UKR Journal of Multidisciplinary Studies (UKRJMS)
Homepage: https://ukrpublisher.com/ukrjms/
Email: submit.ukrpublisher@gmail.com
Volume 1, Issue 5, 2025
ISSN: 3049-432X (Online)

Comparative Evaluation of Machine Learning Techniques for Social Media Sentiment Analysis

Olawuyi, Kayode Timothy[1]; Olabiyisi, Stephen Olatunde[2]; Ismail, Wasiu Oladimeji[3]; Jeremiah, Yetomiwa Sinat[4], Ogunkan, Stella Kehinde[5]; Omotade, Adedotun Lawrence[6]

[1, 2, 3, 4 & 5] *Department of Computer Science Ladoke Akintola University of Technology, Ogbomoso, Oyo state Nigeria*

***Corresponding Author**: Olabiyisi, Stephen Olatunde

DOI: *https://doi.org/10.5281/zenodo.17101522*

Article History

Original Research Article

Received: 20-08-2025

Accepted: 31-08-2025

Published: 11-09-2025

Citation: Olawuyi, Kayode Timothy; Olabiyisi, Stephen Olatunde; Ismail, Wasiu Oladimeji; Jeremiah, Yetomiwa Sinat, Ogunkan, Stella Kehinde.; Omotade, Adedotun Lawrence, (2025), Comparative Evaluation of Machine Learning Techniques for Social Media Sentiment Analysis, UKR Journal of Multidisciplinary Studies (UKRJMS), volume 1(5), 121-129.

Abstract

In the modern era of digital communication, sentiment analysis has emerged as a key research domain within Natural Language Processing (NLP). This study focuses on assessing and comparing the effectiveness of three machine learning techniques Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes (NB) in categorizing sentiments conveyed on social media platforms, using Twitter as the case example. The dataset, obtained from Kaggle through the Kaggle JSON utility, consisted of tweets grouped into three sentiment classes: positive, negative, and neutral. Preprocessing procedures involved text cleaning, tokenization, removal of stop words, and the application of Term Frequency–Inverse Document Frequency (TF-IDF) for feature extraction. From the generated features, the most relevant 6,000 were retained for model training. The three algorithms were implemented in Python within a supervised learning framework, and their effectiveness was assessed using accuracy, precision, recall, and F1-score. The Support Vector Machine (SVM) model recorded an accuracy of 87%, with corresponding precision, recall, and F1-score values of 84%, 86%, and 85%. The Naïve Bayes (NB) classifier achieved 83% accuracy, alongside precision, recall, and F1-scores of 83%, 82%, and 82%, indicating a relatively balanced performance. The Random Forest (RF) model, however, delivered the highest performance, attaining 92% accuracy, 90% precision, 91% recall, and a 91% F1-score. These findings emphasize the strength of Random Forest, especially in addressing class imbalance, positioning it as the most effective technique among the models evaluated. Overall, the results demonstrate that Random Forest offers greater reliability and efficiency for sentiment classification on social media compared to SVM and NB across all evaluation metrics.

Keywords: *Comparative Performance, Social Media Sentiment Analysis, Support Vector Machine (SVM), Naïve Bayes (NB) and Random Forest (RF).*

1. Introduction

The rapid advancement of social media platforms and digital devices has empowered individuals to freely express their opinions and share experiences, thereby contributing substantially to the growth of big data. The advent of these platforms has transformed traditional modes of communication, interaction, and information exchange, creating dynamic spaces for public engagement and debate (Papacharissi, 2010). Platforms such as Twitter, Facebook, and Instagram exemplify this shift, where user-generated content has grown at an unprecedented scale, enabling real-time discussions across diverse social, political, and cultural contexts (Kaplan & Haenlein, 2010).

The ever-increasing volume of social media data presents both opportunities and challenges for understanding public opinion. Public opinion has always served as a crucial compass for navigating societal trends and shaping policy decisions. While traditional methods like surveys, focus groups, and media analysis offer structured data with clear demographics, they can be expensive, time-consuming, and suffer from limitations like potential sampling bias (Cinar, 2018 and Prior, 2018). Social media, on the other hand, offers a constant stream of real-time data at a significantly lower cost (Ohlhausen and Kernbach, 2018).

