Sentiment Analysis of the Conflict Between Meitei and Kuki Communities of Manipur, India Using YouTube Comments: A Comparative Study of Random Forest, Naïve Bayes, SVM, and KNN

MOIRANGTHEM SAJAN SINGH SAJAN
sajanmoirangthem@gmail.com

Assam University

Rajesh Rangappa Aldarthi RAJESH
Assam University

LUCKEY PATHAN LUCKEY
O. P. Jindal Global University

BORNA NATH BORNA
Assam University

Research Article

Keywords:

Posted Date: April 29th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-6511780/v1

Additional Declarations: No competing interests reported.

Abstract

Purpose:

The primary goal of the study is to use YouTube as a data source to examine public opinion about the ethnic conflict or tension between the Meitei and Kuki communities of Manipur, India. It investigates how digital platforms reflect emotional engagement with regional conflict and evaluates machine learning models for sentiment classification in this context.

Methodology:

A dataset of 551 YouTube videos and 124,971 associated comments were collected through Webometric Analyst and preprocessed using natural language processing techniques, including TF-IDF vectorization. Four models for machine learning to classify comments into positive, neutral, or negative attitudes, Support Vector Machines (SVM), Random Forest, Naïve Bayes, and K-Nearest Neighbors (KNN) were trained and evaluated. The model's performance was evaluated using its F1-score, recall, accuracy, and precision.

Findings:

SVM achieved the highest accuracy (93.42%), followed closely by Random Forest (91.59%). Many comments were neutral, while positive sentiments marginally exceeded negative ones. The sentiment distribution suggests a complex emotional response, including empathy and calls for peace. The top 10 most-viewed videos revealed various content types, from news and interviews to political commentary and vlogs, indicating diverse user engagement with the issue.

Originality:

This study presents one of the first applications of machine learning-based sentiment analysis to YouTube content on the Manipur conflict. It contributes original insights to computational social science and digital media research by demonstrating how user-generated video content can assess public sentiment in real time. The findings offer methodological and substantive value, particularly in understanding how digital discourse shapes perceptions of ethnic conflict in India.

Introduction

A specific division of Natural Language Processing (NLP) centred on identifying and categorizing subjective opinions expressed in large amounts of textual data is sentiment analysis. Social media has become a powerful space where people share thoughts, emotions, and stories in real time. With platforms like YouTube gaining massive popularity, especially during times of civil unrest and conflict, these digital arenas provide valuable insights into how the public feels and perceives unfolding events. Sentiment analysis is a NLP technique that detects substantial, meaningful patterns and features in a huge text corpus. It is also called as text mining, subjectivity analysis, opinion mining, or emotion artificial intelligence.

In recent years, YouTube has become a central hub for sharing information, opinions, and narratives about socio-political events. However, due to the sheer volume and diversity of content, extracting meaningful insights from this data remains a methodological challenge. This study applies sentiment analysis techniques to comments from YouTube videos related to the Manipur ethnic violence started from May 2023 to examine how the public responded emotionally to the conflict in real-time.

Manipur, a state in the northeastern region of India, occupies an area of 22,327 square kilometres and is home to a complex demographic composition primarily comprising Meitei, Naga, Kuki, and Meitei Pangal (Manipuri Muslim) communities Royal Thimphu College, Bhutan & Khamrang, (2023). The Meiteis, residing mainly in the Imphal Valley, form the majority, whereas the surrounding Hill Districts are inhabited predominantly by tribal groups such as the Kukis and Nagas Saha Roy, (2024). Historically, the region has been a site of political contention and inter-ethnic strife, rooted in colonial legacies and contemporary identity politics. After merging with India in 1949 under contested circumstances Akoijam (2005); Noni & Sanatomba (2016), Manipur has continued to experience episodes of ethnic conflict, including the Naga-Kuki and Kuki-Paite clashes, primarily influenced by territorial, administrative, and identity-based grievances Oinam (2003).

The most recent and intense outbreak occurred in early May 2023, following events that inflamed existing tensions. On April 20, 2023, the Manipur High Court directed the state government to recommend including the Meitei community under the Scheduled Tribe (ST) category Lakshman (2023). This legal development was perceived as threatening by various tribal organizations, culminating in the ‘Tribal Solidarity March’ organized by the All-Tribal Students’ Union Manipur (ATSUM) on May 3 across ten Hill Districts. While initially peaceful, violence erupted in the Torbung area of Churachandpur district, where an armed mob reportedly attacked Meitei residents Das (2023). This was followed by retaliatory assaults, triggering a widespread and sustained cycle of violence between the Meitei and Kuki communities.

Social media platforms, particularly YouTube, Facebook (Meta), WhatsApp, and Twitter (now X), played a significant role in shaping and disseminating narratives around the conflict. Content ranged from first-hand video documentation to polarizing rhetoric and misinformation, contributing to an emotionally charged and politically fragmented digital discourse. As these platforms increasingly mediate public understanding of events, analyzing the sentiments embedded in such content is essential to comprehend how social media reflects and amplifies real-world intercommunal tensions.

The present study performs a sentiment analysis of YouTube videos and their associated comments on the Manipur violence. We apply NLP-based preprocessing and TF-IDF feature extraction to prepare the data for machine learning classification. In addition, a labelled corpus of 124,971 comments from YouTube was used to train and validate the sentiment classification models, categorizing sentiments into positive, neutral, and negative classes.

Literature Review

This section provides an overview of social media analytics, including the popular sentiment analysis techniques used in social media analytics. Several scholars have conducted sentiment analyses on social media platforms like Twitter and YouTube. These works influence comments, tweets, and other metadata acquired from users' social network profiles or public events that are collected and analyzed to get substantial and intriguing insights.

Bashir (2021) evaluated Twitter's opinion following the Khan Shaykhun Syria Chemical Attack on April 4, 2017. Over 27 days, they collected 13,156 tweets to assess public reactions, indicating that 53.70% expressed unfavourable views while only 12.67% were positive. Using data mining (orange) and visualization tools (Vosviewer) demonstrated Twitter's importance in altering public discourse and creating solidarity during emergencies. Verma et al. (2023) conducted a sentiment analysis study on public responses to the Balasore train tragedy using YouTube video comments, titles, and descriptions. They adopted a quantitative research design with a cross-sectional approach, and YouTube data was extracted via YouTube Data API and Webometric Analyst software. Python-based natural language processing tools were used to classify responses into positive, neutral, and negative sentiments. The findings indicated that most video titles and descriptions were neutral, while comments exhibited a mix of neutral, negative, and some positive sentiments, reflecting diverse public emotions surrounding the incident. Alhur et al. (2022) investigated consumer attitudes toward mHealth applications, focusing on key design aspects. The study employed a four-step methodology: data collection, preprocessing, sentiment analysis using the Valence Aware Dictionary and Sentiment Reasoner (VADER), and thematic analysis through the Latent Dirichlet Allocation (LDA) algorithm via Orange Software. These methods were applied to 836 user reviews of eight mHealth apps available in app stores in Jordan. The findings provide valuable insights for healthcare stakeholders by highlighting positive and negative user experiences, identifying preferred features, and offering recommendations for improving mHealth app design and functionality. Through Twitter discourse analysis, Mir and Sevukan (2024) explored public perceptions of COVID-19 vaccines in India. The study collected tweets spanning January 4 to March 22, 2021, using hashtags #Covid19vaccine and #Coronavirusvaccine. Employing computational methods in Orange software and VOSviewer, the authors categorized sentiment polarity (positive, negative, neutral) and mapped prevalent discussion themes. Results revealed social media's dual function as a vaccine awareness and misinformation propagation platform.

