Comparative Performance Analysis of Selected Machine Learning Techniques for Social Media Sentiment Analysis

Oladipupo, Samuel Adegbite[1], Olabiyisi, Stephen Olatunde[2], Ismaila, Wasiu Oladimeji[3], Oyedele, Adebayo Olalere[4] and Toyobo, Oluwatobi Joel[5]

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

[4]Department of Information systems, University of Portsmouth, United Kingdom

DOI: https://doi.org/10.51584/IJRIAS.2025.100500055

Received: 25 April 2025; Accepted: 06 May 2025; Published: 10 June 2025

ABSTRACT

Social media sentiment analysis plays a crucial role in understanding public opinion and user behavior across platforms. Several techniques have been developed to accurately classify sentiment in social media data. However, these techniques have not been adequately analyzed and compared. Hence, this study investigates the comparative performance of Support Vector Machine (SVM), Logistic Regression (LR) and Long Short-Term Memory (LSTM) in social media sentiment analysis.

Social media data which contains labelled tweet representing different sentiments (positive, negative, neutral) were extracted from *Kaggle.com* using *Kagglejson* tool to facilitate supervised learning tasks. The preprocessing steps involved text normalization, tokenization, stopword removal, and feature extraction using TF-IDF with top 5,000 features selected**.** Next, the three machine learning models – SVM, LR and LSTM were implemented and trained with the preprocessed dataset. Finally, the models were implemented in python, evaluated and compared based on accuracy, precision, recall and F1 score.

The results of the evaluation and comparison indicate that SVM achieved 85% accuracy, 82% precision, 84% recall and 83% F1-score: LR achieved 83% accuracy, 81% precision, 80% recall and 80% F1-score while LSTM achieved 90% accuracy, 88% precision, 89% recall and 89% F1-score

The results demonstrated that LSTM outperformed SVM and LR in terms of accuracy, precision, recall and F1-score highlighting its superior capability in managing imbalanced datasets. LSTM is hereby recommended for social media analysis; this finding underscore the efficiency of LSTM addressing the challenging task in the field of sentiment classification.

Keyword: Comparative Performance, Social Media Sentiment Analysis, Support Vector Machine (SVM), Logistic Regression (LR) and Long Short-Term Memory (LSTM).

INTRODUCTION

The rapid development of social media platforms and various devices has enabled users to share their views, significantly contributing to big data. The emergence of social media has revolutionized how people communicate, interact, and share information. These platforms have become vibrant public squares, fostering real-time discussions on a vast array of topics (Pap8acharissi 2010). With the rise of social media platforms such as Twitter, Facebook, and Instagram, the amount of user-generated content has increased exponentially (Kaplan and 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 subfield of natural language processing, aims to automatically classify text as positive, negative, or neutral (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 Long Short-Term Memory (LSTM) 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 the quest to contribute to knowledge, this study aims to investigate and compare the performance analysis of some machine learning techniques for social media sentiment analysis.

The specific objectives are to:

i. extract relevant twitter datasets of social media posts from Kaggle.com using Kagglejson tool to facilitate supervised learning tasks

ii. implement and train machine learning models (SVM, LR and LSTM) to classify social media data extracted from Kaggle.com using python programming language;

iii. compare the models' performance using accuracy, precision, recall and F1-score as performance matrics.

Related Work

Sentiment Analysis on Twitter using Machine Learning Techniques. This study compared the performance of different machine learning models, such as Naive Bayes, Maximum Entropy, and Support Vector Machines (SVM), on Twitter sentiment analysis. The research uses unigrams, bigrams, and POS (Part-of-Speech) tags as features. The authors found that SVM performs best among the traditional methods due to its ability to handle high-dimensional feature spaces efficiently (Bhayani, and Huang 2015).

Logistic Regression is a fundamental algorithm in Machine Learning, an essential tool in a Computer Scientist's toolkit, and it's widely used in many applications, including Natural Language Processing, Image Classification, and Recommender Systems (Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. 2013, Goodfellow, I., Bengio, Y., and Courville, A. 2016).

