Harnessing Social Media Sentiment for Predictive Insights into the Nigerian Presidential Election

Jelili Olalekan Alao
Department of Computer Science and Digital Technologies
University of East London
London, United Kingdom
u1957684@uel.ac.uk

Mohammad Hossein Amirhosseini
Department of Computer Science and Digital Technologies
University of East London
London, United Kingdom
m.h.amirhosseini@uel.ac.uk

Amin Karami
Department of Computer Science and Digital Technologies
University of East London
London, United Kingdom
a.karami@uel.ac.uk

Seyed Ali Ghorashi
Department of Computer Science and Digital Technologies
University of East London
London, United Kingdom
s.a.ghorashi@uel.ac.uk

Abstract—**Political events are heavily influenced by social media, shaping public opinion and actions. Sentiment analysis of social media content helps policymakers, campaign planners, and analysts understand voter sentiments for informed decision-making. This study performs a comprehensive comparative analysis of traditional machine learning models—Logistic Regression, Random Forest, Naïve Bayes, and SVM—and deep learning models—FFNN, LSTM, and CNN—on tweets collected via the X (formerly Twitter) API regarding the 2023 Nigerian Presidential Election. All models underwent a proper optimisation process and were evaluated using key performance evaluation metrics. Over 1.9 million tweets were collected over eight months. Results show deep learning models outperform traditional ones, with LSTM achieving the highest accuracy (95%), followed by CNN (94%) and FFNN (94%).**

Keywords—Sentiment Analysis, Deep Learning, Machine Learning, Social Media Analysis, Election

I. Introduction

Social media has transformed communication over the past decade through platforms like Twitter, Facebook, and Instagram, driven by widespread internet and smartphone use. Sentiment analysis is crucial for extracting opinions and emotions from text, classifying attitudes as positive, negative, or neutral. It is widely applied in sectors like healthcare, politics, and business to understand public opinion and behaviour. For example, businesses analyse online reviews to gauge customer preferences, while governments assess voter sentiment during elections [25]. According to Dixon [24], 4.26 billion people used social media platforms in 2021, a number projected to reach six billion by 2027. This growth will provide even richer sentiment data, offering valuable insights into consumer behaviours and perceptions.

This study focuses on leveraging machine learning and deep learning techniques for sentiment analysis, particularly in the context of the 2023 Nigerian elections. Traditional lexicon-based and machine learning approaches have limitations in handling unstructured social media data, often requiring extensive feature engineering. In contrast, deep learning methods, which use artificial neural networks, outperform traditional methods by automatically detecting complex language patterns such as negation and nuanced expressions.

The research aims to improve sentiment analysis accuracy for political trends and outcomes, specifically targeting tweets related to Nigerian presidential candidates. It involves (i) conducting a comprehensive literature review on sentiment analysis techniques, (ii) collecting pre- and post-election Twitter data via the Twitter API, and (iii) developing and comparing machine learning and deep learning models using metrics like Accuracy, F1 Score, Recall, and Precision.

A unique contribution of this study is the creation of a large dataset comprising over 1.9 million tweets, offering valuable insights into public sentiment and serving as a benchmark for future research in political studies. This is the first study to comprehensively compare machine learning and deep learning approaches within Nigeria's political landscape.

The research provides practical insights into social media’s role in shaping political events, emphasizing the potential of deep learning to enhance sentiment detection accuracy. It evaluates the real-world performance of these models in capturing public opinion, contributing to more effective electoral predictions. This study is significant for politicians, policymakers, and researchers in computational political science, as it highlights how public sentiment can shape messages that resonate with voters and demonstrates the effectiveness of deep learning in understanding voter sentiment (e.g., negation handling and nuanced expressions).

II. Previous Works

Sentiment analysis of social media platforms, particularly Twitter, has proven to be an effective tool for election prediction, with various studies highlighting its potential. Sharma and Moh [19] analysed Hindi Twitter data during the 2014 Indian elections using supervised machine learning methods like Support Vector Machines (SVM). Their model achieved 78.4% accuracy in predicting a BJP victory, showcasing the promise of this approach while drawing criticism for its focus on Hindi-language data and a specific political context.

Shaikh et al. [17] leveraged labelled Twitter data to predict elections, highlighting the value of sentiment analysis in monitoring political trends but raising concerns about data sequence dependency in supervised learning. Razzaq et al. [15], during the 2013 Pakistan General Elections, classified Twitter sentiments as favourable, neutral, or unfavourable but cautioned against assuming social media behaviour mirrors overall voting trends.

Smailović et al. [20] used a real-time binary SVM classifier for Twitter sentiment analysis during the 2013 Bulgarian parliamentary elections. Despite its success, the study faced limitations due to a small dataset, underscoring the need for larger, more comprehensive data. Similarly, Kumar et al. [10] employed Naive Bayes and SVM algorithms during the 2017 Punjab Legislative Assembly elections, achieving approximately 80% accuracy. They affirmed the utility of sentiment analysis while emphasizing the importance of larger datasets to avoid overgeneralized conclusions.