Methodology

This study collects data at a specific time, focusing on the Manipur violence that started on May 3 2023 and the associated video titles, descriptions, and comments from YouTube.

3.1 Collection of data

The video details and associated comments were extracted using Webometric Analyst software on April 18 2025, using the keyword "Manipur Violence" OR “Manipur Conflict” OR “Meitei and Kuki”.

Subsequently, to extract YouTube video metadata, the Python library googleapiclient.discovery and a custom function get_video_statistics(video_ids, api_key) were used to interact with the YouTube Data API through Webometric Analyst. This enabled the creation of an API client to fetch video statistics, titles, and descriptions. We retrieved YouTube videos (n=551) and video titles, descriptions, and comment lines (n=124,971) for sentiment analysis. The extracted video data were rechecked to ensure that videos are published after May 3 2023. So, 551 video IDs were processed, and comments (including replies) were considered for analysis.

3.2Data Preprocessing

The raw text data was subjected to a structured preprocessing pipeline to reduce noise and standardize input for analysis. The following natural language processing (NLP) techniques were applied:

- Lowercasing: All text was converted to lowercase to ensure uniformity.
- Noise Removal: Non-alphabetic characters, numerical digits, punctuation, and extra whitespace were eliminated.
- Stopword Removal: Common words with limited semantic value (e.g., “is,” “the,” “and”) were removed using NLTK's stopword list.
- Tokenization: Text was segmented into individual tokens using word-level tokenization.
- Lemmatization: Tokens were reduced to their base or dictionary form using WordNetLemmatizer, improving generalization across similar words.

3.3 Sentiment Annotation

Each comment in the dataset was annotated with a sentiment label: positive, neutral, or negative. A combination of manual annotation by trained coders and semi-automated tagging using sentiment lexicons (e.g., VADER, TextBlob) and rule-based heuristics was used to ensure scalability and accuracy. This yielded a labelled dataset suitable for supervised learning.

3.4 Feature Extraction

Term Frequency–Inverse Document Frequency (TF-IDF) vectorization was employed to convert textual data into a format suitable for machine learning algorithms. This technique assigns a weight to each term based on its importance in a given document relative to its occurrence across the entire corpus. The resulting sparse feature matrix served as the input to all classification models.

3.5 Model Training

Four classical machine learning algorithms were implemented for sentiment classification:

- Random Forest (RF): An ensemble learning method aggregating multiple decision trees to reduce variance and improve accuracy.

- Naïve Bayes (NB): A probabilistic model based on Bayes' Theorem, particularly effective for text classification due to its speed and simplicity.
- Support Vector Machine (SVM): A supervised learning algorithm that maximizes the margin between data points of different classes in high-dimensional space.
- K-Nearest Neighbors (KNN): A non-parametric algorithm that assigns labels based on the majority sentiment among the k-nearest labelled instances.

The annotated dataset was split into training 80% and testing 20% subsets to evaluate each model's generalization capability.

3.6Model Evaluation

Model performance was assessed using standard evaluation metrics:

- Accuracy: The proportion of correctly classified samples.
- Precision: The proportion of correctly predicted positive instances among all predicted positives.
- Recall: The proportion of correctly predicted positive instances among all actual positives.
- F1-Score: The harmonic mean of precision and recall, offering a balanced measure.
- Confusion Matrix: A detailed breakdown of correct and incorrect classifications for each sentiment class.

These metrics were computed for each model to enable comparative performance analysis and identify the most suitable algorithm for the dataset.

3.7 Comparative Analysis

Among the four models evaluated, SVM achieved the highest accuracy at 88.99%, followed closely by Random Forest. Naïve Bayes produced consistent results, especially in precision, while KNN performed relatively poorly, likely due to its sensitivity to sparse, high-dimensional data such as that produced by TF-IDF.

Results

Figure 2 shows YouTube video content related to ethnic violence between the Meitei and Kuki communities in Manipur. Initial data from 2023 recorded 92 videos, coinciding with the outbreak of violence. A marked increase of 248 videos was observed in 2024, potentially linked to heightened conflict intensity and expanded coverage by Indian mainstream media outlets. By mid-April 2025, an additional 211 videos had been uploaded within just 3.5 months, more than the total count for 2023. This advocates that the keywords "Manipur Violence", Manipur Conflict or "Meitei and Kuki conflict" were widely used to describe the ethnic confrontation.

4.1 Top 10 Most Viewed YouTube Videos (Among 551) on Manipur Conflict:

This table shows the most-viewed YouTube videos—top 10 out of a dataset of 551—related to the 2023–2024 Manipur violence between the Meitei and Kuki communities. These videos include travel vlogs, news reports, interviews, political commentary, and explainer content. The most-viewed video, a vlog by Bong Travel Mania, garnered nearly 30 million views, showing how non-news content can also gain traction during significant events. News outlets such as *TIMES NOW Navbharat* and *BBC News Hindi* provided victim interviews and on-ground reports, offering insights into the humanitarian aspects of the crisis. Political response was also documented, such as *The Indian Express's* video of Manipur MP *Angomcha Bimol Akoijam* fiercely criticizing the central government in Parliament. These diverse video narratives, measured by millions of views, tens of thousands of likes, and thousands of comments, reflect the emotional and political resonance of the Manipur violence with the Indian public. The analysis highlights how YouTube functions not only as a platform for entertainment but also as a crucial site for civic engagement, awareness-building, and digital witnessing during times of crisis.

4.2 Sentiment Analysis

In the sentiment analysis of 124,971 YouTube comments related to the Meitei-Kuki conflict, many comments were categorized as *neutral* by both the custom polarity scoring method and the VADER sentiment analyzer Figure 3. This predominance of neutral sentiment arises from several factors. Many comments are descriptive, interrogative, or informational, presenting events, asking questions, or offering observations without overt emotional content. Additionally, a large number of comments exhibit *mixed sentiments*, such as criticizing violence while calling for peace or expressing sympathy for victims while denouncing political actors. These blended tones often result in sentiment scores that fall within the neutral classification thresholds used by tools like VADER. Furthermore, sarcasm, irony, or regional slang, common in Indian social media discourse, often go undetected or are misinterpreted by rule-based sentiment tools designed for standard English. The presence of transliterated language (e.g., Hinglish), non-standard spellings, and cultural idioms can also weaken polarity detection.

