AI-Driven Election Prediction Model

Nikhil Venkatachalam
Purdue University
Department of Computer Science
West Lafayette, Indiana, United States

Abstract

Election forecasting has traditionally relied on polling aggregation, historical trends, and expert judgment, which are approaches that often introduce subjective biases and struggle to adapt across different electoral contexts. This completed study presents the development and evaluation of an AI-powered election prediction model that enhances accuracy and adaptability by optimizing factor selection and weighting through machine learning. The model systematically integrates a diverse set of inputs, including election data, macroeconomic indicators, and sociopolitical variables, to identify and quantify key electoral influences while minimizing human bias. Over the course of the project, extensive datasets were collected and preprocessed, followed by the design, training, and refinement of the predictive model. Validation was conducted using historical election data to assess alignment with actual outcomes, with results showing that the AI model can effectively detect emerging voting patterns. Final testing demonstrated that the model can perform adequately in both accuracy and adaptability, though further work is necessitated in the future to ensure reliability. This research contributes to the evolving field of political data science by demonstrating the viability of machine learning as a scalable, systematic tool for election forecasting. The methodology also holds potential for further application to the United States electoral college system, as well as application beyond the United States, offering a framework for predicting elections in other democratic systems and deepening our understanding of global electoral dynamics.

1 Introduction/Motivation

Accurate election forecasting plays a crucial role in political analysis, campaign strategy, and public understanding of democratic processes. Traditional election prediction models rely heavily on polling data, historical trends, and expert judgment, often introducing subjective biases that can undermine the accuracy and adaptability of forecasts. While established models, such as those developed by political analysts like Nate Silver, have demonstrated success in some election cycles, they struggle with consistency due to evolving voter behavior, the unpredictability of political events, and the limitations of conventional data integration methods.

This project aims to address these challenges by developing an AI-driven election prediction model that optimizes factor selection and weighting through machine learning. Unlike traditional approaches, which depend on expert-defined heuristics, this model will leverage AI to systematically analyze and refine data inputs, ensuring greater objectivity and predictive power. By integrating diverse data sources, including polling data, demographic shifts, economic indicators, and social trends, this model aspires to enhance accuracy while maintaining adaptability across different political environments.

The significance of this research extends beyond academic inquiry. A more reliable and dynamic forecasting model has the potential to benefit policymakers, campaign strategists, and the general public by providing clearer insights into electoral dynamics. Additionally, future development of this model can explore the complexities of multi-party systems, allowing for accurate forecasts in democracies that do not conform to the binary red-versus-blue framework prevalent in the United States. In addition, the model can account for structural variables such as the Electoral College, which can decouple popular sentiment from electoral outcomes and pose unique challenges to prediction accuracy.

This framework also positions the model for global applicability. By accommodating the nuances of electoral systems in other countries, such as proportional representation, runoff voting, or coalition governance, the model can be adapted for a wide range of political contexts. Ultimately, this project seeks to set a new standard for election forecasting by merging machine learning methodologies with robust political analysis, mitigating biases, and improving predictive accuracy across diverse electoral landscapes.

2 Results

To collect data and begin training the model, there were some constraints and historical context that needed to be properly considered. Once consideration was done, the model was trained, and results were retrieved.

2.1 Chronological Start Time

When selecting an appropriate chronological starting point for training a model to predict U.S. election outcomes, it is crucial to consider significant political realignments that have reshaped voter behavior. Two major shifts stand out in modern electoral history: the Southern Strategy and the partisan color association of red and blue. These transformations affected voting patterns, party identities, and electoral map dynamics, making them key reference points in determining relevant historical data for predictive modeling.

The Southern Strategy refers to the Republican Party's deliberate efforts, beginning in the 1960s, to attract white voters in the Southern United States by appealing to racial tensions and opposing civil rights advancements. This strategy marked a pivotal shift, as the South transitioned from a Democratic stronghold, known as the "Solid South," to a region increasingly supportive of Republican candidates. The implementation of this strategy led to a significant realignment of party affiliations and voting patterns in the South, reshaping national electoral dynamics [3].

