OPEN ACCESS Freely available online

Research Article

Vol. 1, Issue 2, July-December 2026 | ISSN: 3139-6267

Journal of Artificial Intelligence and Digital Health

Transformer-based Framework for Election Outcome Prediction using Electorate Opinion

Umoh Augustine Uduak[1,5*], Victor Eshiet Ekong[2,5], Temitope Joel Fakiesi[3,5], Philip Asuquo[4,5]

[1]*Department of Information Systems, Faculty of Computing, University of Uyo, Uyo, Nigeria*
[2]*Department of Software Engineering, Faculty of Computing, University of Uyo, Uyo, Nigeria*
[3]*Department of Computer Science, Faculty of Computing, University of Uyo, Uyo, Nigeria*
[4]*Department of computer Engineering, Faculty of Engineering, University of Uyo, Uyo, Nigeria*
[5]*Tetfund Center of Excellence in Computational Intelligence, University of Uyo, Uyo, Nigeria*

***Correspondence to:** Umoh Augustine Uduak[1,5], [1]Department of Information Systems, Faculty of Computing, University of Uyo, Uyo, Nigeria; [5]Tetfund Center of Excellence in Computational Intelligence, University of Uyo, Uyo, Nigeria, E-mail: uduakumoh@uniuyo.edu.ng

Received: July 20, 2026; **Manuscript No:** JAID-26-2523; **Editor Assigned:** July 27, 2026; **PreQc No:** JAID-26-2523 (PQ); **Reviewed:** August 06, 2025; **Revised:** August 06, 2025; **Manuscript No:** JAID-26-2523 (R); **Published:** September 09, 2026

Citation: Uduak UA, Ekong VE, Fakiesi TJ, Asuquo P (2026). Transformer-based Framework for Election Outcome Prediction using Electorate Opinion. J. Artif. Intell. Digit. Health. Vol.1 Iss.2, September (2026), pp:160-169.

ABSTRACT

The accurate prediction of election outcomes plays a pivotal role in enhancing democratic processes by offering insights into voter behaviour and informing strategic decision-making for policymakers, political parties, and media organizations. This study introduces a novel framework leveraging the Robustly Optimized BERT Pretraining Approach (RoBERTa) to predict election results through sentiment analysis of public opinions expressed on X (formerly Twitter). By analyzing nuanced linguistic patterns in social media discourse, the research addresses the challenges of sentiment ambiguity and class imbalance. A case study on the 2023 Nigerian presidential election, focusing on Akwa Ibom State, demonstrates the effectiveness of this framework. The RoBERTa model achieved notable accuracy in predicting election outcomes, highlighting its potential for bridging the gap between online sentiment and real-world electoral results. The model achieved a higher precision, recall, and F1-score for the Negative class, with values of 0.95, 0.97, and 0.96, respectively, indicating exceptional performance in identifying and classifying negative sentiments. This framework underscores the transformative power of transformer-based architectures in electoral studies, offering avenues for more transparent and data-driven decision-making in political analysis.

Keywords: Election; Transformer; Prediction; RoBERTa; Sentiment

INTRODUCTION

The evolution of democratic participation has been significantly influenced by digital transformation, especially with the emergence of social media platforms such as X, Facebook, and Instagram. These platforms provide an expansive data source for analyzing public opinion, making traditional reliance on scheduled campaign events less significant. However, translating online sentiments into accurate election predictions remains challenging due to discrepancies between digital discourse and real-world outcomes. Traditionally, voters relied on scheduled campaign events to hear candidates' manifestos Nowadays, candidates and their supporters are turning more to social media to share their manifestos and aspirations, allowing for broader and more immediate communication [1-2]. This shift has turned social media into a robust data source for analyzing public opinion. However, there is often a discrepancy between online sentiment and actual election outcomes, as candidates with low online ratings can still win due to various influencing factors.

To bridge this gap, it is crucial to understand voters' thoughts and emotions accurately. Computational Intelligence (CI) within Artificial Intelligence (AI) focuses on creating adaptive systems that learn from and adjust to complex, dynamic environments. Inspired by human cognitive processes, CI comprises methods like Neural Networks (NN), Fuzzy Logic (FL), Genetic Algorithms (GA), and Machine Learning (ML), deep learning (DL), and swarm intelligence (SI) [3]. These techniques empower systems to process and analyze data, recognize patterns, and make informed decisions amidst uncertainty or incomplete information. CI applications span optimization problems, data mining, expert systems, and autonomous decision-making, highlighting its capacity to evolve and learn from experience [4-5]. Machine Learning (ML) underpins CI by creating intelligent systems that can learn, reason, and solve problems. ML models are trained to generalize patterns from observed data Thomas, Eyoh, by improving performance with more examples and experiences [6-7].

ML processes are classified into supervised learning (trained on labeled data), unsupervised learning (identifying patterns in unlabeled data), and reinforcement learning (learning through interaction and feedback). Hybrid systems integrating ML with other CI techniques, such as fuzzy logic or evolutionary algorithms, yield robust and adaptive intelligent systems. Deep Learning (DL), a subset of ML, utilizes artificial neural networks with multiple layers to learn hierarchical data representations. Key types of deep neural networks include Convolutional Neural Networks (CNNs) for image-related tasks and Recurrent Neural Networks (RNNs) for sequential data [8-9]. DL's scalability and the concept of transfer learning enhance its efficiency and effectiveness in various applications [10]. Despite challenges like the need for substantial labelled data, DL continues to advance, exemplified by transformer-based models [11].