Interestingly, the dataset also reveals that *positive* sentiment is more prevalent figure 4 than *negative* despite the inherently violent and tragic nature of the event figure 5. This can be understood in several ways. Many users express *hope* and *call for peace*, *solidarity*, or *support for affected communities*, which are linguistically classified as positive even if the underlying context is tragic. For instance, phrases like “We stand with the people of Manipur” or “Peace will prevail” are inherently positive in tone. Additionally, comments that praise the army, humanitarian aid, or political efforts are often scored positively by sentiment tools. This highlights the tone in public discourse where positivity does not always imply the event's approval but reflects *resilience*, *hope*, or *constructive attitudes* in the face of a crisis. The researcher also checked the polarity score; each comment was assigned a polarity score between -1 (most negative) and +1 (most positive) using TextBlob and VADER. A histogram was plotted to visualize sentiment intensity distribution Figure 6. Extreme polarity values were relatively rare, suggesting that most of the discourse was balanced or fact-based rather than emotionally charged.

These findings emphasize the need to contextualize sentiment scores within socio-political realities and linguistic practices, particularly when analyzing user-generated content in conflict scenarios.

The performance of four machine learning models Support Vector Machine (SVM), Random Forest, Naïve Bayes, and K-Nearest Neighbors (KNN) was evaluated for sentiment analysis of YouTube comments. The comments were classified into three sentiment categories: negative, neutral, and positive. The evaluation was based on several metrics: precision, recall, F1-score, and accuracy. The confusion matrices for each model were also analyzed better to understand their classification behaviour across the sentiment categories.

4.3 Model Performance Evaluation.

To assess the effectiveness of different classification algorithms for multiclass sentiment analysis, we evaluated Support Vector Machine (SVM), Random Forest, Naïve Bayes, and k-nearest Neighbors (KNN) on a dataset comprising three sentiment classes: Negative, Neutral, and Positive. Table 1 summarizes the comparative performance of these models in terms of overall accuracy and macro-averaged evaluation metrics (precision, recall, and F1-score), which are particularly informative for datasets with class imbalance table 2.

Table 2 model comparison

Model	Accuracy	Macro Precision	Macro Recall	Macro F1-Score
SVM	0.9342	0.93	0.92	0.92
Random Forest	0.9159	0.91	0.89	0.9
Naïve Bayes	0.8403	0.85	0.8	0.82
KNN	0.545	0.75	0.46	0.43

4.3.1 SVM:

SVM exhibited the strongest overall performance accuracy of 93%, with a macro F1-score of 92% and high class-specific precision and recall. The confusion matrix (Figure 7) demonstrates a clear diagonal dominance, with most predictions aligning with the actual class, especially for Neutral and Positive sentiments. Only minimal confusion is observed between the Negative and Positive classes, a common challenge due to overlapping semantic structures in sentiment-bearing phrases (e.g., sarcasm or subtle polarity shifts).

This robustness can be attributed to the SVM's ability to construct maximally separating hyperplanes in high-dimensional feature space and its effectiveness with sparse, TF-IDF or word embedding-based text representations.

4.3.2 Random Forest:

Random Forest also delivered strong classification performance with 91% accuracy and (macro F1 = 90%), with high recall for Neutral and Positive classes. However, the confusion matrix Figure 8 shows slight overgeneralization between Negative and Positive classes, likely due to correlated features and class overlaps in the feature space.

While Random Forest is generally robust due to its ensemble nature, the presence of multiple weak learners may introduce "majority voting inertia", where ambiguous samples gravitate toward dominant classes (e.g., Positive), especially when the classes are not perfectly balanced or linearly separable.

4.3.3 Naïve Bayes:

Naïve Bayes, though achieving acceptable performance (macro F1 = 82%), displays an explicit limitation in separating Negative and Positive sentiments Figure 8. The model frequently confuses Negative reviews as either Neutral or Positive directs consequence of its strong independence assumption, which fails to capture co-occurring sentiment-bearing phrases or word dependencies (e.g., "not good" vs "good").

This shortcoming is further exacerbated in scenarios where class priors dominate the posterior probability estimation, leading to bias toward more frequent classes such as Neutral.

4.3.4 KNN:

The KNN model shows substantial misclassification, with most samples predicted as Neutral regardless of actual class Figure 9. Specifically, only 450 of 2842 Negative samples are correctly classified, while over 2300 are misclassified as Neutral. A similar trend is observed for Positive samples.

This performance collapse (macro F1 = 43%) stems from KNN's non-parametric, instance-based nature, which makes it particularly vulnerable in high-dimensional vector spaces common in NLP. Without dimensionality reduction or distance-weighted voting, KNN struggles with the curse of dimensionality, leading to poor local neighbourhood quality and an overprediction of the majority class (Neutral in this dataset).

Overall, SVM and Random Forest emerged as the most effective classifiers for this sentiment analysis task, demonstrating high accuracy and balanced performance across all classes.

Findings

This study conducted a large-scale sentiment analysis of 124,971 YouTube comments drawn from 551 videos related to the Meitei and Kuki conflict in Manipur, employing traditional machine learning models such as Support Vector Machine (SVM), Random Forest, Naïve Bayes, and K-Nearest Neighbors (KNN). Among these, SVM achieved the highest classification accuracy at 93.42%, followed by Random Forest at 91.59%, while KNN underperformed, likely due to its lower resilience to high-dimensional and sparse data typical of social media texts. A significant proportion of comments were categorized as neutral, which can be attributed to the mixed emotional tone, descriptive rather than opinionated phrasing, or the linguistic complexity that often characterizes user-generated content. Remarkably, the sentiment distribution revealed a slightly higher proportion of positive sentiments than negative ones, an unexpected trend given the conflict's violent backdrop. This may reflect a digital environment where users were more inclined to share empathetic, hopeful, or solidaristic responses rather than hostility or outrage.

To situate these findings within the broader landscape of conflict-related sentiment analysis, it is important to note that similar patterns have emerged in other geopolitical contexts. For instance, in a study on the Israel-Palestine conflict, Rahardi et al. (2024) reported that SVM outperformed Naïve Bayes, achieving 91% accuracy compared to 85%. This mirrors the results observed in the Manipur conflict, reaffirming SVM's robustness in handling emotionally nuanced and politically sensitive discourse. Likewise, in the context of the Russia-Ukraine conflict, Wadhwani et al. (2023) compared models such as Logistic Regression, Decision Trees, and Extra Trees Classifier. Their study found that the Extra Trees Classifier, when paired with Bag of Words features, achieved an accuracy of 84%, highlighting the significance of feature engineering in maximizing model performance.

Furthermore, an examination of the top 10 most-viewed videos in the dataset revealed a diverse mix of content types, including news reports, expert analyses, interviews, and personal vlogs. This diversity underscores YouTube's growing role not just as an information-sharing platform but as a dynamic space for emotional expression and public discourse, especially during periods of socio-political unrest. Altogether, these insights demonstrate that traditional machine learning models continue to offer reliable and scalable solutions for understanding public sentiment in digitally mediated conflict environments across different cultural and geopolitical settings.

