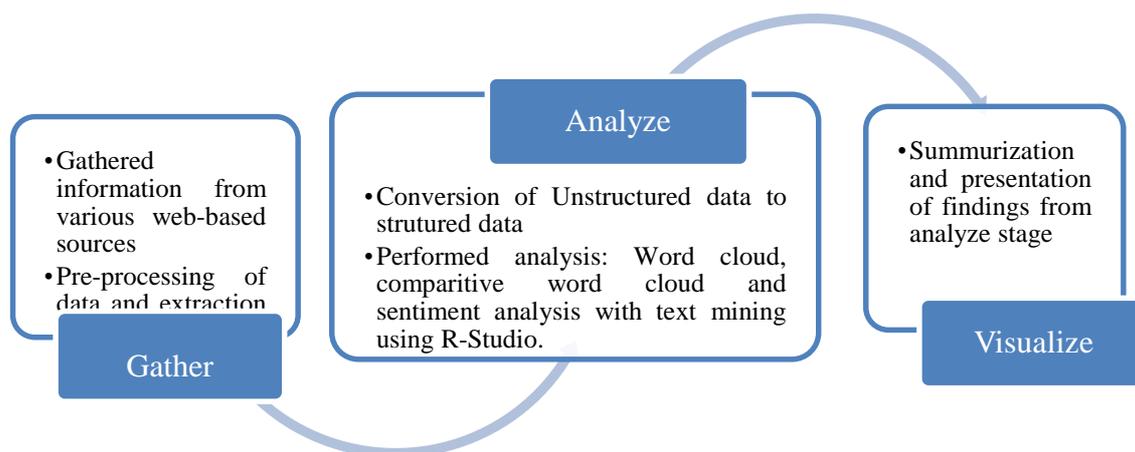

Figure 1. Stages of Web-based Data Mining or Analytics

OBJECTIVES OF STUDY

The main objectives of the current study are as follows:

in the form of diagrams or clouds. Through text clustering, the study provides a topic-oriented analysis technique consisting of related word groups, which strengthens the

data analysis method based on word frequency.

Process of text mining
Text Mining is the procedure of changing unshaped data to shaped data for communicating important information. In this study, R software was used to do data Text Mining. Given next, are the steps used in R packages.

1. Read files using read Lines.
2. Split line to words based on " ".
3. Unlist a single vector of words.
4. Clear vectors of words.
5. Remove numbers and digits.
6. Remove punctuations.
7. Remove white space.
8. Remove special characters.
9. Remove empty vectors.
10. Convert vector of words to a data frame.
11. Clear data frame.
12. Remove stop word.
13. Remove sparse words.
14. Convert vector of words to data frame.
15. Perform operations as required in the study.
16. Plot the desired graph.

Text mining techniques
The various Text Mining techniques used in the current study are as follows:

Word cloud works straightforwardly, i.e., the more the word is in textual data, the bigger it will appear in the word cloud. It is the collection of words depicted in different sizes. The primary purpose of the word cloud is to depict the maximum number of times a word is used in a given text by highlighting it. Text Mining shows what the given text focuses on and highlights are presented in the text. It visualizes the text speech and the tweets that show the most focused terms in the tweets.

Comparison cloud also works like the word cloud and allows to study the difference and similarities between two or more text files or speeches by plotting word cloud of each against one another. It not only plots different word clouds together but plot differences between word usage in the speeches. Commonality cloud complements the comparison cloud and visualizes the common words shown in all documents and their combined frequency across documents. It helps in depicting the conceptual overlap between all the documents.

Sentiment Analysis is also known as Opinion Mining and an inherent language cleaning technique pre-owned to decide whether the data is positive or negative, or neutral. It is frequently performed upon textual data for various purposes like product or service sentiment in customer feedback, recognizing customer wants, evaluating brand performance, etc. Sentiment Analysis assists in processing a large amount of data in a structured and worthwhile way. It can detect complicated situations in real-time, like the crisis of escalating or not so escalating Public Relations (PR) on social media. Models of Sentiment Analysis can assist in instantly recognizing these types of situations so that we can proceed with further corrective actions. So, in the present study, we have also used the Sentiment Analysis to know government sentiments for the last five budgets of the Indian economy through the frequency of keys mainly used to bring the overall effectiveness of the budget.