Mahmood et al. [11] examined elections in Pakistan using Chi-squared Automatic Interaction Detection (CHAID). While their study introduced innovative methods and techniques, it was constrained by a small sample size and the non-disclosure of certain methodologies. Ismail et al. [7] evaluated various Twitter sentiment machine learning classifiers, identifying Multinomial Naive Bayes as the top performer, although results varied depending on the dataset and its structure.

In Indonesia, researchers applied advanced feature selection techniques such as Particle Swarm Optimization (PSO) and Genetic Algorithms to analyse the presidential election. Their study demonstrated very high precision levels, showcasing the efficacy of combining these techniques with machine learning classifiers [9]. For the 2019 Istanbul mayoral elections, Soylu and Baday [21] compared methods including logistic regression, multinomial Naive Bayes, support vector machines, and random forests. They concluded that simpler, traditional approaches could improve election forecasting activities. Moreover, Joseph [8] explored decision tree classifiers to predict the outcomes of Indian general elections based on Twitter sentiment analysis. He stressed the need for an exhaustive evaluation of classifier effectiveness, considering factors such as language diversity and tweet popularity.

Durga Rani et al. [23] explored various machine learning (ML) and deep learning (DL) methods, highlighting the superior ability of DL techniques to extract insights from dynamic social media data on platforms like Twitter. However, the study identified challenges such as high computational requirements and difficulties in processing informal slang commonly used on Twitter. Similarly, Manikandan et al. [12] investigated the influence of informal language on sentiment classification, finding that deep learning methods like LSTM and CNN significantly outperformed traditional techniques.

Building on this, Sathya et al. [16] demonstrated that advanced deep learning models, including BERT and bidirectional networks, effectively manage large and complex Twitter datasets, further confirming the superiority of DL over conventional methods in sentiment classification. Alvi et al. [2] examined a range of approaches, from statistical methods to advanced DL techniques, to analyse real-time public sentiments on Twitter. They highlighted limitations such as representation biases in social networks, which impact sentiment analysis accuracy. Collectively, these studies illustrate the advancements in using DL for sentiment analysis and their relevance in understanding public opinions in the political domain through social media activity.

There is a growing trend in the use of deep learning techniques for sentiment analysis, with numerous studies showcasing their effectiveness. A Bidirectional Long Short-Term Memory (Bi-LSTM) model trained on Twitter data from the 2020 US presidential elections achieved an accuracy of 93.45%, although the study emphasized the need for improved hyperparameter tuning, advanced evaluation metrics, and robust frameworks to ensure reproducibility [18]. Similarly, Hidayatullah et al. [6] utilized Bi-LSTM during the 2019 Indonesian presidential elections, achieving an accuracy of 84.60%, particularly when enhanced with pre-trained word embeddings, highlighting the model's strength.

Comparative research consistently demonstrates the superiority of deep learning models over traditional machine learning approaches. Ali et al. [1] developed a hybrid CNN-LSTM model that achieved 85% accuracy in predicting the 2018 Pakistani general elections. Similarly, Olabanjo et al. [13] applied multiple deep learning models to analyse the Nigerian presidential elections, where they achieved the accuracy of 94%, despite challenges such as time and data constraints.

Additionally, a CNN model incorporating subword information proved effective for morphologically rich languages, although the study stressed the importance of a larger and more diverse corpus to ensure scalability and robustness [14]. These findings highlight the robust performance of deep learning techniques in political sentiment analysis, demonstrating their ability to outperform traditional approaches.

Several studies have explored the advantages and limitations of deep learning approaches in sentiment analysis. Hamed et al. [5] demonstrated that deep learning models, such as CNNs and DNNs, perform effectively but require larger datasets for validation. Similarly, Habimana et al. [4] stressed the importance of using diverse datasets to improve model resilience and strengthen results. However, Bilal et al. [3] highlighted issues such as inadequate documentation and ethical considerations in research practices within this domain.

Despite these challenges, deep learning models, particularly LSTM and CNN, have consistently outperformed traditional approaches, accelerating advancements in automated sentiment analysis [22]. Models like BERT, LSTM, and CNN have shown exceptional potential for election prediction using Twitter sentiment analysis. However, they face challenges such as Twitter's informal language, inherent biases in datasets, and high computational power requirements. These obstacles underscore the need for further improvements to enhance the accuracy and applicability of these methods.

While deep learning has advanced sentiment analysis significantly, gaps remain, particularly in the context of Nigerian elections. One key limitation is the lack of comprehensive comparative studies between deep learning and traditional machine learning techniques, which are essential for evaluating model robustness across diverse conditions. Additionally, existing datasets are often domain-static, focusing on specific timeframes and failing to capture political sentiments before and after significant events. This limitation reduces their ability to analyse evolving election dynamics. Addressing these gaps would improve the adaptability and accuracy of sentiment analysis models for dynamic political landscapes.

III. Methodology

A. Data Collection

The data collection process involved setting up a robust environment for seamless Twitter data extraction, processing, and visualisation. A Twitter Developer Account was obtained to access the Twitter API, requiring registration, approval, and application creation to secure credentials (API key, secret key, Access token, and token secret).

