Sentiment analysis on the Union's Budget 2023: An evaluation based on Youtube comments

Bhagyashree Chetia[1], Ruth Lalthlamuanpuii[2], Yashir Ahamed[3] and Dr Lalngaizauli[4]

[1,2,3,4] Mizoram university

Abstract

Sentiment analysis refers to the technique applied to extract a dataset from social media platforms, with the objective of examining the sentiments expressed in the comments, specifically focusing on the positive or negative elements. The research was conducted with the objective of comprehending the perspective of users on the union budget for the fiscal year 2023-24. The retrieved dataset comprises a total of 47,057 comments generated by 34,323 distinct authors across 332 diverse YouTube channels, within the specified time frame. The comments were assessed using Mozdeh software, which identified the channel that exhibited the highest average positive and average negative attitudes regarding the topic. The study has revealed that there is a greater level of interest among the general public in news channels and the substance presented in their films.

Keywords: Sentiment analysis, Web 2.0, social media, Youtube videos, Union Budget 2023-24 (India), Mozdeh software

1. Introduction

The manner of communicating and sharing information has tremendously changed due to the rapid growth of technology. Earlier methods of dissemination of information have been dominated by different communication tools. The new inventions have led communication to a new dimension where social media plays a vital role. It is a platform that facilitates communication and the dissemination of information at our fingertips. The web 2.0 is a platform that consists of many web services (Google, SNS, Blogging, etc.) and have an immense impact on each and every community. The first known social media site was Six Degree, which was created in 1997 and operated until 2001 (Jones, M, 2015). According to data from January 2023, Facebook has 2,958 million monthly active users, followed by YouTube with 2,514 million users and WhatsApp and Instagram with 2,000 million users each. As a Web 2.0 tool, YouTube provides a platform for common people to create contents and share them with viewers for their views and opinions and for more information. At present, YouTube is found to have more than 2 billion users across the globe and is the second most-used social media platform. As per the report, every minute total 500 the number of YouTube videos are being uploaded; the result from January 2023 shows that India has 467 million users, followed by the United States with 246 million users. India has over 1,200 YouTube channels which have over 1 million subscribers (Divakaran, S 2023).

Nowadays, the comments and viewers' opinions are known to be a very important factor for content creators to know the like or dislike Ness of the people as it plays an important role in the popularity of their content and the creators. With the increasing popularity and adoption by huge audiences, it has gained a huge impact as well as importance. So to make a decision or purchase regarding a service it is much more important to make an analysis; in the current scenario sentiment analysis has gained popularity and importance by executing the mining of people's opinions.

Sentiment analysis is a technique that uses Natural Language Processing (NLP) to determine whether the data being analyzed is positive, neutral, or negative. Back in the 1950s, the concept of sentiment analysis was introduced where it was done and carried out manually with the help of distributing questionnaires. Interviews of people's opinion. With the advancement of ICT and different software's, it has become easier to collect and gather dataset from different social media platform with the help of different social media applications. It also goes beyond this to detect specific moods, urgency, emotions, and intentions. We can collect material and data from websites such as Twitter, Facebook, YouTube, and others by using different software. Data such as images, videos, text, reviews, and others may be scraped (Hillier, W, 2021). With the growing trend and technologies, it has become much easier to deal with the activities related to different social media applications and their related services. Sentiment analysis provides the analysis of human emotions and feelings with different software that is available for the same purpose. It helps the user to know the trends, product analysis as well as customer services and monitoring so that it will help to identify the advantages and drawbacks of any product or service.

Union Budget

A budget is considered as tool to bring a rapid growth and balanced economic growth of a country with equality and social justice. It is referred to as annual financial statement of estimated receipts and expenditure for that particular year. The government presents it on the first day of February so that it could materialised before the beginning of new financial year in April. Union Budget 2023-2024 was presented by Union Finance Minister Nirmala Sitharaman at the Parliament, New Delhi, on 2nd February 2023. A total number of Rs 18,050 crore has been allocated by the Ministry of External Affairs, which is an increase of 4.46% over last year's allocation, which is Rs 17,250 crore.

