ARTICLE

Improving Sentiment Analysis in Election-Based Conversations on Twitter with ElecBERT Language Model

Asif Khan[1], Huaping Zhang[1,*], Nada Boudjellal[2], Arshad Ahmad[3] and Maqbool Khan[3]

[1]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, 100081, China

[2]The Faculty of New Information and Communication Technologies, University Abdel-Hamid Mehri Constantine 2, Constantine, 25000, Algeria

[3]Department of IT and Computer Science, Pak-Austria Fachhochschule: Institute of Applied Sciences and Technology, Haripur, 22620, Pakistan

*Corresponding Author: Huaping Zhang. Email: kevinzhang@bit.edu.cn

ABSTRACT

Sentiment analysis plays a vital role in understanding public opinions and sentiments toward various topics. In recent years, the rise of social media platforms (SMPs) has provided a rich source of data for analyzing public opinions, particularly in the context of election-related conversations. Nevertheless, sentiment analysis of election-related tweets presents unique challenges due to the complex language used, including figurative expressions, sarcasm, and the spread of misinformation. To address these challenges, this paper proposes Election-focused Bidirectional Encoder Representations from Transformers (ElecBERT), a new model for sentiment analysis in the context of election-related tweets. Election-related tweets pose unique challenges for sentiment analysis due to their complex language, sarcasm, and misinformation. ElecBERT is based on the Bidirectional Encoder Representations from Transformers (BERT) language model and is fine-tuned on two datasets: Election-Related Sentiment-Annotated Tweets (ElecSent)-Multi-Languages, containing 5.31 million labeled tweets in multiple languages, and ElecSent-English, containing 4.75 million labeled tweets in English. The model outperforms other machine learning models such as Support Vector Machines (SVM), Naïve Bayes (NB), and eXtreme Gradient Boosting (XGBoost), with an accuracy of 0.9905 and F1-score of 0.9816 on ElecSent-Multi-Languages, and an accuracy of 0.9930 and F1-score of 0.9899 on ElecSent-English. The performance of different models was compared using the 2020 United States (US) Presidential Election as a case study. The ElecBERT-English and ElecBERT-Multi-Languages models outperformed BERTweet, with the ElecBERT-English model achieving a Mean Absolute Error (MAE) of 6.13. This paper presents a valuable contribution to sentiment analysis in the context of election-related tweets, with potential applications in political analysis, social media management, and policymaking.

KEYWORDS

Sentiment analysis; social media; election prediction; machine learning; transformers

1 Introduction

In recent years, social media has emerged as a powerful tool for public discourse, particularly in the context of politics and elections [1]. As a result, sentiment analysis has become a crucial tool for understanding public opinion and sentiment during elections [2]. However, sentiment analysis of election-related tweets poses unique challenges due to the complex nature of political language and the nuances of social dynamics involved [3].

One of the main challenges in sentiment analysis is the lack of dependable and extensive datasets suitable for training machine learning models.

Prior studies have utilized diverse datasets to address this challenge, including 3 million tweets related to the US presidential election [4], a dataset of 38,432,811 tweets from the US 2020 Presidential election [5], and a dataset consisting of 5,299 tweets from the 2022 Philippines national election [6]. Additionally, other datasets have been used for sentiment analysis, such as 1,302,388 tweets from the Ecuadorian presidential elections of 2021 [7] and 50K election-related tweets from the Indian General Election 2019 [8]. Moreover, a study explored the influence of tweet sentiment on opinions and retweet likelihood using datasets focused on various events, including a 2017 demonetization in India dataset (14,940 tweets), a 2016 US election dataset (397,629 tweets), and a 2018 American Music Awards dataset (27,556 tweets) [9].

Furthermore, studies have focused on specific elections, such as Pakistan's general election in 2018 [10], the 2020 US presidential election [11], the Nigeria 2023 presidential election [12], and recent presidential elections in Latin America, utilizing over 65,000 posts from social media platforms [13]. Additional datasets include 9,157 tweets from the 2017 Punjab assembly elections [14], a dataset of 5,000 messages from Twitter and Facebook annotated as neutral/partisan [15], a 100K #Politics dataset [16], and 29,462 tweets related to the West Bengal election in India [17]. Despite the existing datasets utilized for sentiment analysis in elections, there are limitations in terms of their size and comprehensiveness. These datasets may not encompass the wide spectrum of political scenarios or provide a sufficient representation of sentiment variations.

To address the scarcity of dependable datasets, the development of the US Presidential Election Tweets Dataset (UPETD) was undertaken. This dataset comprises 5.3 million election-related tweets and has been labeled with positive, negative, and neutral sentiments using the "Valence Aware Dictionary for sEntiment Reasoning (VADER)" technique. The resulting dataset, named the "ElecSent dataset," serves as a valuable resource for training machine learning models like BERT, enabling more precise and effective sentiment analysis of election-related tweets [18–20].

To further improve the accuracy of sentiment analysis in election-related tweets, this study proposes ElecBERT, a new sentiment analysis model specifically designed for political tweets. ElecBERT is fine-tuned on the ElecSent dataset and utilizes the BERT language model, taking into account the context and nuances of political language and social dynamics for more accurate sentiment analysis.

This study implemented ElecBERT to predict the sentiment of election-related tweets during the US 2020 Presidential Election as a case study. The effectiveness of ElecBERT in analyzing election-related tweets and predicting public sentiment was evaluated by comparing the results with the actual election outcome.

The implications of this study are far-reaching in terms of understanding public opinion in election-related situations. Accurate sentiment analysis can help political campaigns and policymakers to gauge public opinion, identify areas of concern, and design policies accordingly. This study

provides a valuable contribution to sentiment analysis in the context of election-related tweets and has implications for a wide range of applications in the political and social domains.

The main contributions of this study are as follows:

1. ElecSent dataset: 5.3 million tweets related to politics, labeled with a positive, negative, and neutral sentiments.
2. ElecBERT: a new sentiment analysis model specifically designed for political tweets, utilizing the BERT language model and taking into account the complexities of political language and social dynamics for improved accuracy.

The paper is structured as follows: In Section 2, related work is presented. Section 3 introduces the ElecSent dataset and the proposed ElecBERT model methodology. Section 4 describes the experiments and presents the results, followed by a discussion. Finally, the study is concluded in the conclusion section.

2 Related Work

Sentiment analysis on social media data has gained significant attention in recent years due to the increasing importance of understanding public opinion in various domains. Several studies have explored different approaches to sentiment analysis, including rule-based methods such as VADER, and machine learning techniques such as logistic regression, SVMs, and NB. However, the field of sentiment analysis has also seen significant advancements in deep learning-based methods for text analytics. Deep learning techniques, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [21], have demonstrated remarkable performance in this domain. CNNs excel at capturing local features and patterns, while RNNs are effective in modeling sequential dependencies in text data [22].