Afterwards, the association of colors with political parties underwent a transformation. Prior to the 2000 presidential election, there was no consistent color scheme linking red or blue to a specific party; media outlets varied in their representations. However, during the presidential elections of 1992 and 1996, we can begin to see the emergence of what would become the familiar map of "red states" and "blue states." While the color designations were not yet standardized, voting patterns started to show a clearer geographic divide, with the coasts and upper Midwest leaning Democratic and much of the interior leaning Republican [17][18]. During this period, independent candidate Ross Perot also emerged as a massive force in American politics, drawing significant support across the country and disrupting traditional party dynamics, particularly in 1992 when he won nearly 19% of the popular vote [22]. Despite these shifts, the Southern Strategy had not yet fully taken effect; Bill Clinton, a Southern Democrat from Arkansas, was able to win many Southern states in both of his presidential victories, reflecting a transitional moment in regional political alignment. The protracted and contentious 2000 election, particularly the intense focus on Florida's results, ultimately led to the widespread use of red to denote Republican states and blue for Democratic states. This color association became standardized and has since been embedded in political discourse and analysis [9].

In the years following Ross Perot's influential campaigns, third parties in the U.S. continued to appear on ballots but rarely had a significant impact on the red vs. blue state dynamic. While candidates from the Green Party, Libertarian Party, and others occasionally drew modest support, none matched Perot's national presence or ability to disrupt the electoral map [10]. At the same time, political polarization steadily increased, leading to a more rigid division between red and blue states. Exceptions became less common, with fewer states swinging between parties from one election to the next. Over time, notable shifts occurred within this red-blue framework: states like Ohio and Florida, once considered key battlegrounds, became more reliably Republican, while Georgia, traditionally conservative, trended Democratic and became a battleground state in recent elections [7]. These changes reflect broader demographic, cultural, and political realignments that continue to shape the American electoral landscape.

Therefore, to construct a model to predict election results, it is essential to consider key historical shifts that have shaped modern voting patterns. Beginning the training period with the 1992 presidential election provides a strategic starting point, as it marks the emergence of the red-and-blue-state framework (despite the rise of Ross Perot as a major third-party force) and a transitional moment in the Southern Strategy when the South had not yet fully shifted to the Republican column. This period captures the early formation of the electoral landscape that defines current partisan divides, while also offering more consistent and accessible data. Although earlier decades like the 1960s provide valuable context, especially regarding the origins of the Southern Strategy, data limitations make 1992 a more practical and coherent baseline for model development.

2.2 Data Collection

The process of data collection for model training begins with the careful selection and cleaning of relevant datasets to ensure accuracy and reliability. A foundational dataset consists of actual results from past elections, including the year of each contest, to serve as the benchmark for evaluating the model's predictive accuracy [24]. This historical record allows for direct comparisons between predicted outcomes (like poll aggregates [11][27]) and real-world vote margins, providing a basis for fine-tuning model parameters. Once gathered, the data are systematically categorized into distinct groups to facilitate structured analysis, ensuring that different influencing factors, from temporal trends to demographic shifts, are considered both separately and in concert.

One primary category of data includes "base statistics," which capture key political indicators that have historically influenced electoral outcomes. This set comprises poll-predicted margins [13] and voting-eligible-population turnout percentages [31]. Additional structural measures include the party-identification lean margin [12], which quantifies the net partisan tilt among voters, and congressional ideology metrics (both the median score and the polarization gap) to reflect the broader legislative environment [19]. These variables together offer a two-pronged view of both the electorate's short-term sentiment and its deeper, long-term alignments.

Economic indicators form another crucial category, as financial conditions often shape voter preferences. Standard macro-metrics, such as the inflation rate [20], average gas price [25], and the unemployment rate [26], are complemented by urbanization levels (percent urban population) [21]. Rising costs of living or joblessness can erode support for incumbents, while higher urban concentration may correlate with distinct partisan patterns. By incorporating these economic and geographic covariates, the model can assess how fluctuations in both national markets and population-density trends relate to shifts in the Democratic–Republican vote margin.

