



e-ISSN:2582-7219



# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 6, June 2024



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 7.521



6381 907 438



6381 907 438



ijmrset@gmail.com



www.ijmrset.com



# Sentiment Analysis of Indian Elections 2024

Mr. Rohan Suram<sup>1</sup>, Mrs. Chetna Achar<sup>2</sup>

Student, Institute of Computer Science, Mumbai Educational Trust- MET ICS, Mumbai, India<sup>1</sup>

Professor, Institute of Computer Science, Mumbai Educational Trust- MET ICS, Mumbai, India<sup>2</sup>

**ABSTRACT:** this research paper aims to explore the sentiments of Indian voters and individuals interested in Indian elections, as expressed in the comments section of YouTube. The study delves into the requirements, emotions, needs, and challenges voiced by individuals on this subject matter.

**KEYWORDS:** indian elections, india, sentiment analysis, youtube comments, genz

## I. INTRODUCTION

### A. Indian Voters Statistics

In the upcoming Indian election, 986 million individuals are expected to vote, with around 65% of them under 35 and 50% under 25. Currently, 12% of Members of Parliament are below 40. Understanding young voters is crucial as they face unique challenges, such as the impact of the global pandemic on their education, climate anxiety, and a volatile economy. As digital natives, they rely heavily on the internet for information.

Young voters in India are increasingly politically active, including a significant number of first-time voters. Political parties are targeting this demographic via social media, addressing issues like unemployment, poverty, and inflation. This trend mirrors global examples, such as the 2020 US election and Indonesia's recent elections, where young voters played crucial roles.

In India, 77% of Gen Z find voting attractive, with 41% deeming it important. About 46% believe political affiliations are crucial, and 44% value open-mindedness, with 15% unwilling to date someone who doesn't vote. The number of new voters aged 18-19 is expected to rise from 15 million in 2019 to 18 million in 2024.

Digital platforms are vital for political updates among young voters: 40% use Instagram, 26% YouTube, 17% Twitter, and 12% Facebook. Political parties are thus investing in digital campaigns, influencer collaborations, and creating engaging content like reels and podcasts. Young voters consider various factors in their decision-making: 19% focus on party-related issues, 15% on leadership, 10% on family opinions, 14% on youth candidates, and 10% on community or caste. Historically, youth support for political parties has shifted, with the BJP seeing a significant increase in youth votes from 2014 to 2019, while Congress has struggled to maintain youth support.

Primary concerns for young voters include unemployment, poverty, inflation, corruption, and overpopulation. Despite being politically engaged, challenges like voter registration and apathy, especially in urban areas, remain. The upcoming election will test whether the youth's political opinions translate into active participation.

### B. YouTube Statistics

According to the YouTube's potential ad reach totals 2.49 billion users. Users aged between 25 and 34 account for 21.3% of YouTube's user base, making them the largest age group to use the platform. The second largest age group is between ages 35 and 44. Additionally, YouTube is popular among young adults aged between 18 and 24, with this group making up 15.5% of the platform's user base. There are approximately 462 million active YouTube users in India, making it the network's largest audience by country. The United States comes next with 239 million users while 144 million users are from Brazil. [2]

Now the above stats perfectly match with our target audience, as we can conclude that Indians aged from 18-44 does use YouTube and does put their thoughts and remarks in the comments section, which could be an ideal situation to carry out sentiment analysis on it.



### C. Sentiment Analysis Tool

Sentiment analysis, also referred to as opinion mining, is a field that focuses on extracting judgments, reactions, and emotions from text. It finds wide application in areas such as data mining, web mining, and social media analytics, as emotions play a crucial role in evaluating human actions. The main aim of sentiment analysis is to comprehend the opinions of consumers or audiences regarding a specific subject by analysing extensive textual data from various sources. This process involves the utilization of statistical methods, natural language processing (NLP), and machine learning techniques to identify and extract subjective information from text. This chapter provides an overview of sentiment analysis and outlines the research problem associated with this study.