“Sentiment analysis, a branch of natural language processing (NLP), focuses on the automatic categorization of text into positive, negative, or neutral expressions (Pang & Lee, 2008). In recent years, social media sentiment analysis has gained popularity across various applications, including marketing, politics, and healthcare (Gao and Zhang, 2020).

The rise of machine learning has revolutionized sentiment analysis. Techniques like Support Vector Machines (SVMs) and Naive Bayes classifiers can be trained on labelled data sets to automatically identify sentiment in text (Pang and Lee, 2008). More recently, deep learning approaches using Random Forest (RF) networks have shown even greater promise in capturing the complexities of human language and sentiment (Tang *et al*., 2016). However, social media sentiment analysis is not without its limitations. One major challenge is the potential for bias within the data itself. Social media platforms tend to attract specific demographics, and user activity can be influenced by factors like echo chambers and confirmation bias (Bakshy *et al*., 2019). Additionally, the brevity and informal nature of social media communication can pose challenges for accurate sentiment analysis (Stieglitz and Dang-Nguyen, 2018). Furthermore, the algorithms used in sentiment analysis tools are not perfect and can misinterpret sarcasm, irony, and other subtleties of human language (Calvo *et al*., 2020).

In pursuit of contributing to existing knowledge, this study seeks to examine and compare the performance of selected machine learning algorithms for sentiment analysis of social media data. The specific objectives are to:

1. Extract relevant Twitter datasets of social media posts from Kaggle.com using the KaggleJSON tool to support supervised learning tasks.
2. Implement and train machine learning models—Support Vector Machine (SVM), Naïve Bayes (NB), and Random Forest (RF)—to classify the extracted Twitter data using Python programming.
3. Evaluate and compare the performance of the models based on key metrics, namely accuracy, precision, recall, and F1-score.

2. Related Work

Sentiment Analysis on Twitter using Machine Learning Techniques

This research investigated sentiment classification on Twitter by assessing the effectiveness of three machine learning models: Logistic Regression (LR), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) networks. Term Frequency–Inverse Document Frequency (TF-IDF) was employed for feature extraction. The findings revealed that the LSTM model outperformed the conventional algorithms, primarily because of its capability to capture contextual semantics and sequential dependencies within textual data (Oladipupo et al., 2025).

Naïve Bayes (NB) remains one of the foundational algorithms in machine learning and continues to find widespread application across domains such as natural language processing, image recognition, and recommender systems, owing to its computational efficiency and effectiveness (Hosmer et al., 2016).

A major drawback of conventional Recurrent Neural Networks (RNNs) lies in the vanishing gradient problem, where error gradients shrink exponentially as the time interval between relevant inputs increases. This typically occurs when the spectral radius of the recurrent weight matrix falls below one. LSTM networks were developed to mitigate this issue through their specialized gating mechanisms. By maintaining an internal memory state often described as an “error carousel” LSTM units allow error signals to persist and propagate over extended time steps until the gates adaptively regulate them (Greff et al., 2017).

The importance of sentiment analysis also intersects with issues of data security. Data breaches on social media have exposed users’ personal information, enabling identity theft and other cybercrimes. High-profile incidents, such as the Cambridge Analytica scandal, underscored how large-scale misuse of user data can erode public trust in digital platforms (Isaak et al., 2018).

Kumar et al. (2020) introduced a hybrid deep learning framework called ConVNet-SVMBoVW, designed to enhance real-time sentiment analysis.”

The framework utilized a Bag of Visual Words (BoVW) approach trained with SVM to forecast sentiment from visual content, while convolutional neural networks (ConvNets) captured fine-grained features. Their findings demonstrated that the hybrid ConVNet-SVMBoVW model outperformed conventional approaches, showcasing its potential for multimodal sentiment analysis.

Language barriers in social media sentiment analysis challenges that arises when analyzing text data in multiple languages or when the language used is different from the dominant language of the data (Gao *et.al,* 2020).