Furthermore, a recent study by [23] introduced a novel approach that utilizes capsule networks for sentiment analysis, with a specific focus on social media content from platforms like Twitter. The study findings demonstrate the effectiveness of capsule networks in sentiment analysis tasks, particularly when analyzing Twitter data. Another study [24] has introduced a novel neural network model for sentiment analysis, combining multi-head self-attention and character-level embedding. This model effectively tackles the challenges of sentiment word extraction and the out-of-vocabulary problem commonly encountered in existing methods. By employing an encoder-decoder architecture with Bidirectional Long Short-Term Memory (BiLSTM), the model captures contextual semantic information and extracts deeper emotional features, enhancing its ability to analyze sentiment in text.

Domain-specific sentiment analysis, such as election prediction, has also been explored. For instance, in a study [25], the authors employed tools such as Natural Language Toolkit (NLTK), Tweet Natural Language Processing (TweetNLP) toolkit, Scikit-learn, and Statistical Package for the Social Sciences (SPSS) statistical package to predict ideological orientation (conservative or right-leaning, progressive or left-leaning) of 24,900 tweets collected over 9 h during an election, achieving an overall accuracy of 99.8% using Random Forest. Similarly, in [26], Scikit-learn was utilized to analyze 46,705 Greek tweets over 20 days during an election, achieving a Random Forest accuracy of 0.80 and precision values of Negative = 0.74, Neutral = 0.83, and Positive = 1.

In another study [27], Textblob was used to analyze 277,509 tweets from three states (Florida, Ohio, and North Carolina) over a month for sentiment analysis during the election, achieving NB accuracy of over 75%. Furthermore, in [28], the authors employed SVM, NB, and K-Nearest Neighbors (KNN) on 2018 Pakistani election data to predict the support for each political leader.

They found that SVM outperformed other models with an accuracy of 79.89%. Similarly, in [29], SVM with a hybrid (unigram + bigram) was used on 100K tweets during the U.S. Election 2012 and Karnataka (India) Elections 2013, achieving accuracies of 88% and 68%, respectively, using NLTK and Stanford part-of-speech (POS) tagger.

Additionally, study [30] utilized Waikato Environment for Knowledge Analysis (WEKA) to analyze 3,52,730 tweets over a month for sentiment analysis on political parties in India, while study [14] employed the Syuzhet package in R-language, WEKA, and Gephi were used to analyze 9,157 tweets over approximately a month regarding political parties, achieving an SVM accuracy of 78.63%. Furthermore, study [31] utilized KNN to analyze election-related data, achieving an average accuracy of 92.19%.

Moreover, several studies have explored the use of Deep Learning and large language models for sentiment analysis in various domains. For example in [32], the authors analyzed US 2020 Presidential election using BERT and VADER, finding that VADER outperformed BERT. Additionally, in [33], Scikit-learn, NLTK, and VADER were used to analyze 121,594 tweets over two days about a candidate with an SVM accuracy of 0.99. Furthermore, study [34] employed Textblob, OpLexicon (Portuguese sentiment lexicon), and Sentilex (Portuguese sentiment lexicon) to analyze 158,279 tweets over 16 days about a candidate with SVM accuracy of 0.93 and 0.98 for OpLexicon/Sentilex. Moreover, study [35] used Long short-term memory (LSTM) to analyze 3,896 tweets over approximately three months, examining election trends, party, and candidate sentiment analysis, yielding precision = 0.76, recall = 0.75, and F1-score = 0.74.

However, these models often struggle to capture the in-depth nature of political language and social dynamics involved in election-related tweets. Recently, deep learning models such as Transformer-based models have shown remarkable performance in various natural language processing (NLP) tasks, including sentiment analysis. One of the most popular deep learning models for NLP tasks is BERT, which has achieved state-of-the-art performance on several benchmark datasets. However, fine-tuning BERT for specific domains, such as election-related tweets, can improve its performance and make it more effective for sentiment analysis.

Several studies have utilized large language models for different domain-specific tasks. For instance, BERTweet is a pre-trained language model for English Tweets, trained on 850 million Tweets using the Robustly Optimized BERT Pretraining Approach (RoBERTa) pre-training procedure [36]. BERTweet-COVID19 models were pre-trained on a corpus of 23 million COVID-19 English Tweets. It outperforms strong baselines on three Tweet NLP tasks and also achieved good results on several benchmarks, such as SemEval2017 where it achieved 0.732 AvgRec and F_1^{NP} 0.728, and accuracy 0.72, while on SemEval2018 it achieved F_1^{Pos} 0.746 and accuracy 0.782. On WNUT16 it achieved 0.521 F1-score, and on WNUT17 it achieved 0.551 F1-score.

Moreover, PoliBERTweet, a pre-trained language model trained on over 83M US 2020 election-related English tweets [37]. The model is specifically designed to address the nuances of political language and can be used for a variety of downstream tasks such as political misinformation analysis and election public opinion analysis. The authors used a stance detection dataset to check the performance of PoliBERTweet [1]. The F1-score for BIDEN using RoBERTa (RB) is 0.6, RB/P-M is 0.663 TweetEval (TE) is 0.624, TE/P-M is 0.653, BERTweet (BT) is 0.650, BT/P-M is 0.673, poliBERTweet (PoliBERT) 0.708, sentiment knowledge-enhanced pre-training (SKEP) is 0.746, and knowledge-enhanced masked language modeling (KEMLM) is 0.758. While the F1-score for TRUMP using RoBERTa (RB) is 0.771, RB/P-M is 0.779 TweetEval (TE) 0.809, TE/P-M is 0.811, BERTweet

(BT) is 0.828, BT/P-M is 0.831, poliBERTweet (PoliBERT) 0.848, sentiment knowledge-enhanced pre-training (SKEP) is 0.772, and knowledge-enhanced masked language modeling (KEMLM) is 0.788. P-M indicates Poli-Medium.

Numerous other studies utilized large language models for different tasks such as BioBERT, a pre-trained BERT model that has been specifically trained on biomedical text. It can be fine-tuned for various biomedical NLP tasks [38]. FinBERT is a pre-trained BERT model that has been specifically designed for financial text. It can be fine-tuned for various financial NLP tasks [39]. AraBERT is another pre-trained BERT model that has been specifically trained on Arabic text. It can be fine-tuned for various Arabic NLP tasks [19]. ABioNER: A BERT-based model for identifying disease and treatment-named entities in Arabic biomedical text [40]. MEDBERT.de, a pre-trained German BERT model designed specifically for the medical domain. Trained on a large corpus of 4.7 million German medical documents, it achieves state-of-the-art performance on eight different medical benchmarks [41]. MolRoPE-BERT, an end-to-end deep learning framework for molecular property prediction that efficiently encodes the position information of SMILES sequences using Rotary Position Embedding (RoPE). The framework combines a pre-trained BERT model with RoPE for extracting potential molecular substructure information. The model is trained on four million unlabeled drug SMILES and is evaluated on four datasets, demonstrating comparable or superior performance to conventional and state-of-the-art baselines [42].