The conception of transformer models by Vaswani, revolutionized Natural Language Processing (NLP) and extended their influence on fields like computer vision and robotics [12-13]. Transformer models, utilizing self-attention mechanisms, excel in capturing contextual relationships within data intelligence [14]. Transformer models have various types tailored to a specific task. The Bidirectional Encoder Representations from Transformers (BERT) by Delvin, and its variant Robustly optimized Bidirectional Encoder Representations from Transformers (RoBERTa) by Liu, approaches have been pivotals in shaping the landscape of natural language processing (NLP) models [15-16]. Others include DistilBERT, a distilled version for faster deployment and the ALBERTa model which introduces parameter reduction techniques for improved efficiency. Variants like BERT and its robust version RoBERTa have set new benchmarks in NLP. The Vision Transformer (ViT) further demonstrated transformers' versatility by challenging CNNs in image-related tasks.

Recent studies demonstrate the growing application of transformer-based models to sentiment analysis and political discourse mining. Talaat, showed that BERT-based and hybrid transformer architectures can achieve strong performance in social-media sentiment classification, highlighting their ability to capture contextual linguistic patterns [17]. In the electoral domain, Pandiangan and Wijayakusuma, applied BERT to sentiment analysis of social-media discourse surrounding the 2024 Indonesian presidential election, demonstrating the relevance of transformer models for analyzing voters' online opinions [18].

Khatavkar, her extended this line of research through a multilingual transformer-based approach to political tweets, addressing contextual variations and linguistic complexities in political discourse [19]. Similarly, Babu, investigated political sentiment in X (formerly Twitter) comments using transformer models, confirming the effectiveness of contextual language representations for political sentiment classification [20]. More recently, Nguyen and Văn employed XLM-RoBERTa with parameter-efficient fine-tuning for multiclass political sentiment analysis and addressed class imbalance through oversampling and hyperparameter optimization [21].

Collectively, these studies establish the suitability of transformer-based architectures for political sentiment analysis while also revealing persistent challenges related to linguistic ambiguity, multilingual discourse, and imbalanced sentiment classes. RoBERTa, an optimized version of BERT, refines the pre-training process by eliminating the next sentence prediction objective and introducing dynamic masking. This approach enhances RoBERTa's ability to understand a wider range of linguistic patterns and improve generalization. RoBERTa's extensive pre-training on diverse datasets enables it to excel in various NLP tasks, making it a powerful tool for sentiment analysis and election prediction. However, limited attention has been given to integrating these capabilities with election-outcome prediction within the Nigerian context, particularly at the sub-national level.

This study bridges the above-mentioned gap to build on existing transformer-based approaches by applying RoBERTa to X discourse associated with the 2023 Nigerian presidential election in Akwa Ibom State, with the aim of establishing a data-driven relationship between public sentiment and electoral outcomes. Transformer models have redefined Natural Language Processing (NLP) by efficiently capturing contextual relationships within textual data. While existing literature highlights their success in sentiment analysis and classification, this research explores their application in election outcome prediction, emphasizing the unique challenges of sentiment polarity and class imbalance.

MATERIALS AND METHODS

This study employs a transformer-based framework using RoBERTa for predicting election outcomes. The RoBERTa model was chosen for its robust pretraining and ability to handle diverse linguistic patterns. RoBERTa was selected because it provides several improvements over earlier transformer-based architectures such as BERT, DistilBERT, and ALBERT. Unlike the original BERT model, RoBERTa removes the next-sentence prediction objective, employs dynamic masking during training, utilizes substantially larger training corpora, and applies optimized pretraining strategies that improve contextual language understanding. These improvements enable RoBERTa to capture subtle semantic relationships, informal expressions, abbreviations, and contextual dependencies commonly found in social media texts. Since election-related tweets often contain sarcasm, slang, abbreviations, and rapidly evolving political narratives, RoBERTa provides a more robust representation of public opinion than conventional transformer models, making it particularly suitable for sentiment-driven election prediction. The proposed methodology workflow for transformer-based election outcome prediction using RoBERTa and electorate opinion is shown in Figure 1.

Data Collection

For this study, tweets related to the 2023 Nigerian presidential election were collected from November 2022 to January 2023 using Twitter's REST API. The dataset focuses on Akwa Ibom State, with 45,600 tweets initially collected and reduced to 19,600 after preprocessing. Table 1 present sentiment classifications divided into positive, negative, and neutral

categories, maintaining 36 features throughout the analysis. The format, structure, size and composition of the"election_tweet_data" are depicted in Table 2. Table 3 gives the attributes of "Election_Tweet_Data" Dataset. Table 4 presents the data description of the attributes present in the "Election_Tweet_Data" dataset, along with their corresponding data types. Understanding the nature of these attributes is pivotal for conducting meaningful analyses, such as sentiment analysis or user behavior examination, particularly within the context of social media platforms like X (Twitter).

Data Preprocessing

The initial dataset comprised 45,600 tweets collected through the Twitter REST API between November 2022 and January 2023. To ensure the quality and reliability of the dataset, a multi-stage preprocessing pipeline was implemented before model training.

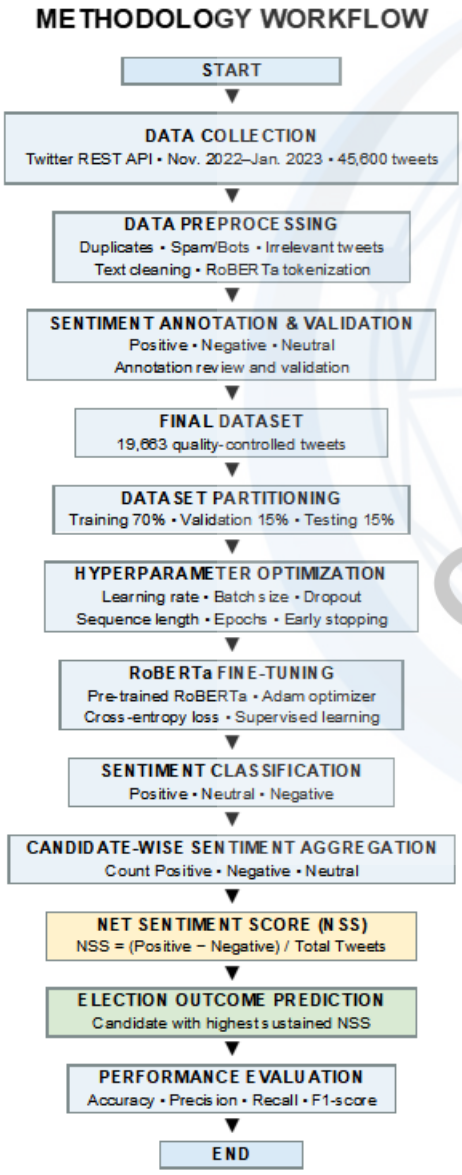

Figure 1: Proposed Methodology Workflow for Transformer-Based Election Outcome Prediction Using Roberta and Electorate Opinion.

