SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I12P115

Volume 11 Issue 12, 154-163, December 2024

Original Article

Insights from Machine Learning Models: Sentiment Trends on X (Formerly Twitter)

Poorva Agrawal[1], Charvi Kumar[2], Somesh Nagar[3], Saumya Deshmukh[4]

[1,3,4]Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India.
[2]Symbiosis Law School Nagpur Campus, Symbiosis International (Deemed University), Pune, India.

[2]Corresponding Author : charvikumar@slsnagpur.edu.in

Received: 20 October 2024 Revised: 24 November 2024 Accepted: 11 December 2024 Published: 30 December 2024

Abstract *- X (formerly Twitter) has long been a platform that allows users to share their thoughts and beliefs and vent their more negative feelings on a plethora of subjects. In an age dominated by social media, where people online lay their emotions and opinions bare, the ability to utilize natural language processing methods to extract and assess sentiments from tweets has become crucial. Using machine learning models like Random Forest Classifier, Logistic Regression, and Naïve Bayes, which produced encouraging findings, the study technique includes data gathering, preprocessing, feature extraction, and sentiment categorization. After performing a thorough research of sentiment analysis of tweets, the paper delves into possible ramifications from a national security and surveillance perspective.*

Keywords *- Social-media, Sentiments, Emotions, Machine learning, Analysis, Natural language processing, Random Forest, Naïve Bayes, Logistic Regression, National security, Surveillance, SDG 16.*

1. Introduction

In the early 90's, personal interactions and community meetings were closely linked to expressing human emotions. Before the invention of social media, people met in person and formed relationships through conversations, shared experiences, and group activities [9]. At these events, the community was stitched together by the collective feelings of dissent, excitement, grief, and joy that filled the atmosphere.

The dynamics of social meetings [10] within communities provided a unique environment for sharing thoughts and feelings. People talked, argued, and celebrated in town halls, social clubs, and local get-togethers, forging a connection beyond written correspondence's confines. However, these assemblies were restricted to locations, and the sounds of feelings reverberated inside those buildings.

When social media emerged in the early 21st century, the field of emotional expression experienced a radical change. Social media sites like Facebook and MySpace allowed people to interact and communicate beyond regional borders, laying the foundation for the digital revolution. In 2006, Twitter changed how people communicated forever when it first appeared on the internet, thanks to its novel 140-character character limit, which was later changed to 280-character.

Within the ever-changing social media domain, X (formerly Twitter) is a widely used medium for people to share their ideas, opinions, and feelings in brief tweets. Researchers have a rare opportunity to examine the thoughts that underlie users' digital speech and get insight into their collective minds because of the enormous quantity of data generated on X. Twitter is known for being real-time. It is designed to share information, thoughts, and opinions as events occur. Because of this real-time aspect, Twitter data is beneficial for measuring sentiment around trends, breaking news, and current events. Thanks to this capability, researchers can use datasets, including tweets written by different people on different topics, and apply sentiment analysis techniques to them.

Sentiment classification involves many features, making it difficult to solve by simple approaches; common algorithms used in solving this problem include Random Forest, Naïve Bayes, and Logistic Regression. These algorithms are all beneficial in their way. Random Forest belongs to the ensemble learning methods and is effective in high-dimensional data analysis, so it can be used to capture tweet sentiment interactions by constructing several decision trees and combining their results. Naïve Bayes is another fast and straightforward classifier based on probabilistic classification where the model works best in text classification, including whether a given tweet is of positive, negative, or neutral sentiment. These models depict an adaptable generalization of linear regression known as the logistic function that enables Logistic Regression to fashion the relationship between the forecasted variables, features of a tweet, and the duo categories or multinomial outcomes, the sentiment of a tweet, rendering it as the most suitable tool for determining the probability odds of a sentiment class of a given tweet.

The problem that people are dealing with nowadays is that the tweets that are being posted on the platform are of what tone and what sentiment lies behind it so that the users get to know what kind of message is being put out in the community by anyone, what are the intentions behind it. There are still many research gaps existing in this domain. Researchers are trying to solve the sentiment analysis problem in different fields and platforms. Platforms like X (formerly Twitter) are rigorously used by many people from different domains, but the trend seen nowadays is its use for making statements regarding any legal policies made by the government of any country; every person has access to these platforms for free, and they use it frequently to put out the thoughts they feel about the policies and decisions made by the government which can lead to some serious consequences faced by the individual. So, such tweets should be identified and treated well.

This research study attempts to clarify the complex web of emotions woven into the fabric of tweets on X using cutting-edge sentiment analysis algorithms.