Beyond political and economic factors, social and perceptual indicators are also considered, as societal trends can heavily influence voter decision-making. Gun-homicide rates per 100,000 people [15] and crime-perception margins (the share of respondents believing crime has increased year-over-year) provide a lens on public safety concerns [4]. Likewise, immigration–approval gaps across party lines (immigration perception margin) capture divergent views on border policy [28]. Although detailed demographic breakdowns (like race, age, education) remain challenging to standardize across cycles, the inclusion of these social indicators helps contextualize underlying voter motivations and refine the model's predictive power.

2.3 Model Training

The final script implements a lot of moving parts. Basically, the script first implements a regularized linear-regression pipeline to predict the Democratic–Republican vote-margin in U.S. elections, treating "year" as just another numeric covariate. First, the data is loaded and four binary "incumbency" indicators (incumbent party and candidate for D and R) are mapped from {Y,N} to {1,0}. Now, all columns, including "year", are selected as predictors X, with the outcome y defined as the Democratic–Republican margin. Because Ridge regression penalizes coefficients according to their scale, each column of X is standardized to zero mean and unit variance via

$$\widetilde{X} = \frac{X - \overline{X}}{\mathrm{sd}(X)}.$$

Ridge-cross-validation (RidgeCV) is then run over a logarithmic grid of 50 candidate penalties $\alpha \in [10^{-2}, 10^{2}]$, using 5-fold CV to choose

$$\hat{\beta} = \arg\min_{\beta} \Big\{ \|y - \widetilde{X}\beta\|_2^2 \; + \; \alpha\, \|\beta\|_2^2 \Big\}.$$

The fitted model yields coefficients (and an intercept) which are printed in a table, and in-sample predictions $\hat{y} = \widetilde{X}\hat{\beta}$ are appended to the dataframe for inspection.

To assess fit over time, the script plots actual vs. predicted margins against "year," allowing one to visually diagnose temporal trends captured by the model (especially the effect of including "year" itself as a regressor). However, standard k-fold CV can leak future information when applied to time-ordered data. Therefore, the final section performs a rolling-origin (walk-forward) time-series CV: for each election $i = 6, \ldots, N$, one trains on all data up to $i - 1$ (re-standardizing within that fold), fits a new RidgeCV, and predicts the ith election. Errors

$$e_i = \hat{y}_i - y_i$$

are collected, and the overall root-mean-square error

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_{\mathrm{test}}} \sum_i e_i^2}$$

is reported to quantify genuine out-of-sample forecasting accuracy. This approach mimics how one would forecast a future election (always training only on past data) and provides a more realistic estimate of predictive performance.

2.4 Model Results: RidgeCV Coefficients

Table 1. Standardized RidgeCV Coefficients for predicting Democratic–Republican margin

Feature	Coefficient
year	0.174245
poll_predicted_margin	1.040194
vep_turnout_percent	-0.385964
incumb_party_d	0.135391
incumb_party_r	-0.135391
incumb_running_d	1.250828
incumb_running_r	-0.562514
incumb_approval	0.243481
party_id_lean_margin	1.535376
congress_med_ideology	-1.221861
congress_ideology_margin	0.582243
inflation_rate	-0.417190
gas_price	-0.821334
unemployment_rate	0.531559
gun_homicide_rate	0.288948
crime_perception	0.162343
immigration_perception_margin	-0.358054
percent_urban	0.347848
intercept	3.155556

The standardized Ridge-CV coefficients (Table 1) seem to reveal that, after penalization, the single strongest structural predictor of the Democratic–Republican vote margin is the underlying partisan lean of the electorate ($\beta_{\text{party\_id\_lean\_margin}} = 1.54$), followed closely by the poll-aggregator's predicted margin ($\beta_{\text{poll\_predicted\_margin}} = 1.04$) and the effect of a running Democratic incumbent ($\beta_{\text{incumb\_running\_d}} = 1.25$). By contrast, holding all else equal, a one-standard-deviation increase in real-time economic distress (like gas prices, $\beta = -0.82$; inflation rate, $\beta = -0.42$) or higher turnout ($\beta = -0.39$) tends to erode Democratic vote share. Ideological signals from Congress are also important: a more Republican median Congress strongly depresses the Democratic margin ($\beta = -1.22$), whereas greater ideological polarization ($\beta = +0.58$) actually boosts it, which may reflect a mobilization effect when parties diverge sharply.