The preprocessing involved the following stages:

1. Removal of duplicate tweets and duplicate user posts.
2. Elimination of retweets that contained no additional user-generated comments.
3. Removal of spam messages, advertisements, promotional tweets, and bot-generated content.
4. Exclusion of tweets unrelated to the Nigerian presidential election using keyword filtering.
5. Removal of incomplete records and tweets containing missing textual information.
6. Elimination of tweets originating outside the geographical scope of the study.
7. Removal of non-English tweets to ensure linguistic consistency.
8. Cleaning textual contents by removing URLs, hashtags, user mentions, HTML symbols, excessive punctuation, and unnecessary white spaces.
9. Conversion of all text to lowercase.
10. Tokenization using the Hugging Face RoBERTa tokenizer.
11. Padding and truncation of token sequences to a fixed maximum sequence length suitable for model training.

After applying these preprocessing procedures, the dataset was reduced from 45,600 to 19,663 high-quality tweets, which formed the final dataset used for sentiment classification. The resulting sentiment distribution consisted of 40.8% positive, 38.8% negative, and 20.4% neutral tweets. These preprocessing procedures substantially reduced noise and improved the quality of textual representations used during model training. The key concept and keywords are shown in Figure 2.

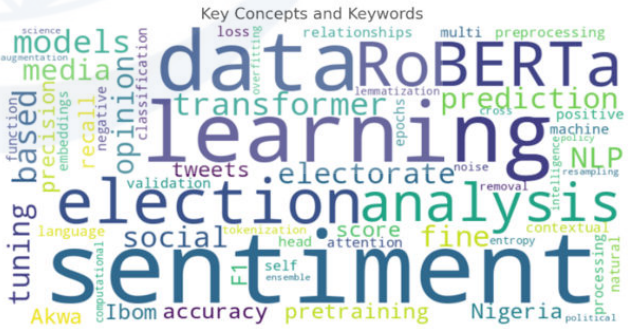

Figure 2: Key Concepts for Election Data Collection

Dataset	Size	Features	Positive Sentiments	Negative Sentiments	Neutral Sentiments
Initial	45,600	36	15,600	14,800	15,200
Final	19,600	36	8,000	7,600	4,000

Table 1: Dataset Sentiment Classifications **Source:** Field Data (2024)

Dataset Name	Dataset Size		Features		
	Rows	Columns	Categorical	Numeric	Meta(text)
election_tweet_data	19663	36	5	18	13

Table 2: Description Statistics for Election_Tweet_Data Dataset **Source:** Field Data (2024)

Attribute Name	Data Type	Description
username	String	Twitter handle of the user posting the tweet.
Text	String	Content of the tweet.
Location	String	Geographical location of the user.
sentiment	String	Sentiment label (positive/negative/neutral).
timestamp	DateTime	Date and time of the tweet.

Table 3: Attributes of "Election_Tweet_Data" Dataset

S/N	Attribute	Data Type	Description
1	username	String	The username of the Twitter user who posted the tweet.
2	name	String	The display name of the Twitter user.
3	place	String	The location from which the tweet was posted.
“	“	“	“
34	time	Object	The time when the tweet was posted.
35	time_zone	String	The timezone in which the tweet was posted.
36	User_id	String	The ID of the Twitter user who posted the tweet.

Table 4: Data Description for Election_Tweet_Data Dataset

Source: Researcher (2024)

Experimental Setup

The methodology involves the following key steps: Data Preprocessing - collected tweets were cleaned to remove noise such as URLs, emojis, and stop words. Tokenization was performed using the Hugging Face's RoBERTa tokenizer to convert text into input token sequences Sentiment Annotation: Each tweet was assigned one of three sentiment labels (Positive, Negative, or Neutral) according to the opinion expressed toward a presidential candidate, political party, or election-related issue. Initially, an automated sentiment annotation procedure was employed to generate preliminary labels using predefined sentiment lexicons and contextual language representations. Subsequently, the annotated tweets were independently reviewed by two researchers with expertise in Natural Language Processing and political discourse analysis. Cases where disagreements occurred were jointly examined until consensus was reached. Tweets expressing support, approval, confidence, or favourable opinions toward a candidate or party were labelled Positive, while tweets expressing criticism, dissatisfaction, rejection, or opposition were labelled Negative. Tweets presenting factual information, balanced opinions, or insufficient emotional polarity were classified as Neutral. This two-stage annotation process improved the reliability and consistency of the labelled dataset before model training.

The Architecture of RoBERTa model is shown in Figure 3. The RoBERTa model pre-trained on diverse textual corpora, was fine-tuned for election sentiment analysis. The key components include input embedding layer; converts tokenized text into dense vectors. Transformer encoder- utilizes multi-head self-attention mechanisms and positional encodings to capture contextual dependencies. Output layer - A classification head applies a softmax activation function to predict sentiment classes. Fine-tuning process was performed. The pre-trained RoBERTa model was fine-tuned on the election dataset using supervised learning.

