Shodh Prabha: A Multidisciplinary Journal
ISSN: 3108-2726 (Online), Volume-2, Issue-3, (May - June) 2026

How Social Media Peoples Perceive 'One Nation One Subscription' Initiative: A Sentiment Analysis

Debolina Biswas[1], Sanu Mondal[2]
[1] Research Scholer, Department of Library and Information Science, Jadavpur University, Kolkata, India
[2] Research Scholer, Department of Library and Information Science, Jadavpur University, Kolkata, India

Abstract

The "One Nation One Subscription" (ONOS) initiative, launched by the Government of India in January 2025, aims to democratize access to scholarly research by providing free access to over 13,000 high-impact journals for approximately 18 million users across 6,400 institutions. This study investigates public sentiment towards ONOS by analyzing Twitter data using sentiment analysis techniques. A total of 137 comments were collected using the hashtag #ONOS, with a focus on posts from the Ministry of Education and ANI, which garnered the highest engagement. The data were preprocessed through cleaning, tokenization, stopword handling, and vectorization using TF-IDF. Sentiment classification was performed using lexicon-based methods (TextBlob and VADER) and machine learning models including Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). The results revealed that 41.6% of the tweets expressed positive sentiment, 52.6% were neutral, and only 5.8% were negative. Linear SVC achieved the highest classification accuracy at 89%, followed by Logistic Regression (86%) and KNN (82%). Word cloud analysis highlighted frequently used terms such as “research, ” “India, ” and “access, ” while sentiment-specific word clouds identified “great, ” “good, ” and “free” as positive indicators, and “useless” and “mean” as negative ones. The findings suggest a predominantly positive or neutral public perception of ONOS, with minimal negative sentiment. This study underscores the utility of sentiment analysis in gauging public opinion on policy initiatives and highlights the potential of social media analytics in informing evidence-based decision-making.

Keywords: Sentiment analysis, ONOS, Twitter data, Natural language processing, Opinion mining

1. Introduction

The swift expansion of social media platforms has revolutionized communication, enabling the effortless exchange of ideas, opinions, and information on an unprecedented scale worldwide. Among these platforms, Twitter has emerged as one of the most impactful channels for real-time information dissemination and public discourse. Its distinctive format, which confines posts to 280 characters, compels users to articulate their thoughts concisely. Twitter has become an invaluable resource for assessing public sentiment on a diverse array of topics, including political events, social movements, marketing initiatives, government initiatives, and public health crises.

Moreover, sentiment analysis enables policymakers to assess public opinion on policy changes and social or academic initiatives, allowing for more informed and responsive governance. According to Albladi (Aish Albladi, 2025) “the rapid growth of social media platforms has transformed how people communicate, making it easier than ever to share ideas, opinions, and information on a large scale. Among these platforms, Twitter has emerged as one of the most influential channels for real-time information dissemination and public discourse. The potential of Twitter to influence real-world events has made it an essential tool for stakeholders across various sectors. Policy makers monitor public opinion on key issues, corporations track consumer sentiment to adapt their branding strategies, and public health officials assess responses to health advisories or outbreaks. This dynamic has made sentiment analysis on Twitter not just a matter of academic curiosity but also a necessary tool for timely and informed decision-making.”

Let us consider a scenario in which we want to analyze whether any government policies satisfy public requirements; we can use sentiment analysis to monitor public opinion. Analyzing public reviews, comments, and feedback to understand the sentiment behind them helps identify areas for improvement and address user concerns, ultimately enhancing user satisfaction. Sentiment analysis is also efficient when there is a large set of unstructured data, and we want to classify that data by automatically tagging it. Social media platforms, such as Twitter, provide a space where users can share their thoughts and opinions, as well as connect, communicate, and contribute to certain topics.