Natural Language Processing (NLP) serves as a bridge between human language and computer language, enabling computers to understand, analyse, and respond to human language inputs. NLP techniques empower machines to grasp the structure, meaning, and context of textual or verbal data. One aspect of NLP involves discerning the sentiment or emotional tone conveyed in text, such as determining whether a user comment is positive, negative, or neutral. While traditional programming involves providing specific instructions to a computer, machine learning takes a different approach. Instead of explicit instructions, machine learning algorithms are fed with data, enabling them to learn patterns and relationships within the data to make predictions or decisions.

This study explores the application of sentiment analysis to YouTube comments on Indian Elections news videos, utilizing TextBlob. By leveraging TextBlob, this research aims to shed light on the strengths and limitations of each approach in capturing subtle nuances of sentiment within the dataset.

## II. METHODOLOGY

This paper aims to get the insights on comments posted by users on various videos on YouTube. These comments on each videos have been collected separately and in an unbiased manner. The methodology is as follows:

### A. Data Collection

The YouTube comments dataset has been created by taking 128 videos of titles which talk about BJP, Congress and other party/election related in roughly equal amount. And around 11485 comments have been collected.

#### i. Data:

##### 1. Title related

BJP	44
Congress	43
Any other/neutral	41

Table 1. Frequencies of news titles with respective to particular parties

##### 2. Comments related:

There are about 11485 comments collected, after filtering, removing values with string length less than 3 and removing duplicate values and emoji texts the final count of comments are 9118. The below is the table of counts comments and its association with party.

BJP	3045
Congress	1872
Any other/neutral	4201

Table 2. Frequencies of comments with respective to particular titles

Table 1 and Table 2 have been derived after performing on EDA on the data.



ii. Columns

Attribute information is as follows:

title : title of youtube videos

text : comments of youtube\_videos

text\_status : comments association with particular party title\_status : title association with particular party sentiment\_text

: sentiment score of comments sentiment\_title : sentiment score of title

polarity\_text : converting text sentiment score into positive, negative or neutral polarity\_title : converting title sentiment score into positive, negative or neutral

B. Data Pre-processing

The dataset will be pre-processed to clean and prepare it for analysis. This will involve removing irrelevant information, handling missing values, and converting the text data into a suitable format for sentiment analysis.

Now, To Pre-process the data, the following steps were taken:

- └ Remove <br> tags from text column
- └ Remove links in the text column
- └ Translating the title column and text column into English using google translate
- └ Remove rows from text which has word length of less than 3 words
- └ Remove emoji's from both title as well as text column
- └ Remove duplicate values from text column

Fig.1 shows the distribution of length of words in 9118 comments.

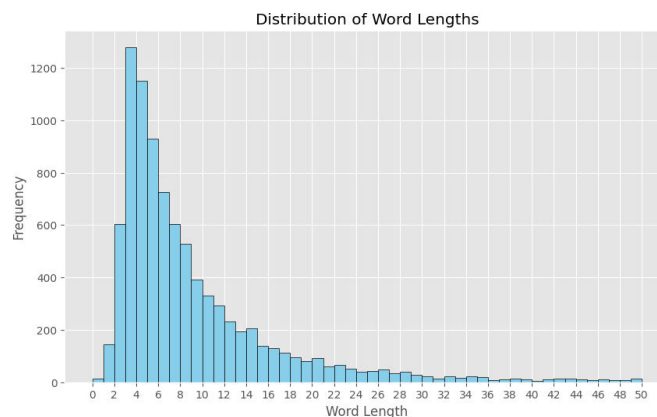


Fig. 1 distribution of word lengths in text data.

There are few procedures which need to be done before we dive into finding insights. Dividing and creating new columns which might help in plotting would help in further analysis of our data.

Below are the few columns derived from the existing columns:

- └ Before we associate text columns to either BJP, Congress or any other/neutral we first need to find all the proper noun and common nouns.
- └ Once all the proper and common noun have been collected, map them individually and manually, whether which word could be associate with which party example:
- └ 'BJP', 'modiji', 'jaishankar', 'yogi', 'pm' with BJP
- └ 'Congress', 'rahul', 'sonia', 'gandhi', 'opposition' with Congress
- └ So the count of BJP words should be taken into BJP\_count and for Congress should be taken into Congress\_count column.
- └ So in new column (new\_count) will contain the difference between (BJP\_count) and (Congress\_count)
- └  $(new\_count) = (BJP\_count) - (Congress\_count)$

There would be 4 condition on the values of (new\_count) which would decide the values for (text\_status), they are as follows:





(new_count)	(text_status)
Less than zero	Congress
Equal to zero	any other/ neutral
More than zero	BJP
NA value	any other/ neutral

Table 3. conditions of mapping comments to particular party

Do the same for title also, map them and enter the result into column (title\_status).