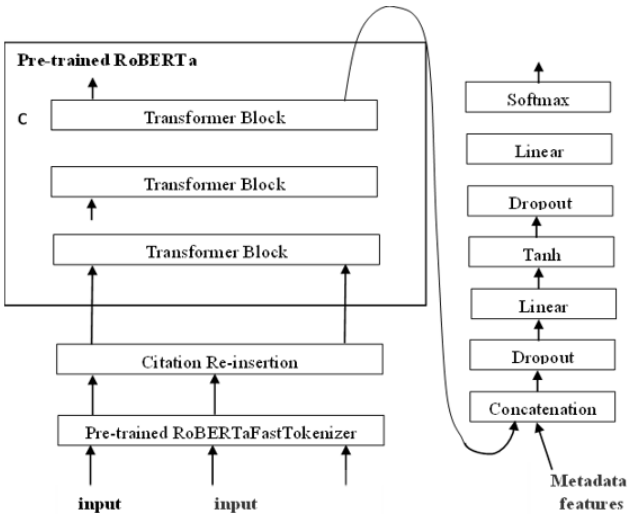

Figure 3: The RoBERTa Architecture for Election Outcome Prediction using Electorate Opinion

Cross-entropy loss was employed as the loss function, optimized using the Adam optimizer with a learning rate of 2e-5. The model was trained over 10 epochs with a batch size of 32. Dropout layers were included to mitigate overfitting. The final preprocessed dataset was randomly partitioned using stratified sampling into 70% training, 15% validation, and 15% testing datasets to preserve the class distribution across all subsets. The training dataset was used for model learning, the validation dataset for hyperparameter optimization and early stopping, while the testing dataset was reserved exclusively for final model evaluation.

Hyperparameter optimization was conducted through systematic experimentation by varying several model parameters, including the learning rate (1×10^{-5}, 2×10^{-5}, and 5×10^{-5}), batch size (16 and 32), dropout probability (0.1 and 0.3), maximum sequence length (128 and 256 tokens), and training epochs (5, 10, 20, and 30). Model performance on the validation dataset was evaluated using Accuracy, Precision, Recall, and F1-score. The optimal configuration consisted of a learning rate of 2×10^{-5}, batch size 32, dropout rate 0.1, maximum sequence length of 128 tokens, and 10 training epochs, producing the best validation performance while minimizing overfitting. Early stopping based on validation loss was implemented to terminate training once no further performance improvement was observed.

i. Input Layer Unit: In the RoBERTa model, the input layer consists of token embeddings that encode the textual input into numerical representations suitable for processing by the model. The input layer typically accepts tokenized text sequences from the domain of election, where each token represents a word or subword in the input text which are then converted into dense vector embeddings using pre-trained embedding matrices. These input embeddings serve as the initial representation of the input text and are subsequently fed into the model's layers to denote the beginning or classification of a sentence for further processing and feature extraction.

ii. Transformer Encoder/Positional Encoding: The core building block of RoBERTa model is the transformer encoder. It consists of multiple layers, each containing a multi-head self-attention mechanism and a feedforward neural network. Transformer encoder- utilizes multi-head self-attention mechanisms along-with positional encodings to acquire contextual relationships. Sine as well as Cosine functions shown in Equations 1 and 2 were adopted to generate different frequencies for the positional encoding (PE), we use sine and cosine functions at varying wavelengths for each position and each dimension i of the word embedding vector dmodel=512

$$PE_{(pos\ 2i)} = sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (1)$$

$$PE_{(pos\ 2i+1)} = cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (2)$$

Starting from the beginning of the word embedding vector with a fixed size of 512, the index begins at i=0i = 0 and ends at i=511i = 511. The sine function is used for even indices, while the cosine function is applied to odd indices.

iii. Multi-Head Self-Attention: Within each transformer layer, the multi-head self-attention mechanism enables the model to weigh the importance of different words in the input sequence based on their context. The attention mechanism is crucial for capturing long-range dependencies. The vector dimension of each word xn in an input sequence is dmodel=512: pe(xn)=[d1=9.09297407e−01, d2=−4.16146845e01,...,d 512= 1. 00000000e+00]. Each word is mapped to all others to determine its contextual relevance within the sequence. Attention is formulated as "Scaled Dot-Product Attention," expressed in Equation 3 and plug Q, K, and V.

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

All matrices share the same dimensions, allowing for a straightforward application of the scaled dot product to compute attention values for each head. The outputs from all 8 heads are then concatenated to form Z. To derive Q, K, and V, the model is trained using the weight matrices Qw, Kw, and Vw, where each has dk=64 columns and dmodel=512 rows. For instance, Q is obtained by performing a dot product between x and Qw, resulting in Q having a dimension of dk=64. We adjust various parameters, including the number of layers, attention heads, dmodel, dk, and other transformer variables, to optimize our model.

iv. Feedforward Neural Network: we employ a feedforward neural network to process the election information, applying non-linear transformations to capture complex patterns and relationships in the dataset. The input of the feedforward network (FFN) is the dmodel = 512 output of the post-LN of the previous sublayer. The optimized and standardized FFN is given in Equation 4.

$$FFN(x) = max(0, xW_1 + b_1)\,W_2 + b_2 \quad (4)$$

Where, $W1$ and $W2$ denote weight matrices, while $b1$ and $b2$ refer to bias vectors. The input (sentiment values) to this system is represented by x, which undergoes these transformations to yield the final output.

v. Layer Normalization and Residual Connections: These components help with the stability and efficient training of the deep neural networks. Each attention sublayer and feedforward sublayer in the Transformer is followed by post-layer normalization (Post-LN). As shown in Equation 5, Post-LN includes an addition function and a layer normalization process, which is further detailed in Equation 6. The addition function handles the residual connections originating from the sublayer's input selection.

$$. Layer\ Normalization\ (x\ + Sublayer(x)) \quad (5)$$

where, Sublayer(x) refers to the sublayer itself, where xx represents the data available at the input stage of Sublayer(x).