R and RStudio were configured as the coding environment, with essential libraries installed. *rtweet* and *twitteR* were used for Twitter API interaction and data extraction. *dplyr* was employed for efficient data manipulation, *ggplot2* was utilised for data visualisation, and *write.csv* was used for exporting data to CSV format. This setup enabled efficient extraction, processing, and visualisation of Twitter data, streamlining the workflow for sentiment analysis.

Authentication with the Twitter API was achieved using the *setup_twitter_oauth()* function from the *twitteR* package, utilising previously obtained credentials. Tweets were targeted using defined search criteria, including relevant keywords, hashtags, and user accounts linked to election candidates. Examples of handles used included *@atiku* (People's Democratic Party), *@peterobi* (Labour Party), and *@officialABAT* (All Progressives Congress). The searchTwitter() function was employed to fetch tweets, with parameters such as tweet count, date range, and search terms ensuring comprehensive data collection.

To ensure dataset balance and representativeness across candidates, a monthly cap of 100,000 tweets per candidate was set, preventing any single candidate or group from dominating the dataset. Over eight months, 1.9 million tweets were collected. Due to computing limitations, a 30% random sample (over 570,000 tweets) was selected for analysis and we made sure that the extracted dataset is balanced. Tweets were converted into a dataframe using the *twListToDF()* function and saved in CSV format via *write.csv()*. The datasets were securely stored on Google Drive for easy access. Data quality was reviewed using tools like *view()* and *head()* in *R*, and the *length(tweets)* function confirmed the dataset's adequacy for analysis.

B. Data Description

The dataset aims to capture public perceptions and discussions about political figures on Twitter during the election period and key events. It enables the classification of sentiments toward candidates as *positive*, *negative*, or *neutral*, reflecting individual opinions. Structured with multiple fields obtained via the Twitter API, the dataset facilitates comprehensive sentiment analysis. Table 1 details the features of the dataset, supporting insights into political sentiments during campaigns.

TABLE I. DATA DESCRIPTION

Feature	Description
text	Contains the actual text of the tweet.
favorited	Value indicating whether the authenticated user has favorited the tweet.
favoriteCount	Number of times the tweet has been favorited by users.
replyToSN	The screen name of the original tweet's author if the tweet is a reply.
created	The date and time when the tweet was created, in UTC.
truncated	Indicates whether the text of the tweet was truncated.
replyToSID	The ID of the original tweet's status if the tweet is a reply.
id	The unique identifier for the tweet.
replyToUID	The user ID of the original tweet's author if the tweet is a reply.
statusSource	Utility used to post the tweet, provided as an HTML-formatted string.
screenName	The screen name of the user who posted the tweet.
retweetCount	Number of times this tweet has been retweeted.
isRetweet	Indicates whether this tweet is a retweet.
retweeted	Boolean indicating whether the authenticated user has retweeted this status.
longitude	The longitude of the tweet's location, if available.
latitude	The latitude of the tweet's location, if available.

C. Data Preprocessing

The data preprocessing phase focuses on refining the dataset for effective sentiment analysis and model training. Unnecessary elements like URLs, mentions, hashtags, and special characters were removed to reduce noise. Text was standardised by converting it to lowercase and handling abbreviations for consistency. Stop words were eliminated to focus on meaningful terms, while stemming and lemmatization normalise word forms, improving classification accuracy.

Text was encoded into numerical representations using methods like *TF-IDF*, *Word2Vec*, or *GloVe* to enable processing by machine learning and deep learning models. Sentiment scores were calculated using tools to classify polarity (positive, negative, or neutral). Processed and labelled datasets were exported as CSV files for further analysis.

The labelled data was loaded into a *pandas* DataFrame for manipulation. Additional cleaning removed special characters, and the dataset was split into training and testing sets (80-20 ratio). Tokenization converted text into integer sequences, which were padded for uniform input length, using *Keras*'s Tokenizer and *pad_sequences*. Sentiment labels were encoded numerically and transformed into categorical format with *Keras*'s *to_categorical*. These steps ensure the dataset is clean, consistent, and ready for accurate model training.

D. Machine Learning and Deep Learning Models

In this work, four traditional machine learing models including Logistic Regression, Random Forest, Naïve Bayes, and SVM, and three deep learning models including FFNN, LSTM, and CNN were implemented. Model training is a critical step where pre-processed data is used to train the models, enabling them to learn patterns and make accurate sentiment predictions.

(i) Logistic Regression: Logistic regression is a simple yet powerful classification algorithm. Text data is transformed into numerical feature vectors using Term Frequency-Inverse Document Frequency (TF-IDF) vectorization, which captures the importance of words. A streamlined pipeline is constructed to standardize the data and implement logistic regression. GridSearchCV is used for hyperparameter tuning to identify optimal parameters, ensuring that the model effectively learns patterns while maintaining computational efficiency.