3. Methodology

This research work focused on evaluating the performance of three machine learning algorithms Support Vector Machine (SVM), Naïve Bayes (NB), and Random Forest

(RF) on a Twitter dataset for sentiment analysis. The methodological steps are outlined below:

i. Data Collection

Dataset. This study utilized the publicly available Twitter Sentiment Analysis Dataset, which comprises labeled tweets categorized into positive, negative, and neutral classes. Its structure makes it suitable for supervised learning tasks. The dataset was accessed and downloaded from **Kaggle.com** using the Kaggle JSON tool.

ii. Data Preprocessing

To enhance the quality of the dataset and ensure effective model training, several preprocessing procedures were applied:

i. **Text Normalization:** All tweets were converted to lowercase, and punctuation marks, special symbols, numerical values, and unnecessary spaces were removed.

ii. **Tokenization:** Each tweet was split into tokens (individual words).

iii. **Stopword Removal:** Commonly used words such as *"the," "is,"* and *"and"* which contribute little to sentiment orientation were excluded.

iv. **Stemming/Lemmatization:** Were reduced to their root or lemma forms (e.g., *"running"* → *"run"*) to minimize feature dimensionality and redundancy.

v. **Feature Extraction:** The cleaned text was converted into numerical feature vectors using the Term Frequency Inverse Document Frequency (TF-IDF) technique, enabling effective processing by machine learning models.

These steps preprocessing standardized the dataset, reduced noise, and improved the efficiency and accuracy of the learning process.

iii. Sentiment Classification

Three machine learning models Support Vector Machine (SVM), Naïve Bayes (NB), and Random Forest (RF) were implemented and trained on the preprocessed dataset. Each model was designed to classify tweets into one of the three sentiment categories: positive, negative, or neutral.

Model Selection

Model Training and Testing

This study evaluated three distinct machine learning techniques for sentiment classification:i. i. **Support Vector Machine (SVM):** SVM is a supervised learning technique commonly applied in classification tasks. Its core principle is to identify the most suitable hyperplane that separates sentiment categories within a high-dimensional feature space, ensuring maximum margin between classes.

ii. **Random Forest (RF):**

Random Forest is an ensemble method that builds multiple decision trees and combines their outcomes to arrive at a final prediction. Its strength lies in handling imbalanced datasets, mitigating overfitting, and lowering variance. These properties make it particularly effective for sentiment analysis, as it enhance stability and predictive accuracy.

iii. **Naïve Bayes (NB):**
NB is a probabilistic model derived from Bayes' theorem, operating under the assumption of conditional independence among features. Despite this simplification, it remains highly effective for text classification, including sentiment analysis, due to its simplicity, computational efficiency, and strong performance in high-dimensional contexts.

iv. **Training and Testing Split:**
The dataset was partitioned into training and testing sets, usually in a 70:30 ratio. The training set was used to develop the models, while the testing set evaluated how well the models generalized to unseen data.

v. **Training Stage:**
During this phase, the SVM, RF, and NB classifiers were fitted with the preprocessed training data. Each model learned relevant patterns within the tweets to classify them into sentiment categories (positive, negative, or neutral).

vi. **Testing Stage:**
The models were then validated using the testing subset. Their effectiveness was measured with key evaluation metrics such as accuracy, precision, recall, and F1-score, which together provided a holistic view of classification performance. (See Figure 3.3 for the workflow representation)

textID	text
f87dea47db	Last session of the day http://twitpic.com/67ezh
96d74cb729	Shanghai is also really exciting (precisely -- skyscrapers galore). Good tweeps in China: (SH) (BJ).
eee518ae67	Recession hit Veronique Branquinho, she has to quit her company, such a shame!
01082688c6	happy bday!
33987a8ee5	http://twitpic.com/4w75p - I like it!!
726e501993	that`s great!! weee!! visitors!
261932614e	I THINK EVERYONE HATES ME ON HERE lol
afa11da83f	soooooo wish i could, but im in school and myspace is completely blocked
e64208b4ef	and within a short time of the last clue all of them
37bcad24ca	What did you get? My day is alright.. haven`t done anything yet. leaving soon to my stepsister though!
24c92644a4	My bike was put on hold...should have known that.... argh total bummer
43b390b336	I checked. We didn`t win
69d6b5d93e	.. and you`re on twitter! Did the tavern bore you that much?
5c1e0b61a1	I`m in VA for the weekend, my youngest son turns 2 tomorrow......it makes me kinda sad, he is getting so big
504e45d9d9	Its coming out the socket I feel like my phones hole is not a virgin. That`s how loose it is... :`(
ae93ad52a0	So hot today =_= don`t like it and i hate my new timetable, having such a bad week
9fce30159a	Miss you
00d5195223	Cramps . . .