A problem with using gradient descent for standard RNNs is that error gradients vanish exponentially quickly with the size of the time lag between important events. This is due to if the spectral radius of is smaller than 1. However, with LSTM units, when error values are back-propagated from the output layer, the error remains in the LSTM unit's cell. This "error carousel" continuously feeds error back to each of the LSTM unit's gates, until they learn to cut off the value (Greff, K., Srivastava, R. K., Koutník, Steunebrink, and Schmidhuber, 2017).

Data breaches on social mediosed personal information, leading to identity theft and other forms of cybercrime. High-profile cases like the Cambridge Analytica scandal highlighted how user data could be misused on a large scale, affecting public trust (Isaak and Hanna, 2018).

Kumar *et al*. (2020) have presented a hybrid deep learning approach named ConVNet-SVMBoVW that dealt with the real-time data for predicting the fine-grained sentiment. In order to measure the hybrid polarity, an aggregation model was developed. Moreover, SVM was used for training the BoVW to forecast the sentiment of visual content. Finally, it was concluded that the suggested ConvNet-SVMBoVW was outperformed by the conventional models.

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, and Li, 2020).

METHODOLOGY

This research focused on the performance evaluation of three different machine learning algorithms Support Vector Machine (SVM), Logistic Regression (LR), and Long Short-Term Memory (LSTM) on a Twitter dataset for sentiment analysis. The following methodology outlines the research approach:

i. Data Collection: The research utilized a publicly available Twitter dataset as the Twitter Sentiment Analysis Dataset. These datasets contain labeled tweets representing different sentiments (positive, negative, neutral), which facilitated supervised learning tasks. The dataset was collected through the Kaggle.com using the Kagglejson tool.

ii. Data Preprocessing: Preprocessing techniques which are text normalization, tokenization, stopword removal and feature extraction were applied to clean and prepare the raw text for model training. Data preprocessing is essential for transforming raw text into a suitable format for machine learning models.

   a) Text Normalization: Converting text to lowercase and removing special characters, punctuation, and extra spaces.

   b) Tokenization: Splitting each tweet into individual words or tokens.

   c) Stopword Removal: Removing common words like "the," "is," and "and" that do not contribute significantly to sentiment analysis.

   d) Stemming/Lemmatization: Reducing words to their root forms (e.g., “running” becomes “run”) to reduce dimensionality.

   e) Feature Extraction: Converting text data into numerical features using techniques like Term Frequency-Inverse Document Frequency (TF-IDF).

   These preprocessing steps standardize the dataset, allowing the machine learning algorithms to be trained effectively.

iii. Sentiment Classification: Three machine learning models SVM, LR, and LSTM were employed to classify the sentiment of the tweets.

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

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

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue V May 2025

RESULTS AND DISCUSSIONS

The dataset utilized in this study underwent rigorous preprocessing to ensure its appropriateness for training and testing the machine learning models. The preprocessing steps involved essential tasks such as text normalization, tokenization, stopword removal, and feature extraction utilizing Term Frequency-Inverse Document Frequency (TF-IDF) methodology.

Summary of Preprocessed Data

Step	Description
Dataset Size	73,996 tweets
Tokenization	Average 15 tokens per tweet
Stopwords Removed	50% of tokens identified as stopwords
Feature Extraction	TF-IDF with top 5,000 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 20,655 tweets labeled as Positive, 18,108 as Neutral, 22,358 as Negative, and 12,875 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.