The digital era has affected the people and their livelihood. It is now become very much easier to get the information and resources with single sign in on the to all available resources on the internet and easily access to it via

smartphones, laptops etc. Now a days, the means for communication and sharing or dissemination of information can be easily and directly done through YouTube. Many news channels have upgraded directly to watch on YouTube. It is available for audiences globally and consists of videos and live-telecast that have been watched from different parts irrespective of time and distance.

2. Literature review

Shevtsov, Alexander (2023), investigated the connection between Twitter and YouTube by analysing the URLs of videos that were posted on both platforms. They searched through close to 20 million tweets and 29,000 comments left on videos hosted on YouTube and analysed with the use of the Vader sentiment analysis model, which is part of the Python NLTK toolkit. **Deori et al. (2021)** studied the characteristics and user sentiments of the videos on YouTube about the software Koha and DSpace by using the Webometrics Analyst for data retrieval and Parallel dots for sentiment analysis. YouTube is a platform for troubleshooting the newest content for users with certain queries and the content creators of Koha and DSpace have been directed to improve the videos with creativity. **Porreca et.al (2020)** studied the Italian YouTube videos on vaccines between the years 2017 and 2018 to understand people's opinions before and after the vaccination campaign. The finding demonstrates that vaccination campaigns conducted on social networks have the potential to be a successful strategy for influencing decisions and health policy as well as combating misconceptions. **Bhuiyan at. al (2018)** experiment on 1000 videos available on YouTube and almost one million YouTube comments were analyzed using a natural language processor, which was found to be the most effective tool for finding the most relevant and popular videos. **Thelwall (2014)** analysed the sentiment and the time series analysis of Twitter corpus relating to the UK Riots of 2011 using Mozdeh software to evaluate the increasing and decreasing interest over time or any constant changes in the broad pattern. It also compared and displayed the graph of the average positive sentiment strength and the

average negative sentiment strength. **Thelwall et al. (2012)** examined the YouTube remarks separated by utilizing arbitrarily chosen 65,536 terms from a bunch of English blogs and Really Simple Syndication (RSS) channels through Webometric Analyst. Utilizing the SentiStrength program, the age, sex, area, length, and sentiments of the remarks were assessed seeing that the majority of the positive remarks were posted by 29 years of age males containing 58 characters and about 23.4% of remarks were the answers of the remarks. The most transcendent classes of recordings that got the greatest remarks were Music, Comedy, and How-to and Style.

From the literature available on the topic of sentiment analysis, it can be understood that currently, it is one of the most popular areas for study in the subject of Library and Information Science. The study is an effort to fill the gap as no studies have been carried out and was found on the sentiment analysis of the Union Budget 2023.

3.   Scope of the study

The scope of the study is limited to the viewers' opinion that is posted on YouTube videos on the topic India Union budget 2023-24. The study is an effort to fill the gap as no studies have been carried out and was found on the sentiment analysis of the comments of Union Budget 2023. It will help to provide the readers with detailed characteristics and information about it and mostly it will provide the people's opinions and thinking on the topic with its effectiveness and impact on society. Further, the study is limited to the videos and comments available from 31/01/2023 to 25/02/2023. The study is also limited to the growing trend of comments, most viewed and commented videos and channels, like and dislike counts, and the sentiments behind the comments irrespective of gender and also without any geographical location restrictions.

3.1 Objectives

1. To find out the growth trend of comments posted on different YouTube channels.
2. To find out the most viewed and commented video topic on YouTube.
3. To evaluate the average and gender-wise sentiments on the comments.
4. To evaluate the sentiment strength of comments on YouTube.

3.2 Methodology

The present study is designed to evaluate YouTube comments on the selected topic. The comments contained in the different videos posted by different YouTube channels were evaluated. The extraction of data was done using the Mozdeh software along with the analysis by entering the query "Union budget 2023-24" and "India budget 2023-24" respectively via the API V3 key. After getting the dataset the video IDs of YouTube videos were identified with the help of Webometric Analyst software and also it was verified manually on YouTube.