Figure 1: A screenshot of the sample structure of the collected Twitter dataset.

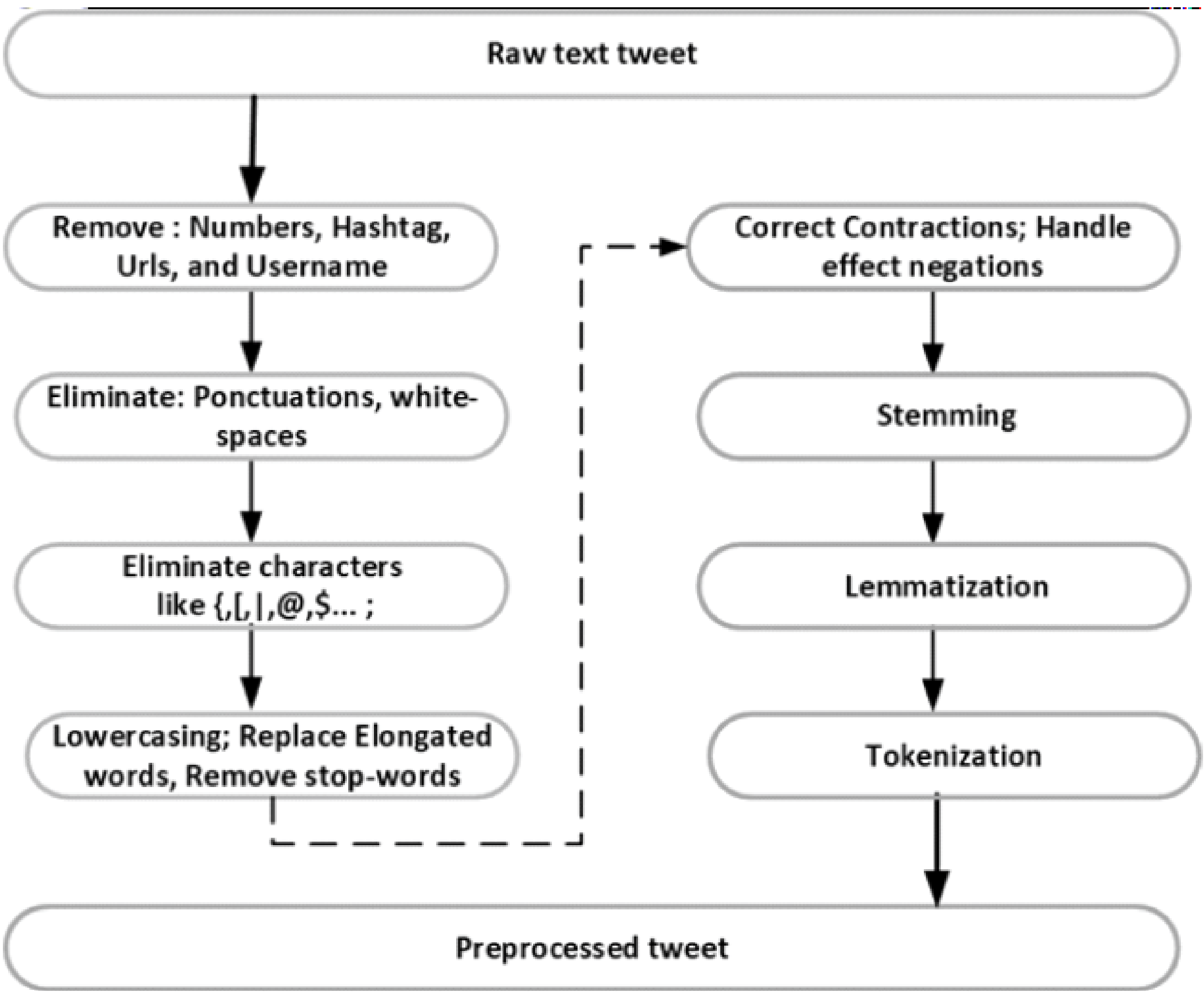

Figure 2: Diagram illustrating the Data Preprocessing

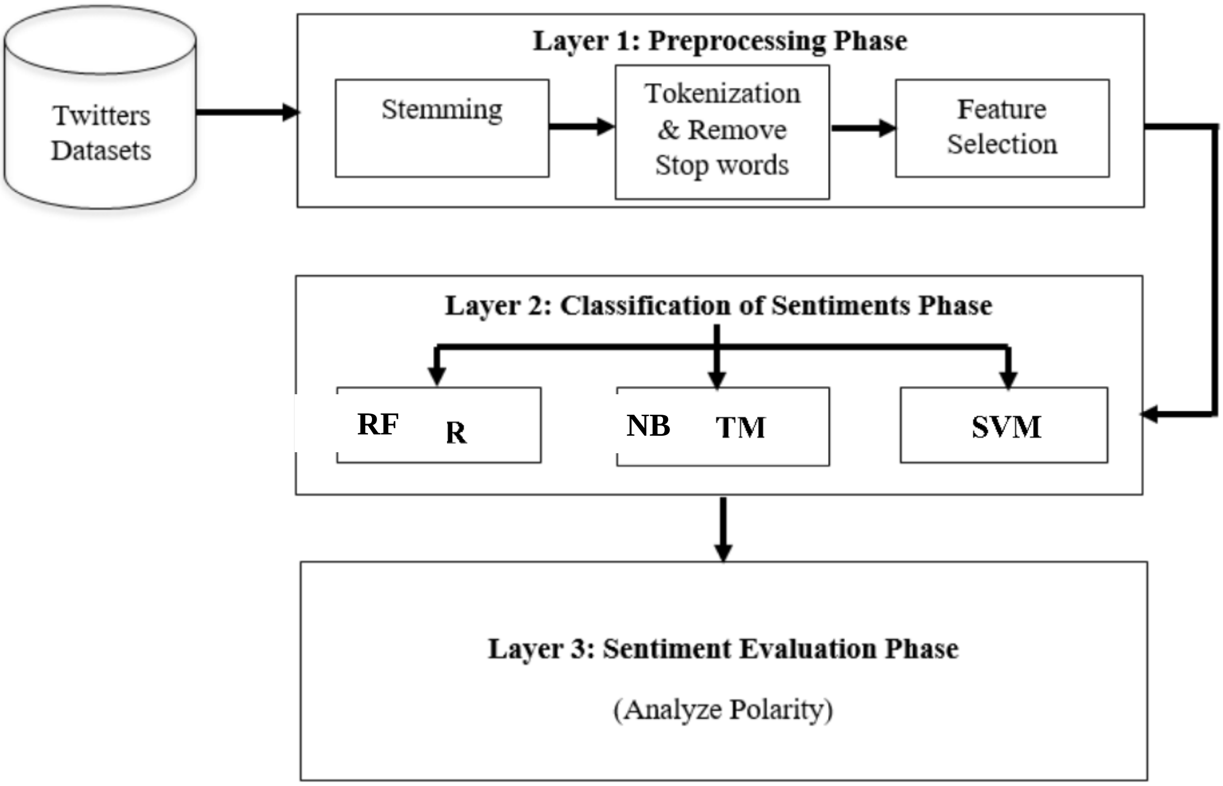

Figure 3: Diagram showing the Model Training and Testing of (RF, NB and SVM)

i. Performance Evaluation: The models' performance was assessed using accuracy, precision, recall, F1-score, and computational efficiency as evaluation metrics.

ii. Comparative Analysis: The strengths, weaknesses, and trade-offs between the three models were compared based on the evaluation metrics.

4. Results and Discussions

The dataset used in this study was subjected to a rigorous preprocessing routine to improve its quality and ensure its appropriateness for machine learning applications. The preprocessing steps involved text normalization, tokenization, and the elimination of stopwords. Following this, feature extraction was carried out using the Term Frequency–Inverse Document Frequency (TF-IDF) technique. This method converted the raw text into structured numerical vectors, thereby making it possible to efficiently train and evaluate the machine learning models employed in the analysis.

Summary of Preprocessed Data

Step	**Description**
Dataset Size	77,996 tweets
Tokenization	Average 15 tokens per tweet
Stopwords Removed	50% of tokens identified as stopwords
Feature Extraction	Top 6,000 TF-IDF features selected

These preprocessing steps were instrumental in cleaning and structuring the data, rendering it more conducive for machine learning tasks. By normalizing the text, we ensured consistency, while tokenization and stopword removal streamlined the input for better efficiency in further processing.