However, to the best of our knowledge, no study has explored the use of BERT specifically for sentiment analysis on election-related tweets, accounting for the complexities of political language and social dynamics. Therefore, this study proposes ElecBERT, a fine-tuned BERT model tailored for sentiment analysis on election-related tweets.

3 Materials and Methods

This section presents the methodology employed in this study, including the ElecSent dataset and the ElecBERT sentiment analysis model.

3.1 Dataset

Twitter using its Application Programming Interface (APIs) allow for data collection. They provide developers with access to the platform's vast collection of public data, including tweets, user profiles, and search results. Numerous tools can be employed to collect tweets, for instance, Twitter-Tap, Tweepy, TWurl, twarc, streamR, TweetMapper, Twitonomy, NodeXL, and Twython. This study utilized the Twitter Search API, which enables the retrieval of tweets based on specific criteria, such as keywords, hashtags, and dates. To collect tweets, the Tweepy Python library was applied, which is a widely used tool for interacting with the Twitter Search API. A substantial number of tweets were collected in JSON (JavaScript Object Notation) format, a lightweight data-interchange format that is both human-readable and machine-parseable. Each tweet contained various attributes, including user_id (the unique identifier of the user who posted the tweet), lang (the language of the tweet), id (the unique identifier of the tweet), created_at (the date and time that the tweet was posted), text (the text of the tweet), coordinates (the geographic coordinates of the tweet, if available), and others.

3.1.1 UPETD Dataset

The UPETD dataset comprised nearly 5.31 million tweets related to Joe Biden, Donald Trump, Democratic, Republican, and USElection2020 (during Timeline 1 (T1): 5th Dec 2019 and 30th Nov

2020, and Timeline 2 (T2): = 1ˢᵗ Aug 2020 to 30ᵗʰ Nov 2020). Table 1 shows the statistics of the UPETD dataset.

Table 1: Statistics for UPETD

Stats	Joe biden	Donald trump	Democratic	Republicans	Election	Total
Multi-languages tweets	1,594,370	835,119	248,988	1,724,198	908,508	5,311,183
English tweets	1,400,843	684,093	245,747	1,637,150	785,469	4,753,302
Collection timeline	T2	T2	T1	T1	T2	–

3.1.2 ElecSent Dataset

The ElecSent dataset is based on the UPETD dataset. Sentiments were assigned to each tweet in the UPETD dataset. Several studies used VADER to classify the tweets and later use other machine learning approaches for sentiment analysis [43–45]. VADER is a lexicon and rule-based sentiment analysis tool specifically designed for social media text. It uses a sentiment lexicon that contains a list of words with their associated sentiment scores (positive, negative, or neutral), along with a set of grammatical rules and heuristics to analyze the sentiment of a given text. This study also employed VADER to assign the sentiments to each tweet in the UPETD dataset. This dataset was subsequently named the ElecSent dataset. The ElecSent dataset is presented in two forms, (i) ElecSent-Multi-Language dataset, and (ii) ElecSent-English. ElecSent-Multi-Language contains tweets in multiple languages, including English, and has a total of 5.31 million tweets. ElecSent-English includes only English-written tweets and has a total of 4.753 million tweets, which is almost 89.5% of the total dataset.

To analyze the distribution of sentiments in the dataset, the percentage of positive, negative, and neutral tweets was calculated. Fig. 1 shows that both ElecSent-Multi-Language and ElecSent-English have a similar distribution, with positive tweets being the highest, followed by negative and neutral tweets. ElecSent-Multi-Language has 41% positive, 32% negative, and 27% neutral tweets, while ElecSent-English has 43% positive, 31% negative, and 26% neutral tweets. ElecSent can be further used in machine learning election prediction models based on the public's sentiment towards various candidates and political parties.

The ElecSent dataset has an imbalanced distribution of labels in both its ElecSent-Multi-Languages and ElecSent-English versions, with the majority of samples being Positive, followed by Negative and Neutral. This imbalance can lead to overfitting issues in models trained on this dataset. To address this problem, this study applied the Synthetic Minority Over-sampling Technique (SMOTE), which creates synthetic samples for the minority classes. with an equal distribution of samples among all three classes [46–48]. Specifically, Fig. 2 shows the balanced dataset that contains 34% Positive, 33% Negative, and 33% Neutral samples.
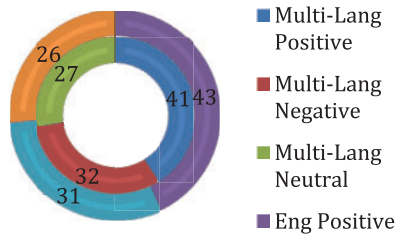
Figure 1: Sentiment distribution of the ElecSent dataset (Multi-Languages & English-only)
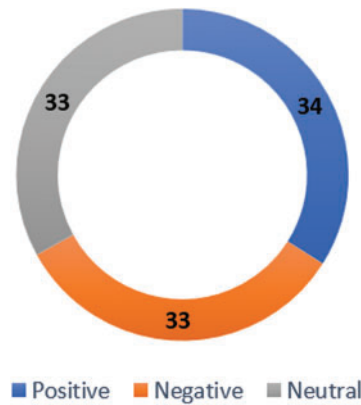

Figure 2: The balanced ElecSent dataset

3.2 Building ElecBERT: Architecture and Fine-Tuning Approach

BERT is a pre-trained language model that is trained on a large corpus of text data, and it is very effective for various natural language processing tasks, including sentiment analysis. The BERT model (bert-base-uncased) was fine-tuned on the ElecSent dataset, represented as D, with labels $L=$ {*positive, negative, neutral*}, and obtained a new model called "ElecBERT". Fig. 3 shows the process of fine-tuning the ElecBERT model. Fine-tuning was performed by minimizing the loss function $L(D, \theta)$, where θ is the parameters of BERT, and the output of the fine-tuned model for an input x is denoted as $f(x; \theta*)$. The statement can be defined as:

Let D denote the "ElecSent" dataset, which consists of a set of tweets labeled as either positive, negative, or neutral. Let X denote the set of input features and Y denotes the corresponding set of labels in D. The dataset is split into two parts: a training set and a validation set, with a ratio of 80:20.

Then, the dataset D can be represented as follows:

$$D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

where each x_i is a feature vector and each y_i is a label in L.