(ii) Random Forest: The random forest algorithm is an ensemble method that builds multiple decision trees and combines their outputs for more accurate and stable predictions. Text data undergoes *TF-IDF* vectorization, followed by hyperparameter tuning using *GridSearchCV* to optimize model performance. Random forest is particularly robust with high-dimensional data, benefiting from its ability to reduce overfitting and enhance prediction accuracy.

(iii) Naïve Bayes: The Multinomial Naïve Bayes algorithm is highly efficient for text classification tasks. It converts textual data into numerical form using TF-IDF vectorization and optimizes the smoothing parameter through *GridSearchCV*. Naïve Bayes is computationally efficient, making it well-suited for large datasets. It effectively identifies optimal alpha values to improve performance when trained with extensive data, delivering fast and reliable results.

(iv) Support Vector Machine (SVM): Support Vector Machines (SVMs) are powerful classifiers designed for high-dimensional spaces, making them ideal for natural language processing tasks like sentiment analysis and document classification. Text data is vectorized using TF-IDF, and GridSearchCV is employed to tune hyperparameters such as the regularization parameter (C) and kernel coefficient (γ). SVMs efficiently handle large feature sets while maintaining high accuracy and memory efficiency, excelling in complex NLP scenarios.

(v) Feed Forward Neural Network (FFNN): A feed-forward neural network is employed for text categorization, consisting of an embedding layer, a flattening layer, and one or more dense layers. Word embeddings are used to represent words in a dense vector space, improving the model's ability to capture semantic relationships. The Adam optimizer is utilized during training, dynamically adjusting the learning rate to enhance convergence and performance. This straightforward architecture is effective for text classification tasks due to its adaptability and simplicity.

(vi) Long Short-Term Memory (LSTM): Long Short-Term Memory (LSTM) networks, a specialized type of recurrent neural network (RNN), excel at sequence prediction tasks by capturing long-term dependencies in sequential data. In this implementation, LSTM layers are combined with embedding layers to process text input, while the Adam optimizer facilitates efficient training. LSTMs are particularly well-suited for tasks where context and order matter, such as text classification, as they preserve information over longer sequences.

(vii) Convolutional Neural Network (CNN): The CNN model captures local dependencies within text by applying 1D convolutions to word embeddings across sentences or documents. These convolutional layers detect patterns, such as phrases or word combinations, indicative of sentiment. Pooling layers (max or average) follow the convolutions to reduce dimensionality and extract the most relevant features, culminating in final classification through dense layers. The Adam optimizer is used during training, ensuring efficient learning. This architecture excels at capturing localized patterns in text data, making it effective for tasks like sentiment analysis.

E. Model Optimization

To optimize model performance, hyperparameter tuning was conducted for both machine learning (ML) and deep learning (DL) models. For ML models, an exhaustive search of parameter values was performed using *GridSearchCV*, which systematically tests various combinations of hyperparameters. This process ensures each model is equipped with its optimal settings. For instance, logistic regression models were evaluated with different values of the regularization parameter *C* through *GridSearchCV*. The method fits multiple models with different hyperparameter values using cross-validation, identifying the configuration that yields the best performance metrics on the validation set, thereby improving prediction accuracy.

For deep learning models, hyperparameter tuning involved adjusting configurations such as the number of layers, units per layer, batch size, epochs, and learning rates. Specific settings for the Adam optimizer were also explored, including a learning rate of 0.001, which was tested and trained to ensure quick convergence and higher accuracy. By systematically testing these settings, the deep learning models achieved superior performance.

In comparison, traditional ML systems typically require fewer hyperparameters for optimization, resulting in optimal performance. Conversely, deep learning models, due to their complexity, demand extensive hyperparameter tuning to achieve desired outcomes. By systematically adjusting these parameters, the models in this project demonstrated improved accuracy and reliability. Table 2 below summarizes the hyperparameter tuning process and configurations applied in this study.

TABLE II. HYPERPARAMETER TUNING FOR EACH MODEL

Model	**Parameter**	**Value**
Logistic Regression	C max_iter solver	0.01 1000 'saga'
Random Forest	n_estimators max_depth min_samples_split Bootstrap	150 None 2 True
Naive Bayes	Alpha	0.1
Support Vector Machine (SVM)	C gamma kernel	1 0.1 'rbf'
Feed Forward Neural Network (FFNN)	embedding_dim units_dense_1 units_dense_2 optimizer learning_rate batch_size epochs	100 128 64 Adam 0.001 32 50

Long Short-Term Memory (LSTM)	embedding_dim units_lstm optimizer learning_rate batch_size epoches	100 128 Adam 0.001 32 50
Convolutional Neural Network (CNN)	embedding_dim filters kernel_size pool_size units_dense optimizer learning_rate bach_size epoches	100 128 5 2 128 Adam 0.001 32 50

F. Model Evaluation

Using key metrics such as accuracy, precision, recall, and F1-score, we assessed each model's performance and ability to classify sentiments effectively. Confusion matrices provided visual insights into correct and incorrect predictions, highlighting areas of underperformance and opportunities for improvement. This comprehensive evaluation allowed for a thorough comparison, identifying the most reliable and accurate models for sentiment analysis.