$$LayerNormalization(v) = \gamma\frac{v-\mu}{\sigma} + \beta \quad (6)$$

here, μ represents the mean of v with a dimension of d, given in Equation 7 as:

$$\mu = \frac{1}{d}\sum_{k=1}^{d} v_k \quad (7)$$

σ represents the standard deviation of v with a dimension of d, illustrated in Equation 8.

$$\sigma^2 = \frac{1}{d}\sum_{k=1}^{d} (v_k - \mu) \quad (8)$$

γ is a scaling parameter, and β is a bias vector.

vi. Task-Specific Heads: Task-specific heads interpret the learned representations for the specific task at hand. The Linear layer transform the hidden representation of the model into a vector of logits, which are unnormalized scores representing the likelihood of each class. The Softmax layer then applies the softmax function to these logits, converting them into probabilities, ensuring that they sum up to 1. These output probabilities represent the model's confidence scores for each class and are used to make predictions or decisions based on the highest probability class.

vii. Fine-tuning RoBERTa Model for Election Prediction: The parameters θ of the pre-trained RoBERTa model are fine-tuned by minimizing the task-specific loss function L(θ) on the dataset D as shown in Equation 9.

$$\theta^* = argmin\theta \sum(x_i,y_i \quad) \in D\ \mathrm{L}(\theta, x_i,y_i) \quad (9)$$

where (xi , yi) are the input-output pairs in the dataset and θ* represents the optimized parameters. The diagram for Fine-tuning RoBERTa model is seen in Figure 4.

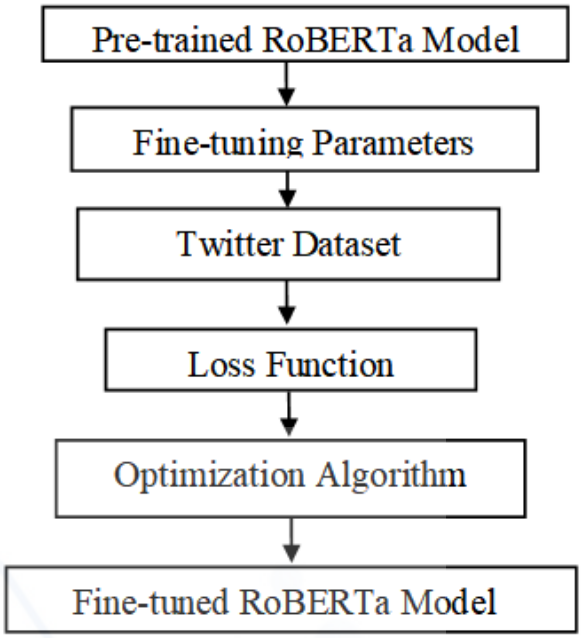

Figure 4: Diagram for Fine-tuning RoBERTa for Election Outcome Prediction using Electorate Opinion

The pre-trained RoBERTa model's parameters are fine-tuned by minimizing a twitter loss function that is optimized using the stochastic gradient descent (SGD) algorithm. Loss function L(θ, xi , yi) is measured the discrepancy between the predicted outputs and the ground truth labels in Equation 10.

$$L(\theta, x_i,y_i) = \mathrm{f}(\mathrm{RoBERTa}(x_i, \theta)) \quad (10)$$

where (xi , θ) denotes the output for input xi with parameters θ, and yi represents the ground truth label. The function f computes the loss, which depends on the specific task (e.g cross entropy loss for classification tasks). The optimization algorithm updates the parameters θ iteratively to minimize the loss function using stochastic gradient descent parameters are updated as Equation 11.

$$\theta_{t+1} = \theta_t - \eta\nabla\mathrm{L}(\theta, x_i,y_i) \quad (11)$$

where η is the learning rate, and ∇L represents the gradient of the loss function with respect to the parameters. The actual value of η, used in this work is 2e-5, and ∇L was optimized during training using the Adam optimizer. The model was trained over 10 epochs with a batch size of 32. Dropout layers were included to mitigate overfitting.

Following sentiment classification, the predicted sentiment labels were aggregated according to the presidential candidates referenced within each tweet. For every candidate, the total numbers of positive, negative, and neutral tweets were computed over the data collection period. A Net Sentiment Score (NSS) was then calculated using

Starting from the beginning of the word embedding vector with a fixed size of 512, the index begins at i=0i = 0 and ends at i=511i = 511. The sine function is used for even indices, while the cosine function is applied to odd indices.

where Positive, Negative, and Neutral represent the respective numbers of classified tweets associated with each candidate.

The Net Sentiment Score was computed periodically to monitor changes in public opinion throughout the election campaign. Candidates exhibiting consistently higher positive sentiment and larger Net Sentiment Scores were predicted to possess stronger electoral support. The candidate with the highest cumulative Net Sentiment Score at the end of the campaign period was identified as the predicted election winner. Finally, these predictions were compared with the officially declared election results to evaluate the effectiveness of the proposed transformer-based election prediction framework.

Evaluation metrics were applied to assess the model performance using accuracy, precision, recall, and F1-score. Separate metrics for positive and negative sentiments provided insights into class-specific performance. Hyperparameters, such as the learning rate and number of attention heads, were tuned to optimize performance. Early stopping criteria were applied to prevent overfitting.

RESULTS AND DISCUSSIONS

In this work, a novel approach to predicting election outcomes using electorate opinions was presented to leverage the capabilities of the RoBERTa model for natural language processing. The model was trained on a comprehensive dataset of public sentiment, expressed through various channels such as social media and surveys, to classify opinions as positive or negative towards candidates and political parties. The implementation was carried out using Python, leveraging on PyTorch and Hugging Face Transformer libraries. Google Colab was used for computational efficiency and data visualization. The results of the study revealed key insights into the performance of the RoBERTa model in sentiment-based election prediction. Pie Chart of Election_Tweet_Data sentiment distribution is given in Figure 4. The pie chart in Figure 4 illustrates the sentiment distribution in the dataset, highlighting the proportions of positive (40.8%), negative (38.8%), and neutral (20.4%) sentiments among the processed tweets. Figure 5 illustrates the sentiment distribution, with a balanced representation of positive and negative classes critical for model fairness. Yet, the neutral class remains underexplored, which could provide additional insights into voter indecision or ambivalence.