Before feeding the text data to BERT, the tweets in the "ElecSent" dataset have undergone pre-processed. The BERT tokenizer is utilized to tokenize the pre-processed text data. The tokenizer adds special tokens like [CLS] and [SEP] to the start and end of each sentence, truncates/pads the sentences to a maximum length of 64 tokens, maps the tokens to their IDs, and creates an attention mask. To prepare the data for training, the training set and validation set are concatenated, and the tokenizer is employed to encode the concatenated data.
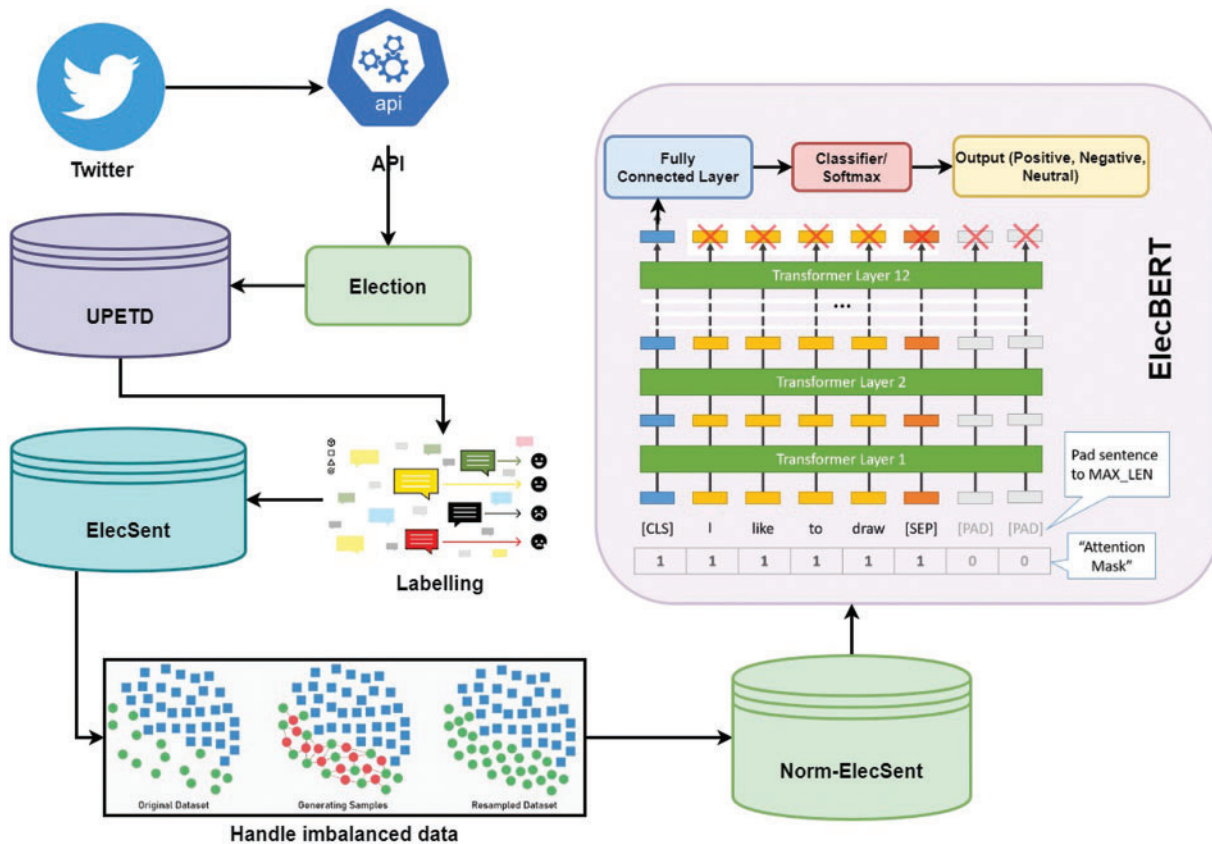
Figure 3: Overview of the ElecBERT model

The BERT model (bert-base-uncased) is then fine-tuned on the ElecSent dataset using cross-entropy loss and the *AdamW* optimizer. The loss function $L(D, \theta)$ is defined as:

$$L(D, \theta) = \Sigma_i = 1 \, to \, n \, l(f(x_i; \theta), y_i)$$

where θ denotes the parameters of BERT, $f(x; \theta)$ is the output of the BERT model for input x with parameters θ, and l is the cross-entropy loss function that measures the discrepancy between the predicted output and the ground-truth label.

The *TensorDataset* and *DataLoader* classes are utilized from the *PyTorch* library to create data loaders for the training set and the validation set. This study uses the *RandomSampler* class to sample the data randomly during training and the *SequentialSampler* class to sample the data sequentially during validation. During training, the training loss, the validation loss, and the F1-score for each label (positive, negative, and neutral) are monitored using the validation set.

The model undergoes training for 6 epochs with a batch size of 32 and a learning rate of *2e-5*. After each epoch, the model is evaluated on the validation set, and metrics such as the validation loss, accuracy, precision, recall, and the F1-score are computed. The resulting model is named *"ElecBERT"*, which is denoted as:

$$ElecBERT = BERT(D, \theta*)$$

where $\theta*$ denotes the optimal parameters obtained after fine-tuning.

4 Experiments and Results

The experiments for SVM, NB, and XGBoost were conducted in Google Colab (python 3.7). Furthermore, for ElecBERT, NVIDIA TITAN XP 12 GB with 128 GB RAM was used.

The proposed ElecBERT model was evaluated through extensive experimentation on two datasets: ElecSent-Multi-Language and ElecSent-English. The former comprises 5.3 million election-related tweets in various languages, while the latter has 4.753 million English-written tweets. The datasets were split into training-validation sets in an 80:20 ratio. The bert-base-uncased pre-trained model was utilized as the initial BERT model. The ElecBERT model was fine-tuned for six epochs with a batch size of 32 and a learning rate of 2e-5. The experiments used the AdamW optimizer with default parameters and the cross-entropy loss function.

The performance metrics for the proposed ElecBERT model are impressive. Specifically, the ElecBERT-Multi-Languages model achieved an accuracy of 0.9905, precision of 0.9813, recall of 0.9819, and an F1-score of 0.9816 during its 5th and 6th epochs. The validation loss at the 6th epoch was 0.140. Comparatively, ElecBERT-English performed better with an accuracy of 0.9930, precision of 0.9906, recall of 0.9893, and an F1-score of 0.9899 during the 6th epoch. Figs. 4a and 4b represents the training metrics of both models. Furthermore, Fig. 5 shows the validation loss for both models. ElecBERT-English may have outperformed ElecBERT-Multi-Languages due to the use of VADER to label the ElecSent dataset. VADER is known to perform better in English than in other languages, and this could have led to a higher quality of labeled data for the ElecBERT-English model to train on, resulting in its better performance on the English dataset. These metrics indicate that ElecBERT has achieved excellent performance on the ElecSent dataset.
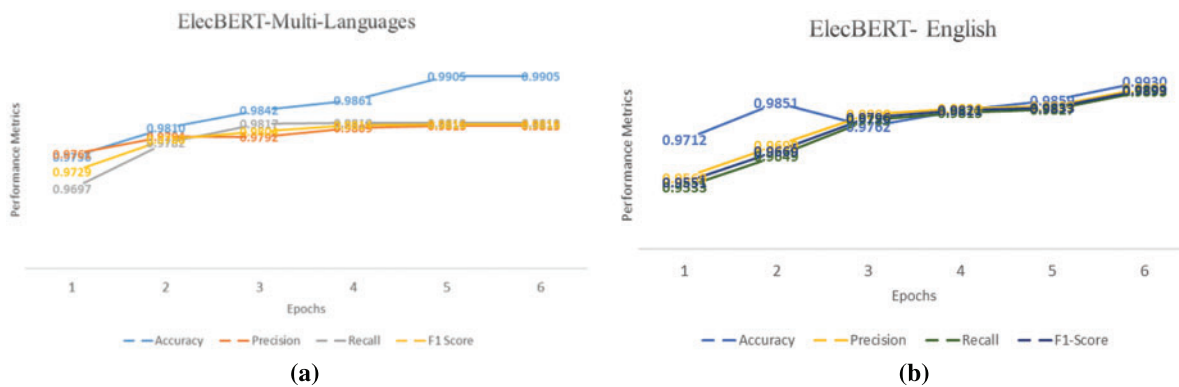