IV. RESULTS

A. Sentiment Analysis

The sentiment analysis of the candidates reveals distinct patterns that offer valuable insights into voter behaviour. The sentiment analysis results are presented in figure 1, and below is a summary of the key findings from the sentiment distribution for each candidate:

(i) OfficialABAT: OfficialABAT received the lowest percentage of positive comments, at 35.5%, indicating potential areas for improvement in terms of campaign promises, personality, or ideas. The proportion of negative comments is also significant, at 24.1%, suggesting a notable segment of the electorate is dissatisfied with or opposed to this candidate. Meanwhile, the neutral sentiment is the largest category, at 40.4%, representing a considerable group of voters who are undecided or indifferent and, therefore, a crucial target for engagement.

(ii) Peter Obi: Peter Obi garnered the highest positive sentiment, at 42.6%, reflecting strong support from voters and an excellent base within the electorate. Negative comments remain relatively low, at 20.6%, demonstrating less disapproval compared to other candidates. However, the neutral sentiment, at 36.8%, suggests that while many voters are inclined to support him, a significant portion of the population remains undecided or apathetic, representing an opportunity for further outreach and persuasion.

(iii) Atiku: Atiku received a fairly good proportion of positive sentiment, at 40.1%, indicating a moderate level of support. Negative comments stand at 22.0%, showing some opposition but not as pronounced as for OfficialABAT. The neutral sentiment, at 37.9%, is slightly lower than Peter Obi’s, suggesting that Atiku's followers may be more committed, though a notable share of voters still remains undecided.

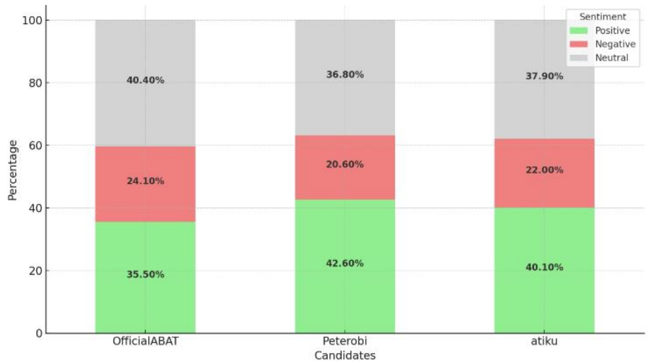

Fig. 1. Sentiment analysis results

These findings highlight the varying levels of voter sentiment toward the candidates, with clear opportunities for each to improve their outreach and engagement strategies. The neutral sentiment category, in particular, represents a significant portion of the electorate and a key focus area for influencing voter behavior.

B. Confusion Matrices

The confusion matrix is an essential tool for evaluating the effectiveness of classification models. It highlights the counts of true negatives (TN), true positives (TP), false negatives (FN), and false positives (FP), offering valuable insights into a model's prediction performance. Specifically, Type I error refers to false positives, while Type II error corresponds to false negatives. The confusion matrices presented in Figure 2, summarize the performance of all models.

Fig. 2. Confusion matrix for all models

According to Figure 2, Logistic regression demonstrated a good performance with fewer false positives (FP) and similar false negatives (FN) compared to Naïve Bayes, reducing Type I errors (FP). However, it still struggled with some positive instances, as indicated by its Type II errors (FN). The random forest algorithm showed balanced performance, with moderate counts of FP (4,523) and FN (1,140), effectively capturing complex data patterns but failing to significantly reduce both error types. Naïve Bayes exhibited high false positives (4,391) and relatively fewer false negatives (1,311), highlighting high specificity but low sensitivity, leading to frequent misclassification of neutral or negative sentiments as positive.

Among machine learning models, SVM delivered the best results, achieving minimal FP (1,215) and FN (822), excelling in correctly identifying neutral, negative, and positive instances with low Type I and Type II error rates. The feed-forward neural network (FFNN) also performed well, with relatively low FP (730) and FN (917). While it faced challenges in detecting all positive instances, its high accuracy and precision demonstrated its effectiveness for sentiment classification.

LSTM achieved outstanding results, recording the lowest FP (530) and FN (761) among all models. Its ability to capture long-term dependencies in textual data significantly enhanced its precision and accuracy, making it the best-performing deep learning model in the study. CNN showed strong results with medium-range FP (579) and FN (1,316). While slightly behind LSTM, CNN effectively captured local patterns but exhibited room for improvement in reducing misclassifications.

C. Evaluation Metrics

The evaluation of various machine learning and deep learning models revealed key insights into their effectiveness in sentiment classification tasks. The models were assessed using crucial metrics, including F1-score, precision, recall, and accuracy, providing a comprehensive understanding of their performance across positive, neutral, and negative sentiments. The summary of model performances is shown in Table 3.