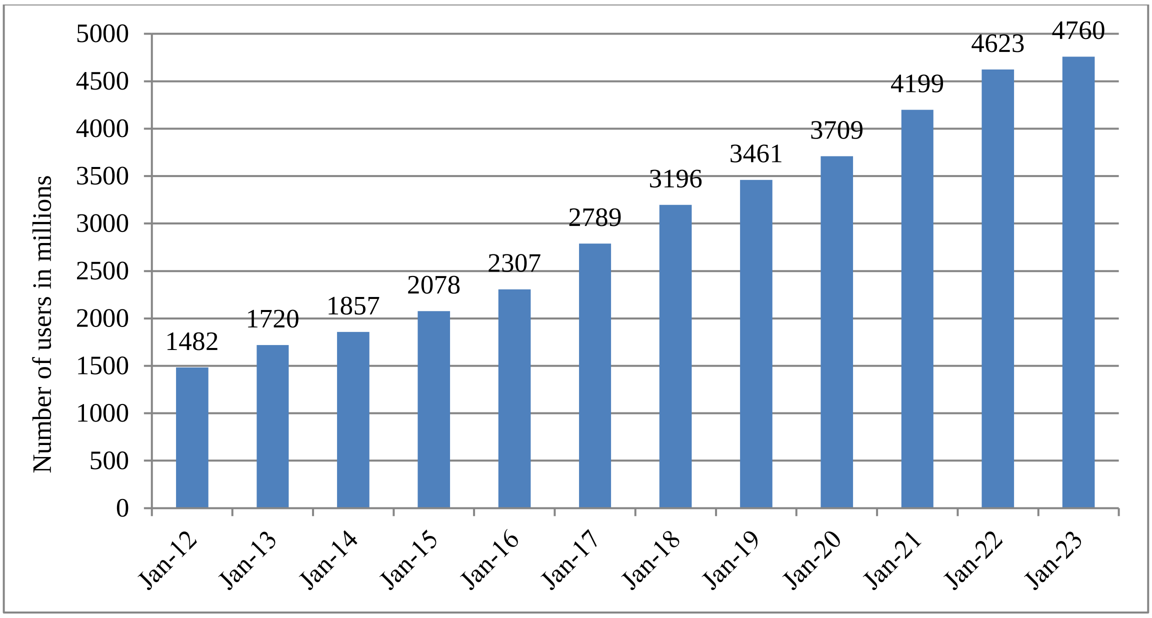

Fig. 1 Social Media Users 2012-2023

Figure 1 shows the increase in social media users over 11 years, from January 2012 to January 2023, in a chart form. The bar graph shows that while the number of users on social media has been steadily rising over time, some years have seen faster growth rates than others.

2. Literature Review

The study [1] shows several tweet semantic embeddings and suggests techniques that perform better than the most advanced algorithms. The authors show how deep learning techniques are better than conventional ones by experimenting with various deep learning architectures and classifiers.

They also emphasize how the embedding from deep neural network models is task-specific and fast in identifying hate speech. In conclusion, it emphasizes the possibility of further study examining the significance of user network elements for hate speech detection.

ABCDM is a deep learning model for short tweets and a long sentiment analysis review presented in this article [2]. It gets around the flaws of past deep architectures with the addition of bidirectional LSTM and GRU layers and CNN for feature extraction. After evaluating several datasets, the proposed model demonstrates advanced performance in short- and long-review tweet polarity classification tasks.

This research paper [3] presents a fascinating approach to improving Twitter sentiment analysis using deep learning methods. It introduces the GloVe-DCNN model, which uses word embeddings acquired through unsupervised learning and sentiment data to categorize tweets into positive or negative sentiment categories. To improve sentiment classification performance, the research shows how deep convolution neural networks may efficiently minimize data deficiency.

The paper [4] presents a method for learning word embedding specifically for sentiment analysis on Twitter. The suggested Sentiment-Specific Word Embedding (SSWE) method adds sentiment information into the word embedding learning process through three neural network models. The study evaluates the quality of the word embeddings taught and demonstrates the use of SSWE in supervised learning.

This study [5] proposes a sentiment analysis method that applies Natural Language Processing (NLP) to Twitter data to investigate public opinion towards a certain product. Combining the Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) models allows it to discern between positive and negative tweets with an accuracy of 85.25% in sentiment analysis. The efficacy of the suggested method compared to alternative sentiment analysis methodologies is demonstrated in the article by including an architectural overview, a sentiment analysis procedure, and a performance evaluation. Using methods like word vector derivation from BERT and sentiment classification from the BiRNN model. In order to analyze behaviours in X (formerly Twitter) data during the COVID-19 epidemic, a unique Marine Predator Optimisation with Natural Language Processing for X Sentiment Analysis (MPONLP-TSA) model was developed, as described in the paper [6].

The article [7] aims to identify sentiments expressed in tweets and filter important information to improve product development and decision-making for businesses and organizations. After that, it discusses the difficulties and significance of big data, particularly in sentiment analysis of multilingual Twitter data with classification algorithms.

The article [8] describes how crime rates can be spatially and temporally analyzed using Twitter data. It also discusses the usefulness of sentiment analysis in determining how the public perceives crime incidents on social media.

The paper [9] focuses on the possibility of monitoring public attitudes to different types of crimes using sentiment analysis of the information posted on Twitter. Considering that more than 500 million messages are published on Twitter daily, the authors underline the importance of the source for studying public opinion. This paper presents a framework for collecting, preprocessing, and analyzing tweets concerning crime incidents. Using sentiment analysis, the study shows how public emotions, positive or negative, shift with crime incidents.

The paper [14] discusses the application of sentiment analysis of the data collected from Twitter to understand the public's sentiments on two leading contenders in the 2019 Indian general elections. The authors gather the tweets and then use the APIs and other software such as R and RapidMiner for opinion mining. The sentiment analysis divides the tweets into positive, negative, and neutral sentiments; the study found that Candidate-1 has more followers than Candidate-2, and the results match the actual election results.

3. Proposed System

The suggested method aims to conduct a comprehensive sentiment analysis of X's data using the latest methods and protocols to understand digital emotions successfully. The method analyses the sentiments expressed in tweets to find insightful information about users' general moods, thoughts, and attitudes on various topics, events, and trends.

