2024 Presidential Election Forecast

line 1: Brett Anwar
line 2: *Computer Science Department*
line 3: *Emory University*
line 4: Atlanta, United States
line 5: brett.anwar@emory.edu

line 1: Xavier Pierce
line 2: *Computer Science Department*
line 3: Emory University
line 4: Atlanta, United States
line 5: xavier.pierce@emory.edu

line 1: Izana Melese
line 2: *Computer Science Department*
line 3: *Emory University*
line 4: Atlanta, United States
line 5: izana.melese@emory.edu

***Abstract*—This project aims to use historical US County Presidential Election returns from 2000 to 2020 alongside socio-economic indicators such as demographics, employment rates, and median household income, to predict the outcomes of the 2024 Presidential election across US counties.**

Keywords—voting, data analysis, election, random forest, swing votes, demographics, linear regression, sentiment analysis

I. Introduction

Predicting election outcomes is a complex challenge that has significant implications for political strategy, governance, and public policy. The motivation behind this project is to derive meaningful patterns and insights from historical election data and socio-economic indicators to forecast future election results. Such predictions can help in understanding the shifting political landscape and the factors that drive electoral preferences. This project involves analyzing a vast array of data, including past election results and various socio-economic indicators, to identify patterns and factors with a significant impact on voting behavior. The ability to accurately predict election outcomes at the county level can provide valuable insights for political campaigns, policy makers, and analysts. It can identify key battleground counties, explain the socio-economic drivers of voting behavior, and develop strategies tailored to the needs and preferences of specific populations.

II. Related Work

After some research, the team recognizes a lack of data mining experiments and/or projects related to predicting election results based solely on historic demographic data.

A. Split Ticket

The most prominent piece of related work is a political analysis website, Split Ticket. Split Ticket uses historic demographic and electoral data to train models used for an election forecast, then making its best attempt at inferring the actual margin for current candidates using current election data. It must be noted that the CEO of this forecaster has highlighted the high potential for bias to "creep into even a code-based approach to elections" [3].

B. Swing Vote Analysis Experimentation

Another particular experiment seeks to predict the election results of the 2024 Indian General Elections. The research calculates what it calls "swing parameters" using techniques like linear regression, Naive Bayes, Random Forest, Time Series, and more [2]. This research identifies "swing parameters" as reasons for swing voting, satisfaction with the current candidate, and current needs; these are parameters and techniques we should potentially include in our own research and experiment.

C. Sentiment Analysis Experimentation

As it stands, most approaches to our selected problem deal in sentiment analysis on social media sites such as Twitter or Facebook. Seeing as many experiments highlight "...the relationship between social media data and political results…" [5], it is likely that our initial proposed approaches, utilizing solely demographic and electoral data analysis, will not suffice for an accurate solution to our problem. If this is the case, we may opt for including some sentiment analysis in our methodologies in an effort to provide an accurate solution.

III. Proposed Approaches

We have contemplated using logistic regression and Random Forest models for predicting the 2024 Presidential election outcomes at the county level.

A. Logistic Regression

Logistic regression for its simplicity and effectiveness in binary classification problems which makes it suitable for predicting a binary outcome (Republican or Democrat).

B. Random Forest

The Random Forest model is better at handling high-dimensional data and reducing the risk of overfitting, providing us with robust alternatives that can incorporate the multitude of factors influencing electoral outcomes.

C. Sentiment Analysis

We must also consider implementing sentiment analysis on Twitter, Reddit, and/or Facebook data to further enhance the model's predictive capability by gauging public sentiment and potential shifts in voter preferences. We still would like to do more research to find the model that is most applicable to our task.

IV. System Design and Implementation

While this step is highly up in the air, we have found a dataset that contains US County Presidential Election Returns from 2000-2020. We plan to integrate other datasets with demographics that are also sorted by county to discern more information about each specific county and its voting trends. We will use python as there are many helpful packages and tools that are simple to integrate and easy to use. We are interested in trying out Orange for component based mining and machine learning.

References

[1] Alvi Q, Ali SF, Ahmed SB, Khan NA, Javed M, Nobanee H. “On the frontiers of Twitter data and sentiment analysis in election prediction: a review.” PeerJ Comput Sci. 2023 Aug 21;9:e1517. doi: 10.7717/peerj-cs.1517. PMID: 37705657; PMCID: PMC10495957.

[2] P. Parida, S. Sinha, A. P. Agrawal and R. Singh Yadav, "Predicting the General Election 2024 using ML and data analytics," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-9, doi: 10.1109/INCET57972.2023.10170638.

[3] R. Leven, “Should we take election forecasts seriously? A computer scientist says yes | CDSS at UC Berkeley,” data.berkeley.edu. https://data.berkeley.edu/news/should-we-take-election-forecasts-seriously-computer-scientist-says-yes (accessed Feb. 19, 2024).

[4] E. Culliford, “How political campaigns use your data,” Reuters, Oct. 12, 2020. Available: https://www.reuters.com/graphics/USA-ELECTION/DATA-VISUAL/yxmvjjgojvr/

[5] Hazim Moawi, “Predicting Voting Behaviors and Election Results Using Digital Trace Data and Twitter,” Social Science Research Network, Jan. 2023, doi: https://doi.org/10.2139/ssrn.4464