TABLE III. PRECISION, RECALL, F1 SCORE AND ACCURACY FOR EACH MODEL

Model	Label	Precision	Recall	F1-Score	Accuracy
RF	Negative	0.80	0.51	0.62	0.76
	Neutral	0.71	0.87	0.78	
	Positive	0.79	0.79	0.79	
LR	Negative	0.78	0.53	0.63	0.75
	Neutral	0.70	0.90	0.78	
	Positive	0.86	0.77	0.81	
SVM	Negative	0.90	0.83	0.86	0.76
	Neutral	0.88	0.93	0.90	
	Positive	0.94	0.94	0.94	
Naïve Bayes	Negative	0.77	0.48	0.59	0.71
	Neutral	0.71	0.78	0.74	
	Positive	0.72	0.81	0.76	
LSTM	Negative	0.95	0.95	0.95	0.95
	Neutral	0.95	0.94	0.94	
	Positive	0.96	0.97	0.96	
CNN	Negative	0.97	0.92	0.94	0.94
	Neutral	0.93	0.93	0.93	
	Positive	0.92	0.98	0.95	
FFNN	Negative	0.93	0.94	0.94	0.94
	Neutral	0.92	0.92	0.92	
	Positive	0.96	0.95	0.96	

According to Table 3, LSTM (Long Short-Term Memory) emerged as the best-performing model overall, achieving an outstanding overall F1-score of 0.95 and a remarkable accuracy of 0.95. Its ability to capture long-term dependencies in textual data enabled it to outperform other architectures. LSTM excelled in positive sentiment classification with a precision of 0.96, indicating minimal misclassification. This performance underscores its suitability for processing textual information where context and sequential dependencies are critical.

Feed-Forward Neural Network (FFNN) also demonstrated exceptional performance, achieving an overall F1-score of 0.94 and an accuracy of 0.94. It maintained high precision and recall across all sentiment classes, excelling in positive sentiment classification with a precision of 0.96. The model's multi-layered architecture enabled it to effectively learn intricate patterns in the data, making it highly suitable for sentiment analysis tasks.

Convolutional Neural Network (CNN) performed comparably well, achieving an overall F1-score of 0.94 and an accuracy of 0.94. The model effectively captured local patterns within sentences through its convolutional layers, enabling accurate classifications across negative, neutral, and positive sentiments. Compared to LSTM and FFNN models, CNN achieved slightly lower precision for positive sentiment classification (0.92) but outperformed in negative sentiment classification with a precision score of 0.97. Although it achieved slightly lower accuracy than LSTM, CNN's robust performance highlights its efficiency and reliability in sentiment classification.

Support Vector Machine (SVM) was the best-performing classic machine learning algorithm, delivering an overall F1-score of 0.90 and an accuracy of 0.76. It performed consistently well across all sentiment classes, excelling in positive sentiment classification with 0.94 precision and recall. The model demonstrated a balanced approach with minimal false negatives and false positives, positioning it as a strong candidate for robust sentiment classification tasks.

Random Forest achieved a simillar accuracy score of 0.76 but a lower overall F1-score of 0.73. Its performance on negative sentiments was weaker, with an F1-score affected by a lower recall rate (0.51) despite a precision of 80%. This highlights the ensemble model's strength in certain sentiment classes but limitations in handling imbalanced distributions.

Logistic Regression demonstrated balanced performance, achieving an overall F1-score of 0.76 and an accuracy of 0.75. The model excelled in predicting positive sentiments with an impressive precision rate of 86%. However, its recall for negative sentiments was relatively low, indicating some misclassification of negative instances. Despite its simplicity, logistic regression showed potential for improvement, particularly in classifying neutral and negative sentiments more effectively.

Naïve Bayes recorded the lowest accuracy among all evaluated models at 0.71 and an overall F1-score of 0.71. While it performed well in predicting positive sentiments, achieving a recall of 0.81, its precision for negative sentiments was relatively low at 0.77. This reflected a higher rate of false positives, attributed to the assumption of feature independence, which often does not hold in textual data.

Overall, deep learning models, particularly LSTM, significantly outperformed traditional machine learning methods in sentiment classification tasks, as reflected in key evaluation metrics such as precision, recall, F1-score, and accuracy. LSTM achieved the best overall performance, leveraging its ability to capture long-term dependencies and sequential context. Among classic machine learning models, SVM stood out with an accuracy of 0.76, overall F1-score of 0.91, and consistent performance across all classes.

These results emphasize the importance of selecting appropriate sentiment analysis architectures to optimize classification performance. Additionally, they highlight the potential for further research into advanced deep learning architectures and hybrid approaches to improve sentiment analysis capabilities in future studies. Figure 3 summarizes the accuracy performance of all evaluated models, providing a visual comparison of their strengths and weaknesses.