Mozdeh is a free software that allows the user to extract datasets from social media platforms such as YouTube, Twitter, Bing, etc. by entering the query or channel ID and also analysing the sentiments behind the tweets, YouTube comments, etc It has features like evaluation of time series, filters in words and date, frequency of communication, mine association, etc. The software is a convenient tool for sentiment analysis of social media content developed and introduced by Mike Thelwell.

Over the specified time period, 34,323 commentators posted a total of 47057 comments across 332 YouTube channels on the specified subject.

4. Interpretation and Result

4.1 Gender-wise distribution of comments

Figure 1 describes the gender variation of the commentators on the videos. The Mozdeh software identifies the viewers of the comments as male, female, and none or unspecified where the user does not want to disclose their gender details. A total of 34324 users commented on the videos, of which 834 were men and 826 were women. It is noticeable that out of the total user's highest number are not interested in revealing their gender identity. However, the male views have the domination over the female views on the topic of discussion. The surge in anonymous users may be due to Google offering users the choice of four genders: male, female, custom, and rather not to say. Research also reveals that the majority of users do not wish to reveal their gender.
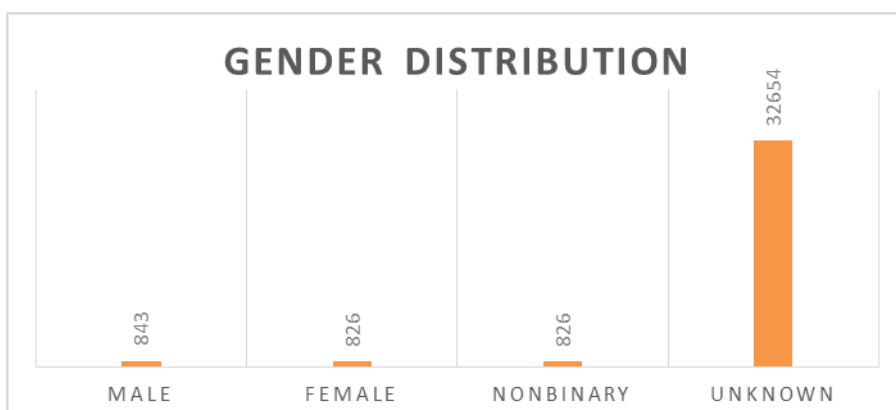

Figure 1. Gender-wise distribution

Source: Author

4.2 Trend of comments

The graph below represents the communication trends on the topic during the study period. It clearly displays the communication trend from the date 30-01-2023 where the users have commented and most of the communications took place during the first commencement of the budget which is on the date 2-02-

2023. It is also noticeable that the variation of comments after the declaration date.
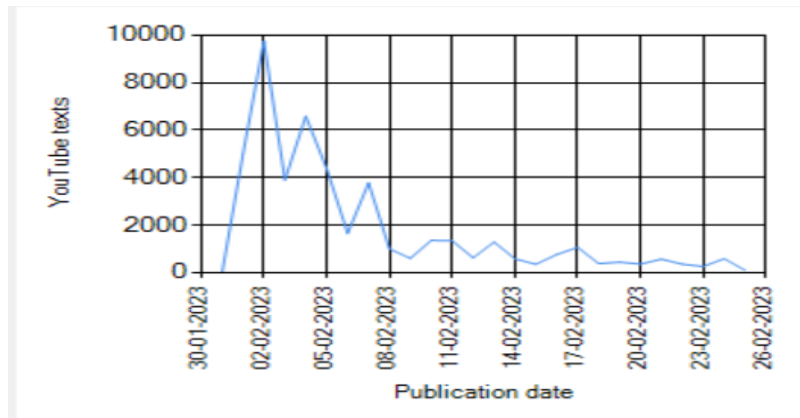
Figure 2.