(new_count)	(title_status)
Less than zero	Congress
Equal to zero	any other/ neutral
More than zero	BJP
NA value	any other/ neutral

Table 4. conditions of mapping titles to a particular party

*D. Sentiment Analysis*

For sentiment analysis, the tool used is textblob, textBlob is a Python library for processing textual data, offering a straightforward interface for tasks like sentiment analysis, part-of-speech tagging, and more. With its simplicity and ease of integration, TextBlob enables researchers to analyse sentiments in text effortlessly. Its sentiment analysis module evaluates polarity and subjectivity, providing insights into the emotional tone of text. Additionally, TextBlob offers customizable lexicons, enhancing its adaptability to different contexts and languages. Overall, TextBlob serves as a versatile tool for NLP tasks, facilitating the exploration of sentiments in textual data across various domains with efficiency and accuracy.

We carry out sentiment analysis on text as well as title, the scores of which are stored in sentiment\_text and sentiment\_title column respectively.

Once the scores are stored, we would map them into positive,negative or neutral into column polarity\_text and polarity\_title from sentiment\_text and sentiment\_titles respectively.

The conditions for mapping for both are as follows :

sentiment_text/sentiment_title	polarity_text/polarity_title
-1 < s < 0	Negative
s == 0	Neutral
0 < s < 1	Positive

Table 5. Conditions of mapping sentiment score to polarity

Remove the stop words from (text\_column) and tokenize them into a column named (cleaned\_tokens)



*E. Plots and Insights*

There are total 9118 comments after all the filtrations  
The comments distribution based on (text\_status) is as follows:

Any other/neutral	4201	46%
BJP	3025	33%
Congress	1872	20%

Table 6. Frequency of comments titles with respective to particular party with percentage share

The comments distribution based on (title\_status) is as follows:

Any other/neutral	41	32%
BJP	44	34%
Congress	43	33%

Table 7. Frequencies of news titles with respective to particular parties with percentage share

The sentiment distribution based on text is as follows:

neutral	4617	50%
positive	3108	34%
negative	1396	15%

Table 8. sentiment distribution based on comment(text) and percentage share

The sentiment distribution based on title is as follows:

neutral	70	54%
positive	39	30%
negative	19	14%

Table 9. sentiment distribution based on title and percentage share

Let us now individually dive into News titles\_status and find insights about it.

1. News title with BJP i.e. (title\_status = BJP)

There are total of 44 news title which talk about BJP in their title

It contains about total 3140 comments of 9118Of which 60% (1905/3140) people talk neutral, which in some sense could be considered positive25% (801/3140) talk positive about BJP13% (433/3140) talk negatively about BJP.Of neutral i.e. of 1905, the word 'BJP' occurs 266 times and 'modi' comes 248 times but this does not conclude any sentiment.



Sentiment Distribution of BJP News Titles based on merged\_pol\_title\_status

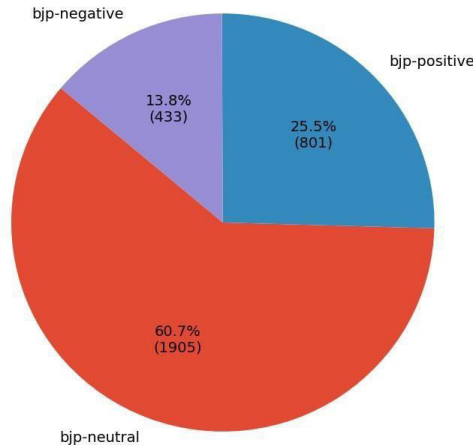


Fig. 2 pie chart of sentiment distribution of BJP News Titles

Out of 248 'BJP', 94 times 'Congress' word appears. So we can take it as neutral. The word 'vote' appears 76 times which clouds mean a little positive if word associated with BJP comes in that sentence or negative if word associated with Congress comes in that sentence. But words 'BJP' comes 27 times, Congress 13 times and 'modi' 4 times. Now, words associated with 'vote' as well as 'BJP' are 'ji', 'rajput' and 'modi', which are BJP related terms. So we can consider these 27 comments as positive too. But with Congress we can consider negative. There are words 'BJP', 'remove', 'save', 'country' appearing 7 times, which can be considered negative sentiment for BJP. So if we manually do a deep analysis on neutral data, we might be able to find more insight, but as of this research paper we are not concerned about the deep analysis.