“Text analysis can be used to determine the sentiments and attitudes of certain target groups. Much of the available literature focuses on texts in English, but there is a growing interest in multi-lingual analysis” (Qi & Shabrina, 2023). Text analysis can be performed by extracting subjective comments on a certain topic using different sentiments, such as Positive, Negative, and Neutral (Srinagesh & Arun, 2020). One of the current topics of interest is the Coronavirus (Covid-19), which was initially identified in late 2019. The swift global spread of Covid-19 has impacted numerous countries, resulting in lifestyle changes, such as wearing masks on public transport and practicing social distancing. Sentiment analysis can be applied to social media data to investigate shifts in people's behavior, emotions, and opinions. This can be done by categorizing the spread of Covid-19 into three phases and examining negative sentiments towards the virus using topic modeling and feature extraction (Boon-Itt & Skunkan, 2020). Earlier research collected tweets using specific hashtags (#) to organize content by topics such as “#stayathome” and “#socialdistancing” to assess their frequency. Twitter provides raw, unfiltered feedback on audience opinions about offerings. By examining how people discuss policies on Twitter, you can gain insights into public perception and your target audience.

Based on the above analysis, this study focuses on public opinion about ‘ONE NATION ONE SUBSCRIPTION’ initiative by govt of India, The One Nation One Subscription (ONOS) initiative, introduced by the Government of India in January 2025, represents a pivotal advancement in democratizing access to scholarly research throughout the nation. By securing national licenses with 30 prominent international publishers, including Elsevier, Springer Nature, Wiley, and Taylor & Francis, ONOS affords complimentary access to over 13,000 high-impact

journals for approximately 18 million students, researchers, and faculty members across 6,400 government-managed higher-education and research institutions. This initiative holds particular significance in India's evolving digital and educational landscape. This aligns seamlessly with the National Education Policy 2020, which underscores the critical role of research and innovation in higher education. By dismantling financial and logistical barriers to academic resources, ONOS ensures that institutions in Tier 2 and Tier 3 cities are afforded the same opportunities as their metropolitan counterparts, thereby promoting equitable educational and research prospects nationwide. The initiative's digital implementation via the Information and Library Network (INFLIBNET) guarantees user-friendly access, further amplifying its impact on the community. In conclusion, ONOS is a groundbreaking initiative that not only narrows the knowledge divide among India's educational institutions but also propels the nation towards a more inclusive and research-centric future, thereby unifying various consortia and institutional subscriptions. In addition to the above, it is also planned to have a consolidated central funding to support Indian authors to pay for APCs for good quality OA journals (Qi & Shabrina, 2023). We propose a method that integrates text-data mining and sentiment analysis. This approach aims to extract key information from Twitter to gain a deeper understanding of consumer sentiments.

2. Objective

- How do Twitter users perceive the 'One Nation One Subscription' initiative?

3. Methodology

Individuals may have varying perspectives on the One Nation One Subscription (ONOS) initiative. Some may choose to express their views or critiques of this initiative by commenting and reviewing social networking platforms. These contributions assist decision-makers in assessing initiatives and discerning the significance of various attributes by synthesizing insights from the collective opinions of the public. The advancement of text mining techniques has facilitated the automatic evaluation and summarization of these comments. The proposed opinion mining methodology begins with the extraction of text posts from networking sites. Initially, a search was conducted using the hashtag #ONOS to locate relevant content, after which we compiled posts pertaining to #ONOS. Our focus extends to posts from the Ministry of Education of India and ANI, as these have garnered the highest volume of user comments at the time of writing. To collect data from Twitter, we employed a web scraping technique that enabled us to extract information directly from the platform. Specifically, we utilize an AI-powered tool known as "Twitter Comment Scraper." This software is adept at efficiently gathering comments and other pertinent data from Twitter, thereby streamlining our research efforts. Once we have amassed the data, we will undertake a comprehensive analysis, with particular emphasis on sentiment analysis. For this purpose, we harnessed Python programming and utilized Jupyter Notebook, a versatile platform that allows us to analyze and visualize the data effectively.

3.1 Data Collection

Initially, a search process was conducted using the hashtag #ONOS to explore the content, leading us to discover that during the data collection period, the Ministry of Education of India

and the ANI news channel garnered the highest number of comments. Consequently, we compiled all the comments made by the users. To extract data from Twitter, we employed a web scraping technique that enabled direct information retrieval from the platform. Specifically, we are utilizing an AI-powered tool known as "Twitter Comment Scraper." This software was designed to efficiently gather comments and other pertinent data from Twitter, thereby facilitating our research. Ultimately, we collected 144 rows of data.