5.0 Limitations

The sentiment analysis is limited to three classes (positive, neutral, negative) and does not capture more nuanced emotions like fear, anger, or hope. The study focuses solely on English and Hinglish comments; regional language content may have been excluded due to tool limitations. Sarcasm, irony, and cultural references often go undetected by sentiment tools like VADER and TextBlob, leading to possible misclassification. The analysis does not account for the influence of YouTube's recommendation algorithm, which may shape comment visibility and content popularity.

Recommendations

Future research could explore multilingual sentiment models that better handle regional languages and transliterated scripts common in Indian digital discourse. Incorporating deep learning models (e.g., BERT, RoBERTa) may enhance performance in detecting complex sentiment and emotion. A qualitative discourse analysis of a sample of comments could complement the quantitative findings and provide richer contextual insights. Further research can examine the role of recommendation algorithms and platform design in influencing narrative exposure and audience engagement during conflicts.

Originality of the Study

This study is among the first to apply machine learning-based sentiment analysis to YouTube comments explicitly related to the 2023–2024 Meitei–Kuki conflict in Manipur. Combining real-time social media data with multiple classifiers and incorporating metadata and comment-level insights, the research contributes an original methodology for analyzing regional conflict narratives through digital platforms. It also integrates computational and socio-political perspectives, bridging a gap between data science and conflict studies.

Conclusion

According to this study, YouTube is a crucial medium for documenting public opinion and conversation amid sociopolitical disputes. The promise of computational techniques in comprehending digital public opinion is demonstrated by this study's application of sentiment analysis to user-generated content. SVM and Random Forest outperformed all other classification models examined, yielding reliable results for unstructured, multilingual data. Additionally, the analysis demonstrates that emotional expression during conflict is complex and frequently combines pleas for peace, empathy, and criticism. By providing a way to track sentiment in real-time and comprehend how digital platforms reflect persistent regional tensions

References

1. Akoijam, B. (2005). *Another 9/11, another act of terror: the 'embedded disorder' of the*
2. *AFSPA*. 'Sarai Reader 2005: Bare Acts'. CSDS, 481-491**.**
3. Alhur, M. S., Alshamari, S., Oláh, J., & Aldreabi, H. (2022). Unsupervised Machine Learning to Identify Positive and Negative Themes in Jordanian mHealth Apps: *International Journal of E-Services and Mobile Applications*, *14*(1), 1–21. https://doi.org/10.4018/IJESMA.313950
4. Arambam, N. & Kangujam S. (2016). *Colonialism and Resistance: Society and State in Manipur.* Indian books and Periodicals
5. Bashir, S. (2021). Twitter chirps for Syrian people: Sentiment analysis of tweets related to Syria Chemical Attack. *International Journal of Disaster Risk Reduction*.
6. Das, P., 2023. *The Unfolding Kuki–Meitei Conflict in Manipur*, Manohar Parrikar Institute for Defence Studies and Analyses. India. Retrieved https://coilink.org/20.500.12592/d6q2qw
7. 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*, 09732586211049232. https://doi.org/10.1177/09732586211049232
8. Lee, C. S., Osop, H., Goh, D. H.-L., & Kelni, G. (2017). Making sense of comments on YouTube educational videos: A self-directed learning perspective. *Online Information Review*, *41*(5), 611–625. https://doi.org/10.1108/OIR-09-2016-0274

9. McGarry, K. (2023). Analyzing Social Media Data Using Sentiment Mining and Bigram Analysis for the Recommendation of YouTube Videos. *Information*, *14*(7), 408. https://doi.org/10.3390/info14070408
10. Mir, A. A., & Sevukan, R. (2024). Sentiment analysis of Indian Tweets about Covid-19 vaccines. *Journal of Information Science*, *50*(5), 1308–1320. https://doi.org/10.1177/01655515221118049
11. Oinam, B. (2003). Patterns of Ethnic Conflict in the North-East: A Study on Manipur. *Economic and Political Weekly*, *38*(21), 2031–2037. https://www.jstor.org/stable/4413594
12. Rahardi, M., Riamrizal Arsya, M. R., Aminuddin, A., Abdulloh, F. F., Miah, M. B. A., & Pratikno, H. (2024). *Sentiment Analysis of Israeli-Palestinian Conflict on Indonesian Tweets Using Machine Learning*. 59–64. https://doi.org/10.1109/icecos63900.2024.10791149
13. Rajkumari, B., Ingudam, D., Yengkokpam, C., Oinam, G., & Yumnam, R. (2024). Psychological effects of Manipur violence among the internally displaced persons residing in relief camps across Imphal valley of Manipur – A cross-sectional study. *Journal of Family Medicine and Primary Care*, *13*(10), 4354. https://doi.org/10.4103/jfmpc.jfmpc_341_24
14. Royal Thimphu College, Bhutan, & Khamrang, L. (2023). Manipur Mayhem: A Latent Enmity, Generalized Violence and State's Apathy. *Journal of Management & Public Policy*, *15*(1), 4–15. https://doi.org/10.47914/jmpp.2023.v15i1.001
15. Saha Roy, S. (2024). The grammar of a hunger strike: Nonviolence and biopolitics in Manipur, India. *Journal of the Royal Anthropological Institute*, *30*(4), 1027–1045. https://doi.org/10.1111/1467-9655.14148
16. Singar, S. (2023). Identity Struggles: Unraveling the Manipur's Ethnic Turmoil. *Indian Journal of Integrated Research in Law*, *3*, 1. https://heinonline.org/HOL/Page?handle=hein.journals/injloitd4&id=2185
17. Singh, R., & Tiwari, A. (2021). YouTube Comments Sentiment Analysis. *International Journal of Scientific Research in Engineering and Management (IJSREM)*, *5*(5). https://www.researchgate.net/publication/351351202
18. Sinha, A., Rout, B., Mohanty, S., Mishra, S. R., Mohapatra, H., & Dey, S. (2024). Exploring Sentiments in the Russia-Ukraine Conflict: A Comparative Analysis of KNN, Decision Tree And Logistic Regression Machine Learning Classifiers. *Procedia Computer Science*, *235*, 1068–1076. https://doi.org/10.1016/j.procs.2024.04.101
19. Verma, N. K., Nazim, M., & Verma, M. K. (2023). Unveiled Public Sentiments on Balasore (India) Train Tragedy: An analysis based on YouTube videos' comments. *World Digital Library*.*16*(2), 97–118. http://dx.doi.org/10.18329/09757597/2023/16208
20. Wadhwani, G. K., Varshney, P. K., Gupta, A., & Kumar, S. (2023). Sentiment Analysis and Comprehensive Evaluation of Supervised Machine Learning Models Using Twitter Data on Russia–Ukraine War. *SN Computer Science*, *4*(4). https://doi.org/10.1007/s42979-023-01790-5

Table 1

Table 1 is available in the Supplementary Files section.