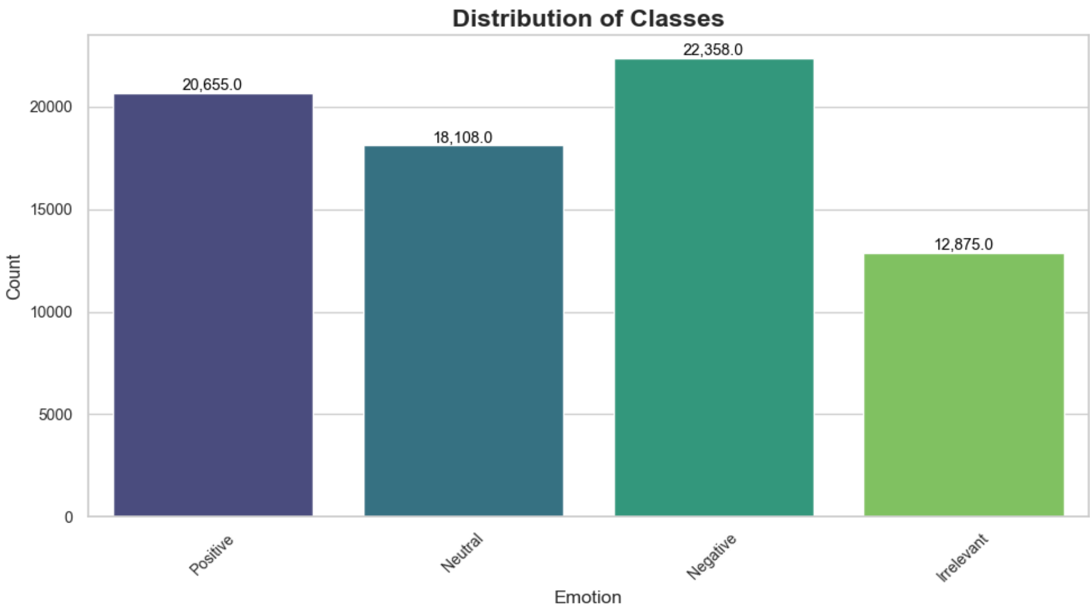

Figure 4.1: Distribution of Dataset into Classes

The performance of the three models was evaluated based on metrics such as accuracy, precision, recall, F1-score, and computational efficiency. Each model was trained and tested using a 70-30 split of the dataset. A detailed examination of the results for each model reveals crucial insights into their respective performance capabilities.

The SVM model operates by identifying the hyperplane that best separates the dataset into distinct classes. The objective function maximized is the margin between data points of different classes, as expressed in the relevant equations. This approach allows SVM to exhibit a high degree of effectiveness in classification tasks under various conditions.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue V May 2025

Support Vector Machine Performance Metric

Metric	Value
Accuracy	0.85
Precision	0.82
Recall	0.84
F1-Score	0.83
Computation Time	10 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.

Logistic Regression predicts probabilities by mapping input features through the sigmoid function, transforming the linear combination of weights and features into a probability score suitable for classification. This model, while straightforward, holds unique advantages.

Logistic Regression Performance Metrics

Metric	Value
Precision	0.81
Precision	0.81
Recall	0.80
F1-Score	0.80
Computation Time	5 minutes

Logistic Regression 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 LSTM 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 LSTM operations underscore its depth of capability.

LSTM Performance Metrics

Metric	Value
Accuracy	0.90
Precision	0.88
Recall	0.89
F1-Score	0.89
Computation Time	54 minutes

The LSTM 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	0.85	0.82	0.84	0.83	10 seconds
Logistic Regression	0.83	0.81	0.80	0.80	5 seconds
LSTM	0.90	0.88	0.89	0.89	60 seconds

Key Observations:

i. Accuracy: LSTM emerged as the most accurate model with a score of 0.90, closely followed by SVM at 0.85.

ii. Precision and Recall: LSTM outperformed both SVM and Logistic Regression in terms of precision and recall, highlighting its superior capability in managing imbalanced datasets.

iii. Computation Time: While Logistic Regression was optimal for speed, LSTM's extensive computational demands render it the least suitable for real-time analysis.

