

Sentiment-Driven Election Prediction using Multilingual Social Media Analytics

Ranojit Palit¹, Kanishka Nayak², Nikhil Pateria³, Manmeet Kaur Anand⁴, Kushagra Shekhar⁵,
and Nitya Bhardwaj⁶

¹Department of Computer Science and Engineering, Manipal University, Jaipur, India

^{2,3,4}Department of Artificial Intelligence and Machine Learning, Manipal University, Jaipur,
India

⁵Department of Software Engineering (Artificial Intelligence), Torrens University Australia,

⁶Department of Computer Science and Engineering, Manipal University Jaipur

¹ranojit.229301014@muj.manipal.edu, ²kanishka.2427010142@muj.manipal.edu,

³nikhil.pateria@jaipur.manipal.edu, ⁴manmeetanand2424@gmail.com, ⁵kushagrashekhar@gmail.com,

⁶nitya.2502051225@muj.manipal.edu

Abstract

Elections influence public policy priorities and social governance. Traditional forecasting methods, such as opinion polling, are subject to sampling biases and delayed feedback. Media coverage still drives most political conversations, but social media has taken precedence, with many users now more comfortable arguing from their phones or other electronic devices [10,18]. We present a cloud-scalable Global Election Prediction System (GEPS) that combines multilingual social media sentiment analysis with machine learning to estimate election trends [16,24]. In this GEPS, natural language processing is employed to normalise posts and classify them into three sentiment groups: positive, negative, and neutral. The three groups summarise sentiment and can be aggregated regionally and over time to track political momentum [2,6]. Logistic Regression, Random Forest, and LSTM models were evaluated, with LSTM being the most accurate forecasting model. In particular, LSTM was the most useful, as it can capture language cues based on context [5,15,23]. We broadly conclude that digital public opinion is useful, often providing early indicators of election outcomes and helping analysts, journalists, and campaigns in the decision-making process [19,20].

Keywords: Election prediction, Sentiment analysis, Machine learning, social media analytics, Political forecasting

1 Introduction

Elections determine representation and the trajectory of policies. Therefore, the ability to measure and forecast voter sentiment around significant events—like debates, protests, scandals, and policy statements—is of tangible utility[9,20]. Traditional forecasting pipelines rely on face-to-face or telephone surveys, turnout models, and economic indicators. While informative, these methods impose structural limitations: sample bias and overall non-response; the time lag between data collection and publication; and challenges in tracking rapid opinion shifts. At the same time, social media platforms (e.g., Twitter, Facebook, YouTube, Reddit, Instagram) have emerged as a primary venue for political discussion, serving as real-time, high-volume, multilingual streams of text [10,18].

Extracting reliable signals from these streams of text is non-trivial, as posts contain significant noise, are often code-switched or multilingual, and are frequently filled with sarcasm, figurative language, coordinated effort, or automation. However, several advances in natural language processing and machine learning now make it feasible to convert these raw data streams into structured measures of public mood [5,7,15]. For example, polarity analysis provides a useful estimate of support and opposition that can be aggregated locally and regionally over time to indicate momentum for parties or candidates[2,6]. This paper presents a comprehensive architecture called the Global Election Prediction System (GEPS). It uses

multilingual political discussions collected across countries and languages, then processes the text with natural language processing tools to classify sentiment and visualise the results through interactive dashboards. GEPS is designed to be deployable across elections in any country, regardless of language or culture [16,24]. The major contributions of this paper are: (i) a cloud-based framework that also enables data collection and analysis at multiple regions, (ii) a polarity prediction framework that evaluates both traditional machine learning and deep learning approaches, (iii) interactive visual dashboards that help users explore election trends in real time, and (iv), an evaluation framework including standard performance metrics (accuracy, precision, recall, and F1-score) from a dataset that encompasses a multi-country election collection over the 2022 to 2024 period.

1.1 Background

Polling reveals images of public opinion but suffers from coverage bias and rising non-response[20]. Shocks—such as news investigations and viral videos—can shift attitudes in hours, not days or weeks[18]. Near-real-time online discourse provides a mechanism for observing shifting public attitudes, but it is rife with issues of validity, representativeness, and ethical handling [10,12]. The basic point is not to replace polling with social media, but to complement traditional polling methods with greater temporal granularity and contextual traces that can be used in qualitative and quantitative assessments[19,20].

1.2 Objectives

The study's goal is to: (1) develop an end-to-end multilingual pipeline for sentiment analysis for an election; (2) assess different machine learning models for sentiment classification; (3) produce temporal sentiment indices and connect them with election phases; and (4) provide analysts and media teams with interpretable dashboards.