The dataset was classified into four distinct categories: Positive, Neutral, Negative, and Irrelevant. The distribution of these classes, as shown in Figure 4.1, highlights the balanced representation of sentiments within the dataset. Specifically, the dataset comprises 24,665 tweets labeled as Positive, 19,613 as Neutral, 22,853 as Negative, and 12,865 as Irrelevant. This classification ensures a diverse range of sentiments, enabling the development of robust machine learning models capable of effectively capturing the nuances of sentiment analysis.

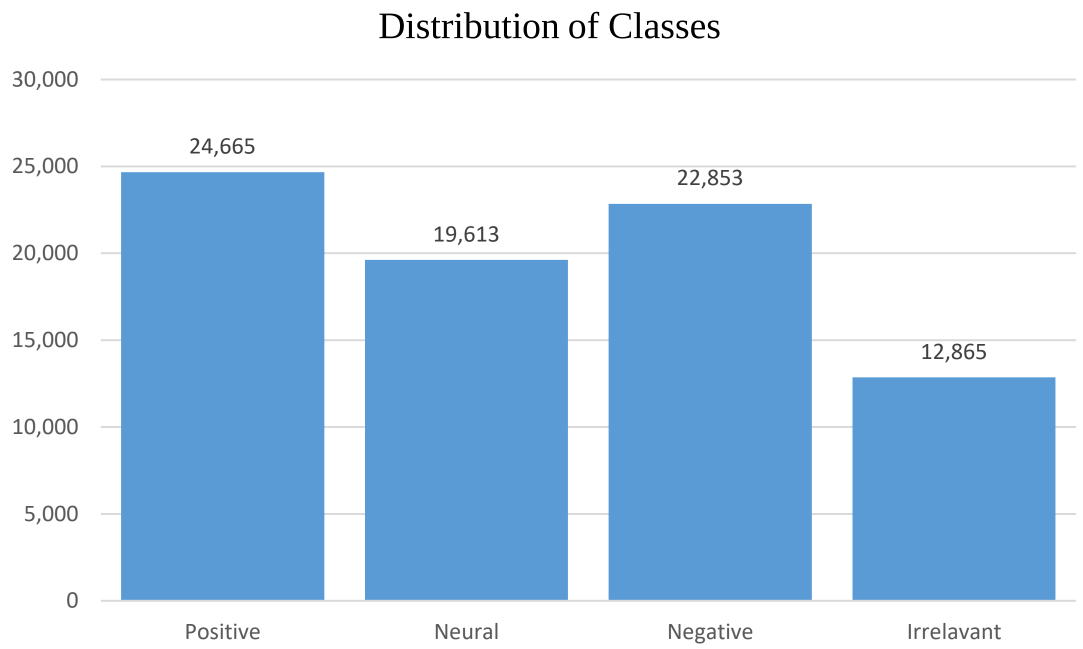

Figure 4: Distribution of Dataset into Classes

The performance of the three machine learning algorithms was evaluated using accuracy, precision, recall, F1-score, and computational efficiency as the main performance indicators. The dataset was divided into training and testing subsets in a 70:30 ratio, where the training portion was used to develop the models, while the testing portion assessed their ability to generalize to unseen data. This evaluation strategy offered valuable insights into the comparative strengths and weaknesses of each algorithm.

The Support Vector Machine (SVM) classifier functions by identifying the optimal hyperplane that best separates data points across different classes. Its central goal is to maximize the margin between class boundaries, which improves generalization and reduces the risk of misclassification. Because of this rigorous mathematical foundation, SVM has consistently demonstrated strong performance and reliability, making it particularly effective for high-dimensional datasets, such as those commonly encountered in text mining and sentiment analysis tasks.

Support Vector Machine Performance Metrics

Metric	**Value**
Accuracy	0.87
Precision	0.84
Recall	0.86
F1-Score	0.83
Computation Time	11 minutes

The SVM model achieved reliable performance with balanced precision and recall values, demonstrating its effectiveness in sentiment classification. However, an important consideration is the model's computational cost, which significantly escalates with larger datasets due to its quadratic complexity during the optimization process.