3.2 Data Preprocessing

Owing to the substantial workload of the data collection task in this study, frequent operations during the collection process may have resulted in some missing values and duplicate data. Additionally, due to the diverse nature of online shopping review text information, there may be a portion of data with no value. To ensure the accuracy of subsequent analysis results, it is necessary to clean the above invalid data, transforming them into high-quality data that can be efficiently utilized, and avoiding any impact on the analysis results. The specific data cleaning rules and steps were as follows:

- **Remove Null Values:**

Exclude data with other information such as comment time but with empty comment content or entirely empty entries. Such comments were excluded because frequent visits and data crawling triggered the anti-crawling mechanism of the shopping website, causing the relevant information not to be displayed.

- **Eliminate Duplicate Data:**

Exclude multiple comments with the same content when the time interval is short or when the comments have the same timestamp. The reasons for such duplicate comments include improper operations during repeated crawling, consumers copying others' comments for positive reviews, cashback, and rebate discounts, as well as duplicates generated by merchants engaging in order brushing.

- **Sentence Segmentation:**

Sentence segmentation, refers to the process of using tokenization techniques to break down a text into individual words or sequences of words, with separation achieved through spaces. The purpose of sentence segmentation is to break down a block of text into individual sentences, making the text more manageable and facilitating further linguistic analysis or processing of the text. For example, the sentence "I love pizza!" becomes ["I", "love", "pizza", "!"]. This step enables models to analyze text at the word level and is foundational for further text processing. Sentence segmentation serves as the foundation and prerequisite for in-depth text mining. Text data can be transformed into a mathematical vector format after tokenization, facilitating subsequent analysis (Srinivasan & Dyer, 2021).

- **Stopword Handling:**

The handling of stopwords is crucial in the preprocessing phase of sentiment analysis. Stopwords

such as "the, " "is, " and "and" typically lack substantial meaning, and their elimination can sharpen the focus on significant sentiment-related terms. However, indiscriminate removal may adversely affect the analysis, as certain stopwords, including "not, " "no, " and "never, " are vital for assessing sentiment polarity. For example, the phrase "not good" conveys negativity; omitting "not" could lead to a positive interpretation.

4. Sentiment Analysis Technique

- **Lexicon based sentiment calculation**

In our project, we opted for a lexicon-based approach to circumvent the need to generate labeled data. This methodology necessitates the computation of the semantic orientation of words found within the text (Rajman & Besançon, 1998)A significant advantage of employing a lexicon-based strategy lies in its inherent simplicity and the ease with which it can be modified by a human. By utilizing this technique, we can effectively capture and categorize the semantic orientation as neutral, positive, or negative. From a definitional standpoint, sentiment analysis serves as a means to extract subjectivity and polarity from text, while semantic orientation assesses both the polarity and intensity of the content (Szabolcsi, 2004). In this context, adjectives and adverbs are instrumental in revealing the semantic orientation of the text (Jain, Seeja, & Jindal, 2021),specifically in our case, a tweet. We employed a well-known Python package, TextBlob, which facilitates intricate operations and analyses of tweet data. Tweets were represented in a numerical format, with TextBlob assigning individual scores to each tweet. Ultimately, the sentiment of the tweets was derived through a pooling operation that averaged all the sentiment scores.

To enable the Machine Learning Algorithm, it is essential to convert the text into numerical representations, a procedure referred to as text vectorization. Various vectorizers are available; however, the researcher has opted for TF-IDF, as opposed to CountVectorizer, which merely counts the frequency of words in the document and thus tends to favor more common terms. In contrast, TF-IDF not only counts word occurrences but also takes into account their overall significance. Consequently, TF-IDF identifies words that are crucial to the document, even if they are not frequently mentioned. Once the data were vectorized, they were divided into training and test sets. For this study, an 80:20 split was utilized, allocating 80% of the data for model training and 20% for testing. The random state for the split is set at 42 to ensure consistency with each re-execution of the division.

$$TF\text{-}IDF(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \cdot \log\left(\frac{N}{1+n_t}\right)$$

The machine learning models used in this study were logistic regression, K-nearest neighbors (KNN), and linear SVC. The accuracy of the models and their evaluation are discussed further in the following section.

- **NLTK (VADER lexicon)**

is the most widely used Python library for conducting sentiment analysis. It contains several

lexicon sources and processing libraries that can be used for removing punctuation and numerals, converting capitalized words to lower case, tokening, filtering emoticons and stop words, and lemmatizing. An essential NLP analysis, it contains almost all the functions required for both the preprocessing and analysis phases.