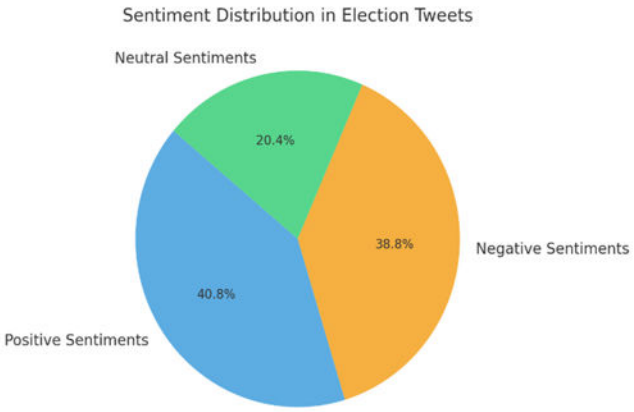

Figure 5: Pie Chart of Sentiment Distribution

Figure 6 gives the pie chart representation of continuous and categorical columns. This depiction reveals that continuous values constitute 52.8%, while categorical values account for 47.2% in the pie chart. Table 5 give model performance across epochs. Table 4 presents the RoBERTa model's performance across different epochs, showing trends in training loss, validation loss, accuracy, F1-score, precision, and recall. From Table 5, the RoBERTa model achieved its highest accuracy of 92.26% at Epoch 3, indicating robust early learning. However, a decline in accuracy and F1-scores was observed in later epochs, highlighting challenges of overfitting and diminishing generalization.

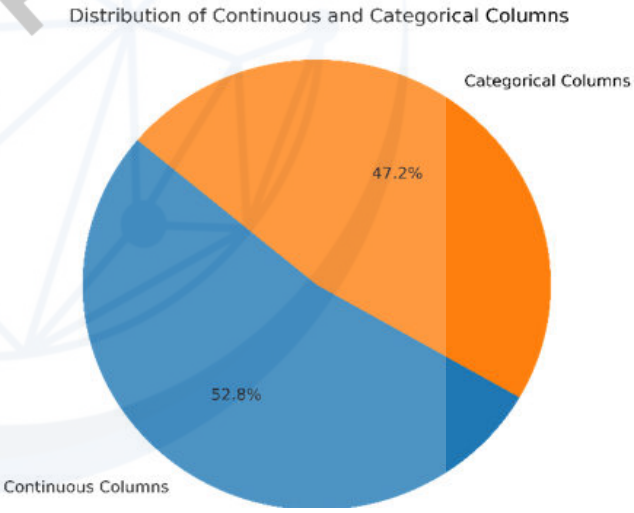

Figure 6: Pie Chart Representation of Continuous and Categorical Columns

Epoch	Training Loss	Validation Loss	Accuracy (%)	F1 Score	Precision	Recall
3	0.16	0.235	92.26	0.55	0.64	0.49
10	0.342	0.307	90.41	0.19	0.56	0.11
35	0.22	0.246	90	0.54	0.49	0.59
100	0.334	0.258	88.82	0.18	0.32	0.12

Table 5: Model Performance across Epochs

The accuracy and loss trends across epochs plot are presented in Figure 7. This illustrates the trends in accuracy, training loss,

and validation loss across different epochs, showing stable accuracy with slight fluctuations and consistent low loss values throughout the training process. Table 6 gives RoBERTa model's election prediction analysis across different epochs interval (Observations). The Table summarize the prediction analysis of the RoBERTa model across different epochs. Figure 8 gives the plots of both the training and validation loss, as well as the accuracy, F1 score, precision, and recall across the epochs. The training and validation losses are shown on the left y-axis, while the accuracy, F1 score, precision, and recall are shown on the right y-axis. The sentiment classification metrics result is presented in Table 7 and graphically represented in Figure 9. The bar chart illustrates the sentiment classification metrics for the "Positive" and "Negative" classes. Precision, recall, and F1 score are represented for each class. This visualization highlights the performance differences across the metrics and classes.

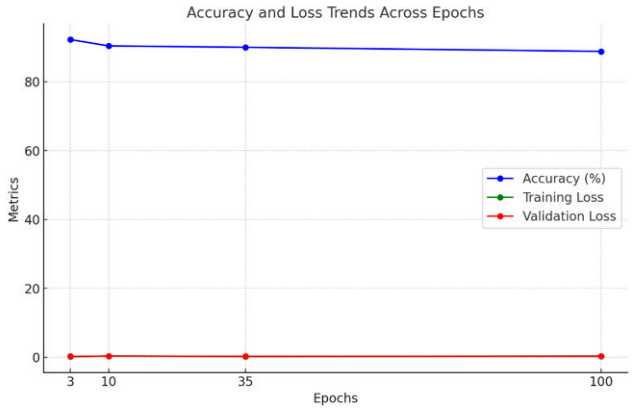

Figure 7: Accuracy and Loss Trends Across Epochs

Epoch	Training Loss	Validation Loss	Accuracy (%)	F1 Score	Precision	Recall
Epoch 3	0.16	0.235632	92.26	0.551929	0.639175	0.48564
Epoch 10	0.3423	0.307732	90.41	0.186957	0.558442	0.112272
Epoch 35	0.2203	0.246416	90	0.535714	0.492341	0.587467
Epoch 100	0.3347	0.258542	88.82	0.177358	0.319728	0.122715

Table 6: RoBERTa model's prediction analysis across different epochs interval (Observations):

Class	Precision	Recall	F1 Score
Positive	0.64	0.49	0.55
Negative	0.95	0.97	0.96

Table 7: Sentiment Classification Metrics

Figure 8: The plots of both the training and validation loss, as well as the accuracy, F1 score, precision, and recall across the epochs.