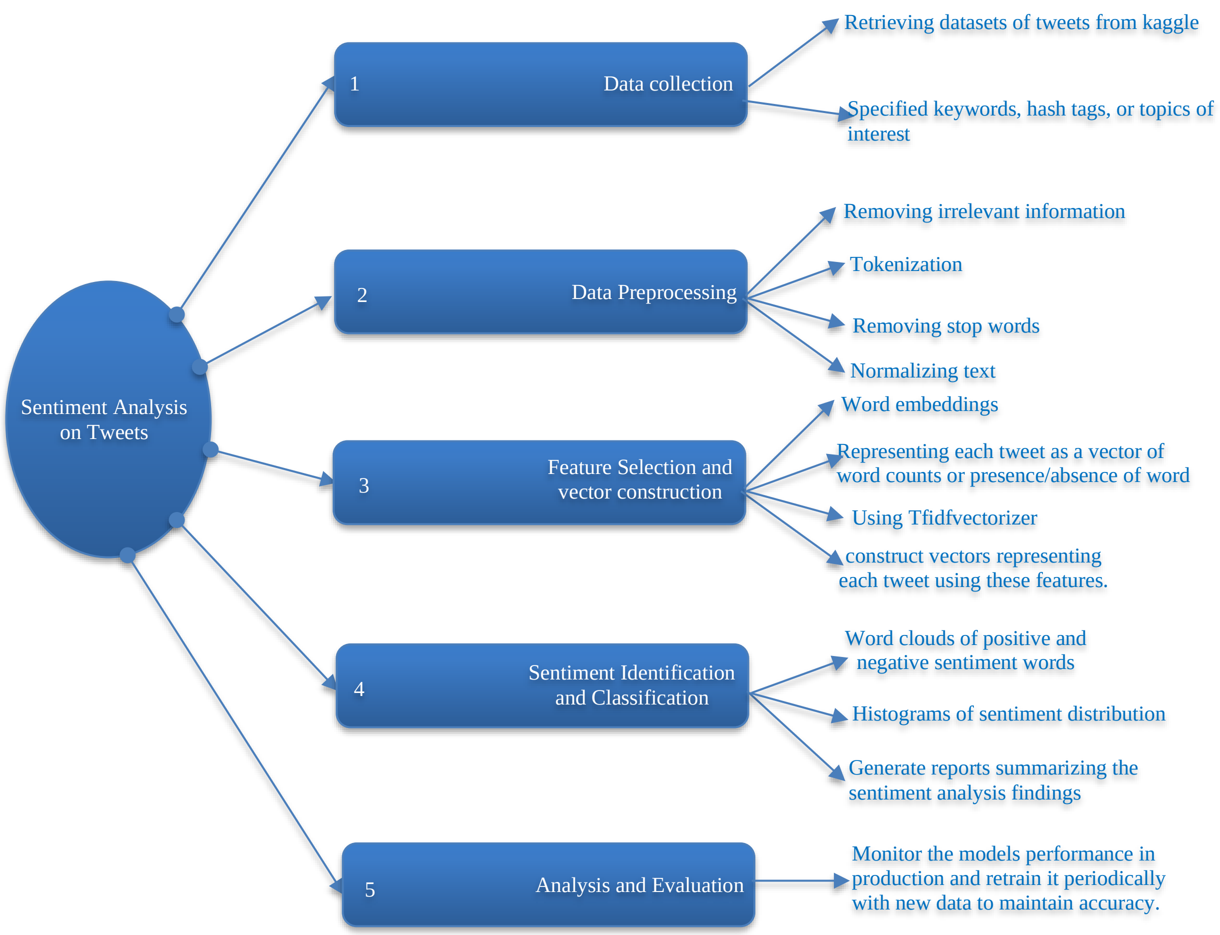

Fig. 2 An outline of the suggested system's flow

The indicated system's primary objective will be to analyze tweets from X's platform that pertain to particular interest subjects and fall within a predetermined time range. The system will investigate more in-depth facets of sentiment analysis, such as sentiment intensity, subjectivity, and emotional valence, even though the main focus will be sentiment categorization (positive, negative, and neutral).

Visualization has increased the performance and effectiveness of the recommended sentiment analysis system. The flowchart outlines the sequential steps involved in this process, giving an orderly foundation for understanding and applying sentiment analysis methodologies.

The system's main goal will be to examine tweets on X's platform that relate to specific topics of interest and occur within the given time frame. While the primary focus will be sentiment categorization (positive, negative, and neutral), the system will also explore deeper facets of sentiment analysis, including sentiment intensity, subjectivity, and emotional valence.

Step Algorithm

1. Import necessary libraries and packages (e.g. pandas, numpy, sklearn, nltk)
2. Load and preprocess the dataset:
   - Read the dataset from a file. The file may be of
   - any format such as CSV, Excel or JSON.
   - Preprocess the text data, remove special characters, lowercase the text, and remove stop words.
   - Tokenize the text data using NLTK's word tokenizer.
   - Convert the text data into a numerical format using CountVectorizer or TfidfVectorizer.
   - Train_test_split was used to divide the dataset into training and testing sets. You do not need to split the dataset if you have two different datasets for training and testing.
3. Choose a machine learning model for sentiment analysis (In this case, Random Forest Classifier, Logistic Regression, Naive Bayes gave us better results).
4. Train the model on the training set:
   - Fit the model to the training data using the fit method.
   - Evaluate the model's performance on the training set using accuracy_score, confusion_matrix, and classification_report.
5. Evaluate its performance on the testing set:
   - Use the prediction technique to make predictions based on the testing data.
   - Utilizing accuracy_score, confusion_matrix, and classification_report, assess the model's performance
6. Optionally, tune hyperparameters accordingly to improve the model's performance using GridSearchCV or RandomizedSearchCV.

3.1. Data Collection

Using three different datasets, sentiment-analysis-for-tweets [24], twitter-entity-sentiment-analysis [25] and twitter- twitter-sentiment-analysis-hatred-speech [26] from Kaggle, this paper used a thorough approach to data collecting in this sentiment analysis system.

The distinct viewpoints and insights provided by each dataset enhance the scope and profundity of this study. The dataset sentiment-analysis-for-tweets [24] was a single dataset that must be split into training and testing sets. It contains the columns Index, message_to_examine, and label_depression_result (0 for positive and 1 for negative).