1.3 Scope and Assumptions

This study focuses on text posts and comments drawn from publicly available APIs and feeds [10,18]. We do not draw any individual-level inferences from data; all indicators are summed at the individual level. The framework focuses on estimating party/candidate momentum and does not attempt vote-share regressions or seat predictions [19]. Ethical guardrails included rate limiting, anonymising user handles in any outputs from the analysis, and compliance with opt-out policies for platform terms and use[12].

2 Ease of Use

2.1 System Design for Usability

GEPS has been developed with a user-centric perspective, so that both technical and non-technical analysts can access it with ease [26,27,28]. The interface employs familiar layout styles and structures the workflow into three simple steps: collect, analyse, and report. Tooltips, guided prompts, and example views will enhance the analytics process and make the transition to the data product easier [26,27]. Each chart also provides the date of data collection, the language present in the data, and the model version, adding transparency.

GEPS runs in a web browser and does not require installation. Long-duration tasks, such as retraining models, run in the background with a progress status[26]. Accessibility for GEPS includes keyboard navigation, high-contrast display options, screen readers, and adjustable text sizes, enabling a wide spectrum of users [26,28].

2.2 User Interaction Features

Interactive dashboards expose: (i) polarity distributions (positive/neutral/negative) over time; (ii) candidate vs. candidate comparisons; (iii) geo-aggregated heatmaps by region; and (iv)

key-event overlays (debates, manifestos)[19,27]. Power users can compose filters (language, country, party, hashtag) and store views for recurring reports. Keyboard shortcuts (e.g., Ctrl+U to upload a CSV of tracked keywords; Ctrl+R to refresh analysis) reduce menu traversal. A help layer bundles searchable FAQs, short videos, and glossary definitions (e.g., “F1-score”, “macro vs. micro averaging”)[26,27]. Error messages are actionable and precise, indicating inputs, limits, and next steps[28].

3 Conceptual Framework

3.1 System Integration

The architecture has a layered structure, with, from top to bottom, ingestion, preprocessing, modelling, and presentation [16]. The ingestion layer connects directly to the platform APIs and collects public posts that match election-related queries [10,18]. The pre-processing layer performs language detection, normalisation (handling URLs, emojis, and stop words), tokenisation, and, optionally, machine translation for cross-lingual unification. The message bus connects all stages and enables horizontal scalability [1,24]. The modelling layer operates several sentiment classifiers that contain versioned artefacts [5]. The presentation layer renders metrics, charts, and alerts using a time-series data store[27].

3.2 Machine Learning Models

We developed a baseline Logistic Regression model using TF-IDF features, a Random Forest model incorporating character- and word-level n-grams along with metadata features (such as the presence of question marks or exclamation marks), and an LSTM model that ingests word-piece embeddings [1,5,15,23]. The LSTM model is better suited to handling informal syntax and sequential dependencies in text data [23]. The models were trained using stratified train/test splits with early stopping to prevent overfitting[5]. Hyperparameter tuning was performed using either grid search or Bayesian optimisation, depending on the available computational budget [5,15].

3.3 Presentation and Decision Support

System outputs have been structured to be user-friendly[26,27]. Each sentiment chart displays confidence ranges and includes tooltips with example posts (user information removed) to provide context for any sharp changes[19]. Analysts can set alerts to monitor changes in public opinion over hours or days. Comparison panels allow users to compare opinion trends across different political events to see how particular regional or group sentiment differs from others [20]. There are also options to export, allowing users to save charts as images and downloadable data as CSVs for additional analysis[27].

3.4 Software Engineering Principles

The code respects modular boundaries and ensures validity through unit and integration testing using well-documented APIs. Security schemes include API access via OAuth, encrypted secrets, and role-based access management. Quality attributes are mapped to ISO/IEC 25010 and documented in accordance with IEEE guidelines. Containerization (e.g., Docker) and infrastructure-as-code support deployment reproducibility, and continuous integration pipelines are used for linting, testing, and vulnerability scanning [5].

4 Literature Review

4.1 Evolution of Social Media in Election Prediction

The growth of online communication platforms has transformed how political attitudes are expressed, shared, and even constructed[9,10]. Early election research focused on investigating the relationship between how often a candidate or political party was mentioned

in the digital conversation and polling outcomes[9]. Scholars assumed that if a political figure was widely discussed on social media, it could be interpreted as public interest or even support. Though these studies offered a simple directional signal for many core research questions, none provided insight into emotional tone or context. In the hopes of improving interpretability, researchers began adopting a lexicon-based approach and generalising it to sentiment analysis, effectively scoring political texts using existing positive and negative dictionaries [2,6,13,22]. The advantage of a lexicon-based approach was its straightforwardness; however, such methods were not always sufficient to convey certain linguistic features, including sarcasm, negation, slang, figurative language, and/or multilingualism. As online political discourse evolved to include memes and humour-based critiques, hybrid engagements, and rapid code-switching, these constraints became much more salient. Another major revolution followed a subsequent round of research on supervised machine learning approaches[17]. Classifiers such as Naive Bayes, Support Vector Machines, or Logistic Regression were trained using lexicon-based approaches or specific textual data [1,15].