When the title talks about BJP, let's see what people talk about in the comments.

Out of 3140, 1329 i.e. 42% if people talk about BJP, including positive, negative and neutral sentiments. Out of 3140, only 385 talks about Congress which is only 12%. And the rest i.e. 1435 which is 45% talks about neutral things. But on further analysis of each word and deep study we could be able to bifurcate some of the statements into BJP or Congress.

Sentiment Distribution of BJP News Titles based on merged\_pol\_title\_status

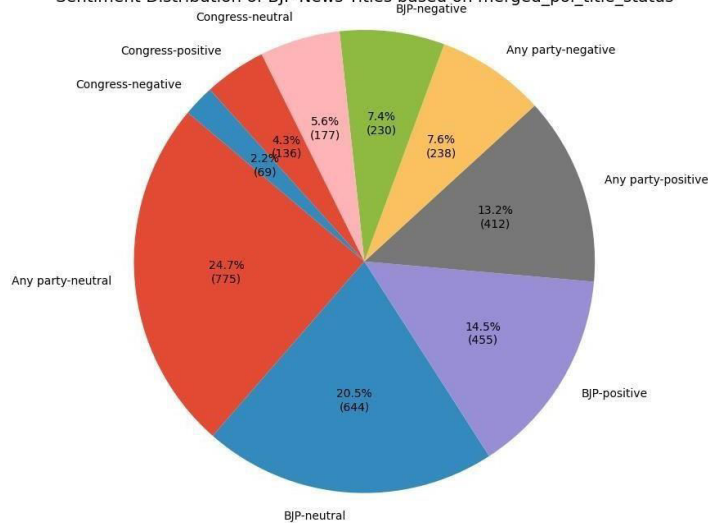


Fig. 3 pie chart of in depth sentiment distribution of BJP News Titles

Individually speaking let's see in tabular format what percentage of people's comment to what topic they are talking



BJP as News Titles			
BJP	Positive	452	1320
	Negative	208	
	Neutral	640	
Congress	Positive	137	385
	Negative	69	
	Neutral	179	
Any Party/neutral	Positive	415	1435
	Negative	242	
	Neutral	776	

Table 10. Frequencies of sentiments with respective to BJP as news titles along with respective party comments and their shares in numbers

In Percentage format:

BJP as News Titles				
News Title	Sentiment	Percentage of whole	Percentage of News Title	Percentage of Total
BJP	Positive	14%	34%	42%
	Negative	07%	17%	
	Neutral	20%	48%	
Congress	Positive	04%	35%	12%
	Negative	02%	17%	
	Neutral	05%	46%	
Any Party/neutral	Positive	13%	28%	45%
	Negative	07%	16%	
	Neutral	24%	54%	

Table 11. Percentage distribution of sentiments with respective to BJP as news titles along with respective party comments and their shares in percentage

One thing has to be kept in mind that neutral data can be further analysed on deep analysis manually.

2. News title with Congress i.e. (title\_status = Congress)

There are total of 41 news title which talk about Congress in their title. It contains about total 2852 comments of 9118 Of which 50%(1408/2852) people talk neutral, which in some sense could be considered positive 35%(1010/2852) talk positive about Congress





15.2 % ( 433/2852) talk negatively about Congress. Speaking of neutral comments in Congress. The word ‘Congress’ appears 336 times, which is further associated with ‘party’ 44 times and ‘vote’ 43 times. There can be no conclusion from the word ‘party’, so let’s ignore it and dive into the word ‘vote’ which can make a positive or negative sense.