By converting the linear combination of weights and features into a probability score appropriate for classification, Naïve Bayes predicts probabilities by mapping input features through the sigmoid function.

Metric	**Value**
Precision	0.83
Precision	0.83
Recall	0.82
F1-Score	0.82
Computation Time	6 minutes

Random Forest demonstrated competitive performance, emerging as the fastest among the three models. Despite its advantages, it struggled with capturing non-linear relationships, limiting its efficacy in discerning the nuances of sentiment embedded within the dataset. Thus, while it remains an effective tool for simpler applications, more complex patterns can elude its analytical grasp.

The RF model, a specialized form of recurrent neural network, excels in managing sequential data by capturing temporal dependencies and contextual nuances within text. The key equations governing RF operations underscore its depth of capability.

Random Forest Performance Metrics

Metric	**Value**
Accuracy	0.92
Precision	0.90
Recall	0.91
F1-Score	0.91
Computation Time	55 minutes

The RF model showcased exceptional performance across all metrics, attributable to its proficiency in recognizing context and sequential dependencies. Nonetheless, this strength comes with a downside: its computational cost is significantly higher than that of the other models, which can make it less feasible for real-time applications, particularly on systems with constrained resources.

Comparative Analysis of Models

The comparative performance of the three models is summarized in Table 4.5, with Figure 4.1 providing a visual representation of their respective metrics.

Comparative Performance Metrics

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Computation Time**
SVM	0.87	0.84	0.86	0.85	11 seconds
NB	0.83	0.83	0.82	0.82	6 seconds
RF	0.92	0.90	0.91	0.91	55 seconds

Key Observations:

i. Accuracy: RF emerged as the most accurate model with a score of 0.92, closely followed by SVM at 0.87.
ii. Precision and Recall: RF outperformed both SVM and NB in terms of precision and recall, highlighting its superior capability in managing imbalanced datasets.
iii. Computation Time: While NB was optimal for speed, RF's extensive computational demands render it the least suitable for real-time analysis.

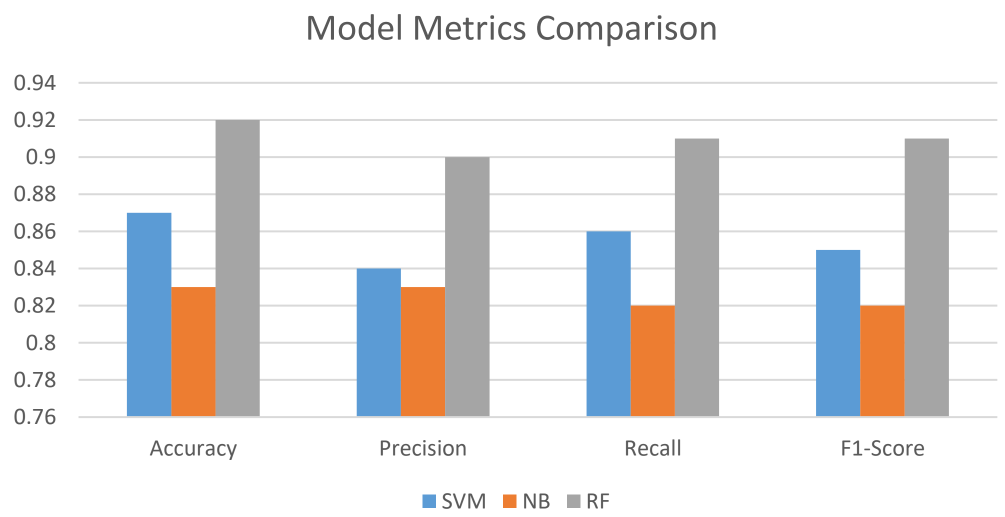

Figure 5: Performance Metrics Comparison

5. Conclusion

This study examined how well three machine learning algorithms Random Forest (RF), Naïve Bayes (NB), and Support Vector Machine (SVM) classified sentiments from a sizable Twitter dataset. The comparative analysis provided insightful information about each model's applicability for various sentiment analysis applications by highlighting its unique advantages and disadvantages.