The compound score for each tweet is obtained by summing the scores of each word, adjusted according to predefined rules, and

Then, it was normalized to be between - 1 and + 1 as follows:

$$x = \frac{s}{\sqrt{s^2+a}}$$

where x denotes the sum of the scores of the words contained in each tweet and α denotes the normalization constant.

- **TextBlob Python library**

TextBlob is inspired by the NLTK library. Although it offers fewer functions than its predecessor, it compensates with a more accessible syntax and maintains an acceptable average level of precision. The part-of-speech tagging (POS) feature warrants further examination: in the context of complex texts, TextBlob efficiently classifies sentence elements with minimal computational complexity, distinguishing nouns, verbs, adverbs, conjunctions, and others. Specifically, TextBlob identifies lexical entities through its proprietary library and tagging phrases, subsequently employing a POS tool to automatically assign tags to words within a sentence. To determine the sentiment of an individual word, TextBlob calculates the average from multiple entries for that same word. Nevertheless, when analyzing tweets or other content with low linguistic complexity, this advantage becomes negligible.

5. Sentiment Analysis Model

Support Vector Machines (SVM) and Logistic Regression (LR) and KNN are three of the most effective and widely employed techniques for sentiment analysis, attributable to their robustness, accuracy, and proficiency in managing high-dimensional textual data. Their adoption is well supported in sentiment analysis tasks by empirical evidence from numerous studies.

- **K-Nearest Neighbors (KNN)**

“The K-nearest neighbors (KNN) algorithm is a supervised machine learning technique applicable to both classification and regression tasks; however, it is predominantly employed for classification in industrial settings. The core concept of KNN involves identifying the k-nearest data points to a specific test data point and utilizing these neighbors to make predictions. The parameter k, which signifies the number of neighbors to consider, is a hyperparameter that requires fine-tuning. In classification scenarios, KNN assigns the test data point to the class that is most common among the k-nearest neighbors, that is, the class with the majority of neighbors is chosen. For regression tasks, KNN predicts the value of the test data point by averaging the values of the k-nearest neighbors. The choice of distance metric, which measures the similarity

between data points, is crucial for the effectiveness of the algorithm. Commonly used distance metrics include the Euclidean, Manhattan, and Minkowski distances." (Tutorialpoint, n.d.)

- **Support Vector Machine (SVM)**

The Support Vector Machine (SVM) is a prevalent supervised learning algorithm for sentiment analysis, and is renowned for its exceptional accuracy and efficacy in text classification tasks. In this context, SVM categorizes textual data, such as reviews or social media content, into sentiment classifications, including positive, negative, or neutral. It operates by identifying the optimal hyperplane that distinguishes data points of different classes within a high-dimensional space. For textual data, features are typically extracted using techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) or word embeddings. SVM excels at managing high-dimensional and sparse datasets, a common characteristic in natural language processing.Numerous studies have validated the effectiveness of SVM in sentiment classification; for instance, (Bo Pang, 2002) revealed that SVM surpassed other classifiers such as Naïve Bayes and Maximum Entropy in accurately classifying movie reviews based on sentiment polarity.

- **Logistic Regression**

Logistic regression is a prevalent machine learning approach for sentiment analysis and is celebrated for its simplicity, efficiency, and effectiveness in binary classification tasks. In this domain, the primary objective is often to categorize textual data, such as reviews or social media posts, into positive or negative sentiments. Logistic regression estimates the probability. Both models are capable of handling large-scale datasets and perform well in binary sentiment classification. Their effectiveness has been consistently validated in comparative studies, making them reliable choices for sentiment analysis (Ahmed Hassan Yousef, 2014)