Comment

trends on

Youtube

Source: Author

4.3 Top viewed videos

The list of the most popular videos, together with the channel name, is displayed in Table 1 below. According to the analysis, the educational channel StudyIQ IAS has the video "Union Budget 2023–24 | Complete Analysis & Highlights | UPSC Economy #Budget Economic" with the most views, with 1963097. The video "Nirmala Sitharaman In Aap Ki Adalat: Rajat Sharma " from IndiaTV, which has 1770651 views, is the one that follows it. The top five videos on YouTube that have received the most views are listed below. The fact that the budget becomes a hot topic for discussion among the public as soon as it is made public or revealed to them is one of the crucial considerations. It is crucial to study this subject because it covers all relevant areas and issues, as well as both new and old schemes and how they are implemented, as a country's economy is founded on its budget. The vast majority of the viewers of the educational channel StudyIQ IAS are students and UPSC candidates. The channel offers a thorough study of the budget and numerous related issues, as well as the benefits and drawbacks of the recently constituted budget. As a

result, the candidates receive the right assistance with the subject, and the educator's accurate analysis aids in their understanding of the budget and its associated problems.

Sl.No	Channel Name	Title of the video	Total views
1	StudyIQ IAS	Union Budget 2023-24	Complete Analysis & Highlights	UPSC Economy #Budget Economic	1963097
2	IndiaTV	Nirmala Sitharaman In Aap Ki Adalat: Rajat Sharma के तीखे सवालों का निर्मला सीतारमण ने दिया जवाब	1770651
3	V6 News Telugu	No Tax On Income Up To Rs.7 Lakh Under New Tax Regime, Says Nirmala Sitharaman	V6 News	964209
4	Dr. Vivek Bindra: Motivational Speaker	Budget 2023	दुनिया का सबसे आसान Analysis	Nirmala Sitharaman	Narendra Modi	941258
5	Zee News	Nirmala Sitharaman On Adani Live: SBI के बाद वित्तमंत्री के बयान से तहलका!	Breaking News		928999

Table 1. List of the top watched videos

Source: Authors

4.4 Top commented videos

Table 2 is a compilation of the most popular YouTube videos pertaining to the topic under investigation. The video titled "UNION BUDGET INDIA 2023-24

FOR JAMMU & KASHMIR | PAKISTANI REACTION ON INDIA REAL ENTERTAINMENT TV" uploaded by Real Entertainment TV received the greatest number of comments, amounting to a total of 7284, which accounts for 15.47% of the overall comments received. Below is a compilation of channel videos that have received the greatest number of comments during a given time period. The Real Entertainment TV channel has garnered significant attention with its most popular videos, such as "UNION BUDGET INDIA 2023-24 FOR JAMMU & KASHMIR | PAKISTANI REACTION ON INDIA REAL ENTERTAINMENT TV," which has received 7284 comments. Another noteworthy video is "200 CRORE FOR AFGHANISTAN IN INDIA'S UNION BUDGET 2023-24 | PAKISTANI REACTION ON," which has generated 3106 comments. Additionally, the video titled "INDIA INCREASED DEFENCE BUDGET FOR 2023 | PAK." has also garnered significant attention. The videos centre around the region of Jammu & Kashmir and the reaction of the Kashmiri populace to India's Union Budget for the fiscal year 2023–2024, which is one of the several consequences analyzed in relation to Union Budgets. The videos also addressed the topic of Afghanistan's receipt of 200 crores from the federal budget, along with the prevailing Pakistani public sentiment over this issue. The allocation of funds towards Defense expenditure has emerged as a topic of deliberation within the Pakistani budgetary discourse. The free food program of India has garnered significant attention among individuals in Pakistan, who are discussing it with other programs implemented in the union budget. This is due to the persistent and urgent nature of the food crisis in Pakistan.

Sl. No	Channel Title	Title of the video	Total Comments
1	Real entertainment tv	UNION BUDGET INDIA 2023-24 FOR JAMMU & KASHMIR	PAKISTANI REACTION ON INDIA REAL ENTERTAINMENT TV	7284
2	Ribaha Imran	Kashmiri's On 35581CR For Kashmir India Budget 2023-24: PAK PM Sensational FREEZE KASHMIR Plot Claim	4071
3	Real	200 CRORE FOR AFGHANISTAN IN	3106

	entertainment tv	INDIA'S UNION BUDGET 2023-24 ▢ | PAKISTANI REACTION ON	
4	Real entertainment tv	INDIA INCREASED DEFENCE BUDGET FOR 2023 | PAKISTANI PUBLIC REACTION ON INDIA REAL ENTERTAINMENT TV	2879
5	Real entertainment tv	INDIA'S FREE FOOD SCHEME IN BUDGET 2023-24 | FOOD CRISIS IN PAKISTAN | PAK REACTION ON INDIA REA TV	2127