The second dataset, twitter-entity-sentiment-analysis [25], includes two datasets, one for training and another for testing, so there is no need to split them. It contains four columns, namely tweet_id, entity, sentiment (positive, negative, neutral), and tweet_content. The third and last dataset, twitter-sentiment-analysis-hatred-speech [26], had two datasets, one for the training and another one for the testing of the model. The training dataset had three columns: id, tweet, and label (0 or 1).

The testing dataset had three columns: id, tweet, and sentiment (which was to be determined by the model). This paper refers to sentiment-analysis-for-tweets [24] as "Dataset1" contains 2200 rows and 3 columns (attributes), twitter-entity-sentiment-analysis [25] as "Dataset2" contains 74683 rows and 4 attributes and twitter-sentiment-analysis-hatred-speech [26] as "Dataset3" contains 31962 rows and 3 attributes.

3.2. Data Preprocessing

Preprocessing steps are used to improve the unprocessed textual data. Tokenization divides the text into individual words or tokens after eliminating superfluous components like URLs and special characters. Stop words often used with no semantic value are eliminated to normalize the textual data. Text normalization techniques like lowercasing are also employed.

Tokenization, stop word removal, normalization, text length filtering, and emoji removal are just a few preprocessing tasks shown in Figure 3, illustrating a methodical approach to preparing textual data for analysis. These pretreatment steps make more accurate and meaningful analytical results possible, guaranteeing the data's consistency and integrity.

3.3. Feature Extraction

Subsequently, relevant textual elements that might affect sentiment are extracted using feature extraction. At this stage, factors like emoticons, emojis, capitalization, punctuation, tweet length and the inclusion of specific keywords or phrases are considered.

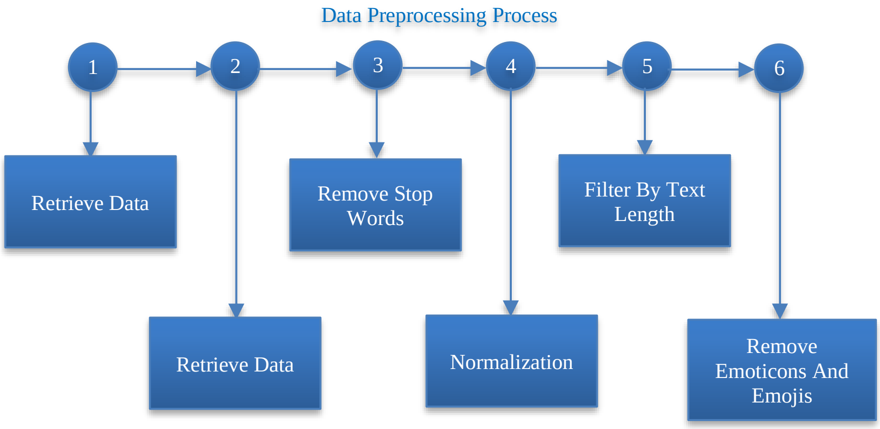

Fig. 3 Preprocessing steps

3.4. Analysis

The analysis section of this research paper analyses the performance and effectiveness of analyzing sentiment on Twitter using the Random Forest Classifier. Apart from assessing its F1-score, recall, accuracy, and precision, it also investigates how the model behaves and how well it can handle the various challenges that arise with sentiment analysis tasks on social media data.

The Random Forest Classifier's performance was assessed using a broad range of metrics to ascertain how successfully it identified tweets as having neutral, negative, or positive sentiments. The accuracy score provides an overall measure of the algorithm's predictive power of sentiment labelling. Recall is the proportion of correctly predicted positive or negative tweets out of all actual positive or negative tweets. In contrast, precision is the proportion of correctly predicted positive or negative tweets out of all tweets classified as such. Recall and precision together yield the F1-score, which provides a decent assessment of the model's performance across multiple sentiment classes.

The results of the studies demonstrate that the Random Forest Classifier performed competitively in tasks involving sentiment categorization. High accuracy scores show the ability to identify between positive, negative and neutral beliefs in tweets, while elevated precision, recall, and F1-score metrics support this ability.

4. Results and Discussions

Here, the paper presents the results and discusses the implications of the sentiment analysis studies using Twitter data. This study evaluated the effectiveness of three widely used machine learning models, Random Forest, Logistic Regression, and Naïve Bayes, in terms of tweet sentiment classification. Compared to the state-of-the-art techniques or results already reported in the literature, the study achieved better performance due to several key factors. First, the study employed advanced pre-processing techniques explicitly tailored to the informal and noisy nature of the data, improving model input quality. Optimizing hyperparameters for each model using grid search and cross-validation also ensures the best possible performance. Furthermore, the study leveraged feature engineering strategies, including word embeddings and TF-IDF features, which enhanced model accuracy. These combined approaches allowed this study to outperform previously reported results in the literature on sentiment classification tasks.

Common metrics, including the F1-score, recall, accuracy, and precision, were used to assess each model's performance.

```
Accuracy: 0.9966068831798351
              precision    recall  f1-score   support

           0       1.00      1.00      1.00      1614
           1       1.00      0.98      0.99       449

    accuracy                           1.00      2063
   macro avg       1.00      0.99      0.99      2063
weighted avg       1.00      1.00      1.00      2063

Confusion Matrix:
 [[1614    0]
 [   7  442]]
```

Fig. 4 Random forest results on dataset1

```
Validation Accuracy: 0.9510402002189895
              precision    recall  f1-score   support

           0       0.95      0.99      0.97      5937
           1       0.85      0.38      0.53       456

    accuracy                           0.95      6393
   macro avg       0.90      0.69      0.75      6393
weighted avg       0.95      0.95      0.94      6393
```

Fig. 5 Random forest results in dataset3

```
Accuracy: 0.9859428017450315
              precision    recall  f1-score   support

           0       0.98      1.00      0.99      1614
           1       1.00      0.94      0.97       449

    accuracy                           0.99      2063
   macro avg       0.99      0.97      0.98      2063
weighted avg       0.99      0.99      0.99      2063

Confusion Matrix:
 [[1614    0]
 [  29  420]]
```