Thus, as discussed, we have used these primary forms of visualization to display the present study results. We have used tables and charts like histogram and bar chart diagram to display the study outcomes effectively.

ANALYSIS AND INTERPRETATION
Word cloud analysis
A word cloud is a pictorial representation of useful information extracted from any data/document file. The more a word is used in a file, the bigger it will be in the word cloud. The size of words in the word cloud reflects the number of times it is used in a file. It is a technique to extract powerful words from textual data or blogs, or any database. Word clouds are executed to compare two different files/documents. It can also be applied to find similarities between two or more documents. A word cloud can be used to extract information from critical data. Word clouds can be made based on calculation rather than appearances. There are three ways to categorize data into the word cloud, namely, frequency, categorization, and mixed.

Individual Frequency Word Cloud and Word Cloud2
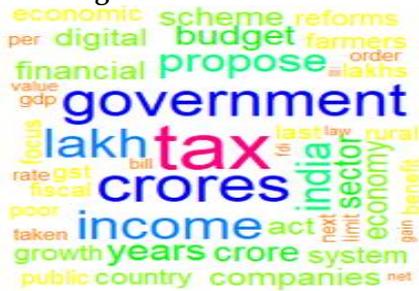
- **Budget 2017-2018**

Figure 2: Word Cloud of Budget 2017-2018

Figure 3. Word Cloud2 of Budget 2017-2018

- Finance Minister Arun Jaitley presented the Union Budget of 2017. The primary thing that impacted the economy during that period was demonetization. It was a bold and decisive measure to lead to higher GDP growth and help the economy grow.

- Figures 1 and 3 depicted that in the budget, the importance given to the agriculture sector and farmers assured the protection against losses due to natural calamities. Funds of NABARD also increased in the budget by 40,000 crores. Estimate of taking around 5 lakh farm ponds under the MGNREGA scheme. Schemes like Pradhan Mantri Gram Sadak Yojana, MGNREGA, and many others were launched. The government also focused on uplifting "poor people" out of poverty. Focus on infrastructure was also given for the overall growth of the country. Many Foreign Direct Investment (FDI) policy reforms were also done in the budget for inviting foreign investors, which further increased the country's GDP. Digital India *Abhiyaan*/program was also introduced by unleashing the BHIM UPI application for cash-less transactions. Moreover, several changes were also made in taxation policies. Tax rates of companies and personal income were also altered.

- **Budget 2018-2019**

Figure 4. Word Cloud of Budget 2018-2019

Figure 5. Word Cloud 2 of Budget 2018-2019

- Figures 4 and 5 showed that the government proposed the GDP growth from 6.3% to 7.2–7.5% in the second quarter of the year 2017. The investment of 14.34 lakh crores is planned to be spent on creating infrastructure in rural areas for the year 2017 also. A total of 1,200 crores is allocated for health and wellness centers. National Health Protection Scheme was also launched for low-income families. The total funds of Rs 5.97 lakh crores were also allocated for infrastructure development which would help in the growth of the economy. For the education sector, the government proposed to set up twenty-four new government medical colleges and hospitals. More focus was given to providing digital facilities in the educational sector. Approval of making 35,000-kilometer roads was given in the budget as the development of roads is also necessary for the economy's growth. United India Assurance Co. Ltd, National Insurance Co. Ltd. and Oriental India Insurance Co. Ltd. proposed a merger into a single insurance entity. Many tax policies changed in favor of the Ministry of Micro, Small and Medium Enterprises

(MSME) companies. MSME companies contribute a large amount of proportion to the growth of the economy. Pradhan Mantri Yojana was also extended till 2020. The government proposed a 10% tax on equity-oriented mutual funds' allocated income.