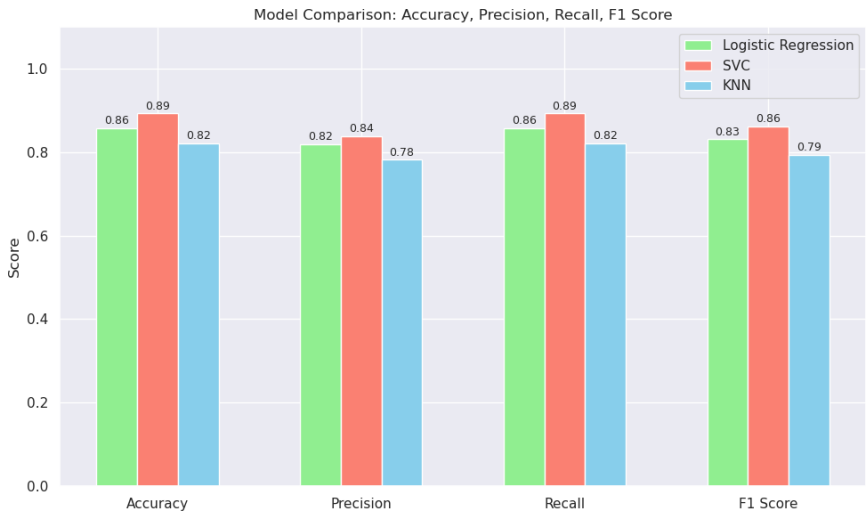

Fig. 1

From Fig. 1, it is evident that the Linear SVC achieved the highest accuracy at 89%, while Logistic Regression and KNN recorded accuracies of 86% and 82%, respectively. Subsequently, we turned our attention to Precision, Recall, and F1-Score. Precision is a metric that reflects the accuracy of the model in predicting positive outcomes, indicating the proportion of positive predictions made by the model that are indeed accurate. In contrast, Recall denotes the fraction of actual positives that were correctly identified. The F1-Score represents the weighted average of both Precision and Recall. In the context of this research, Precision serves as a valuable measure, as the researcher aims to ascertain whether the review is genuinely negative or positive.

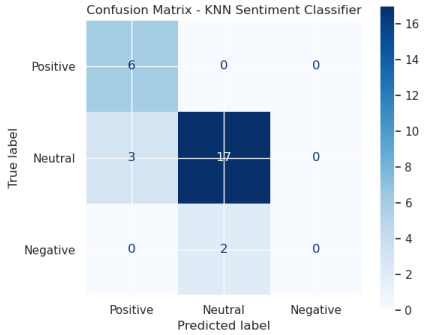

Fig. 2

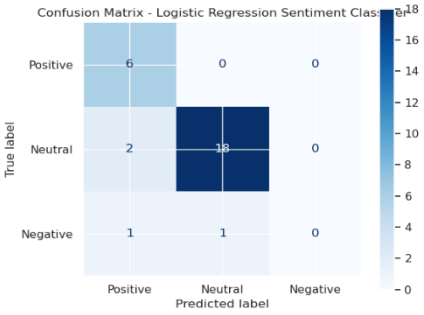

Fig. 3

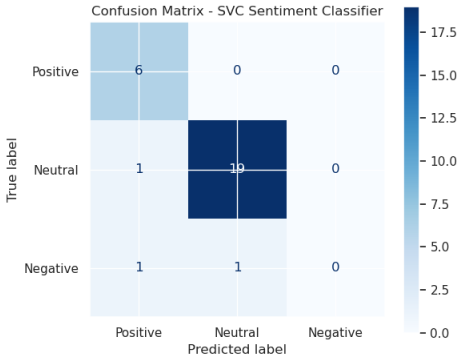

Figure 4

Figure 2 illustrates the confusion matrix for the KNN algorithm employing the TFIDF Vectorizer (Ashwin Sanjay Neogi, 2021).The true label signifies the actual sentiment of the tweet, whereas the predicted label denotes the anticipated sentiment. The confusion matrix indicates that 6 + 17 + 0 = 23 tweets were accurately predicted in alignment with their true sentiments. Furthermore, it reveals that 3 + 2 = 5 tweets were misclassified. Overall, 83% of the tweets were correctly classified according to their true labels, while the remaining were inaccurately predicted in terms of sentiment. Similarly, Fig. 3 presents the confusion matrix for Logistic Regression utilizing the TF-IDF vectorizer. From the confusion matrix, we observe that 6 + 18 = 24 tweets were correctly

identified, while 4 tweets were misclassified. Likewise, Fig. 4 displays the confusion matrix for the SVC with the TF-IDF vectorizer. The matrix indicates that 6 + 18 = 25 tweets were accurately predicted, whereas 3 tweets were classified incorrectly.