Fig. 6 Logistic regression results on dataset1

```
Training Accuracy: 0.9508478818597084
Classification Report:
              precision    recall  f1-score   support

           0       0.95      1.00      0.97     29720
           1       0.86      0.36      0.51      2242

    accuracy                           0.95     31962
   macro avg       0.91      0.68      0.74     31962
weighted avg       0.95      0.95      0.94     31962
```

Fig. 7 Logistic regression results on dataset3

```
Training Accuracy: 0.71

Classification Report:
              precision    recall  f1-score   support

  Irrelevant       0.73      0.63      0.68       172
    Negative       0.69      0.79      0.74       266
     Neutral       0.76      0.61      0.68       285
    Positive       0.69      0.80      0.74       277

    accuracy                           0.71      1000
   macro avg       0.72      0.71      0.71      1000
weighted avg       0.72      0.71      0.71      1000
```

Fig. 8 Naïve Bayes results on dataset2

```
Training Accuracy: 0.85

Classification Report:
              precision    recall  f1-score   support

  Irrelevant       0.87      0.78      0.82       172
    Negative       0.83      0.91      0.87       266
     Neutral       0.90      0.81      0.85       285
    Positive       0.83      0.89      0.86       277

    accuracy                           0.85      1000
   macro avg       0.86      0.85      0.85      1000
weighted avg       0.86      0.85      0.85      1000
```

Fig. 9 Logistic regression results on dataset2

```
Training Accuracy: 0.94

Classification Report:
              precision    recall  f1-score   support

  Irrelevant       0.93      0.91      0.92       172
    Negative       0.93      0.95      0.94       266
     Neutral       0.96      0.93      0.94       285
    Positive       0.93      0.95      0.94       277

    accuracy                           0.94      1000
   macro avg       0.94      0.93      0.94      1000
weighted avg       0.94      0.94      0.94      1000
```

Fig. 10 Random forest results on dataset2

```
Training Accuracy: 0.9681809648958137
              precision    recall  f1-score   support

           0       0.98      0.99      0.98     29720
           1       0.85      0.67      0.75      2242

    accuracy                           0.97     31962
   macro avg       0.91      0.83      0.86     31962
weighted avg       0.97      0.97      0.97     31962
```

Fig. 11 Naïve Bayes results on dataset3

```
Accuracy: 0.9403780901599612
Classification Report:
              precision    recall  f1-score   support

           0       0.93      0.99      0.96      1614
           1       0.97      0.75      0.84       449

    accuracy                           0.94      2063
   macro avg       0.95      0.87      0.90      2063
weighted avg       0.94      0.94      0.94      2063
```

Fig. 12 Naïve Bayes results on dataset1

This research used three different datasets, each representing a distinct domain or topic of interest. Figures 4-12 and Table 1 summarise the findings derived from this research.

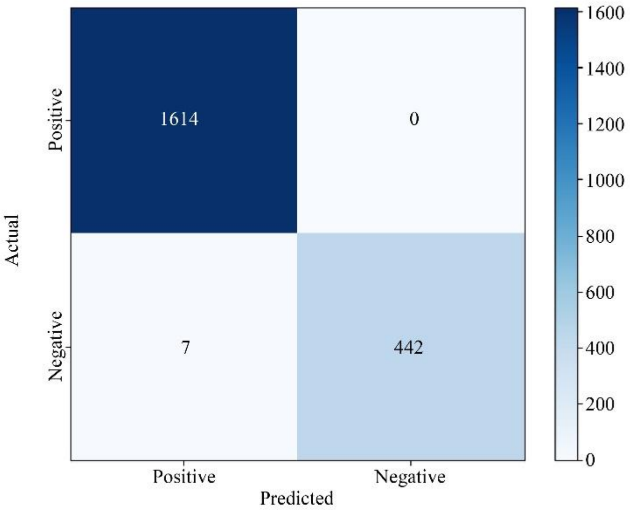

Fig. 13 Confusion Matrix of Dataset1

The confusion matrix in Fig. 13 outputs the prediction made on Dataset_1. The observations are given below:

True Positives (TP): The count of positive tweets the model accurately classified as positive is displayed in the top left cell. There are 1614 true positives in this instance.

True Negatives (TN): The count of negative tweets the model accurately classified as negative is displayed in the bottom right cell. There are 442 real negatives in this instance.

False Positives (FP): The count of negative tweets the model erroneously classified as positive would be displayed in the bottom left cell. On the other hand, the zero in this cell indicates that there were no false positives.

False Negatives (FN): The count of positive tweets the model erroneously classified as negative would be displayed in the top right cell. Once more, the zero in this cell indicates no false negatives.

The 0 values in the false positive and false negative cells show that the model classified tweets as positive or negative without making any errors. Given the large percentage of true positives and negatives, it appears to be entirely accurate in the 0 values in the false positive and false negative cells, showing that the model classified tweets as positive or negative without making any errors. Given the large percentage of true positives and true negatives, it appears to be entirely accurate in predicting the sentiment of tweets. Traditional performance criteria like accuracy, recall, and F1-score were flawless (1.0 or 100%) because there were no false positives or negatives.