4.2 Current Trends and Case Studies

Recent case studies indicate that online political sentiment may provide a reliable measure of electoral momentum across various political contexts [19,20]. In the context of United States presidential elections, research shows strong correlations between favourable sentiment and shifts in candidate approval after televised debates[19]. This has also been observed in the Indian general elections, where large political rallies and regional-level speeches led to observable sentiment shifts on Twitter and Facebook [16]. Even in Brazil, digital public opinion played a substantial role in public opinion formation in the run-up to elections, affecting public attitudes months before election results during anti-corruption protests[18].

Across several nations, sustained upward spikes in sentiment were typically associated with rising voter approval or increased polling strength; however, sharper upward spikes in negative sentiment were associated with controversies, differences of opinion on issues, or public scandals [19,20]. Nonetheless, these relationships are not entirely linear, and the magnitude of effects may depend on the individual media echo chambers of opinion, the level of political polarisation present, or the communication conventions of specific social media platforms [11,18]. To improve predictive stability, researchers have begun combining online sentiment indicators with secondary data sources (e.g., news headline sentiment, search engine trends, event calendars, or individual demographics)[12,16,21]. By combining multiple data sources, researchers can reduce the “noise” within specific platforms, helping ensure predictions are more robust. Researchers are also combining multiple predictive approaches[21].

4.3 Recent Developments (2023–2024)

The latest developments in election prediction research have focused on improving model accuracy, mitigating misinformation concerns, and investigating cross-lingual generalisation [7,8,24]. Models based on transformer technology, such as mBERT and XLM-R, have advanced cross-lingual opinion classification by transferring contextual language representations across languages [7, 8]. This allows models to be trained in one language and transferred via learning into another language with less annotated data [24].

Developing automated techniques for detecting coordinated manipulation efforts represents another important area of research, including bot networks, paid campaigns, and misinformation spreading[12,18]. Automated or misleading accounts can be used to distinguish between misleading signals and authentic public opinion indicators to depict legitimate public opinion. This is significant in a polarised election environment where

narrative manipulation is expected. Domain adaptation has started to become an area of focus. Since no election is entirely similar to another (each election has its unique cultural and political context), many models will not generalise effectively from one election to another[19]. Researchers are now experimenting with methods such as zero-shot transfer learning and weak supervision to adapt models to new areas of interest with minimal manual labelling effort [7,24]. Lastly, a human-in-the-loop approach is an emerging trend in which automated models flag complex or ambiguous posts for manual review (e.g., posts that contain sarcasm, humour, or idioms)[17]. This hybrid framework improves classification accuracy without relying entirely on manual coding for all instances.

5 Methodology

5.1 System Design

The Global Election Prediction System (GEPS) employs an efficient, modular pipeline that processes large volumes of multilingual online political data [10,16]. The data workflow consists of five distinct stages. In the first stage, data are collected via platform APIs and lists of election-related keywords and hashtags. The hashtags are regularly updated as campaigns unfold to capture the dynamic, trending political discourse.

In the second stage, the collected posts are pre-processed. In this step, text is cleaned by identifying the post's language, removing URLs and emojis, converting text to lowercase, removing stop words, and (as needed) stemming or lemmatising [1,6]. As necessary, multilingual data can be brought into a common format via translation [24]. Posts that switch between languages are retained to respect meaning.

In the third stage, machine learning classifiers classify each post as positive, negative, or neutral[2,15,23]. Sentiment analysis is conducted at the sentence level to gain a richer insight into the emotional cues within the text. In the fourth stage, the results are aggregated to group-level opinion indicators. These indicators show public opinion trends by country, region, political party, or candidate, which supports comparisons across political contexts.

The fifth and final stage is the visualisation of the results.

5.2 Ground Truth Data

Establishing consistently reliable ground truth labels is critical because online political discourse is often informal, context-dependent, and may even involve sarcasm [17]. GEPS addresses this issue by implementing a hybrid labelling approach to create a training dataset that best supports consistency [2,6]. The initial labels are generated using sentiment lexicon tools such as VADER and TextBlob, which assign polarity labels based on keyword scores [2,16]. The emotional polarity markers or classification labels serve as an approximate starting point; however, they are insufficient for deep learning models[15].