Figure 4: (a) ElecBERT-Multi-Languages | Evaluation Metrics. (b) ElecBERT-English | Evaluation Metrics
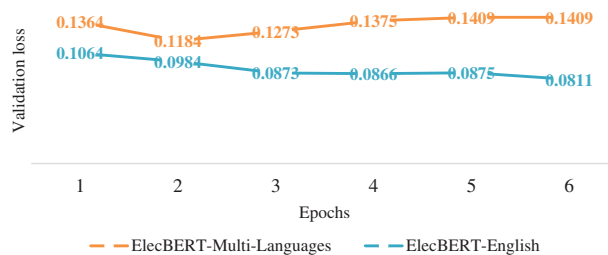

Figure 5: Validation loss for ElecBERT-Multi-Languages and ElecBERT-English

In addition, the experiments on XGBoost, SVM, and NB were conducted using the ElecSent dataset. Interestingly SVM and Naive Bayes (NB) achieved lower F1-scores 0.802 and 0.826, and XGBoost achieved 0.8964 F1-score. Fig. 6 shows the evaluation matric, F1-score for the ElecBERT model, as well as XGBoost, SVM, and NB. This suggests that ElecBERT was able to capture the nuances of sentiment in political tweets better than the traditional machine learning models, leading to superior performance on this task.
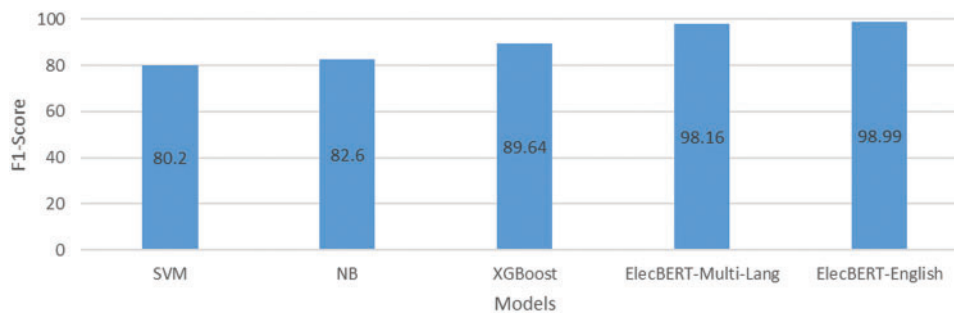

Figure 6: F1-score | ElecBERT-Multi-Lang, ElecBERT-English, SVM, XGBoost, and NB

4.1 Leveraging ElecBERT on the 2020 US Presidential Election

This section presents a case study using ElecBERT to analyze sentiment in election-related tweets during the US 2020 Presidential Election and predict election outcomes. The study aims to explore the effectiveness of ElecBERT in predicting public sentiment and election outcomes. The results were compared with the actual election outcomes to evaluate the performance of the model.

4.1.1 Data

The data in this study was gathered from December 2019 to November 2020 using hashtags related to the Democratic and Republican Parties (#Democratic, #TheDemocrats, #GOP, and #Republican). The dataset consists of 1,637,150 tweets from Republican Party and 245,757 tweets from Democratic Party.

4.1.2 Analyzing the Elections

Fig. 7 displays sentiment analysis results for tweets about the Democratic and Republican Parties using three different language models: BERTweet, ElecBERT-Multi-Lang, and ElecBERT-English. The majority of tweets in all categories are classified as neutral sentiments towards both Democratic and Republican politicians. This is likely because political tweets often contain objective statements of fact or news updates, which may not express a clear sentiment towards a particular politician. However, the results also indicate that both ElecBERT-Multi-Lang and ElecBERT-English are more effective than BERTweet in identifying positive sentiment towards both Democratic and Republican politicians. For example, ElecBERT-Multi-Lang had the highest percentage of positive sentiment classification for Republican politicians at 42.36%, which is significantly higher than BERTweet's 7.94% positive sentiment classification. Similarly, ElecBERT-English had the highest percentage of positive sentiment classification for Democratic politicians at 39.62%, which is also significantly higher than BERTweet's 13.43% positive sentiment classification. On the other hand, BERTweet had the highest percentage of negative sentiment classification for both Democratic and Republican politicians. For instance,

BERTweet classified 44.67% of Republican tweets as negative sentiment, which is considerably higher than ElecBERT-Multi-Lang's 33.36% negative sentiment classification.
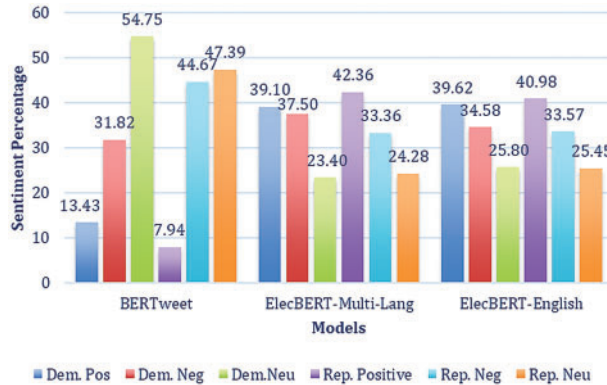

Figure 7: Sentiment analysis for democratic and republican parties using BERTweet, ElecBERT-Multi-Lang, and ElecBERT-English

In addition, two equations (Eqs. (1) and (2)) were utilized in this study to forecast the vote share for each political party. The approach involved assuming that positive sentiments expressed towards the Democratic Party and negative sentiments expressed towards the Republican Party represent support for the Democratic Party, and conversely for the Republican Party.

$$Vote\ share\ for\ Democratic = \frac{Dem.Pos. + Rep.Neg.}{Dem.Pos + Dem.Neg. + Rep.Pos. + Rep.Neg.} \tag{1}$$

$$Vote\ share\ for\ Republican = \frac{Rep.Pos. + DemNeg.}{Dem.Pos + Dem.Neg. + Rep.Pos. + Rep.Neg.} \tag{2}$$

Table 2 presents the vote share percentages for BERTweet, ElecBERT-Multi-Lang, and ElecBERT-English, as well as the actual US Election 2020 results. Additionally, the table includes the normalized Democratic and Republican results. The reason for providing the normalized results is that the study only focused on the Republican and Democratic parties and excluded tweets about other political parties. Therefore, the sum of the vote share percentages for the two parties in the actual US Election 2020 results does not add up to 100. To address this, the actual results were normalized to add up to 100, which facilitates comparison with the results obtained using the three language models.