Among the three, the Random Forest (RF) model achieved the best overall performance, recording an accuracy of 92%, with precision and recall values of 0.90 and 0.91, respectively. Its ability to capture complex patterns and contextual nuances in textual data underscores its robustness for sentiment analysis tasks. Nevertheless, the relatively high computational cost associated with RF presents a limitation. Such characteristics make the algorithm highly relevant for real-time or resource-constrained applications, where both processing speed and minimal resource consumption are essential. This ensures that the model can be deployed effectively in environments such as mobile devices, embedded systems, or online monitoring platforms without compromising accuracy. Future work could explore hybrid or deep learning approaches that balance predictive accuracy with computational efficiency, thereby enhancing the practical deployment of sentiment analysis models in real-world scenarios.

The **Naïve Bayes (NB)** classifier, in contrast, proved to be the fastest, with a computation time of just **6 seconds**, but its accuracy (83%) was comparatively lower. Its inability to handle non-linear and complex data patterns limits its effectiveness in sophisticated analyses, though it remains useful for simpler applications requiring rapid processing.

With an accuracy of 85%, the Support Vector Machine (SVM) offered a moderate balance between precision and reliability. Despite this, its computational complexity and scalability challenges restrict its effectiveness in large-scale data environments.

Overall, the findings emphasize that **model selection should depend on application requirements**: RF is best suited for accuracy-critical tasks, SVM for balanced performance, and NB for scenarios prioritizing speed. This research thus provides a foundation for future studies aimed at improving efficiency and extending sentiment analysis across diverse domains.

REFERENCES

1. Oladipupo,S., Olabiyisi S., Ismail., W. (2025). Survey on Ssentiment Analysis using Twitter Dataset. International Journal of Research and Innovation in Applied Science (IJRIAS), Volume X issues V/app 601-607.
2. Bakshy, E., Rosenstein, S., Matias, C., Mislove, A., & Zittrain, J. (2019). The mechanics of filter bubbles. Science, 365(6448), 764-771.
3. Balahur, A., Turchi, M., and Steinberger, R. (2012). *Multilingual Sentiment Analysis for Social Media Applications: A Survey*. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, 1-15.
4. Calvo, R. A., Arias, M., and López, M. I. (2020). Sentiment analysis in social networks: Recent progress and challenges. Information Processing & Management, 57(6), 102439.
5. Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). *Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling*. arXiv preprint arXiv:1412.3555.
6. Gao, X., and Zhang, Y. (2020). Social media sentiment analysis: A survey. IEEE Transactions on Knowledge and Data Engineering, 32(1), 101-114.
7. Ghosh, A., and Veale, T. (2016). *Fracking Sarcasm using Neural Network*. Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis, 161-169.
8. Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
9. Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., and Schmidhuber, J. (2017). *LSTM: A Search Space Odyssey*. IEEE Transactions on Neural Networks and Learning Systems, 28(10), 2222-2232.
10. Huang, X., and Zhang, Y. (2020). *A Semi-supervised Framework for Social Media Topic Detection using Contextual Information*. IEEE Transactions on Computational Social Systems, 7(5), 1164-1175.
11. Isaak, J., and Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer, 51*(8), 56-59.
12. Kaplan, A. M., and Haenlein, M. (2010). Users of the world, unite! The chaSSSllenges and opportunities of social media. Business Horizons, 53(1), 59-68.
13. Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., and Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin, 140*(4), 1073.
14. Kumar, A., and Jaiswal, A. (2020). *Systematic literature review of sentiment analysis on Twitter using soft computing techniques*. Concurrency and Computation: Practice and Experience, 32(1), e5107.

15. Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

16. Liu, C., Fang, F., Lin, X., Cai, T., Tan, X., Liu, J., and Lu, X. (2021). Improving sentiment analysis accuracy with emoji embedding. *Journal of Safety Science and Resilience*, *2*(4), 246-252..

17. Pang, B., and Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135. https://www.nowpublishers.com/article/DownloadSummary/INR-011

18. Papacharissi, Z. (2010). Affective publics: Sentiment analysis for social media. Oxford University Press.

19. Stieglitz, S., and Dang-Nguyen, L. A. (2018). Sentiment analysis in social media. Annual Review of Information Science and Technology, 52(1), 179-238. https://www.annualreviews.org/doi/10.1146/annurev-statistics-030718-105242

20. Vannucci, A., Flannery, K. M., and Ohannessian, C. M. (2017). Social media use and anxiety in emergi*Journal of Affective Disorders*, *207*, 163-166.