Smaller "control" coefficients seem to be in accordance with political-science intuition. The symmetric signs on incumb_party_d (+0.14) and incumb_party_r (−0.14) simply reflect the two-dummy encoding, and the positive intercept ($\approx$ 3.16) indicates a modest pro-Democratic baseline when all predictors are at their means. Moderately sized positive effects for incumbent approval ($\beta = +0.24$), urbanization ($\beta = +0.35$), unemployment rate ($\beta = +0.53$), and crime-related perceptions ($\beta = +0.16$–0.29) suggest nuanced trade-offs: some "bad news" indicators paradoxically favor Democrats, possibly because of attribution effects or interactions not captured here. In sum, the penalized estimates balance goodness-of-fit (like polls and partisanship) with tightness, shrinking weaker or collinear effects toward zero while preserving the dominant role of structural partisan and polling variables.

2.5 Model Results: Display and Drawbacks

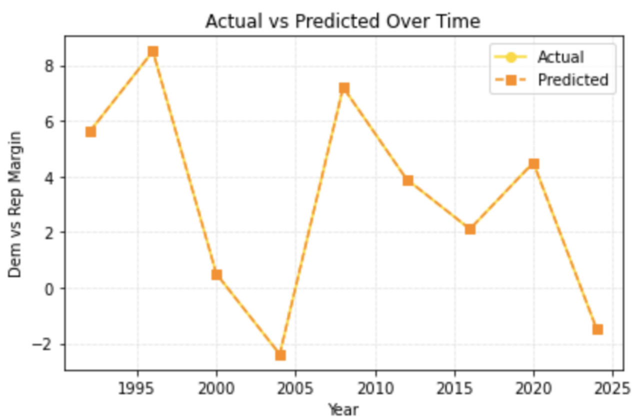

Figure 1. Actual vs. predicted Democratic–Republican vote margins over time.

As can be observed, the fitted model's in-sample predictions (Figure 1) almost perfectly reproduce the observed Democratic–Republican margins: errors never exceed a few thousandths of a percentage point (e.g. 2024: predicted −1.4967 vs. actual −1.5000). This near-identity stems from the dominant weight placed on the poll aggregate and underlying partisan lean (which together explain the lion's share of variance) as well as the inclusion of "year" to capture any residual linear trend. While the small residuals imply that economic and incumbency controls only fine-tune the forecast, the extreme accuracy also highlights a risk of tautology: using poll_predicted_margin almost amounts to benchmarking against itself. In practice, truly out-of-sample validation (as in the rolling-origin CV) is essential to guard against overfitting and to assess whether the modest contributions of secondary covariates hold up when predicting genuinely unseen elections.

However, the rolling-origin CV (Table 2) reveals that, regrettably, our ostensibly perfect in-sample fit does not carry over to genuine out-of-sample forecasting. When training on elections up to 2008, 2012, 2016, and 2020 and then predicting the next cycle, the model's errors balloon to −1.75 points in 2016, −3.34 points in 2020, and +4.02 points in 2024, yielding an overall RMSE of 2.757. This underperformance, particularly the large swing in 2024, suggests the RidgeCV weights (notably on polls and partisan lean) fail to capture shifting dynamics or nonlinear shocks between cycles. Therefore, we must acknowledge that our linear, penalized framework, while tightened, may be overly rigid for the evolving nature of electoral behavior, underscoring the need for richer temporal models or interaction terms in future work.

Table 2. Rolling-Origin CV Results

Train Yr	Test Yr	Obs. Margin	Pred. Margin	Error
2008	2012	3.90	4.07	0.17
2012	2016	2.10	0.35	-1.75
2016	2020	4.50	1.16	-3.34
2020	2024	-1.50	2.52	4.02

2.6 Model Reflection

The model does exhibit several strengths that merit some recognition. First, by standardizing all predictors to zero mean and unit variance, it places variables measured on disparate scales, such as "year," "gas_price," and "percent_urban", on an equal footing, ensuring that coefficient magnitudes genuinely reflect relative importance rather than arbitrary units. Second, the use of Ridge regularization (ℓ_2-penalty) systematically shrinks noisy or highly collinear coefficients toward zero, thereby reducing overfitting and improving numerical stability when predictors like "incumbent approval" and "poll_predicted_margin" may themselves be correlated. This balance of bias and variance is especially valuable in electoral contexts, where polls can be imprecise and economic indicators volatile.