Table 2. List of top commented videos

Source: Authors

4.5 Top channels sharing most videos

Figure 3 illustrates the compilation of prominent channels that have exhibited the highest frequency of video sharing throughout the designated research timeframe pertaining to the chosen subject matter. Among the 332 channels in consideration, the following is a compilation of the top 10 channels that possess the greatest number of shared videos. According to the graph, Sansad TV (35) is the leading YouTube channel in terms of video sharing, with CNN-News18 (33) ranking second. In contrast, the Nationalist Hub channel has seen the lowest frequency of video sharing.
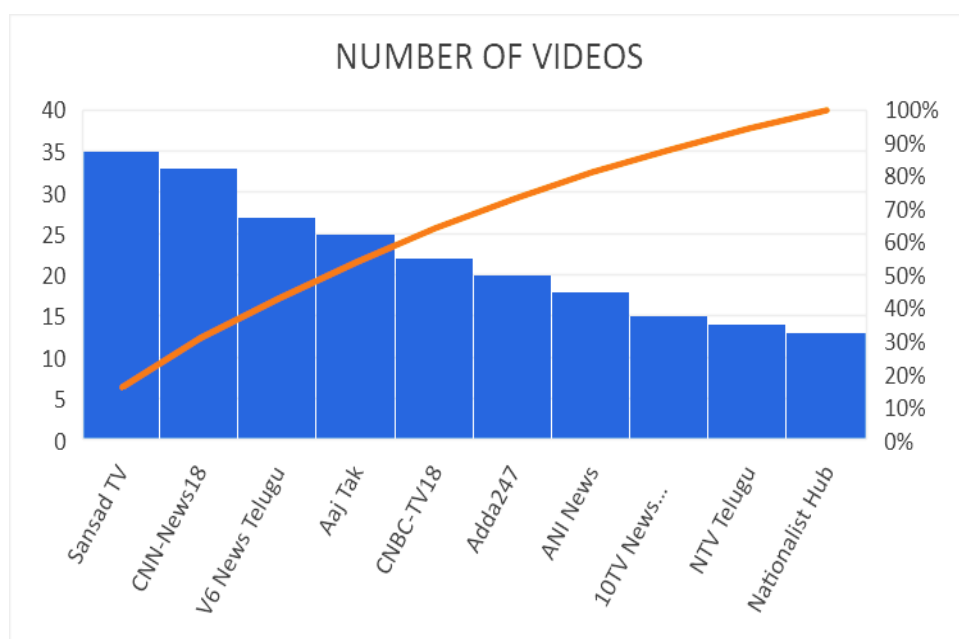
Figure 3. List of top channels with the most shared video

Source: Author

4.6 The Average Sentiment Analysis Result

Table 3 presents the gender-specific distribution of participants included in the study. The Mozdeh programme categorises the audience of YouTube channels into three groups: male, female, and none. The "none" category indicates that the commentator has not disclosed their gender information. Mozdeh is a tool used to assess the attitudes of viewers on a five-point scale, ranging from positive (1-5) to negative (-1 to -5). It then calculates the mean positive and mean negative sentiments over all five levels, along with 95% confidence intervals. The following table presents the mean positive and negative sentiments of 47,057 comments, as assessed by the Mozdeh software. The software categorizes the attitudes expressed in the comments into five distinct levels, encompassing both positive (ranging from 1 to 5) and negative (ranging from -1 to -5) sentiments. It then calculates the average positive and negative sentiments over all five levels. The software provides confidence intervals at a 95% level of confidence. The findings of the study indicate that the mean positive score is significantly elevated across all categories. The rate of females is shown to be rather high at 1.4628, followed by males at 1.4410. In contrast, the mean values for positive and negative aspects are 1.4285 and 1.3156, respectively.