Results and Discussion

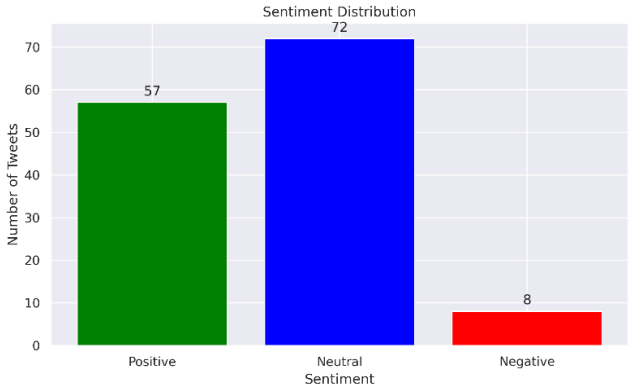

Figure 5

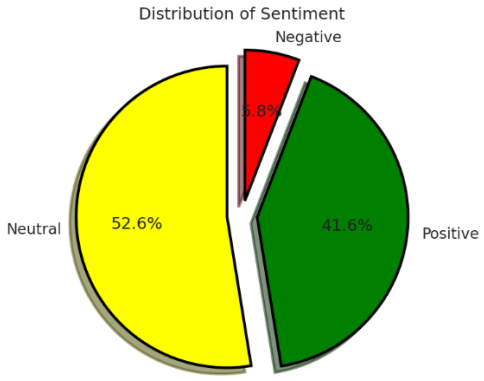

Figure 6

Fig.5 shows some that 137 comments were retrieved (related to One Nation One Subscription) from Twitter post. Classification results demonstrate that our code retrieved 57 positive, 72 neutral and 8 negative opinions among 137 comments. Classification results are graphically represented using Pie Chart in Fig.6 By this we can clearly see that 41.6% result was positive, 52.6% neutral and 5.8% negative opinions. we can clearly see that overall sentiment trends related to 'One Nation One Subscription' mostly neutral or positive, there are very less comment where people give negative opinion related to 'One Nation One Subscription'.

Fig. 7 show the scatterplot of positive negative and neutral tweet. red dots are the positive comment. blue represent neutral and Green negative comments. We can also observe that extreme level of positive and negative are very less. Most of the comment between -0.4 to +0.6.

Fig. 7

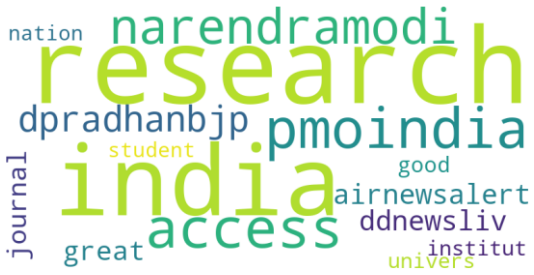

Figure 8

Figure 9

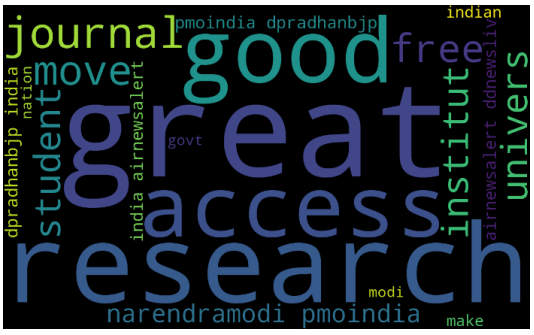

Figure 10

Fig. 8 shows the word cloud obtained from the cleaned tweets. Word Cloud is a pictorial representation of commonly used words in a particular dataset. We provided our dataset of cleaned tweets to the model to generate this word cloud (Heimerl, Lohmann, Lange, & Ertl, 2014). The entire word cloud represents the top 20 frequently used words like research, india, pmoindia, access, narendramodi, great, good, free. The words with a larger font occur more commonly than the words with a smaller font. Fig. 9 represent Positive word. For example 'great', 'good', 'free' and Fig. 10 represent negative word for example 'useless', 'mean'