- **Budget 2019-2020**

Figure 6. Word Cloud of Budget 2019-2020

Figure 7. Word Cloud 2 of Budget 2019-2020

- Figures 6 and 7 depicted that the main focus of budget 2019–2020 was on taxation. The proposed GDP was estimated to be 12% in 2019–2020. Many changes were proposed in this budget regarding the SEBI Act, CGST Act, RBI Act, and PMLA Act. The surcharge was also increased to 25% for people/individuals having earnings between 2 crores and 5 crores, and for people having earnings above Rs 5 crores were increased to 37%. Many exemptions were added for individuals to use while calculating their income. Duty on customs was increased from 10% to 12.5% on gold and precious metals. Public sector banks provided Rs 70,000 crores for recapitalization purposes. The government planned to increase

borrowing through foreign currency. Focus on developing infrastructure was also done by launching Bharatmala project Phase 2 to develop state highways. Insurance intermediaries were permitted with 100% FDI. Investment by foreign investors in Public Sector Undertakings (PSUs) were also increased. Moreover, Pradhan Mantri Gram Sadak Yojana, Matsya Sampada Yojana, Karam Yogi Maandhan Scheme and MUDRA Scheme were also introduced.

- **Budget 2020-2021**

Figure 8. Word Cloud of Budget 2020-2021

Figure 9. Word Cloud 2 of Budget 2020-2021

- Figures 8 and 9 showed that a new tax regime was announced and people were given a choice to pay taxes according to the new rates or old rates. Over 70 deductions were removed from the tax deductions. The government also proposed doubling farmers' income by 2022. Women were given a chance to run the village storage scheme. Further, the allocated budget of Rs 69,000 crores was granted to the health sector. The government projected to set up new hospitals. The transport infrastructure section was also highlighted in the budget. Many schemes were introduced to facilitate the overall development and growth of the economy. The theme of this budget was to provide a clean, sound, corruption-free financial sector as well as policy-driven governance. **Budget 2021-2022**

Figure 10. Word Cloud of Budget 2021-2022

Figure 11. Word Cloud 2 of Budget 2021-2022

Figures 10 and 11 showed that when the budget *2021–2022* was declared, COVID-19 had already impacted the economy in very negative terms. This budget was prepared in keeping in mind the growth of the economy. The main motive of this budget was to raise the GDP of the economy. The government did not propose any new taxation policies in this budget. The government exempted pension earnings and interest income received by senior citizens. The government made an excessive investment in technology and took one more step towards Digital India. The government allocated Rs 35,000 crores for the COVID-19 vaccine. To be supplied in a non-public-public partnership mode, seven port tasks plan to guide Indian shipping costs with subsidy. The government also allocated Rs 1,000 crores to the Solar Energy Corporation of India. Also, FDI in the insurance sector was increased from 49% to 74% in coverage companies. The government projected recapitalization of public sector banks by allotting Rs 20,000 crores. Refunds increased to relieve the burden of small taxpayers. The dividend was exempted from Tax Deduction Scheme (TDS). The government was investing in the manufacturing of solar panels, cells. "Make in India" was promoted by the

government after introducing many schemes in favor of small manufacturers.

Figure 12. Comparative Word Cloud

Figure 13. Common Cloud of 5 Budgets

After concluding the clouds (like the individual and highest frequency clouds), we have also visualized the comparative cloud (Figures 12 and 13) of five different year budgets. Comparative cloud highlights the most used words with a comparison to a different budget.

Sentiment Analysis
Performing SA through web search is not difficult. The entire method can be alienated into the subsequent phases (Figure 14). Figure 14 helps in thoughtful how we can complete a simple SA for web-based data with the support of the R programming language. Firstly, to perform Sentiment Analysis, data is retrieved from the official website of the Indian budget for the years *2017–2018, 2018–2019, 2019–2020, 2020–2021*, and *2021–2022*. The next step is to fetch web searched data into the R console. In Sentiment Analysis, opinion search and retrieval functions are performed to know the sentiments of the government for the last five-year budgets of the Indian economy. In opinion search, two key tasks were accomplished, similar to regular web search: obtaining suitable documents to the requester and ranking the

retrieved sentences. It is generally performed on textual data to find the intentions behind the formation of data. Positive and negative sentiments are detected by doing Sentiment Analysis on any data. It worked for monitoring and understanding sentiments. With the help of Sentiment Analysis, we can find whether data is very positive, optimistic, neutral, negative, and very negative.