From Table 6 and Figure 7, Epoch 3 shows the highest accuracy and relatively balanced precision and recall. Epoch 10 exhibits a significant drop in F1 score, especially for the POSITIVE class, with low recall. Epoch 35 demonstrates a slight recovery in F1 score, but with fluctuating precision and recall. Epoch 100 indicates a general decline in prediction performance, suggesting possible overfitting or degradation in model generalization over time. The results from the RoBERTa model across different epochs present valuable insights into the classification and prediction performance of the model, particularly in the context of sentiment analysis.

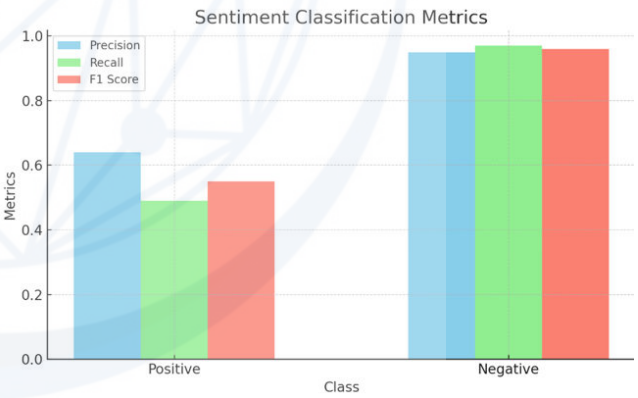

Figure 9: The bar chart illustrates the sentiment classification metrics for the "Positive" and "Negative" classes

Analysis of Epoch 3 on performance metrics indicate that the model exhibits an accuracy of 92.26%, with an F1 score of 0.5519. The precision and recall metrics for this epoch show a moderate balance, with precision at 0.6392 and recall at 0.4856. Classification report indicate that the overall accuracy is high at 92%. For the NEGATIVE class, the precision, recall, and F1-score are robust at 0.95, 0.97, and 0.96 respectively. However, the POSITIVE class struggles, with precision at 0.64, recall at 0.49, and an F1-score of 0.55. This indicates that while the model is proficient in predicting negative sentiments, it has difficulty accurately identifying positive sentiments. Analysis of Epoch 10 on performance metrics indicates that the accuracy slightly drops to 90.41%, and the F1 score for the POSITIVE

class remains low at 0.1869. The precision and recall for this class are 0.5584 and 0.1123, respectively, which indicates a significant imbalance. Classification report shows that the NEGATIVE class continues to perform well with an F1-score of 0.95. However, the POSITIVE class still underperforms, with a significant drop in recall to 0.11, leading to an F1-score of 0.19. This highlights a challenge in the model's ability to generalize well across both sentiment classes, particularly for the POSITIVE class.

Analysis of Epoch 35 and Epoch 100 on the performance metrics shows that the analysis of later epochs (35 and 100) reflects a progressive deterioration in the model's capacity to predict POSITIVE class effectively, indicated by the low F1 scores and a gradual decrease in accuracy, reaching 88.82% by Epoch 100. It is observed that the training loss remains relatively low, but the validation loss fluctuates, suggesting potential overfitting. The model's decreasing recall and precision for the POSITIVE class signal a need for additional tuning or possibly a different approach to class balancing or model architecture. Specifically, it is observed that the model consistently excelled in predicting negative sentiments, with an F1-score peaking at 0.96. This suggests strong feature extraction for this class, likely due to more distinguishable patterns in negative language. Performance for positive sentiments lagged behind, with an F1-score of 0.55, particularly in later epochs, reflecting the model's struggle to accurately predict positive sentiments. This discrepancy points to the need for targeted interventions such as data augmentation or weighted loss functions to address class imbalance.

Generally, the accuracy decreases slightly over time, with precision and recall for the POSITIVE class showing notable declines as the epochs progress, suggesting potential overfitting or class imbalance issues. Training over multiple epochs revealed diminishing returns, with metrics like precision and recall showing instability beyond Epoch 10. This emphasizes the importance of early stopping and dynamic learning rate adjustments during training Analysis of sentiment classification revealed that the model excels in predicting negative sentiments (F1-score: 0.95), while performance for positive sentiments remains suboptimal. Table 7 and Figure 8 present the performance metrics for the sentiment classification model across two classes: Positive and Negative. The model achieved a higher precision, recall, and F1-score for the Negative class, with values of 0.95, 0.97, and 0.96, respectively, indicating exceptional performance in identifying and classifying negative sentiments. In contrast, the Positive class demonstrated lower metrics, with a precision of 0.64, a recall of 0.49, and an F1-score of 0.55, highlighting a need for improvement in the model's ability to accurately classify positive sentiments. This disparity suggests an imbalance in class performance and emphasizes the importance of addressing it to enhance the overall effectiveness of the model. The incorporation of dropout layers and adaptive learning rates helped mitigate overfitting to some extent but was insufficient to sustain performance across extended training epochs. Hyperparameter tuning and additional regularization techniques remain areas for further exploration. The framework's ability to translate online sentiment into actionable election predictions underscores its utility for political strategists and analysts. However, the discrepancy between online discourse and actual voter behaviour requires additional contextual analysis, such as geographical or demographic segmentation.

CONCLUSION AND FUTURE RESEARCH AREA

This study highlights the potential of transformer-based architectures like RoBERTa for sentiment-driven election prediction, offering significant contributions to political analysis. The model demonstrated robust performance in predicting negative sentiments but faced challenges with positive sentiment classification, reflecting a need for targeted improvements. Key findings indicate that early learning stages yield high accuracy, but overfitting becomes a concern with extended training. The RoBERTa model demonstrates strong performance in detecting negative sentiments while struggles with positive sentiment prediction across the epochs in this study. Despite high overall accuracy, the disparity between the NEGATIVE and POSITIVE class performances highlights an area for improvement for future works, possibly through advanced techniques like data augmentation, re-sampling, or fine-tuning with more balanced datasets.