Table

Category	Average positive	Average negative
Average	1.4285	1.3156
Male	1.4410	1.4089
Female	1.4628	1.3058
Total comments		47057

3.

Gender-wise sentiment

4.7 Sentiment Strength of comments

The graph presented below displays the sentiment strength of the user across five distinct levels, namely None (average), weak, moderate, strong, and extremely strong. The Mozdeh software assesses the intensity of sentiments by analyzing the emotional content derived from the input data. Additionally, the user's mood on the subject of study can be indicative of its strength or weakness. The picture demonstrates a notable disparity between the intensity of negative sentiment and positive sentiment.
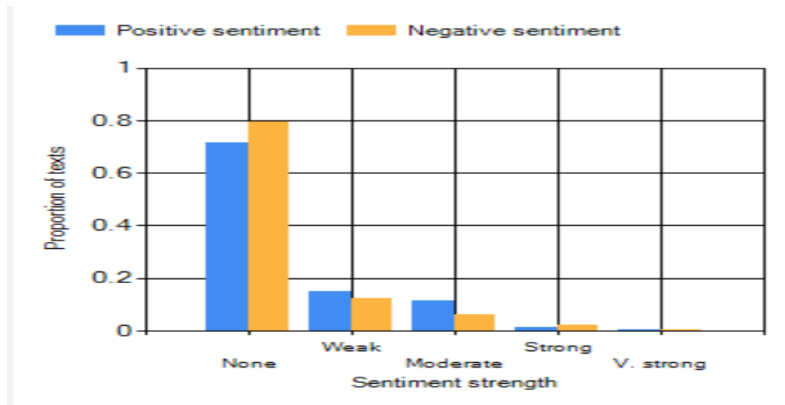
Figure 4. Sentiment Strength of comments

Source: Authors

4.8 Time series illustration of comments during the time period

The time series graph is commonly used to identify trends, illustrating the fluctuation of comments over the duration of the study. The Mozdeh software aimed to accurately capture and analyse the remarks made throughout each day, from its inception to its conclusion. The graph presented depicts the temporal distribution of good and negative attitudes observed on the day of February 2nd, 2023, during the period of highest communication activity.
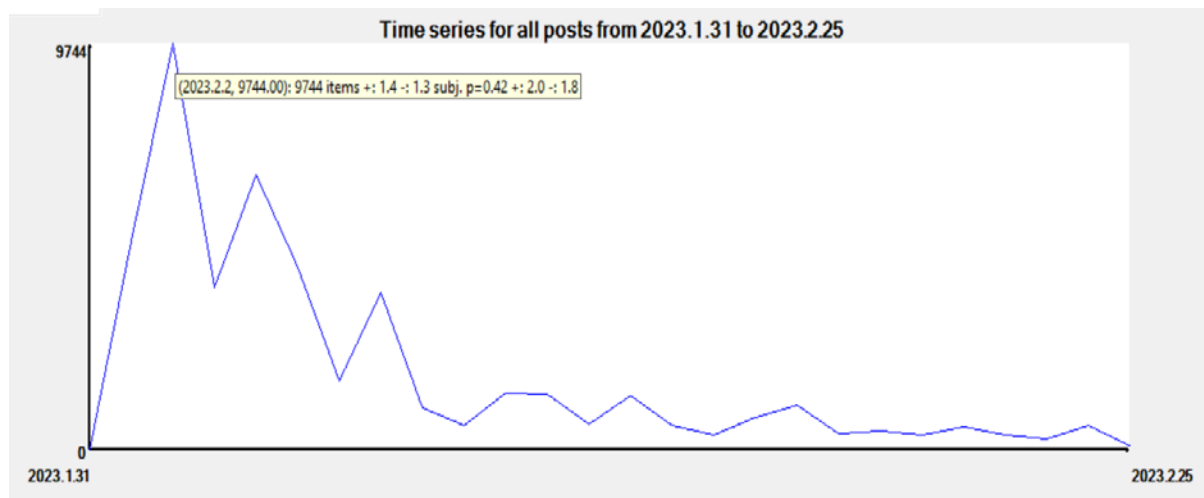
Figure 5. Time series illustration of comments during the time period