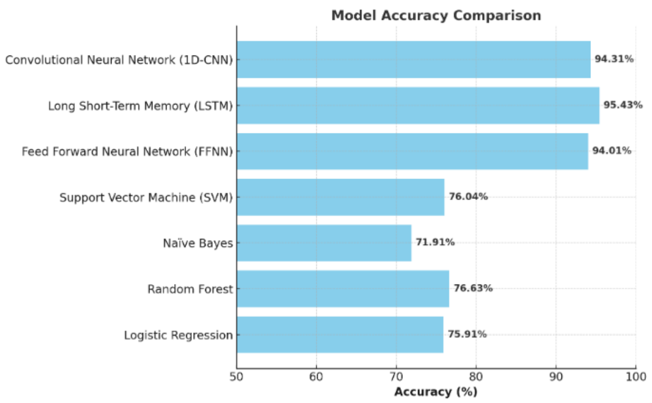

Fig. 3. Comparison of model performance based on their accuracy

V. Discussion

This research addresses several gaps identified in the literature review. One major gap is the limited comparison of deep learning architectures with traditional machine learning models. This study bridges that gap by evaluating a variety of models, including Naïve Bayes, logistic regression, Support Vector Machine (SVM), random forest, Long Short-Term Memory (LSTM), 1D Convolutional Neural Network (1D-CNN), and feed-forward neural network (FFNN). By doing so, it identifies the most effective methods for sentiment analysis in the context of political elections.

Another notable gap was the lack of datasets reflecting how political sentiments evolve over time. To address this, data was collected from tweets surrounding the Nigerian 2023 elections, capturing sentiments before and after significant political events. This temporal dimension enhances the investigation by revealing how public opinion shifts in response to political developments over time.

According to the results, the most effective methods for sentiment analysis are the LSTM and CNN models, outperforming all other approaches in terms of accuracy. These deep learning models excel due to their ability to capture complex patterns and relationships in text data, which often contains diverse expressions, such as emojis and abbreviations, frequently found in social media posts. LSTM, in particular, demonstrates strong performance with sequential data due to its robustness over time. It can interpret subtle changes in word order, ensuring accurate predictions. This finding aligns with previous research emphasizing the importance of temporal dependencies in large volumes of unstructured text data. Similarly, the 1D-CNN model effectively extracts key features within text through convolution operations, enhancing sentiment classification. While FFNN performed relatively well, its simpler structure limited its ability to capture intricate relationships among words, making it less effective compared to LSTM and CNN.

Among the traditional machine learning models, random forest was the best performer. However, it is evident that deep learning models significantly outshine conventional approaches, highlighting the advancements made in sentiment analysis techniques.

A. *Implications for Political Sentiment Analysis*

The findings of this study hold significant value for political analysts, campaign strategists, and policymakers. Advanced deep learning models provide valuable insights into public sentiments, enabling decision-makers to craft more targeted campaign messages, make informed choices, and predict election outcomes with greater precision. By using deep learning models, analysts can achieve a more comprehensive understanding of voter sentiment, identifying key issues that resonate with voters and optimizing election strategies accordingly.

The ability to monitor real-time social media data offers an invaluable perspective on public opinion. In dynamic political environments where sentiments can shift rapidly due to events or developments, this capability becomes even more critical. Additionally, this research demonstrates the potential of sentiment analysis as a tool for predicting election outcomes. Traditional polling methods, while useful, often exclude individuals who do not participate in conventional surveys. Social media sentiment analysis complements these methods, capturing the voices of those who might otherwise be overlooked.

B. *Limitations and Future Directions*

The focus on Twitter data, while rich in sentiment, does not fully represent the entire electorate. Future research should incorporate data from other platforms, such as Facebook, to ensure a more comprehensive analysis. Another area for improvement is the exploration of how sentiments evolve across different stages of the election cycle. Political events can significantly influence public opinion, and capturing these shifts over time would provide deeper insights. Although this study compared several models, there is room to include more emerging techniques in sentiment analysis. Combining and hybridizing these methods could yield even better results than relying on a single approach.

VI. Conclusion

This research highlights the effectiveness of deep learning models for sentiment analysis within the political context of the 2023 Nigerian election. Through comparative analysis, it was found that traditional machine learning methods underperform compared to deep learning models in accurately detecting social media sentiment. These findings are particularly significant for political stakeholders, as they enable a deeper understanding of public opinions, real-time sentiment tracking, and more precise predictions of electoral outcomes.

Furthermore, focusing exclusively on Twitter data fails to capture the sentiments of the broader electorate. Future studies should incorporate multiple social media platforms to achieve a more comprehensive analysis. Additionally, this study did not extensively explore the dynamic nature of emotions over time, leaving an opportunity for future researchers to investigate shifts in public attitudes during various stages of the election process. Despite these constraints, the study makes significant contributions by addressing key gaps, such as lack of large dataset in the context of politics and the limited comparative analysis between deep learning architectures and traditional machine learning models for setiment analysis. It also provides a custom dataset and insights specific to the Nigerian political landscape.

Moving forward, future research should evaluate various models on multilingual datasets while considering language-specific features during training. These factors are likely to play critical roles in enhancing the accuracy and applicability of sentiment analysis in diverse political and cultural contexts.

REFERENCES

[1] Ali, T., Rehman, M., Javed, A., & Khan, A. (2022). Deep Learning-Based Election Results Prediction Using Twitter Activity. *Journal of Information Science and Engineering, 38*(2), 135-142.

[2] Alvi, F., Iqbal, Z., & Rehman, A. (2023). On the Frontiers of Twitter Data and Sentiment Analysis in Election Prediction. *Journal of Information Science and Engineering, 49*(3), 205-220.