Additionally, the hybrid inclusion of both real-time signals (poll aggregates, approval ratings) and deeper structural factors (party lean, ideological polarization, urbanization) allows the model to capture both short-term shifts and long-term partisan backdrops. Such a two-pronged strategy aligns with contemporary political-science theory, which emphasizes the joint roles of "fundamentals" and "campaign dynamics" in shaping electoral outcomes. Finally, the straightforward interpretability of linear coefficients, paired with a transparent cross-validation scheme, means that researchers can readily assess which features drive predictions, and policymakers or journalists can gain clear, data-driven insights without recourse to "black-box" algorithms.

3 Plans For Potential Future Work

First and foremost, the model outlined in this report is at its very preliminary stages. Given that knowledge, several avenues for future work could bolster both the robustness and realism of this forecasting framework. First, introducing interaction terms, especially between economic indicators (like unemployment × inflation) and incumbency status, could capture conditional effects that a purely additive model overlooks. Second, exploring nonlinear or semi-parametric extensions (like Gaussian processes, generalized additive models, or tree-based learners) may better accommodate regime changes and volatility in political dynamics. Third, a hierarchical Bayesian formulation could explicitly model state- or demographic-level variation and poll-forecast uncertainty, yielding calibrated predictive intervals rather than point estimates alone. Fourth, time-varying coefficients or state-space approaches would allow key effects (such as the weight on polls versus structural lean) to adapt endogenously over multiple election cycles. Finally, expanding cross-validation to block-bootstrap or more frequent rolling windows, and incorporating exogenous shock indicators (like major events, campaign spending spikes), would provide a more nuanced assessment of out-of-sample performance and guard against overconfidence in fixed-coefficient forecasts.

Additionally, the foundational framework developed in this project is not limited to the context of U.S. elections. While it was initially designed within a predominantly two-party, United States-based, popular vote political system, the model should be extended in the future so that it will be applicable in a wide array of electoral contexts. It should be adapted to simulate multi-party competition, account for coalition dynamics, and reflect the broader political diversity observed in parliamentary systems. This can be particularly valuable for studying elections in countries with more fragmented party systems, such as those in Western Europe, where coalition governments are common.

The model should also be modified to incorporate complex institutional structures that influence how votes are translated into power. For example, within the United States, the Electoral College system necessitates nuanced, state-by-state modeling to accurately forecast electoral outcomes. A model that aligns state-level vote shares with electoral vote projections can provide a clearer picture than national popular vote forecasts alone.

Future work can also open the door to international applications. From analyzing proportional representation in countries like Germany or Spain, to understanding ranked-choice voting in Australia, or presidential runoff systems in countries like Brazil or France, this model can offer a scalable, data-informed tool for comparative electoral analysis in the future. Ultimately, there is a very preliminary framework here to advance global political science research and contribute to a deeper understanding of evolving democratic processes around the world.

4 Related Work

Related work to election prediction using AI is mostly related to other forms of election prediction and other ways AI can be used in elections.

4.1 Sentiment Analysis/Social Media

Sentiment analysis, or opinion mining, is widely used to gauge public opinion by analyzing text data, particularly from social media. In election prediction, this method helps assess voter sentiment by examining posts, tweets, and comments. Social media platforms have become critical spaces for political discourse, providing researchers with extensive data to analyze voter preferences and predict election outcomes [6].

Political parties and candidates increasingly rely on social media for campaigns, prompting researchers to assess various computational approaches for election forecasting. A study focusing on Twitter data emphasized the need for classifying research based on methodology, technology, and citation trends to better understand progress and challenges in the field [2]. Machine learning models have also been applied, as seen in research on Pakistan's 2018 general election, which used sentiment analysis to evaluate election results and fairness [1].