Source: Author

5. Discussion and conclusion

YouTube is a widely utilised global platform that hosts an extensive collection of videos including a diverse range of subjects, including but not limited to entertainment, instructional content, educational materials, gaming, and numerous more categories. It is often believed that conducting an analysis of the comments posted on YouTube can yield valuable insights into the perspectives and thought processes of ordinary individuals, as well as shed light on the influence these viewpoints have on the chosen subject matter. In order to assess the efficacy of the viewer's perspective, it is necessary to get an understanding of its effectiveness. This study was conducted to gain insights into the perspectives of users regarding the union budget for the fiscal year 2023-24, which was announced on February 2nd, 2023. In contemporary times, there exists a prevailing and widely embraced phenomenon whereby the economic stability of a nation is intricately linked to the financial allocations made by the incumbent administration. The analysis of comments on the YouTube platform was conducted using the Mozdeh software. The findings indicate a positive impact, with a total of 47,057 comments contributed by 34,323 writers across 332 distinct YouTube channels within the specified time frame. Upon doing an analysis of the comments, it is evident that the video with the most viewership was disseminated by StudyIQ IAS, an instructional channel, amassing a cumulative total of 1,963,097 views. The video titled "UNION BUDGET INDIA 2023-24 FOR JAMMU & KASHMIR | PAKISTANI REACTION ON INDIA REAL ENTERTAINMENT TV" from Real Entertainment TV garnered the most amount of recorded comments, totaling 7284 (15.47%) of the total comments. During the examination of gender distribution, it was observed that male commenters (843) constituted the majority in the gender category, whereas female commenters (826) were comparatively fewer in number. Mozdeh conducted an analysis of the average sentiments of the posts, utilizing a

confidence level of 95%. The results indicated that the sentiment score for positive posts was 1.4285, while negative posts had a sentiment score of 1.3156. On February 2, 2023, the subject matter garnered significant attention and engagement from commenters on the YouTube platform, thereafter experiencing a decline in popularity over time. The findings of the analysis indicate that a significant majority of user comments were directed towards our channels, while a limited number of other channels have provided coverage on the subject matter of the Union budget for the fiscal year 2023-24. In addition to the aforementioned channels, there are also educational channels that serve the aim of distributing videos pertaining to academic subjects and other competitive examinations.

References

Shevtsov, A., Oikonomidou, M., Antonakaki, D., Pratikakis, P., & Ioannidis, S. (2023). What Tweets and YouTube comments have in common? Sentiment and graph analysis on data related to US elections 2020. *Plos one*, *18*(1), e0270542. What Tweets and YouTube comments have in common? Sentiment and graph analysis on data related to US elections 2020 | PLOS ONE

Ceci, L., (2023). Youtube – Statistics & Facts. Retrieved from YouTube: Statistics & Data | Statista

Laine, Amanda., (2022). The first social media platform: the history of social media. Retrieved from The First Social Media Platform: The History Of Social Media (publer.io)

Deori, M., Kumar, V. and Verma, M.K. (2022), "What news sparks interest on YouTube? A study of news content uploaded by India's top five Hindi news networks", Online Information Review, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/OIR-01-2022-0007

Singh, R., & Tiwari, A. (2021). YouTube comments sentiment analysis. *International Journal of Scientific Research in Engineering and Management (IJSREM)*. YOUTUBE-COMMENTS-SENTIMENT-ANALYSIS.pdf (researchgate.net)

Deori, M., Kumar, V., & Verma, M. K. (2021). Analysis of YouTube video contents on Koha and DSpace, and sentiment analysis of viewers' comments. Library Hi Tech. Advance online publication. https://doi.org/10.1108/ lht-12-2020-0323

Deori, M., Verma, M. K., & Kumar, V. (2021). Sentiment Analysis of Users' Comments on Indian Hindi News Channels Using Mozdeh: An Evaluation Based on YouTube Videos. *Journal of Creative Communications.* https://doi.org/10.1177/09732586211049232

Katsurai, M., & Joo, S. (2021). Adoption of data mining methods in the discipline of library and information science. *Journal of Library and Information Studies*, *19*(1), 1–17. https://doi.org/10.6182/jlis.202106_19(1).001