In this study, the sentimental analysis is done on the budget speeches to study the emotions of the government. Data for sentiment analysis was distributed in categories like anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative and positive.

decision regarding the demonetization of money. The announcement that the world's fastest-growing major economy has released a budget that will please everybody except politicians who have been defending "anger" sentiment. The administration made a historic action on November 8, 2016, with far-reaching economic ramifications. With immediate effect, the two highest denomination notes, Rs 500 and Rs 1,000, were "demonetized," meaning they were no longer legal money save for a few specific purposes.

2. *Anticipation*: This emotion shows the excitement level of people. The percentage

Table 1. Sentiment Analysis

Year	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise	Trust	Negative	Positive
2017	76	284	33	154	191	145	64	474	332	730
	3.06%	11.44%	1.33%	6.20%	7.69%	5.84%	2.58%	19.09%	13.37%	29.40%
2018	53	239	40	175	189	97	54	451	307	734
	2.27%	10.22%	1.71%	7.48%	8.08%	4.15%	2.31%	19.28%	13.13%	31.38%
2019	49	264	30	134	182	114	66	436	341	766
	2.06%	11.08%	1.26%	5.63%	7.64%	4.79%	2.77%	18.30%	14.32%	32.16%
2020	53	267	27	177	230	168	83	571	396	938.
	1.82%	9.18%	0.93%	6.08%	7.90%	5.77%	2.85%	19.62%	13.61%	32.23%
2021	49	226	26	111	164	125	55	426	317	644
	2.29%	10.55%	1.21%	5.18%	7.65%	5.83%	2.57%	19.88%	14.79%	30.05%
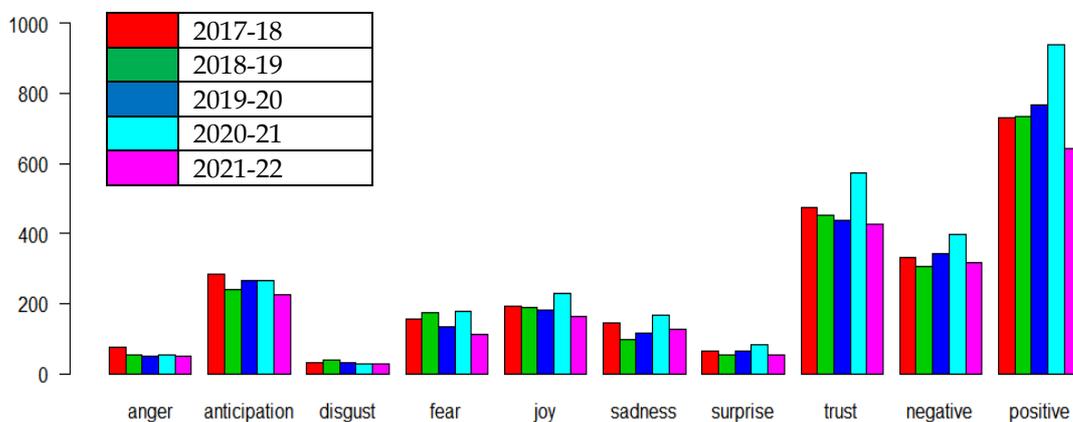
Figure 14. Sentiment Analysis Bar Chart

After the extraction of Sentiment Analysis from Table 1, the bar chart was plotted as shown in Figure 14. The Sentiment Analysis draws results based on ten different emotions. Thus, in this budget, the following conclusions were drawn:

1. *Anger:* Anger is one of the pervasive and dangerous emotions. The maximum percentage of anger appears in budget *2017–2018*. It was due to the government's

of anticipation in all the five budgets was between 9% and 12%. The discussed emotion reminds us that, as it had happened after the two World Wars, there are signs that the political, economic, and strategic relations in the post COVID world are changing. This moment in history is the dawn of a new era in which India is well-poised to be the land of promise and hope.

141

3. *Disgust*: The emotion of disgust communicates the role of judgment and habits for making a choice. The disgusted emotion causes to reject things and it is significantly less in all the five budgets. Its percentage lies between 0.5 and 2. The reason for there being less percentage of "disgust" emotion was because of more spending on infrastructure, tax relief, encouragement to startups, etc.