Sentiment Distribution of Congress News Titles based on merged\_pol\_title\_status

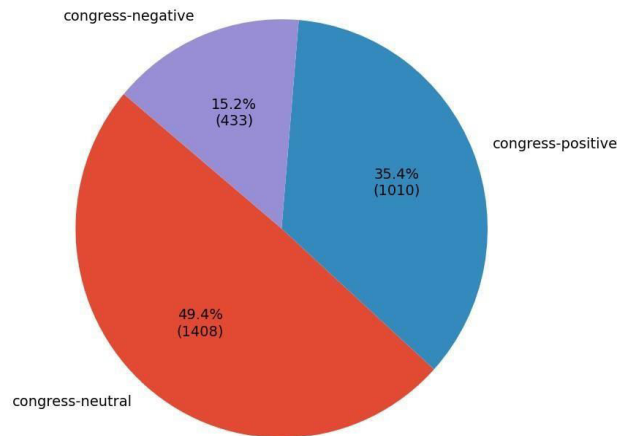


Fig. 4 Pie chart of sentiment distribution of Congress News Titles

But the word ‘vote’ appears with ‘BJP’ 8 times and ‘counts’ 7 times. So there are sentences with vote which contains both Congress and BJP, It’s better to take them as neutral only. The second word that appears the most is ‘raahul’ which is further associated with ‘gandhi’ 99 times, ‘time’ 15 times and ‘vote’ 9 times. This vote word can lead to positive sentiment if analyzed further, The 3rd most appeared word is ‘BJP’, which appeared 175 times, which is associated with word ‘vote’ 19 times and ‘remove’ 17 times. This could boost up the positive sentiment for Congress if these ‘remove’ is directly attached to ‘BJP’ itself. Again, we can’t conclude which of the comments are positive or negative until a deep analysis of it.

When the title talks about Congress, let’s see what people talk about in the comments. Out of 2852, 968 i.e. 33% if people talk about Congress, including positive, negative and neutral sentiments. Out of 2852, only 744 talked about BJP which is only 26%. And the rest i.e. 1140 which is 39% talks about neutral things. But on further analysis of each word and deep study we could be able to bifurcate some of the statements into BJP or Congress.

Sentiment Distribution of Congress News Titles based on merged\_pol\_title\_status

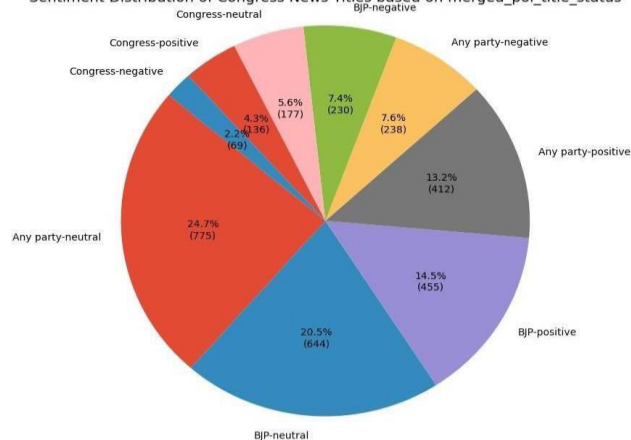


Fig. 5 pie chart of in depth sentiment distribution of Congress News Titles

Individually speaking let’s see in tabular format what percentage of people’s comment to what topic they are talking



Congress News Titles			
BJP	Positive	253	744
	Negative	139	
	Neutral	352	
Congress	Positive	384	970
	Negative	125	
	Neutral	461	
Congress News Titles			
BJP	Positive	253	744
	Negative	139	
	Neutral	352	
Any Party/neutral	Positive	382	1134
	Negative	182	
	Neutral	570	

Table 12. Frequencies of sentiments with respective to Congress as news titles along with respective party comments and their shares in numbers

In Percentage format

Congress as News Titles				
NewsTitle	Sentiment	Percentage of whole	Percentage of News Title	Percentage of Total
BJP	Positive	08%	34%	42%
	Negative	04%	18%	
	Neutral	12%	47%	
Congress	Positive	13%	39%	12%
	Negative	04%	12%	
	Neutral	16%	47%	
Any Party/neutral	Positive	13%	33%	45%
	Negative	06%	16%	
	Neutral	20%	50%	