Mozdeh. (2020). YouTube Comment Term Frequency Comparison (CTFC) method. Retrieved from http://mozdeh. wlv.ac.uk/YouTubeCTFC.html

Mozdeh. (2020). YouTube Comment Term Frequency Comparison (CTFC) method. Retrieved from http://mozdeh. wlv.ac.uk/YouTubeCTFC.html

Novendri, R., Callista, A. S., Pratama, D. N., & Puspita, C. E. (2020). Sentiment analysis of YouTube movie trailer comments using naïve bayes. *Bulletin of Computer Science and Electrical Engineering*, *1*(1), 26-32. View of Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes (bcsee.org)

Parabhoi, L. (2019). Twitter sentiment analysis: A case study of ten university libraries. 12th International CALIBER (pp. 179–191).

Porreca, A., Scozzari, F., & Di Nicola, M. (2020). Using text mining and sentiment analysis to analyse YouTube Italian videos concerning vaccination. *BMC Public Health*, *20*(1), 1-9. Using text mining and sentiment analysis to analyse YouTube Italian videos concerning vaccination | BMC Public Health | Full Text (biomedcentral.com).

Farooq, T., & Shabaz, M. (2019). Revenue generation from recommendation system using sentimental analysis. International Journal of Innovative Technology and Exploring Engineering, 8(6), 351–354.

Parabhoi, L. (2019). Twitter sentiment analysis: A case study of ten university libraries. 12th International CALIBER (pp. 179–191).

Thelwall M (2018) Social Media Analytics for YouTube Comments: Potential and Limitations. International Journal of Social Research Methodology 21 (3): 303–16.

Thelwall, M. (2018). Social media analytics for YouTube comments: Potential and limitations. International Journal of Social Research Methodology, 21(3), 303–316. https://doi.org/10.1080/13645579.2017.1381821

Thelwall, M. (2018). Social web text analytics with Mozdeh. Mozdeh, 1–35. http://mozdeh.wlv.ac.uk/resources/ SocialWebResearchWithMozdeh.pdf

Bhuiyan, H., Ara, J., Bardhan, R., & Islam, M. R. (2017, September). Retrieving YouTube video by sentiment analysis on user comment. In *2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (pp. 474-478). IEEE. Retrieving YouTube video by sentiment analysis on user comment | IEEE Conference Publication | IEEE Xplore

Tanesab, F. I., Sembiring, I., & Purnomo, H. D. (2017). Sentiment analysis model based on Youtube comment using support vector machine. International Journal of Computer Science and Software Engineering (IJCSSE), 6(8), 180–185. http://ijcsse.org/published/volume6/issue8/p2-V6I8.pdf

Kim, E. H. J., Jeong, Y. K., Kim, Y., Kang, K. Y., & Song, M. (2015). Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. Journal of Information Science, 42(6), 763–781. https://doi. org/10.1177/0165551515608733

Hornikx, J., & Hendriks, B. (2015). Consumer tweets about brands: A content analysis of sentiment tweets about goods and services. Journal of Creative Communications, 10(2). https://journals.sagepub.com/doi/ full/10.1177/0973258615597406

Jones, Matthew., (2015, June 16). The complete history of social media: A timeline of the invention of online networking. Retrieved from History of social media: The Invention of Online Networking (historycooperative.org)

Thelwall, M. (2014). Sentiment analysis and time series with Twitter. In K. Weller, A. Bruns, J. Burgess, & M. Mahrt (Eds.), Twitter and Society (pp. 83–95). Peter Lang.

Thelwall, M., Sud, P., & Vis, F. (2012). Commenting on YouTube videos: From Guatemalan rock to El Big Bang. Journal of the American Society for Information Science and Technology, 63(3), 616–629. https://doi. org/10.1002/asi.21679

Thelwall, M. (2014). Sentiment analysis and time series with Twitter. In K. Weller, A. Bruns, J. Burgess, & M. Mahrt (Eds.), Twitter and Society (pp. 83–95). Peter Lang.Thelwall,