Despite its potential, sentiment analysis for election prediction faces significant challenges. A systematic review found that traditional approaches relying on volume and sentiment analysis often yield inconsistent results, whereas regression models trained with polling data tend to be more accurate [5]. Additionally, the vast amount of social media data introduces issues like noise, bias, and misinformation, complicating analysis. Researchers have highlighted these challenges and proposed improvements, including better data processing and advanced machine learning techniques, to enhance predictive accuracy [14].

While sentiment analysis is a valuable tool for election forecasting, its reliability depends on methodology, data quality, and integration with traditional prediction models. Combining sentiment analysis with established electoral forecasting techniques could improve the accuracy and robustness of election predictions.

4.2 Other Election Prediction Systems

In general, modern election forecasting combines a range of methods, blending traditional political science approaches with newer, data-driven innovations. Some of the most commonly used models are based on political and economic fundamentals, such as the state of the national economy and presidential approval ratings. These structural models treat elections as referendums on past performance and often perform well when forecasting over longer periods. At the same time, poll-based models, especially those using techniques like multilevel regression with poststratification (MRP), have become increasingly influential, particularly closer to Election Day when voter sentiment stabilizes [23].

In addition to fundamentals and polling, more recent forecasting strategies include prediction markets and citizen-based models. Prediction markets rely on the financial stakes and collective judgment of traders to anticipate outcomes, while citizen forecasts aggregate public expectations under the theory that group predictions tend to cancel out individual biases. Machine learning and artificial intelligence have also entered the picture, using digital data like social media activity to estimate voter behavior (as discussed previously). Models like PollyVote even combine several of these methods into a single forecast to improve accuracy and reduce individual model biases [23].

Despite these advancements, there are important distinctions between forecasting and scientific prediction. Forecasting aims to predict specific, near-term events, like a particular election outcome, whereas scientific prediction seeks to explain general patterns over time. Forecasts often include probability estimates that are easily misunderstood by the public, leading to skepticism, especially when results fall within predicted margins of error but appear "wrong" in hindsight. The real value of forecasting lies in its ability to test theories against unknown future outcomes, offering a meaningful way to evaluate and refine political models [8].

4.3 AI In Elections

Other than the prediction of elections, AI is increasingly shaping elections in other ways, offering advantages in campaign efficiency, voter engagement, and misinformation detection while also posing risks related to disinformation, voter manipulation, and cybersecurity threats. AI tools streamline processes like voter registration and election monitoring, but concerns about transparency, accountability, and fairness remain significant [16][30].

One major concern is AI's role in political messaging and disinformation. AI-driven targeting allows campaigns to reach voters more effectively, but it also enables the spread of misleading narratives, deepfakes, and psychological profiling that can manipulate public opinion. The unchecked use of AI-generated content in political advertising threatens electoral integrity, making regulation a priority [32].

Governments are responding to these risks through various policies. South Korea has banned deepfakes in political campaigns for 90 days before elections, while Brazil has enacted restrictions on AI-generated political propaganda. Several U.S. states require disclosures for AI-generated election content, and the EU's AI Act aims to ensure transparency in AI systems that influence democratic processes. Meanwhile, tech companies have implemented AI detection, labeling, and misinformation-mitigation strategies, but enforcement remains inconsistent [29].

Despite its challenges, AI has the potential to improve electoral processes if properly regulated. Thus, striking a balance between leveraging AI's benefits and mitigating its risks is essential to maintaining trust in democratic systems in the digital age.

5 Conclusion

This project aims to develop a more accurate, adaptable, and data-driven approach to election forecasting by leveraging artificial intelligence and machine learning. Traditional models, while effective in certain contexts, often struggle with subjectivity, inconsistent performance across different election cycles, and limited integration of diverse data sources. By optimizing factor selection and weighting through AI, this model seeks to address these challenges, offering a more systematic and unbiased method of predicting electoral outcomes.

Through careful data collection, model training, and iterative refinement, the project has worked to enhance predictive accuracy while ensuring robustness across varying political environments. By incorporating a wide range of influencing factors, including polling data, economic indicators, and demographic trends, the AI-driven approach provides a more comprehensive framework for understanding voter behavior. Results suggest that this model can perform with adequate success at predicting larger trends and voting patterns, but needs further refinement to demonstrate reliability.