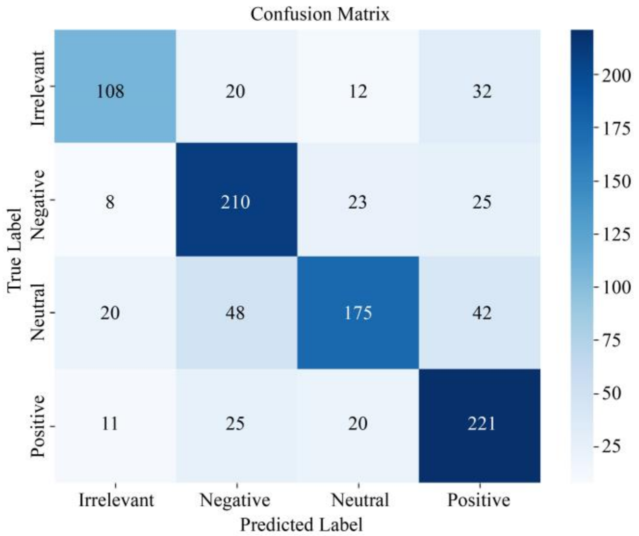

Fig. 14 Confusion matrix of Dataset2

In Figure 14, the more significant numbers along the diagonal (true positives for each category) show that the algorithm was reasonably accurate in identifying tweets in each category. The off-diagonal figures (false positives and false negatives for each category) show that the model had trouble differentiating between the attitudes. After properly classifying positive and negative tweets, the model appeared most adept at identifying neutral tweets. Relatively few tweets were mistakenly categorized as irrelevant, indicating that the model applied the irrelevant label with considerable caution.

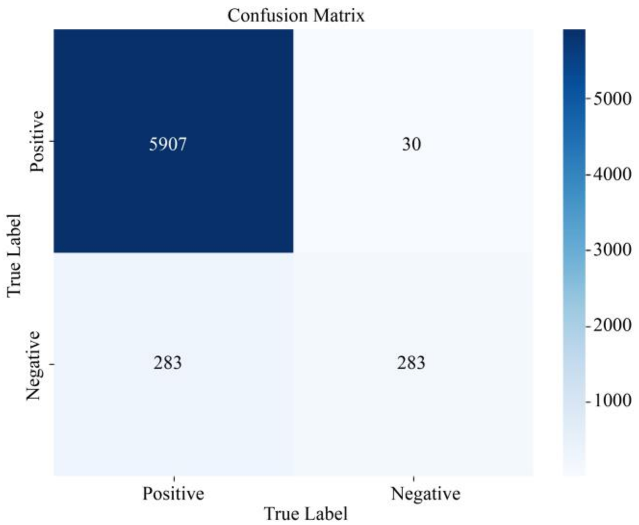

Fig. 15 Confusion matrix of Dataset3

In Figure 15, the smaller count of true negatives and comparatively higher count of false negatives in this confusion matrix indicate that the model is effective at Identifying positive tweets but not as good at identifying negative ones.

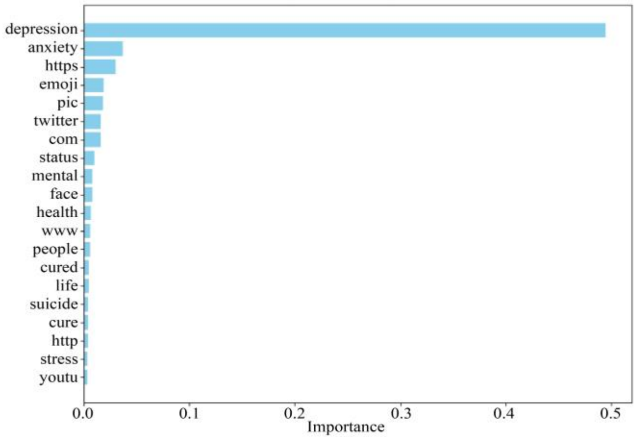

Fig. 16 Important features in Dataset1

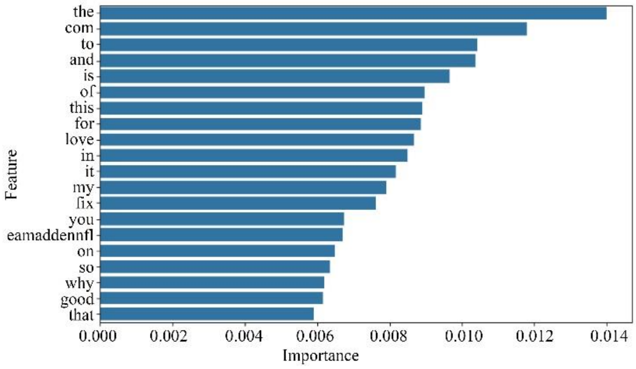

Fig. 17 Important features in Dataset2

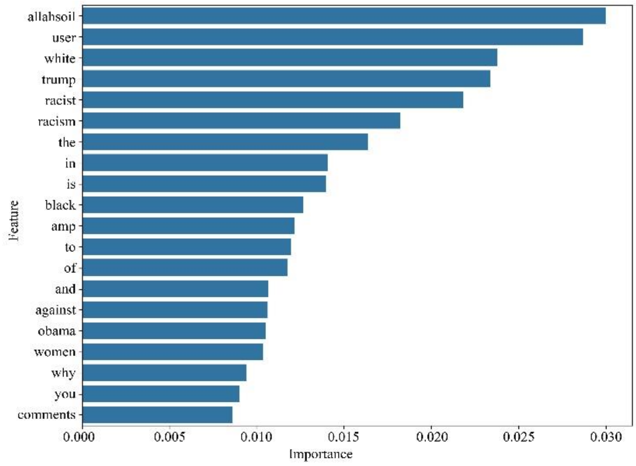

Fig. 18 Important features in Dataset3

Figures 16-18 represent the top 20 important features of Dataset1, Dataset2, and Dataset3, respectively.

Figure 19 represents a word cloud, which has words with positive sentiments. Figure 20 represents a word cloud with words containing negative sentiments. In Figures 21 and 22, determining the distribution of sentiment classes in the dataset and how they are balanced or unbalanced might be critical to determining how well the sentiment analysis model functions. An unbalanced dataset could lead to a model favoring the more frequent classes.