Figures

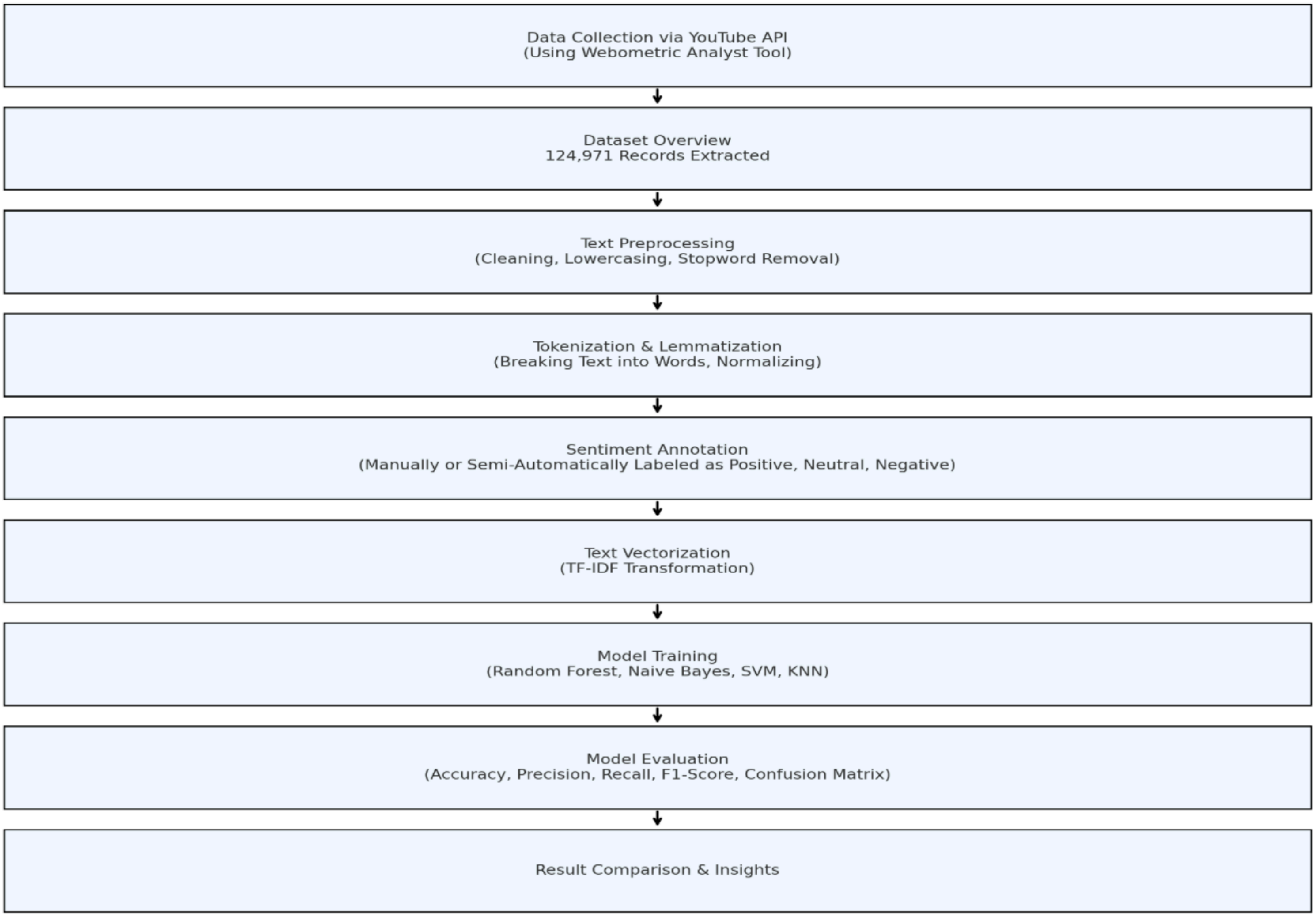


Figure 1

Flow chart for the sentiment analysis

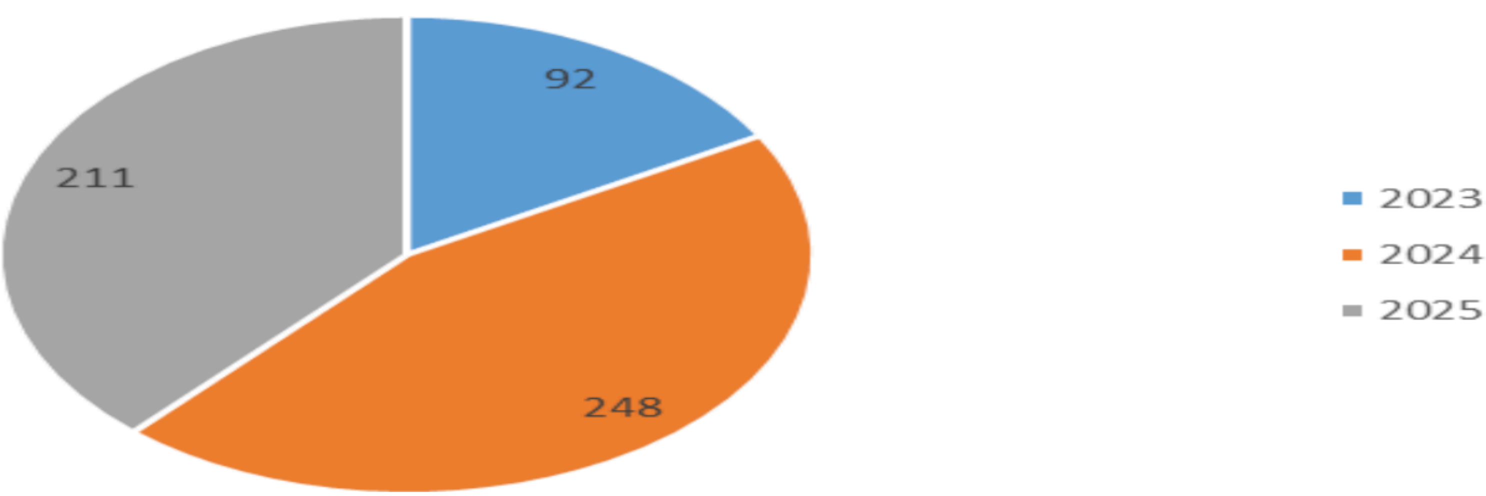


Figure 2

Number of videos published