To improve label quality, a representative sample of posts will be manually reviewed by human annotators [17]. The annotators are provided with explicit guidelines for handling sarcasm, slang, abbreviations, negations, re-shared posts, and memes [17]. If the annotators disagree on the assigned label for a given post, they would discuss it, resolve the disagreement, and reach a consensus. To assess labelling consistency, Cohen's kappa is used to measure interrater reliability [1]. Additionally, the dataset intentionally includes posts from different languages and regions to avoid training a model for only one language or cultural style [24].

Once the labels have been finalised, the dataset is split into training, validation, and test sets by grouping posts based on the same election timeline phases, to avoid exposing the model to similar time-based posting patterns [5].

5.3 Modelling Approaches

Three families of machine learning models were implemented and assessed to compare their effectiveness in sentiment classification [1, 5, 15]. Logistic Regression with TF-IDF features served as the baseline model, offering the advantages of interpretability and fast inference[1]. It used weighted word frequencies to relate sentiment labels to post content, but failed to capture deeper contextual meaning or word order.

The Random Forest classifier provided improved expressiveness by modelling non-linear relationships between features[21]. Random Forest classifiers built a collection of decision trees using random sub-sampling of both data and features, thereby capturing stylistic cues such as repeated punctuation, capitalisation, and sentiment-heavy phrases. However, Random Forest models operated solely on token-level features without utilising semantic structure.

The Long Short-Term Memory (LSTM) neural network represented the deep learning methodology[5,15,23]. LSTMs were specifically structured for sequential data processing and treated input text as a sequence, enabling the model to capture sentence context, emotional nuance, sarcasm cues, and idiomatic expressions. The model used pre-trained word embeddings to initialise contextual representations based on semantic word similarity [3,4]. Furthermore, dropout layers were incorporated into the model architecture to minimise overfitting, while class imbalance was addressed through weighted loss terms or focal loss[5].

Hyperparameters of the model architecture, including embedding dimension, sequence length, learning rate, batch size, and hidden units, were tuned using grid search or Bayesian optimisation to achieve optimal model performance [5,15].

5.4 Evaluation Metrics

To evaluate how well (or poorly) each classification model performed, several evaluation measures were used [1]. Accuracy provides a broad indicator of the percentage of correctly predicted class labels; however, in the presence of class imbalance, it can be deceptive. Thus, both precision and recall were reported to provide a clearer sense of the model's ability to classify positive and negative sentiments correctly and to avoid misclassifying neutral sentiment. The F1-score, or the harmonic mean of precision and recall, balances those successes and failures to provide a snapshot across the class labels. Macro-averaging calculated precision, recall, and F1-score across all polarity classes equally. Examination of the confusion matrices was conducted to investigate common misclassification patterns (ex, negative posts classified as neutral posts)[1]. ROC curves and AUC scores were generated to assess the discriminative power of the classification models [5]. While examining static performance was important, temporal robustness was investigated by computing rolling-window F1-scores to see how the different campaigns were performing as they rolled out, providing important insights on sentiment as it developed over time.

5.5 Tools and Platforms

The technological stack comprised Python 3.10, scikit-learn for machine learning, TensorFlow/Keras for deep learning, and spaCy/NLTK for NLP [5,7,24]. There was a document store and a time-series database for data storage, and the dashboards used a modern JavaScript framework with REST APIs to serve the data. The services ran on a cloud platform using containers, and, where necessary, were auto-scaled and monitored.

Model	Accuracy	F1-score
-------	----------	----------

Logistic Regression	0.721	0.700
Random Forest	0.803	0.780
LSTM (proposed)	0.884	0.860

Table 1. Model performance on the held-out test set.

5.6 Dataset Description

This study utilises a dataset of approximately 250,000 publicly available social media posts collected during national electoral cycles spanning 2022 through 2024 [16,19]. To achieve globally diverse representation across election studies, posts were collected from five different countries - the United States, India, Brazil, the UK, and South Korea[19,24]. Posts were in various languages, including Hindi, English, Spanish, Portuguese, and Korean, and, crucially, represent, in aggregate, the multilingualism of political discourse on social media platforms[24]. The dataset contains several elements for each post, including the text content, the timestamp, and the language in which the post was published (determined via automated language detection). The dataset also contains a few additional derived features, such as indicators for sentiment category or word frequency [1,6]. The dataset has been constructed to examine online political discourse while safeguarding privacy by removing all identifiable user information. There are no usernames, profile information, or personal identifiers included in any of the analyses or outputs of the project [10,18].