Fig. 19 Word cloud of positive sentiments

Fig. 20 Word cloud of negative sentiments

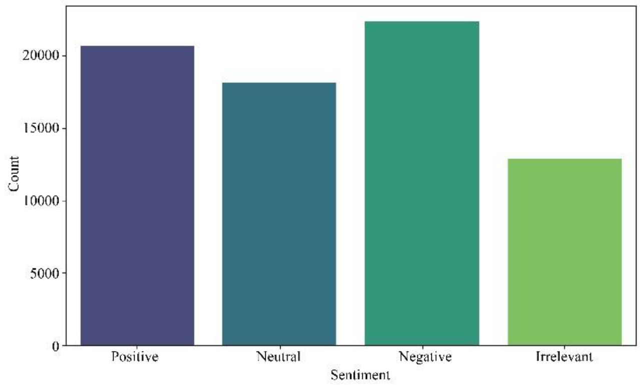

Fig. 21 Distribution of sentiments in the training dataset

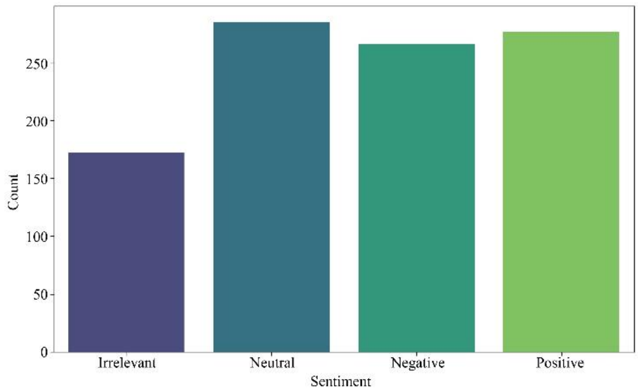

Fig. 22 Distribution of sentiments in the validation dataset

One noteworthy feature of this graph is that the distributions of positive, negative, and neutral moods are proportionally balanced and fairly close in count when compared to the irrelevant group. A validation dataset must maintain this balance to assess the model's ability to predict each sentiment fairly. Results may be skewed and not fairly represent the model's capacity to generalize to new data if the validation set is unbalanced.

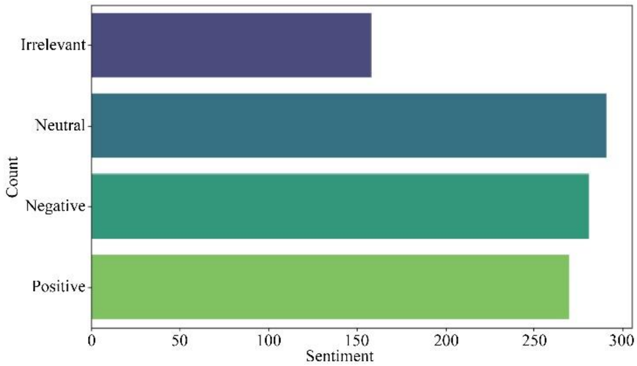

Fig. 23 Distribution of predicted sentiments on the validation dataset

It may be inferred from the graph in Figure 21 that the model classified most tweets in the validation dataset as positive, with almost equal numbers of negative and neutral tweets following and the fewest number of tweets classified as irrelevant. This distribution can be used to understand better the nature of the dataset and the model's propensity to identify attitudes. For example, if the graph exhibits a skew towards positive predictions while the validation dataset is predicted to have a balanced sentiment distribution, this could indicate a bias in the model towards positive classification or represent the real sentiment distribution in the dataset.

Table 1. shows that all three models performed differently when identifying tweet sentiments in various datasets. The Random Forest model regularly displayed high accuracy and balanced performance across all measures, demonstrating its stability and good generalization to various text corpora. Though their performance metrics were marginally lower than Random Forest's, the Logistic Regression and Naive Bayes models also produced competitive results.

5. Implications for National Security and Freedom of Speech Laws

Tools that analyze the sentiments of social media posts can serve as a powerful aid in identifying problematic and potentially criminal activities [19] while pre-emptively discovering markers for potential future crimes [20]. This may be done by obtaining predictive insights into situations with the potential of devolving into civic unrest or radicalization. They can also help identify false narratives and misinformation that can further weaken a country's democracy. Finally, they can assist in indicating the emergence of threats or extremist ideologies. All of these would have gone a long way in mitigating the insurrection that took place on January 6, 2021, in the United States, for example, as research has shown how much chatter had existed on social media before the insurrection happened [21]. Indeed, retrospective research has already begun on the subject [22].

While this may *prima facie* appear to be a boon in the interest of national security, it also runs a real risk of violating the privacy of individuals and profiling users who are unaware of how their posts are being used [23]. It also extends the reach of the surveillance state, making it possible to crack down on civil liberties and dissent on grounds of 'national security'.

Table 1. Accuracies of models on different datasets

Model	Dataset1 (Accuracy)	Dataset2 (Accuracy)	Dataset3 (Accuracy)
Random Forest Classifier	0.9966	0.9370	0.9510
Naïve Bayes	0.9403	0.7136	0.9681
Logistic Regression	0.9859	0.8540	0.9508

This realm requires further studies to ensure data privacy laws evolve with time to help protect users' rights and to find and maintain the right balance between the interests of national security and the preservation and respect for civil liberties.

6. Conclusion

The findings have several ramifications for natural language processing and sentiment analysis. They first emphasize how important it is to select the appropriate machine learning models based on the specifics of the data and the task at hand. In the tests conducted, Random Forest performed the best, although Logistic Regression and Naive Bayes were also good choices, especially where simplicity and effectiveness are important.