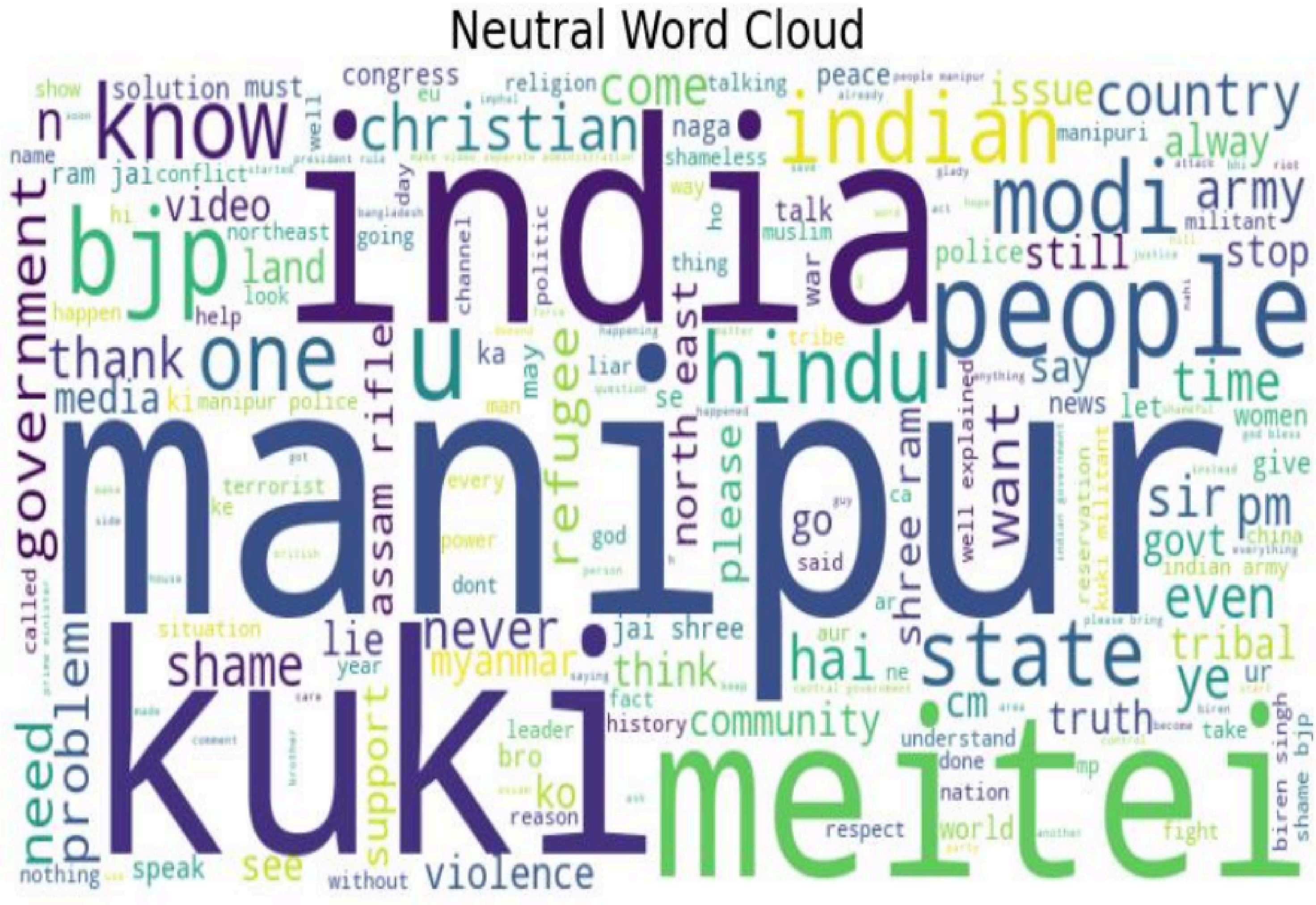


Figure 3

word cloud for neutral sentiment of comments

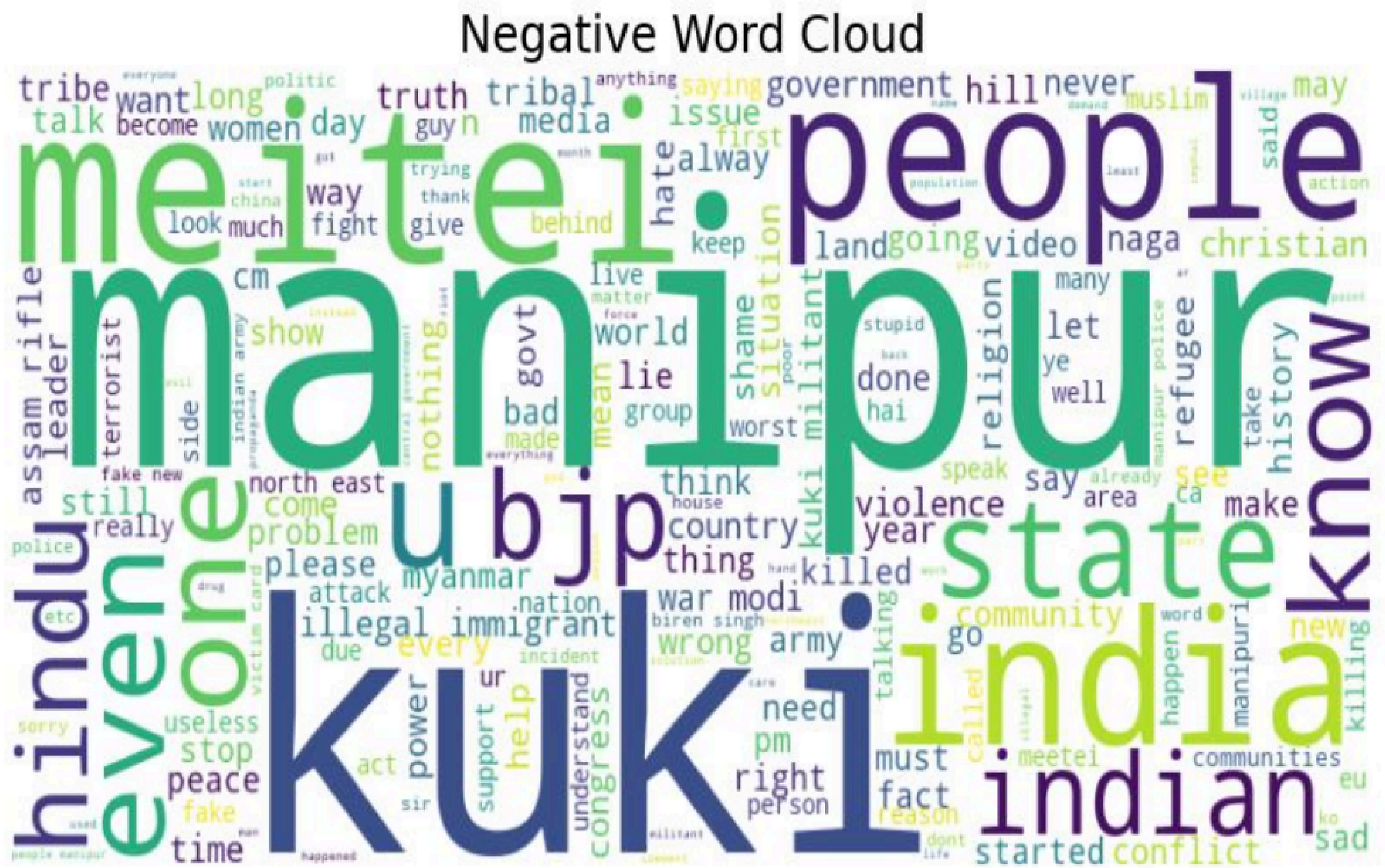


Figure 4

Word cloud for negative sentiment of comments

Figure 5