Table 2: Vote shares for BERTweet, ElecBERT-Multi-Lang, ElecBERT-English, and actual US Election 2020 results along with normalized results

Election (Prediction/Actual)	Democratic	Republicans	Democratic normalized	Republican normalized
BERTweet	78.25	21.75	78.25	21.75
ElecBERT-Multi-Languages	45.98	54.02	45.98	54.02
ElecBERT-English	46.12	53.88	46.12	53.88
US presidential election 2020	51.4	46.9	52.28	47.72

The results show that BERTweet predicted a significantly higher vote share for the Democratic Party (78.25%) than for the Republican Party (21.75%). In contrast, both ElecBERT-Multi-Lang and ElecBERT-English predicted a higher vote share for the Republican Party (54.02% and 53.88%, respectively) than for the Democratic Party (45.98% and 46.12%, respectively). These results indicate that the two ElecBERT models were more accurate in predicting the actual vote share distribution between the two parties. The actual US Presidential Election 2020 results indicate that the Democratic Party received 51.4% of the vote share, while the Republican Party received 46.9%. However, when normalized to 100, the Democratic and Republican results are 52.28% and 47.72%, respectively. This normalization facilitates comparison with the results obtained using the three language models.

The Mean absolute error (MAE) and root mean square error (RMSE) was used to compare the predicted vote shares with the actual results of the US Presidential Election 2020. MAE measures the average absolute difference between the predicted and actual values, while RMSE measures the square root of the average squared difference between the predicted and actual values. The lower the MAE and RMSE values, the closer the predicted results are to the actual election results. The MAE and RMSE values for BERTweet, ElecBERT-Multi-Lang, and ElecBERT-English were calculated using Eqs. (3) and (4), respectively.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |Predicted_i - Actual_i| \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}} \tag{4}$$

In Table 3, both ElecBERT-Multi-Languages and ElecBERT-English outperform BERTweet in terms of MAE and RMSE. This suggests that the two models are better at predicting the vote share for the two major parties in the US election. In particular, the MAE for Democratic and Republican parties for both ElecBERT-Multi-Languages and ElecBERT-English is significantly lower than BERTweet. For instance, the MAE for Democratic party prediction using ElecBERT-Multi-Languages and ElecBERT-English are 5.42 and 5.28, respectively, while the MAE for BERTweet is 26.85. Similarly, the MAEs for Republican party prediction using ElecBERT-Multi-Languages and ElecBERT-English are 7.12 and 6.98, respectively, while the MAE for BERTweet is 25.15.

Table 3: MAE, and RMSE for BERTweet, ElecBERT-Multi-Languages, and ElecBERT-English

Model	MAE Dem.	MAE rep.	MAE	RMSE	MAE normalized
BERTweet	26.85	25.15	26	26.014	25.98
ElecBERT-Multi-Languages	5.42	7.12	6.27	6.327	6.34
ElecBERT-English	5.28	6.98	6.13	6.189	6.16

Moreover, the RMSE values for both ElecBERT-Multi-Languages and ElecBERT-English are also significantly lower than BERTweet for both Democratic and Republican parties, indicating that the predicted vote shares are closer to the actual vote shares. Finally, the MAE Normalized value in the table shows the average difference between the predicted and actual normalized vote share, which is equal to 100 for both parties. Here too, ElecBERT-English models perform better than ElecBERT-Multi-Languages BERTweet, with lower MAE values.

On the whole, the results suggest that ElecBERT models, both ElecBERT-Multi-Lang and ElecBERT-English, can perform well in analyzing election-related tweets and predicting election outcomes. These models outperformed BERTweet in terms of sentiment analysis and vote share prediction. Additionally, the MAE and RMSE values indicate that the ElecBERT models have a lower prediction error than BERTweet, especially when it comes to the Democratic party's vote share prediction. Therefore, it is reasonable to assume that ElecBERT models can help analyze and predict future elections by analyzing large volumes of social media data.

4.2 Practical Usage of the ElecBERT

The proposed ElecBERT model has several practical applications in the field of natural language processing (NLP) and sentiment analysis. Here are some potential practical usages of ElecBERT:

1. Sentiment Analysis: ElecBERT can be utilized for sentiment analysis tasks related to election-related tweets. By leveraging its fine-tuned knowledge of a large corpus of election tweets, ElecBERT can effectively classify the sentiment of new, unseen election-related tweets as positive, negative, or neutral. This can provide valuable insights into public opinion, sentiment trends, and the overall sentiment surrounding political candidates and election events.

2. Election Monitoring: With its ability to analyze sentiment, ElecBERT can be used for real-time monitoring of elections. By processing a stream of tweets in real time, ElecBERT can help gauge the sentiment of the public towards candidates, parties, or specific election issues. This can be valuable for political campaigns, media outlets, and researchers seeking to understand public sentiment and adjust their strategies accordingly.

3. Social Media Analytics: ElecBERT can contribute to social media analytics by providing a deep understanding of election-related conversations happening on platforms like Twitter. By applying ElecBERT to large volumes of election tweets, analysts can identify emerging topics, detect patterns, and gain insights into voter behavior, public opinion, and the sentiment dynamics throughout an election campaign.

4. Opinion Mining: ElecBERT can assist in extracting and analyzing opinions expressed in election tweets. By leveraging its fine-tuned language understanding capabilities, ElecBERT can help identify and categorize different aspects of political discourse, such as policy issues, candidate attributes, or sentiment towards specific campaign promises. This can support opinion-mining tasks and provide a nuanced understanding of voter opinions.

5. Election Prediction: With its fine-tuned knowledge of election-related tweets, ElecBERT can potentially contribute to election outcome prediction models. By analyzing sentiment patterns, trends, and public opinion expressed in tweets, ElecBERT can provide additional insights to complement traditional polling methods, enabling more accurate predictions of election results.

6. Social Listening and Crisis Management: During elections, social media can be a breeding ground for misinformation, rumors, and crises. ElecBERT can be used as a tool for social listening and crisis management by monitoring election-related conversations on platforms like Twitter. It can help identify potentially problematic content, detect the spread of misinformation, and provide real-time sentiment analysis to assist in managing and addressing crises effectively.