The declining F1 scores and recall for the POSITIVE class as training progresses suggest that adjustments in the training process or model hyperparameters could enhance the model's generalization capabilities, particularly in sentiment analysis tasks. By leveraging the RoBERTa model's capabilities in handling large-scale textual data and capturing nuanced sentiments, this framework aims to provide accurate and timely predictions of election outcomes based on real-time public opinion

The proposed framework underscores the importance of leveraging social media data to bridge the gap between online discourse and real-world outcomes. By refining these methodologies, the research offers a path toward more transparent, accurate, and actionable electoral predictions that can serve as valuable tools for political strategists, media organizations, and policymakers. The performance highlights its potential for providing actionable insights to political stakeholders. However, challenges such as class imbalance and overfitting were noted, particularly in predicting positive sentiments. Future research will integrate text, images, and geospatial data for richer analyses. Oversampling can be used to enhance model generalization, and these will provide deeper insights into voter behaviour.

ETHICAL CONSIDERATIONS AND SAFEGUARDS

This study adhered to ethical standards by safeguarding data privacy and anonymity in line with the General Data Protection Regulation and Nigeria's Data Protection Regulation, ensuring that no personally identifiable information from Twitter data was exposed. Also, data authenticity was ensured by filtering bots, spam, and disinformation, thereby preserving the integrity of the electoral discourse analyzed.

ETHICS APPROVAL

The study was conducted in compliance with institutional research ethics guidelines.

PRIVACY SAFEGUARDS

All data were collected from publicly available tweets in line with Twitter's Developer Policy. No personally identifiable information (PII) was stored; only anonymized tweet IDs and metadata were retained. Data were securely stored on encrypted drives and used solely for academic purposes

FUNDING

This research is supported by TetFund through Institutional Based Research grant 2025 on the project, "High-Precision Sentiment-Based Election Prediction in Emerging Democracies: A Hybrid Intelligence Approach.

CONFLICTS OF INTEREST

The authors declare no conflict of interest. The funders had no role in the study's design; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

REFERENCES

1. Volkens A, Krause W, Lehmann P, Matthieß T, Merz N, Regel S and Weßels B (2017). The Manifesto Data Collection. Manifesto Project (MRG/CMP/MARPOR). Version 2019b. Berlin: Social Science Research Center Berlin (WZB). Bron, 1(1990), 15.
2. Kokab ST, Asghar S, Naz S. Transformer-based deep learning models for the sentiment analysis of social media data. Array. 2022;14:100157.
3. Ekong VE, Asuquo DE, Ejodamen PU. Towards an Analytic Framework for Twitter influencers in Boko Haram Discourse. InProc. of the International Conference on Innovative Systems for Digital Economy 2021 (pp. 204-212).
4. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training.
5. Khetani V, Gandhi Y, Bhattacharya S, Ajani SN, Limkar S. Cross-domain analysis of ML and DL: evaluating their impact in diverse domains. International Journal of Intelligent Systems and Applications in Engineering. 2023;11(7s):253-62.
6. Thomas KK, Anil SP, Ebin Kuriakose NG. Sentiment analysis in product reviews using natural language processing and machine learning. Int J Inf Syst Comput Sci. 2019.
7. Eyoh IJ, Umoh UA, Inyang UG, Eyoh JE. Derivative-Based Learning of Interval Type-2 Intuitionistic Fuzzy Logic Systems for Noisy Regression Problems: IJ Eyoh et al. International journal of fuzzy systems. 2020;22(3):1007-19.
8. Adke A, Ghorpade S, Chaudhari R, Patil S. Navigating the confluence of machine learning with deep learning: Unveiling cnns, layer configurations, activation functions, and real-world utilizations.
9. Banerjee I, Ling Y, Chen MC, Hasan SA, Langlotz CP et al. Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. Artificial intelligence in medicine. 2019;97:79-88.
10. Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning. Cambridge: MIT press; 2016.
11. Qian Q, Huang M, Lei J, Zhu X. Linguistically regularized LSTM for sentiment classification. InProceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) 2017 (pp. 1679-1689).
12. Zhai X, Kolesnikov A, Houlsby N, Beyer L. Scaling vision transformers. In2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2022(pp. 1204-1213). IEEE.
13. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Advances in neural information processing systems. 2017;30.
14. Liu M, Ren S, Ma S, Jiao J, Chen Y, Wang Z, Song W. Gated transformer networks for multivariate time series classification. arXiv preprint arXiv:2103.14438. 2021 Mar 26.
15. Devlin J, Chang MW, Lee K, Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. InProceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) 2019 (pp. 4171-4186).
16. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D et al. Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv: 1907.11692. 2019.
17. Talaat AS. Sentiment analysis classification system using hybrid BERT models. Journal of Big Data. 2023;10(1):110.
18. Pandiangan LR, Wijayakusuma IL. Analisis sentimen para kandidat Pilpres 2024 dengan model bahasa BERT. E-Jurnal Matematika. 2024;13(4):253.
19. Khatavkar V, Petkar S, Vaidya AS. Multilingual Transformer Contextual Embedding Model for Political Tweets Analysis. Cureus Journals. 2025;2(1).
20. Babu TA, Ratul MM, Aftahee S, Hossain J, Hoque MM. CUET_NetworkSociety@ DravidianLangTech 2025: A Transformer-Driven Approach to Political Sentiment Analysis of Tamil X (Twitter) Comments. InProceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages 2025 (pp. 536-542). Association for Computational Linguistics.
21. Nguyen DK, Văn TĐ. PhucNguyen@ DravidianLangTech 2026: Political Multiclass Sentiment Analysis with XLM-RoBERTa and Low-Rank Adaptation. InProceedings of the Sixth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages 2026 (pp. 321-325).