Word cloud for positive sentiment of comments

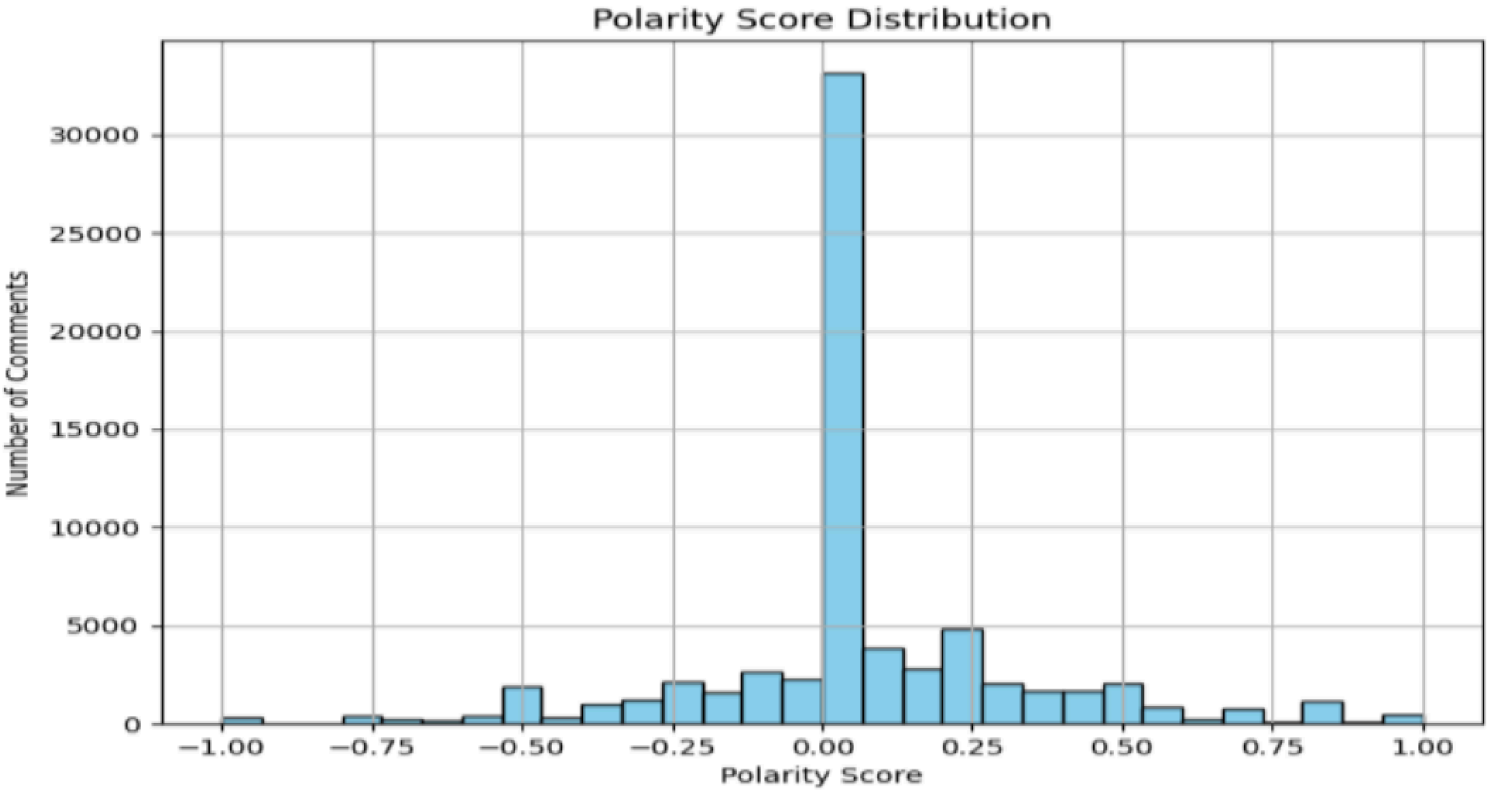


Figure 6

Polarity score of the sentiments

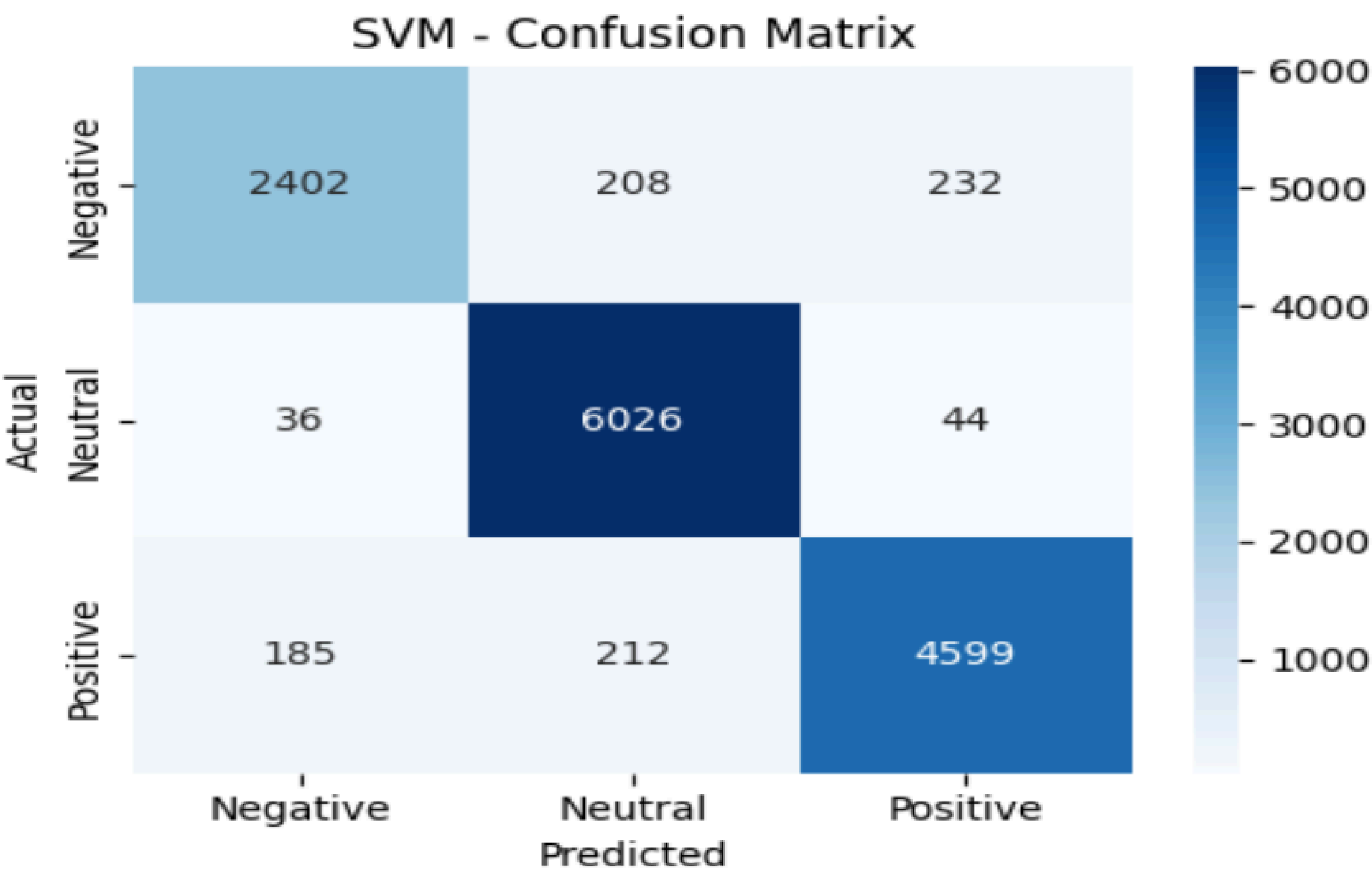
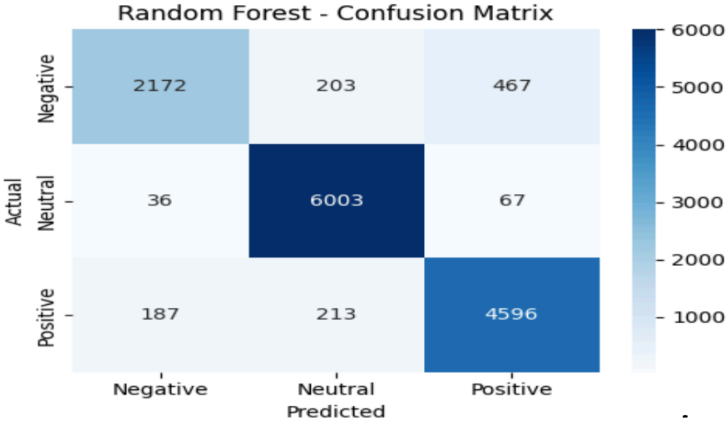