5 Conclusion and Future Work

This paper presented ElecBERT, a new model for sentiment analysis in the context of election-related tweets. The model was fine-tuned on two datasets: ElecSent-Multi-Languages, containing 5.31 million labeled tweets in multiple languages, and ElecSent-English, containing 4.75 million labeled tweets in English. The ElecSent dataset is labeled (positive, negative, and neutral) using VADER. Notably, ElecBERT showcased superior performance when compared to SVM, NB, and XGBoost, achieving an accuracy of 0.9905 and an F1-score of 0.9816 on ElecSent-Multi-Languages, as well as an accuracy of 0.9930, and an F1-score of 0.9899 on ElecSent-English. Furthermore, this study conducted a comprehensive analysis of the 2020 US Presidential Election as a case study, comparing the performance of different models. Among them, both the ElecBERT-English and ElecBERT-Multi-Languages models outperformed BERTweet, with the ElecBERT-English model achieving an MAE of 6.13. This paper presents a valuable contribution to sentiment analysis in the context of election-related tweets, with potential applications in political analysis, social media management, and policymaking.

The ElecBERT model was trained on the 2020 US Presidential Election data only, and its performance on other elections or political events may vary. Additionally, the sentiment labels for the ElecSent dataset were generated using VADER, an automated tool, without manual human reviewing and verification. In the future, more data from other elections should be added to make the model more robust and generalizable. Moreover, using other pre-trained models like PoliBERT can be explored to further improve the accuracy of sentiment analysis on election-related tweets. Finally, expanding the model to incorporate more complex features of political languages, such as sarcasm and irony, could lead to a more nuanced understanding of election-related sentiment on social media.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization: A.K., N.B.; methodology: A.K., N.B., and H.Z; software: A.K., and N.B.; validation: A.K., and N.B.; formal analysis: H.Z., N.B., A.A., and M.K.; investigation: A.K., N.B., A.A., and M.K.; resources: A.K., H.Z., and N.B.; data curation: A.K., and N.B.; writing—original draft preparation: A.K.; writing—review and editing: A.K., H.Z, N.B., A.A, and M.K.; visualization: A.K., and N.B.; supervision: H.Z.; project administration: H.Z.; funding acquisition: H.Z. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data used in this study is available at "https://doi.org/10.57967/hf/0813".

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   R. Rizk, D. Rizk, F. Rizk and S. Hsu, "280 characters to the White House: Predicting 2020 U.S. presidential elections from Twitter data," *Computational and Mathematical Organization Theory*, 2023. https://doi.org/10.1007/s10588-023-09376-5

[2]   A. Oussous, Z. Boulouard and B. F. Zahra, "Prediction and analysis of Moroccan elections using sentiment analysis," *AI and IoT for Sustainable Development in Emerging Countries: Challenges and Opportunities*, vol. 105, pp. 597–609, 2022.

[3]   P. Chauhan, N. Sharma and G. Sikka, "The emergence of social media data and sentiment analysis in election prediction," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 2, pp. 2601–2627, 2021.

[4]   L. Hagemann and O. Abramova, "Crafting audience engagement in social media conversations: Evidence from the U.S. 2020 presidential elections," in *Proc. of the 55th Hawaii Int. Conf. on System Sciences*, Hawaii, United States, pp. 1–10, 2022.

[5]   H. N. Chaudhry, Y. Javed, F. Kulsoom, Z. Mehmood, Z. I. Khan *et al.,* "Sentiment analysis of before and after elections: Twitter data of U.S. election 2020," *Electronics*, vol. 10, no. 17, pp. 2082, 2021.

[6]   R. E. Demillo, G. Solano and N. Oco, "Philippine national elections 2022: Voter preferences and topics of discussion on Twitter," in *5th Int. Conf. on Artificial Intelligence in Information and Communication, ICAIIC*, Bali, Indonesia, pp. 724–729, 2023.

[7]   S. Lopez-Fierro, C. Chiriboga-Calderon and R. Pacheco-Villamar, "If it looks, retweets and follows like a troll; Is it a troll?: Targeting the 2021 ecuadorian presidential elections trolls," in *Proc. 2021 IEEE Int. Conf. on Big Data*, Florida, USA, pp. 2503–2509, 2021.

[8]   S. Chatterjee and S. Gupta, "Incremental real-time learning framework for sentiment classification: Indian general election 2019, a case study," in *2021 IEEE 6th Int. Conf. on Big Data Analytics, ICBDA 2021*, Xiamen, China, pp. 198–203, 2021.

[9]   A. Ahmad, E. Furey and J. Blue, "An investigation into the impact and predictability of emotional polarity on the virality of Twitter tweets," in *32nd Irish Signals and Systems Conf., ISSC 202*, Athlone, Ireland, pp. 1–5, 2021.

[10]   H. Ali, H. Farman, H. Yar, Z. Khan, S. Habib *et al.,* "Deep learning-based election results prediction using Twitter activity," *Soft Computing*, vol. 26, no. 16, pp. 7535–7543, 2022.

[11]   A. Karami, S. B. Clark, A. Mackenzie, D. Lee, M. Zhu *et al.,* "2020 U.S. presidential election in swing states: Gender differences in Twitter conversations," *International Journal of Information Management Data Insights*, vol. 2, no. 2, pp. 100097, 2020.

[12]   O. Olabanjo, A. Wusu, O. Afisi, M. Asokere, R. Padonu *et al.,* "From twitter to Aso-Rock: A sentiment analysis framework for understanding Nigeria 2023 presidential election," *Heliyon*, vol. 9, no. 5, pp. e16085, 2023.

[13]   K. Brito and P. J. L. Adeodato, "Machine learning for predicting elections in Latin America based on social media engagement and polls," *Government Information Quarterly*, vol. 40, no. 1, pp. 101782, 2023.

[14]   P. Singh, Y. K. Dwivedi, K. S. Kahlon, A. Pathania and R. S. Sawhney, "Can Twitter analytics predict election outcome? An insight from 2017 Punjab assembly elections," *Government Information Quarterly*, vol. 37, no. 2, pp. 101444, 2020.

[15]   Kaggle, "Political social media posts," 2023. https://www.kaggle.com/datasets/crowdflower/political-social-media-posts

[16]   Kaggle, "Global political tweets," 2023. https://www.kaggle.com/datasets/kaushiksuresh147/political-tweets

[17]   A. Chakraborty and N. Mukherjee, "Analysis and mining of an election-based network using large-scale Twitter data: A retrospective study," *Social Network Analysis and Mining*, vol. 13, no. 1, pp. 74, 2023.

[18]   L. Zhao, L. Li, X. Zheng and J. Zhang, "A BERT based sentiment analysis and key entity detection approach for online financial texts," in *Proc. of the 2021 IEEE 24th Int. Conf. on Computer Supported Cooperative Work in Design, CSCWD*, Dalian, China, pp. 1233–1238, 2021.

[19] W. Antoun, F. Baly and H. Hajj, "AraBERT: Transformer-based model for Arabic language understanding," in *Proc. of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, Marseille, France, pp. 9–15, 2020.

[20] D. A. Kristiyanti, A. H. Umam, M. Wahyudi, R. Amin and L. Marlinda, "Comparison of SVM naïve bayes algorithm for sentiment analysis toward West Java Governor candidate period 2018-2023 based on public opinion on Twitter," in *2018 6th Int. Conf. on Cyber and IT Service Management, CITSM 2018*, Parapat, Indonesia, pp. 1–6, 2019.