Table 13. Percentage distribution of sentiments with respective to BJP as news titles along with respective party comments and their shares in percentage



3. News title with any party/neutral i.e. (title\_status = neutral)

There are total of 41 news title which talk about Congress in their title 4201 talks about neutrality let's see how many with title neutrality has been. Of which 30.6% (956/3126) people talk positivity, which in some sense could be considered positive. 10.2% (332/3126) talk negatively about any other/neutral. 58.8% (588/3126) talk neutral about any other/neutral. Speaking of neutral comments in any other/neutral. The word 'Congress' appears 336 times, which is further associated with 'party' 44 times and 'vote' 43 times. There can be no conclusion from the word 'party', so let's ignore it and dive into the word 'vote' which can make a positive or negative sense.

Sentiment Distribution of any party News Titles based on merged\_pol\_title\_status

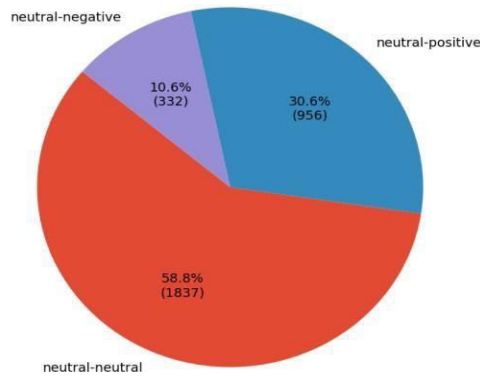


Fig. 6 Pie chart of sentiment distribution of any party News Titles

Let have word association for all three sentiments in neutral titles.

In the analysis of word appearances and their frequencies, the positive sentiment category reveals that "Modi" appears 266 times, "Long" 180 times, "India" 180 times, "Live" 173 times, and "BJP" 156 times. Notably, two of the top five words are associated with the BJP party, with "live" and "long" frequently appearing alongside "India" 79 and 74 times respectively. In the negative sentiment category, the word "People" appears 71 times, "Modi" 63 times, "BJP" 63 times, "Vote" 57 times, and "India" 45 times. Here, two of the top five words are again linked to the BJP party, but in a negative context. Lastly, in the neutral sentiment category, "BJP" appears 137 times, "India" 96 times, and "Vote" 84 times, with "BJP" being associated with "vote" and "remove" 17 and 9 times respectively. This comprehensive analysis highlights the recurring association of BJP-related terms across varying sentiments.

When the title talks about neutral, let's see what people talk about in the comments. Out of 3126, 1636 i.e. 52% of people talk about neutrality, including positive, negative and neutral sentiments. Out of 3126, only 518 talked about Congress which is only 16%.

And the rest i.e. 981 which is 31% talks about BJP. But on further analysis of each word and deep study we could be able to bifurcate some of the statements into BJP or Congress, in the neutral section.

Sentiment Distribution of neutral News Titles based on merged\_pol\_title\_status

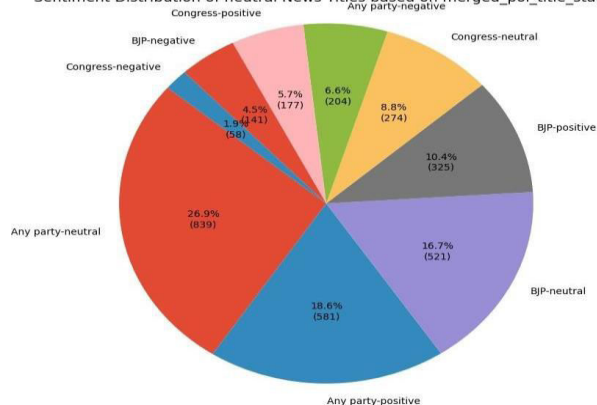


Fig. 7 pie chart of in depth sentiment distribution of neutral news Titles



Individually speaking let's see in tabular format what percentage of people's comment to what topic they are talking

Neutral/any party/other as News Titles			
BJP	Positive	323	981
	Negative	142	
	Neutral	516	
Congress	Positive	182	518
	Negative	60	
	Neutral	2561	
Any Party/neutral	Positive	583	1636
	Negative	204	
	Neutral	849	