4. *Fear*: The emotion of fear is one of the most common emotions. The financial years *2018–2019* have the most significant fear because this budget was implemented after demonetization, and people feared an increase in tax rates.

5. *Joy:* The emotion of joy communicates the level of happiness in the government and people's minds. Years *2018–2019* budget speech contains the most significant percentage of joy and this was because black money was surfaced due to the positive impact of demonetization.

6. *Sadness*: The sadness was visible in the *2021–2022* budget. It was due to COVID-19, which impacted the economy severely.

7. *Surprise*: The emotion of surprise occurs due to sudden or unexpected decisions. The surprise percentage was around 2 in every budget taken in the study. Moreover, there was not much surprise in all the other budgets.

8. *Trust*: The Indian government wanted to build trust in people's minds in every budget, and the percentage of trust was around 19-20 in all budgets. It is because the government tries to make a budget in favor of people from every sector.

9. *Negative/Positive emotions:* Negative emotion is depicted by showing the number of negative terms employed in the budget and positive emotion shows the positivity of words used by the government in the same budget. More positive words appeared in all the speeches as compared to negative words. The government focused on tax rate slab, infrastructure, education, and health facilities, which positively impacted people's mindsets.

FINDINGS

The following are the findings of the study:

- Based on word cloud, the government focuses on the following words (Table 2):

Table 2. Main Words of Each Budget

2017	2018	2019	2020	2021
Tax, Digital India	Tax, Infrastructure, Health, Education sector	Tax, Investment, Financial Sector	Tax, Income	Infrastructure, Health, Tax, Duty, Financial sector

- Based on Sentiment Analysis, the following sentiments depicts which year budgets show a high level of those sentiments (Table 3):

Table 3. Highest Level of Emotion in Budgets

ANGER	2017
ANTICIPATION	2017
DISGUST	2018
FEAR	2020
JOY	2020
SADNESS	2020
SURPRISE	2020
TRUST	2020
NEGATIVE	2020
POSITIVE	2020

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

The Text Mining analysis, Budget Analysis becomes one of the more manageable tasks. Using text analytics, it becomes easier to communicate and compare the different budgets together. Like in the word cloud, it was depicted that the government's main focus was on "Digital India", where the government took the step of demonetization. Likewise, word cloud depicts all-year budgets highlight.

The comparative cloud shows the comparison between all-year budgets. For example, the 2017 budget focused on the conversion of black money as visualized in the speech of the then Finance Minister i.e., Arun Jaitley. He stated that the Union Budget 2017 demonstrates the government's commitment to combating black money, corruption, and other issues. He also mentioned that in the 2017 budget, nationalistic measures will be

avoided in lieu of investment activities with a multiplier effect. Rather than raising tax rates, the finance minister has attempted to raise surplus money via greater tax compliance. The study analysis also showed that the 2018 budget focused on different sectors of the economy simultaneously. Farmers, rural population, youth, digital economy, public service, financial sector, poor and underprivileged, cautious fiscal management, infrastructure and tax administration were among the 10 main topics mentioned by the finance minister in his Budget statement on February 1, 2017. Expect the *2018-19* budget to place a strong emphasis on these topics, as well as some new ones. The 2019 budget focused on the financial sector more like encouraging startups. The Narendra Modi government announcement that "Union Budget *2019-20*: Startups want angel tax to be abolished, corporate tax to be reduced, make capital more accessible" is also depicted that the main discussion of 2019 concentrated on the acceleration of the momentum on technological innovation, and focus on training and skilling initiatives for entrepreneurs, and all of which will cement India's position as a hotbed for startups. The 2020 budget focused on tax rates like changes in the tax rate slabs and the 2021 main budget focus is on "Aatmanirbhar Bharat". The conclusion is that the government tries to bring something new in every budget and focus more on tax rates.