However, this research done has broad implications for political analysis, policymaking, and campaign strategy by offering a framework for a more reliable election forecasting model that can inform decisions, boost voter engagement, and deepen the study of electoral trends. The model can allow for future extensions beyond the U.S. two-party system, accommodating multi-party competition, coalition dynamics, and institutional nuances like the Electoral College. Furthermore, the model can enable global applications, supporting analysis of systems such as proportional representation in Europe, ranked-choice voting in Australia, and presidential runoffs in Latin America, fostering a deeper, data-driven understanding of political dynamics worldwide.

Although challenges remain even beyond this, including the unpredictability of late-stage campaign developments, potential biases in training data, and public skepticism toward algorithmic predictions, this project represents a not insignificant advancement in the field of election forecasting. By combining advanced machine learning techniques with rigorous political analysis, it establishes a scalable and globally relevant foundation for future innovations in predictive modeling. Ultimately, this research highlights the transformative potential of AI in electoral studies, offering new insights and methodologies for forecasting democratic outcomes with greater precision and reliability across political systems.

References

[1] Haider Ali, Haleem Farman, Hikmat Yar, Zahid Khan, Shabana Habib, and Adel Ammar. 2022. Deep learning-based election results prediction using Twitter activity. *Soft Computing* 26, 16 (1 August 2022), 7535–7543. https://doi.org/10.1007/s00500-021-06569-5

[2] Quratulain Alvi, Syed Farooq Ali, Sheikh Bilal Ahmed, Nadeem Ahmad Khan, Mazhar Javed, and Haitham Nobanee. 2023. On the frontiers of Twitter data and sentiment analysis in election prediction: a review. *PeerJ Computer Science* 9 (21 August 2023), e1517. https://doi.org/10.7717/peerj-cs.1517

[3] L. Sue Baugh. 2024. Southern strategy. *Encyclopaedia Britannica*. https://www.britannica.com/topic/Southern-strategy Accessed 13 March 2025.

[4] Megan Brenan. 2025. Smaller Majorities Say Crime in U.S. Is Serious, Increasing. https://news.gallup.com/poll/652763/smaller-majorities-say-crime-serious-increasing.aspx

[5] K. D. S. Brito, R. L. C. S. Filho, and P. J. L. Adeodato. 2021. A Systematic Review of Predicting Elections Based on Social Media Data: Research Challenges and Future Directions. *IEEE Transactions on Computational Social Systems* 8, 4 (August 2021), 819–843. https://doi.org/10.1109/TCSS.2021.3063660

[6] Priyavrat Chauhan, Nonita Sharma, and Geeta Sikka. 2021. The emergence of social media data and sentiment analysis in election prediction. *Journal of Ambient Intelligence and Humanized Computing* 12, 2 (February 2021), 2601–2627. https://doi.org/10.1007/s12652-020-02423-y

[7] Camila Domonoske. 2024. Which States Are Swing States? Here Are the Results to Watch This Year. https://www.npr.org/2024/11/05/g-s1-32673/which-states-swing-states-2024-presidential-election Accessed: 2025-04-15.

[8] Keith Dowding. 2021. Why Forecast? The Value of Forecasting to Political Science. *PS: Political Science & Politics* 54, 1 (2021), 104–106. https://doi.org/10.1017/S104909652000133X Web.

[9] Jodi Enda and Meilan Solly. 2024. When Republicans Became 'Red' and Democrats Became 'Blue'. *Smithsonian Magazine*. https://www.smithsonianmag.com/history/when-republicans-became-red-democrats-became-blue-104176297/ Accessed 13 March 2025.

[10] Mary Claire Evans. 2025. Support for a Third Political Party in the U.S. Dips to 58%. https://news.gallup.com/poll/651278/support-third-political-party-dips.aspx Accessed: 2025-04-15.