Table 13. Frequencies of sentiments with respective to neutral/any party/ other as news titles along with respective party comments and their shares in numbers

In Percentage format

Neutral/any party/other as News Titles				
NewsTitle	Sentiment	Percentage of whole	Percentage per News Title	Percentage of Total
BJP	Positive	10%	32%	31%
	Negative	04%	14%	
	Neutral	16%	52%	
Congress	Positive	05%	35%	16%
	Negative	01%	11%	
	Neutral	08%	53%	
Any Party/neutral	Positive	18%	35%	52%
	Negative	06%	12%	
	Neutral	27%	51%	

Table 14. Percentage distribution of sentiments with respective to any party/other/neutral as news titles along with respective party comments and their shares in percentage



### III. CONCLUSION

- With news titles as BJP it can be seen that only 12% people talk about Congress but vice versa tells a different story.
- With news titled as Congress it can be seen 26% people talk about BJP.
- With titles as BJP, people talking about positive and neutral about Congress is just 4% and 5% respectively, which sums up to 9%
- When titles as Congress, people talking about positive and neutral about positive and neutral about BJP is 8% and 12% respectively which sums up to 20%.
- In titles which are neutral, people talk about neutral things up to 52%.
- But if compared to BJP and Congress, there's a whopping 15% difference between BJP and Congress.
- The neutral comments in BJP as well as Congress titles have almost the same shares i.e. 45% for BJP and 39% for Congress.
- If compared the positive, negative and neutral comments, percentage ratio of Congress in title as well as Congress. The ratio is quite equal i.e. 14:5:18 respectively.
- Same goes for BJP, the positive negative and neutral comment percentage ratio of BJP in title named as Congress. The ratio is quite equal 10:5:14.
- If we consider , from Table 10 and Table 11, BJP-positive = BJP - positive + BJP-neutral + Congress-negative
- Then 37% people talk positively about the BJP party in YouTube titles which are termed around BJP.
- And if we do the same for BJP-negative it is 17%.i.e 17% people talk negative about BJP. which can be considered positive for Congress
- If we consider, Table 12 and Table 13 ,Congress-positive = Congress-positive + Congress-neutral + BJP-negative
- Then 34% people talk positively about the Congress party in YouTube titles which are termed around Congress.
- And if we do the same for Congress-negative it is 25%.i.e 25% people talk negative about Congress. Which can be considered positive for BJP.

### REFERENCES

- [1]Firstpost. (2024, April 7). India Elections: Who Will the Youth Vote For? | Between the Lines with Palki Sharma [Video]. Retrieved from YouTube: <https://www.youtube.com/watch?v=syB3EhpwayY>
- [2]Zote, J. (n.d.). 25 YouTube Stats Marketers Should Know in 2024 [Updated]. Retrieved from Sprout Social: <https://sproutsocial.com/insights/youtube-stats/>
- [3]Times of India. (n.d.). Polls About Future, But Young Not Interested. Retrieved from [https://timesofindia.indiatimes.com/india/polls-about-future-but-young-not-interested/articleshow/109049324.cms?utm\\_source=whatsapp&utm\\_medium=social&utm\\_campaign=TOIArticleshowicon](https://timesofindia.indiatimes.com/india/polls-about-future-but-young-not-interested/articleshow/109049324.cms?utm_source=whatsapp&utm_medium=social&utm_campaign=TOIArticleshowicon)
- [4]Khare, A., Gangwar, A., Singh, S., & Prakash, S. (n.d.). Sentiment Analysis and Sarcasm Detection in Indian General Election Tweets. In Research Advances in Intelligent Computing. Taylor & Francis. Retrieved from <https://www.taylorfrancis.com/chapters/edit/10.1201/9781003320340-20/sentiment-analysis-sarcasm-detection-indian-general-election-tweets-arpit-khare-amisha-gangwar-sudhakar-singh-shiv-prakash>
- [5]Khurana, D., Koli, A., Khatter, K., et al. (2023). Natural language processing: state of the art, current trends and challenges. Multimedia Tools and Applications, 82, 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>





INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | [ijmrset@gmail.com](mailto:ijmrset@gmail.com) |

[www.ijmrset.com](http://www.ijmrset.com)