Thus, from this discussion, it is clear that the Sentiment Analysis nowadays becomes one of the significant analyses to check and communicate the emotions or opinions of the public towards anything. As the current study concluded inclusive sentiments of the Indian budget for the years *2017–2018, 2018–2019, 2019–2020, 2020–2021*, and *2021–2022*. Furthermore, the government needs to consider the sentiments reflected by the public for the budgets. Future research can incorporate the public Sentiment Analysis for the Indian budget. Further, it has been concluded from the study that the emotion of anger was high in the 2017 budget because of demonetization, while other sentiments or emotions were high for the 2020 budget because of change in the tax rate slabs. Thus, public opinions are also crucial for the successful communication and implementation of the budget. It is also

recommended that the government focus on all sectors equally so that a comparison of budgets from different years may be performed easily. Moreover, it is critical to incorporate Sentiment Analysis as part of the budget implementation process for success, as it is significant to understand and communicate the public's feelings about the budget to policymakers.

REFERENCES

Aggarwal, C. C., & Zhai, C. (Eds.). (2012). Mining text data. *Springer Science and Business Media.*

Aggarwal, Charu C., & Cheng Xiang Zhai. (2012). An introduction to text mining. In *Mining text data*, Springer, Boston, MA, 1-10.

Arya, K. K., & Verma, M. (2021). Contemporary 143ollywood's tryst with Bourgeois romances of hindi cinema. *Journal of Content, Community & Communication*. 13(1), 308-315.

Dey, L., & Haque, S. M. (2009). Opinion mining from noisy text data. *International Journal on Document Analysis and Recognition (IJDAR)*, 12(3), 205-226.

Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). Large-scale sentiment analysis for news and blogs. *ICWSM*, 7(21), 219-222.

Guha, S., Joshi, A., & Varma, V. (2015). Sentibase: Sentiment analysis in twitter on a budget. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, 590-594.

Jain, V. (2013). Prediction of movie success using sentiment analysis of tweets. *The International Journal of Soft Computing and Software Engineering*, 3(3), 308-313.

Kaur, R., Kaur, R., Singh, M., & Ranjan, S. (2020). Twitter sentiment analysis of the Indian union budget.

Kaurav, R. P. S., Narula, S., Baber, R., & Tiwari, P. (2021). Theoretical extensions of the New Education Policy 2020 using Twitter mining. *Journal of Content, Community & Communication*. 13(1), 16-26.

Kaurav, R. P. S., Suresh, K. G., Narula, S., & Baber, R. (2020). New education policy: Qualitative (contents) analysis and Twitter mining (sentiment analysis). *Journal of Content, Community and Communication*, 12(1), 4-13.

Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.

Miner, G., Elder IV, J., Fast, A., Hill, T., Nisbet, R., & Delen, D. (2012). Practical text mining and statistical analysis for non-structured text data applications. *Academic Press*.

O'Connor, B., Bala Subramanyan, R., Routledge, B., & Smith, N. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth International AAAI Conference on Web and Social Media,* 4(1).

O'Connor, Brendan, Ramnath Bala Subramanyan, Bryan R. Routledge, & Noah A. Smith. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth International AAAI Conference on Weblogs and Social Media*.

Park W J & Yu K Y. (2015). Spatial clustering analysis based on text mining of location-based social media data. *Journal of the Korean Society for Geospatial Information Science*, 23(2), 89-96.

Proksch, S. O., Lowe, W., Wäckerle, J., & Soroka, S. (2019). Multilingual sentiment analysis: A new approach to measuring conflict in legislative speeches. *Legislative Studies Quarterly*, 44(1), 97-131.

Sharma, R., & Gupta, S. (2021). Bharat towards Atmanirbharta: A Twitter based analysis using NVIVO. *Journal of Content, Community and Communication*, 58-65.

Shakeel, M., & Karwal, V. (2016). Lexicon-based sentiment analysis of Indian union budget 2016–17. In *2016, International Conference on Signal Processing and Communication (ICSC)*, 299-302, IEEE.

Singh, P., Sawhney, R. S., & Kahlon, K. S. (2018). Sentiment analysis of demonetization of 500 and 1000-rupee banknotes by Indian government. *ICT Express*, 4(3), 124-129.

Syaifudin, Y. W., & Puspitasari, D. (2017). Twitter data mining for sentiment analysis on people's feedback against government public policy. *MATTER: International Journal of Science and Technology*, 3(1), 110-122.

https://www.rbi.org.in/Scripts/Publications View.aspx?id=20359

***