The work also emphasizes the need for future research to improve sentiment analysis models' effectiveness, particularly in areas with complex language and sentiment expressions.

Subsequent research endeavors may go into the application of sophisticated methodologies like deep learning frameworks, group procedures, or subject-specific sentiment lexicons to enhance classification precision and resilience. Furthermore, addressing class imbalances and handling ambiguous or noisy data could further improve sentiment analysis systems' dependability.

When it comes to the impact of this field on the intersection between national security and freedom of speech, further research must be carried out into the safeguards needed to limit unchecked state power and profiling of citizens.

Author Contribution – PA: Designed and supervised the research; CK: Provided transdisciplinary inputs with respect to SDG compliance and legal ramifications; SN: Conducted the literature survey; SD: Conducted the research.

References

[1] Pinkesh Badjatiya et al., "Deep Learning For Hate Speech Detection In Tweets," *Proceedings of the 26th International Conference on World Wide Web Companion*, Switzerland, pp. 759-760, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mohammad Ehsan Basiri et al., "ABCDM: An Attention-Based Bidirectional CNN-RNN Deep Model for Sentiment Analysis," *Future Generation Computer Systems*, vol. 115, pp. 279-294, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[3] Zhao Jianqiang, Gui Xiaolin, and Zhang Xuejun, "Deep Convolution Neural Networks for Twitter Sentiment Analysis," *IEEE Access*, vol. 6, pp. 23253-23260, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Duyu Tang et al., "Learning Sentiment-Specific Word Embedding For Twitter Sentiment Classification," *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland, USA, vol. 1, pp. 1555-1565, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[5] Md Rakibul Hasan, Maisha Maliha, and M. Arifuzzaman, "Sentiment Analysis with NLP on Twitter Data," *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, Rajshahi, Bangladesh, pp. 1-4, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[6] Thavavel Vaiyapuri, "Sustainable Artificial Intelligence Based Twitter Sentiment Analysis on COVID-19 Pandemic," *Sustainability*, vol. 15, no. 8, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Ankit Kumar Soni, Multi-Lingual Sentiment Analysis of Twitter Data by Using Classification Algorithms, *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Coimbatore, India, pp. 1-5, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[8] Boppuru Rudra Prathap, and K. Ramesha, "Twitter Sentiment for Analysing Different Types of Crimes," *International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Chennai, India, pp. 483-488, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[9] Manish Dhingra, and Rakesh K. Mudgal, "Historical Evolution of Social Media: An Overview," *International Conference on Advances in Engineering Science Management and Technology (ICAESMT)*, Uttaranchal University, Dehradun, India, pp. 1-8, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[10] Thomas Aichner et al., "Twenty-Five Years of Social Media: A Review of Social Media Applications and Definitions from 1994 to 2019," *Cyberpsychology, Behavior, and Social Networking*, vol. 24, no. 4, pp. 215-222, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] C. Hutto, and Eric Gilbert, "Vader: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, pp. 216-225, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[12] Simranpreet Kaur, Pallavi Kaul, and Pooya Moradian Zadeh, “Monitoring the Dynamics of Emotions During COVID-19 Using Twitter Data,” *Procedia Computer Science*, vol. 177, pp. 423-430, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Ankita Sharma, and Udayan Ghose, “Sentimental Analysis of Twitter Data With Respect to General Elections in India,” *Procedia Computer Science*, vol. 173, pp. 325-334, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[14] Da Li et al., “HEMOS: A Novel Deep Learning-Based Fine-Grained Humor Detecting Method for Sentiment Analysis of Social Media,” *Information Processing and Management*, vol. 57, no. 6, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Muhammad Asif et al., “Sentiment Analysis of Extremism in Social Media from Textual Information,” *Telematics and Informatics*, vol. 48, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] Malliga Subramanian et al., “A Survey On Hate Speech Detection And Sentiment Analysis Using Machine Learning And Deep Learning Models,” *Alexandria Engineering Journal*, vol. 80, pp. 110-121, 2023 [CrossRef] [Google Scholar] [Publisher Link]

[17] Anisha P. Rodrigues et al., “Retracted:: Real-Time Twitter Spam Detection and Sentiment Analysis Using Machine Learning and Deep Learning Techniques,” *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1-15, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[18] T.M. Saravanan et al., “Machine Learning Algorithm for Sentiment Analysis on Tweets,” *2nd International Conference on Automation, Computing and Renewable Systems (ICACRS)*, Pudukkottai, India, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Boppuru Rudra Prathap, “Geospatial Crime Analysis and Forecasting with Machine Learning Techniques,” *Artificial Intelligence and Machine Learning for EDGE Computing*, pp. 87-102, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Joan Donovan, “How Networked Incitement Fueled the January 6 Capitol Insurrection, Scientific American, 2024. [Online]. Available: https://www.scientificamerican.com/article/jan-6-was-an-example-of-networked-incitement/

[21] Loris Belcastro et al., “Analyzing Voter Behavior on Social Media During The 2020 US Presidential Election Campaign,” *Social Network Analysis and Mining*, vol. 12, no. 1, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[22] Livio Bioglio, and Ruggero G. Pensa, “Analysis and Classification of Privacy-Sensitive Content in Social Media Posts,” *EPJ Data Science*, vol. 11, no. 1, pp. 1-24, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Shinigami, Sentimental Analysis for Tweets, Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/gargmanas/sentimental-analysis-for-tweets

[24] Passionate-Nlp, Twitter - Entity Sentiment Analysis, Kaggle, 2021. [Online]. Available: https://www.kaggle.com/datasets/jp797498e/twitter-entity-sentiment-analysis/data

[25] Ali Toosi, Twitter Sentiment Analysis, Kaggle, 2019. [Online]. Available: https://www.kaggle.com/datasets/arkhoshghalb/twitter-sentiment-analysis-hatred-speech