Figure 7

Confusion Matrix - Random Forest.



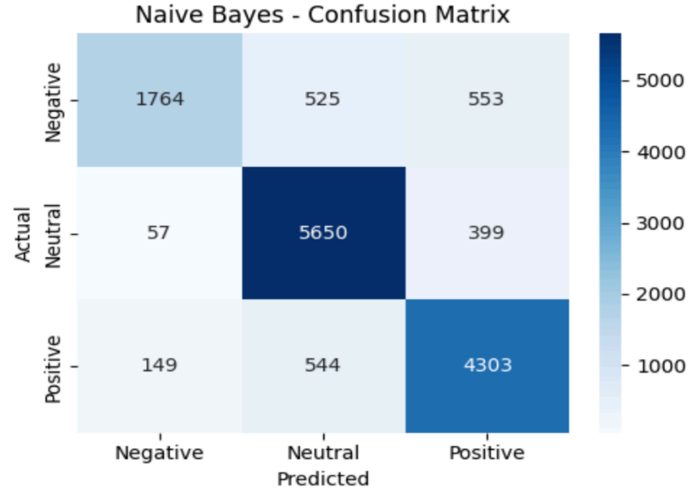


Figure 8

Confusion Matrix - Random Forest

Confusion Matrix - Naïve Bayes.

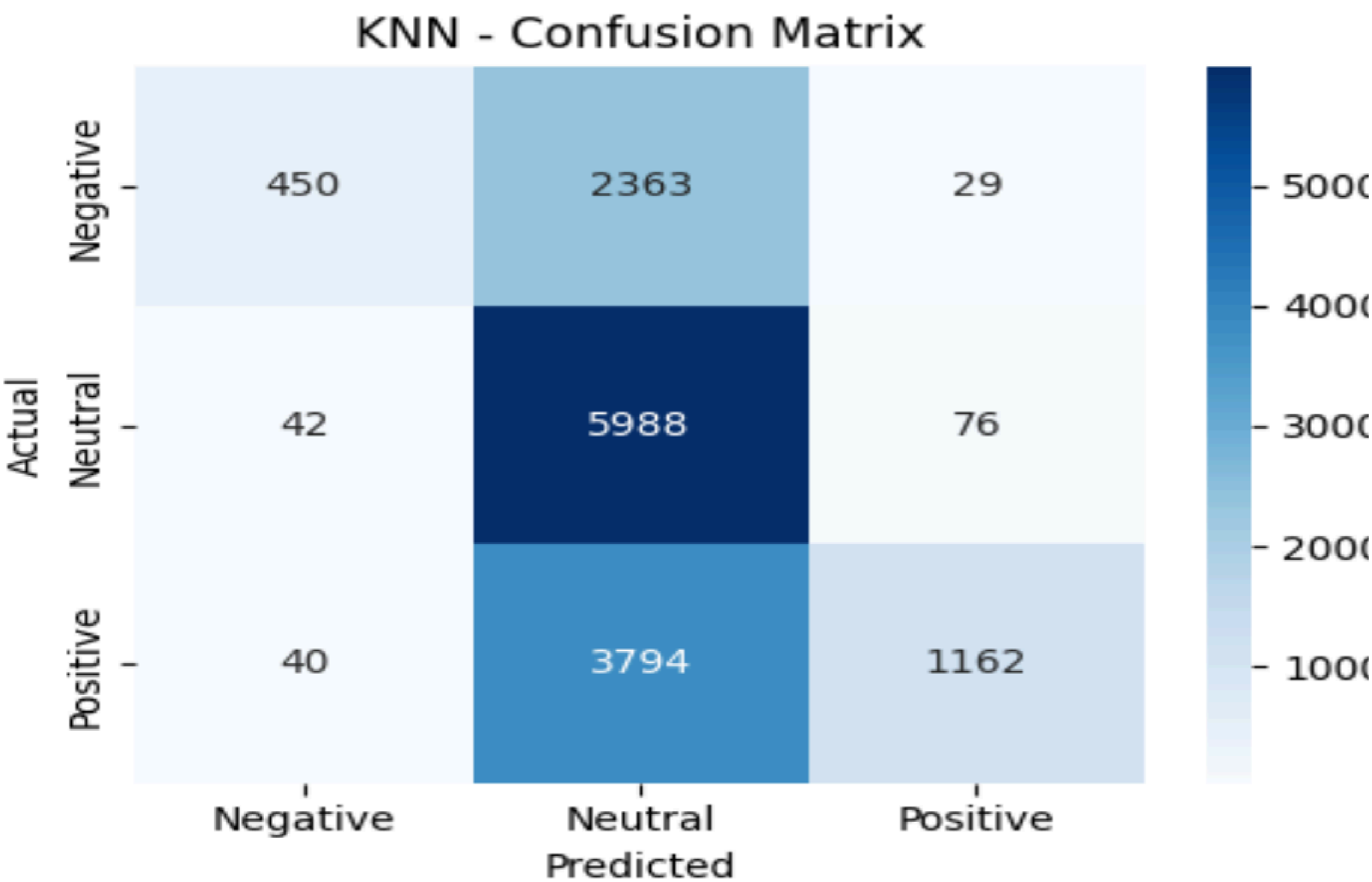


Figure 9

Confusion Matrix - KNN.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Table1Top10MostViewedYouTubeVideos.docx