6 Results and Discussion

Table 1 summarises classification performance. The LSTM achieves the strongest overall results, reflecting its advantage in modelling sequential dependencies and informal syntax[5,15,23]. Qualitatively, positive sentiment peaks aligned with campaign momentum surges, whereas negative clusters coincided with controversies or policy pushback[19,20]. Temporal alignment around debates showed asymmetric reactions across regions, underscoring the importance of multilingual coverage[24].

Observations. (1) Aggregated opinion indicators are useful for modelling emerging changes in public opinion; (2) multilingual processing is important to mitigate geographic and cultural biases; (3) model performance declines for sarcastic/social media posts, which motivates the exploration of combining state-of-the-art transformers for multilingual text analysis with sarcasm detection and ensemble sentiment models. Although these opinion signals are informative, they are not a replacement for traditional polling and should be treated as an additional factor in considering turnout and structural elements.

References

- [1] Aggarwal, C. C.: *Machine Learning for Text*. Springer (2018).
- [2] Hutto, C., Gilbert, E.: VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In: ICWSM (2014).
- [3] Mikolov, T., Chen, K., Corrado, G., Dean, J.: Distributed Representations of Words and Phrases and Their Compositionality. In: NIPS (2013).
- [4] Pennington, J., Socher, R., Manning, C.: GloVe: Global Vectors for Word Representation. In: EMNLP (2014).

- [5] Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press (2016).
- [6] Liu, B.: *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers (2012).
- [7] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of DeepBidirectional Transformers. NAACL (2019).
- [8] Vaswani, A. et al.: Attention Is All You Need. In: NeurIPS (2017).
- [9] Tumasjan, A., Sprenger, T., Sandner, P., Weppe, I.: Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment. In: ICWSM (2010).
- [10] Stieglitz, S., Dang-Xuan, L.: Social Media and Political Communication. *Journal of Information Technology* 28(4), 215–236 (2013).
- [11] V.K.Mishra, K. Sharma, V. Sharma, G. Shrivastava. "A Machine Learning based approach to detect fake news in social media." In 2023, the 5th International Conference on Advances in Computing, Communication, Control, and Networking (ICAC3N), pp. 1029-1036. IEEE, 2023.
- [12] Wang, W., et al.: Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31(2), 211–236 (2017).
- [13] Saif, H., He, Y., Alani, H.: Semantic Sentiment Analysis of Twitter. In: ISWC (2012).
- [14] Pak, A., Paroubek, P.: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In: LREC (2010).
- [15] Hassan, A., Mahmood, A.: Deep Learning Approaches for Sentiment Analysis: A Survey. *IEEE Computational Intelligence Magazine* 13(2), 81–97 (2018).
- [16] Singh, R., Sharma, V.: Social Media Data Analytics for Election Prediction. *IEEEAccess* 9, 67045–67057 (2021).
- [17] Joshi, A., Sharma, A., Bhattacharyya, P.: Sarcasm Detection Using Neural Attention Models. In: ACL (2017).
- [18] Borge-Holthoefer, J., et al.: Content and Network Dynamics in Political Discussions on Social Media. *EPJ Data Science* 4(1), 1–15 (2015).
- [19] Zimmermann, M., et al.: Predicting Elections with Social Media: A Cross-Country Study. *Information Sciences* 588, 209–224 (2022).
- [20] Llewellyn, C., Cram, L.: Social Media and Public Opinion in Elections. *Political Studies Review* 19(2), 147–170 (2021).
- [21] Gupta, K., Raj, S.: Machine Learning Techniques for Election Outcome Prediction. *Journal of Information and Intelligence Systems* 12(4), 250–268 (2022).
- [22] Kouloumpis, E., Wilson, T., Moore, J.: Twitter Sentiment Analysis: The Good, The Bad and The OMG! In: ICWSM (2011).
- [23] Zhao, W., et al.: Deep Learning for Short Text Sentiment Classification. *AppliedSciences* 9(16), 3388 (2019).
- [24] Sharma, S., et al.: Multilingual NLP for Social Media Political Analysis. *ACMTALLIP* 20(3), 1–19 (2021).
- [25] Kumar, A., Sebastian, T.: Sentiment Analysis on Twitter Data Using Machine Learning. *IJSR* 4(1), 5–9 (2020).

- [26] Nielsen, J.: *Usability Engineering*. Morgan Kaufmann (1994).
- [27] Shneiderman, B., Plaisant, C.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson (2016).
- [28] ISO 9241-11: Ergonomics of Human-System Interaction — Usability: Definitions and Concepts. International Organisation for Standardisation (2018).