Sentiment Trends

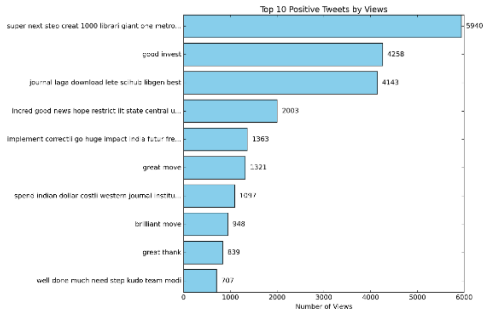

Figure 11

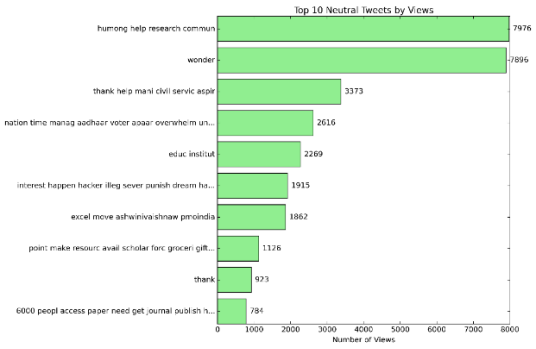

Figure 12

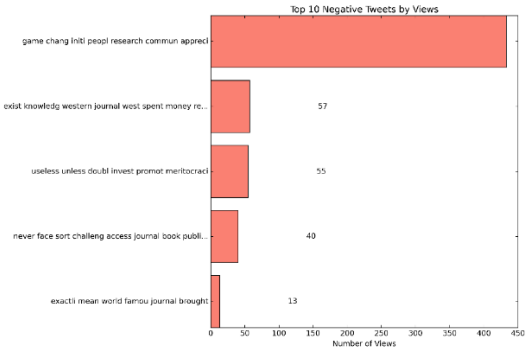

Figure 13

Fig.11 shows the top 10 comments by view in respect of Positive comments.

- “Super next step creates 1000 library giant one metro”
- “Implement correctly go huge impact India future free high quality education cheap internet access”
- “Brilliant move, great move”
- “Empower student foster growth among research enthusiast”
- “Help research commune”

The feedback provided in the earlier comments indicates a widespread perception that this initiative by the government is a commendable step forward. Many individuals express optimism that this program will enhance researchers' access to a wealth of research journals, ultimately fostering a more vibrant research environment. This access is expected to invigorate scholarly activities, allowing for greater collaboration and innovation in various fields. However, it is important to acknowledge that there are also some negative remarks from social media users, as illustrated in Fig. 13, which reflect a more critical viewpoint on the initiative.

- “never face sort challenge access journal book public base theory standard never heard anyone say particular document pertain study research whatever assign avail bring exist knowledge western journal west spent money research compare India take benefit research also provieds create journal well”

Above negative comments we observe that people are want more money input in the research activity, like any western or European country.

Conclusion

In conclusion, while the study offers valuable insights, its findings must be interpreted with caution due to several limitations. Constraints related to sample size, data quality, and geographic scope may affect the generalizability and robustness of the results. Temporal limitations further restrict the ability to capture long-term trends, while reliance on self-reported data introduces potential biases. Methodologically, the cross-sectional design limits causal inference, and the use of non-standardized or potentially unreliable measurement tools may impact the validity of the findings. Inadequate control of confounding variables and possible violations of statistical assumptions also pose challenges to the accuracy and depth of the analysis. These limitations highlight the need for future research employing more rigorous designs, larger and more representative samples, validated instruments, and advanced analytical techniques to strengthen the evidence base and enhance the applicability of the findings across diverse contexts.

Traditionally, sentiment analysis has concentrated mainly on the evaluation of text through the use of natural language processing (NLP). However, recent advancements in technology, alongside the explosion of big data, have ushered in a remarkable evolution within this field. The rise of big data analytics and deep learning technologies has not only tackled the conventional

challenges faced in sentiment analysis but has also had a profound impact on Web 2.0 applications. In today's landscape, sentiment analysis has become indispensable for evaluating and interpreting product feedback, thereby enabling well-informed decisions across various sectors such as marketing, healthcare, banking, and issues related to social or political contexts. This paper delves into the sentiments, emotions, and intentions that surround the 'One Nation One Subscription' initiative, drawing insights from comments gathered from a dataset comprising 136 responses. Our findings reveal that a significant majority of users convey a notably positive perspective regarding the government's initiative, expressing optimism that it will bolster the research community. This initiative holds the potential to improve access to essential academic journals that are frequently costly and challenging to acquire, ultimately promoting the advancement of academic excellence. Limitation of the study is data were collected over a short period, they may not capture long-term trends or seasonal variations, limiting the ability to draw conclusions about sustained effects or pattern. Multi-Platform Sentiment Analysis, this study is limited to Twitter, but future research could include additional social media platforms such as Facebook, YouTube, Instagram, LinkedIn, or Reddit.