[21] T. V. Cherian, G. J. L. Paulraj, I. J. Jebadurai and J. Jebadurai, "Experimental comparative analysis on convolutional neural network (CNN) and recurrent neural network (RNN) on aspect-level sentiment analysis," in *4th EAI Int. Conf. on Big Data Innovation for Sustainable Cognitive Computing*, Coimbatore, India, pp. 17–27, 2023.

[22] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria and U. R. Acharya, "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis," *Future Generation Computer Systems*, vol. 115, no. 3, pp. 279–294, 2021.

[23] P. Demotte, K. Wijegunarathna, D. Meedeniya and I. Perera, "Enhanced sentiment extraction architecture for social media content analysis using capsule networks," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 8665–8690, 2023.

[24] H. Xia, C. Ding and Y. Liu, "Sentiment analysis model based on self-attention and character-level embedding," *IEEE Access*, vol. 8, pp. 184614–184620, 2020.

[25] R. C. Prati and E. Said-Hung, "Predicting the ideological orientation during the Spanish 24M elections in Twitter using machine learning," *AI and Society*, vol. 34, no. 3, pp. 589–598, 2019.

[26] D. Beleveslis, C. Tjortjis, D. Psaradelis and D. Nikoglou, "A hybrid method for sentiment analysis of election related tweets," in *4th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conf., SEEDA-CECNSM 2019*, Piraeus, Greece, pp. 1–6, 2019.

[27] L. Oikonomou and C. Tjortjis, "A method for predicting the winner of the USA presidential elections using data extracted from Twitter," in *South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conf., SEEDA_CECNSM*, Kastoria, Greece, pp. 1–8, 2018.

[28] A. Khan, H. Zhang, J. Shang, N. Boudjellal, A. Ahmad *et al.,* "Predicting politician's supporters' network on Twitter using social network analysis and semantic analysis," *Scientific Programming*, vol. 2020, no. 1, pp. 9353120, 2020.

[29] M. Anjaria and R. M. R. Guddeti, "A novel sentiment analysis of social networks using supervised learning," *Social Network Analysis and Mining*, vol. 4, no. 1, pp. 1–15, 2014.

[30] R. Srivastava, M. P. S. Bhatia, H. Kumar and S. Jain, "Analyzing Delhi assembly election 2015 using textual content of social network," in *ACM Int. Conf. Proc. Series*, New York, NY, USA, pp. 78–85, 2015.

[31] S. Sharma and N. P. Shetty, "Determining the popularity of political parties using Twitter sentiment analysis," in *Proc. of the 6th Int. Conf. on FICTA*, Odisha, India, pp. 21–29, 2018.

[32] A. Khan, H. Zhang, N. Boudjellal, L. Dai, A. Ahmad *et al.,* "A comparative study between rule-based and transformer-based election prediction approaches: 2020 US presidential election as a use case," in *Database and Expert Systems Applications-DEXA 2022 Workshops*, Vienna, Austria, pp. 32–43, 2022.

[33] K. Jaidka, S. Ahmed, M. Skoric and M. Hilbert, "Predicting elections from social media: A three-country, three-method comparative study," *Asian Journal of Communication*, vol. 29, no. 3, pp. 252–273, 2019.

[34] B. Justino Garcia Praciano, J. P. Carvalho Lustosa Da Costa, J. P. Abreu Maranhao, F. L. Lopes de Mendonca, R. T. De Sousa Junior *et al.,* "Spatio-temporal trend analysis of the Brazilian elections based on Twitter data," in *IEEE Int. Conf. on Data Mining Workshops, ICDMW*, Beijing, China, pp. 1355–1360, 2019.

[35] M. Z. Ansari, M. B. Aziz, M. O. Siddiqui, H. Mehra and K. P. Singh, "Analysis of political sentiment orientations on Twitter," *Procedia Computer Science*, vol. 167, no. 1, pp. 1821–1828, 2020.

[36] D. Q. Nguyen, T. Vu and A. Tuan Nguyen, "BERTweet: A pre-trained language model for English Tweets," in *Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 9–14, 2020. https://aclanthology.org/2020.emnlp-demos.0/

[37] K. Kawintiranon and L. Singh, "PoliBERTweet: A pre-trained language model for analyzing political content on twitter," in *2022 Language Resources and Evaluation Conf., LREC*, Marseille, France, pp. 7360–7367, 2022.

[38] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim *et al.,* "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, 2020.

[39] A. H. Huang, H. Wang and Y. Yang, "FinBERT: A large language model for extracting information from financial text∗," *Contemporary Accounting Research*, vol. 40, no. 2, pp. 806–841, 2022.

[40] N. Boudjellal, H. Zhang, A. Khan, A. Ahmad, R. Naseem *et al.,* "ABioNER: A bert-based model for Arabic biomedical named-entity recognition," *Complexity*, vol. 2021, pp. 1–6, 2021.

[41] K. K. Bressem, J. M. Papaioannou, P. Grundmann, F. Borchert, L. C. Adams *et al.,* "MEDBERT.de: A comprehensive german bert model for the medical domain," arXiv:2303.08179, 2023.

[42] Y. Liu, R. Zhang, T. Li, J. Jiang, J. Ma *et al.,* "MolRoPE-BERT: An enhanced molecular representation with rotary position embedding for molecular property prediction," *Journal of Molecular Graphics and Modelling*, vol. 118, no. 5, pp. 108344, 2023.

[43] M. Mujahid, E. Lee, F. Rustam, P. B. Washington, S. Ullah *et al.,* "Sentiment analysis and topic modeling on tweets about online education during COVID-19," *Applied Sciences*, vol. 11, no. 18, pp. 8438, 2021.

[44] A. Borg and M. Boldt, "Using VADER sentiment and SVM for predicting customer response sentiment," *Expert Systems with Applications*, vol. 162, pp. 113746, 2020.

[45] A. Vohra and R. Garg, "Deep learning based sentiment analysis of public perception of working from home through tweets," *Journal of Intelligent Information Systems*, vol. 60, no. 1, pp. 255–274, 2022.

[46] N. V. Chawla, K. W. Bowyer, L. O. Hall and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[47] M. R. Pribadi, D. Manongga, H. D. Purnomo, I. Setyawan and Hendry, "Sentiment analysis of the pedulilindungi on google play using the random forest algorithm with SMOTE," in *2022 Int. Seminar on Intelligent Technology and Its Applications: Advanced Innovations of Electrical Systems for Humanity, ISITIA, 2022-Proc.*, Surabaya, Indonesia, pp. 115–119, 2022.

[48] W. Jiang, K. Zhou, C. Xiong, G. Du, C. Ou *et al.,* "KSCB: A novel unsupervised method for text sentiment analysis," *Applied Intelligence*, vol. 53, no. 1, pp. 301–311, 2023.