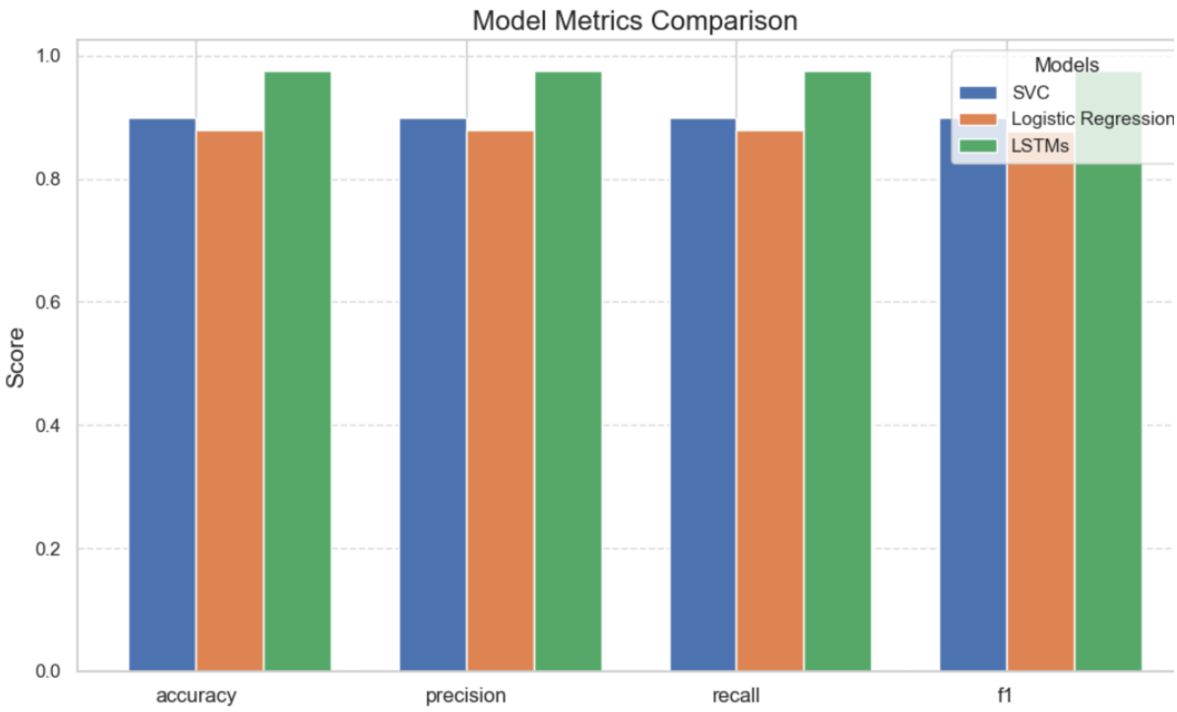

Figure 4.2: Performance Metrics Comparison

CONCLUSION

This study evaluated three machine learning models Support Vector Machine (SVM), Logistic Regression (LR), and Long Short-Term Memory (LSTM) for sentiment classification on a large Twitter dataset. The findings highlight each model's strengths and limitations, offering practical guidance for their selection based on specific use cases. The LSTM model achieved the highest accuracy (90%) and strong precision (0.88) and recall (0.89), demonstrating its ability to capture temporal and contextual nuances in text. However, its high computational cost limits its feasibility for real-time applications. Logistic Regression, while the fastest model with a computation time of 5 seconds, showed lower accuracy (83%) due to its inability to handle complex, non-linear patterns, making it suitable for simpler tasks requiring quick results. SVM provided balanced performance with an accuracy of 85% but struggled with scalability for larger datasets. These results underscore the importance of selecting models based on application needs. LSTM is ideal for tasks requiring high accuracy, SVM for balanced performance, and LR for rapid analysis. This research offers a foundation for future work in optimizing model efficiency and expanding their use in diverse contexts.

REFERENCES

1. Bakshy, E., Rosenstein, S., Matias, C., Mislove, A., & Zittrain, J. (2019). The mechanics of filter bubbles. Science, 365(6448), 764-771.
2. 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.
3. 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.
4. 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.
5. Gao, X., and Zhang, Y. (2020). Social media sentiment analysis: A survey. IEEE Transactions on Knowledge and Data Engineering, 32(1), 101-114.
6. 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.
7. Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning. MIT Press.
8. 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.
9. 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.
10. Isaak, J., and Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer, 51(8), 56-59.
11. Kaplan, A. M., and Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. Business Horizons, 53(1), 59-68.
12. 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.
13. 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.
14. Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.
15. 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..
16. 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
17. Papacharissi, Z. (2010). Affective publics: Sentiment analysis for social media. Oxford University Press.
18. 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
19. Vannucci, A., Flannery, K. M., and Ohannessian, C. M. (2017). Social media use and anxiety in emergiJournal of Affective Disorders, 207, 163-166.