Declarations

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Bibliography

Ahmed Hassan Yousef, W. M. (2014). Sentiment Analysis Algorithms and Applications: A Survey. *Ain Shams Engineering Journal*, 20. doi:10.1016/j.asej.2014.04.011

Aish Albladi, M. I. (2025). Sentiment Analysis of Twitter Data Using NLP Models: A Comprehensive Review. *IEEE Access*, 25. doi:10.1109/ACCESS.2025.3541494

Ashwin Sanjay Neogi, K. A. (2021). Sentiment analysis and classification of Indian farmers' protest using. *International Journal of Information Management Data Insights*, 21. doi:doi.org/10.1016/j.jjimei.2021.100019

Bo Pang, L. L. (2002). Thumbs up? Sentiment classification using machine learning techniques. (p. 7). Association for Computational Linguistics. doi:10.3115/1118693.1118704

Boon-Itt, S., & Skunkan, Y. (2020). Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study. *JMIR Public Health Surveill, 6*, 12. doi:10.2196/21978

Chen, S. R. (2007). Yahoo! For Amazon: Sentiment Extraction from Small Talk on the Web. *INFORMS*, 14. Retrieved from https://www.jstor.org/stable/20122297

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*, 18. doi:doi.org/10.1108/LHT-12-2020-0323

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word Cloud Explorer: Text Analytics Based on Word Clouds. IEEE. doi:10.1109/HICSS.2014.231

Jain, S., Seeja, K., & Jindal, R. (2021). A fuzzy ontology framework in information retrieval using semantic query expansion. *1*, p. 15. International Journal of Information Management Data Insights. doi:https://doi.org/10.1016/j.jjimei.2021.100009

Lee, B. P. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends® in Information Retrieval*. doi:dx.doi.org/10.1561/1500000011

Liu, B. (2022). *Sentiment Analysis and Opinion Mining.* Springer Cham. doi:doi.org/10.1007/978-3-031-02145-9

Luciano Barbosa, J. F. (2010). Robust Sentiment Detection on Twitter from Biased and Noisy Data. *Coling 2010 Organizing Committee*, 8. Retrieved from https://aclanthology.org/C10-2005/

Pascual, F. (2022, july 7). *Hugging Face.* Retrieved from huggingface.co: https://huggingface.co/blog/sentiment-analysis-twitter

Priyanshu Kumar, S. A. (2022, march). CONSUMER BEHAVIOUR ON ELECTRIC VEHICLES. *International Research Journal of Modernization in Engineering Technology and Science, 4*(3), 10. Retrieved from https://www.irjmets.com/uploadedfiles/paper/issue_3_march_2022/19823/final/fin_irjmets1647283559.pdf

Qi, Y., & Shabrina, Z. (2023). *Social Network Analysis and Mining, 13*(1), 31. doi:10.1007/s13278-023-01030-x

Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach. *Social Network Analysis and Mining*, 14. doi:doi.org/10.1007/s13278-023-01030-x

Rajman, M., & Besançon, R. (1998). Text Mining - Knowledge extraction from unstructured textual data. *Advances in Data Science and Classification*, (p. 8). doi:https://doi.org/10.1007/978-3-642-72253-0_64

Srinagesh, A., & Arun, K. (2020). Multi-lingual Twitter sentiment analysis using machine learning. *International Journal of Electrical and Computer Engineering (IJECE)*, 10. doi:10.11591/ijece.v10i6.pp5992-6000

Srinivasan, S., & Dyer, C. (2021). Better Chinese Sentence Segmentation with Reinforcement Learning. *ACL Anthology*, 9. doi:10.18653/v1/2021.findings-acl.25

Szabolcsi, A. (2004). Positive Polarity – Negative Polarity. *22*, p. 43. Natural Language & Linguistic Theory. doi:https://doi.org/10.1023/B:NALA.0000015791.00288.43

Tutorialpoint. (n.d.). Retrieved from https://www.tutorialspoint.com/: https://www.tutorialspoint.com/machine_learning/machine_learning_knn_nearest_neighbors.htm