[11] Gallup. 2024. Gallup Presidential Election Trial-Heat Trends, 1936–2008. https://news.gallup.com/poll/110548/gallup-presidential-election-trialheat-trends-19362004.aspx#4

[12] Gallup. 2025. Party Affiliation. https://news.gallup.com/poll/15370/party-affiliation.aspx

[13] Gallup. 2025. Presidential Approval Ratings – Gallup Historical Statistics and Trends. https://news.gallup.com/poll/116677/presidential-approval-ratings-gallup-historical-statistics-trends.aspx

[14] Abhishek Gaur and Dharmendra Kumar Yadav. 2025. A comprehensive analysis of forecasting elections using social media text. *Multimedia Tools and Applications* (27 January 2025). https://doi.org/10.1007/s11042-024-20528-w

[15] John Gramlich. 2025. What the data says about gun deaths in the U.S. Pew Research Center. https://pewrsr.ch/448q4hU

[16] Segun Kehinde, Busola Simon-Ilogho, Kemi Kehinde, and Tola Kehinde. 2024. Exploring the Impact of AI on Voter Confidence and Election Information in 2024. *Qeios* (08 July 2024). https://doi.org/10.32388/UT898Q

[17] Michael Levy. 2025. United States Presidential Election of 1992. *Encyclopædia Britannica*, Encyclopædia Britannica, inc.. https://www.britannica.com/event/United-States-presidential-election-of-1992 Accessed 15 Apr. 2025.

[18] Michael Levy. 2025. United States Presidential Election of 1996. *Encyclopædia Britannica*, Encyclopædia Britannica, inc.. https://www.britannica.com/event/United-States-presidential-election-of-1996 Accessed 15 Apr. 2025.

[19] Jeffrey B. Lewis, Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet. 2025. Voteview: Congressional Roll-Call Votes Database. https://voteview.com/

[20] Macrotrends. 2025. U.S. Inflation Rate 1960-2025. Retrieved from Macrotrends. https://www.macrotrends.net/global-metrics/countries/USA/united-states/inflation-rate-cpi

[21] Macrotrends. 2025. U.S. Urban Population 1960–2025. https://www.macrotrends.net/global-metrics/countries/USA/united-states/urban-population

[22] Suzanne McGee. 2025. How Billionaire Ross Perot Brought Populism Back to Presidential Politics. *History.Com*, A&E Television Networks. https://www.history.com/articles/ross-perot-populist-1992-election-changed-politics

[23] Philippe Mongrain and Mary Stegmaier. 2024. Introduction to Forecasting the 2024 US Elections. *PS: Political Science & Politics* (2024), 1–8. https://doi.org/10.1017/S1049096524001008 Web.

[24] The Editors of Encyclopaedia Britannica. 2025. United States Presidential Election Results. *Encyclopaedia Britannica*. https://www.britannica.com/topic/United-States-Presidential-Election-Results-1788863 Accessed 13 March 2025.

[25] U.S. Bureau of Labor Statistics. 2025. Average Price: Gasoline, Unleaded Regular (Cost per Gallon/3.785 Liters) in U.S. City Average [APU000074714]. Retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/APU000074714

[26] U.S. Bureau of Labor Statistics. 2025. Unemployment Rate [UNRATE]. Retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/UNRATE

[27] RealClearPolitics. 2025. RealClearPolitics Polling Archive. https://www.realclearpolling.com/historical-polling-archive

[28] Lydia Saad. 2025. Americans Still Value Immigration, but Have Concerns. https://news.gallup.com/poll/508520/americans-value-immigration-concerns.aspx

[29] Carol Soon and Samantha Quek. 2024. *Safeguarding Elections from Threats Posed by Artificial Intelligence.* Technical Report IPS Working Papers No. 56. Lee Kuan Yew School of Public Policy. 1–52 pages. https://doi.org/10.25818/chkz-en87 Unpublished.

[30] Prof. K. Stevenson. 2024. Artificial Intelligence: A Double-Edged Sword In Elections. *Educational Administration: Theory and Practice* 30, 4 (13 April 2024), 1660–1667. https://doi.org/10.53555/kuey.v30i4.1724

[31] UCSB. 2025. Voter Turnout in Presidential Elections. https://www.presidency.ucsb.edu/statistics/data/voter-turnout-in-presidential-elections

[32] Chen Yu. 2024. *How Will AI Steal Our Elections?* Technical Report un7ev. Center for Open Science. https://doi.org/10.31219/osf.io/un7ev