[3] Bilal, M. et al. (2019). 2018 Pakistan General Election: Understanding the Predictive Power of Social Media. *12th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics, MACS 2018 - Proceedings*, 1-6. doi: 10.1109/MACS.2018.8628445.

[4] Habimana, O., Mpotu, T., & Nduwayezu, E. (2020). Sentiment Analysis Using Deep Learning Approaches: An Overview. *International Journal of Data Mining & Knowledge Management Process, 10*(1), 1-10.

[5] Hamed, H. A., Ezzat, A., & Hefny, H. (2020). A Survey of Sentiment Analysis Methods. *Journal of Artificial Intelligence Research, 67*, 1-33.

[6] Hidayatullah, A., Cahyaningtyas, F., & Hakim, A. (2021). Deep Learning for Sentiment Analysis of Indonesian Presidential Election Tweets. *Journal of Information Processing Systems, 17*(4), 714-724.

[7] Ismail, N., Harous, S., & Belkhouche, B. (2016). A Comparative Study of Machine Learning Classifiers for Twitter Sentiment Analysis. *Proceedings of the International Conference on Innovations in Information Technology (IIT)*, 1-6.

[8] Joseph, T. (2019). Decision Tree Classifier for Predicting the Outcome of Indian General Elections Based on Twitter Sentiment Analysis. *International Journal of Computer Applications, 182*(21), 25-30.

[9] Kristiyanti, T., Nugroho, A. S., Wahyuni, S., & Hutama, R. F. (2019). Sentiment Analysis Using SVM with Particle Swarm Optimization and Genetic Algorithms for Feature Selection. *Procedia Computer Science, 157*, 240-247.

[10] Kumar Singh, D., Kumar Gupta, S., & Mohan Singh, A. (2017). Sentiment Analysis of Twitter Data Regarding 2017 Punjab Legislative Assembly Election. *Procedia Computer Science, 115*, 346-353.

[11] Mahmood, A., Qazi, Z., Afzal, M., & Ali, A. (2013). Mining Twitter Big Data to Forecast 2013 Pakistan Election Winner. *International Journal of Computer Science Issues, 10*(1), 11-15.

[12] Manikandan, J., Srinivasan, R., & Venkatesh, S. (2023). Twitter Sentiment Analysis Using Various Machine Learning Techniques. *Journal of Big Data, 10*(2), 87-101.

[13] Olabanjo, O., Adekanbi, A., & Aderemi, A. (2022). From Twitter to Aso-Rock: A Natural Language Processing Spotlight for Understanding Nigeria 2023 Presidential Election. *Journal of Information Science and Engineering, 40*(3), 189-197.

[14] Pota, M., Sapienza, A., & Greco, S. (2018). A Subword-based Deep Learning Approach for Sentiment Analysis of Political Tweets. *Journal of Information Technology and Politics, 15*(2), 140-157.

[15] Razzaq, A., Qamar, A. M., & Bilal, H. S. (2014). Prediction and Analysis of 2013 Pakistan General Elections Using Sentiment Analysis. *International Journal of Computer Applications, 96*(25), 1-6.

[16] Sathya, T., Thulasimani, K., & DevaSurithi, R. (2023). A Study on Twitter Sentiment Analysis Using Deep Learning. *Journal of Information Technology, 58*(2), 91-110.

[17] Shaikh et al. (2016). Election Result Prediction Using Twitter Sentiment Analysis. *International Journal of Computer Applications, 135*(9), 20-25.

[18] Shah, D., & Li, Y. (2021). Twitter Based Sentiment Analysis of Each Presidential Candidate Using Long Short-Term Memory. *Journal of Big Data, 8*(1), 101-110.

[19] Sharma, R., & Moh, T.-S. (2016). Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter. *International Journal of Computer Applications, 134*(16), 1-6.

[20] Smailović, J., Grčar, M., Lavrač, N., & Žnidaršič, M. (2015). Monitoring the Twitter Mood during the Bulgarian Elections. *Journal of Information and Telecommunication, 4*(2), 1-15.

[21] Soylu, A., & Baday, A. (2020). Predicting Istanbul Mayoral Election Outcomes Using Twitter Data. *Journal of Information Science and Engineering, 36*(2), 187-196.

[22] Umarani, J., Juliana, J., & Deepa, P. (2021). Sentiment Analysis Using Various Machine Learning and Deep Learning Techniques. *Journal of Big Data, 8*(1), 115-126.

[23] Durga Rani, G., Sridhar, V., J., & Sai Prasad, M. (2024). Sentiment analysis on Twitter using ML and DL techniques. Journal of Artificial Intelligence Research, 75(1), 1-20.

[24] Dixon, S. (2022). Number of worldwide social network users 2027. *Statista*. Retrieved from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

[25] Lagrari, F.-E., & Elkettani,Y. (2021). Traditional and Deep Learning Approached for Sentiment Analysis: A Survey. Advances in Science, Technology and Engineering Systems Journal, 6(5), 1-7.