Alícia Chimeno Sarabia
Silvia Ferrer Vázquez December 6, 2024

US Election Prediction

Advanced Machine Learning Project Proposal

1 Problem Proposal

1.1 Objective

Predicting election results is a complex and often controversial subject in today's society. In 2016, the United States experienced widespread surprise with Donald Trump's victory, leading to significant disappointment in statisticians when their forecasts failed to align with the actual outcome. Many forecasting had projected a decisive win for Hillary Clinton, only for the results to tell a different story. The *fivethirtyeight* gave Clinton a 71% of victory [1], the *The New York Times* gave a 85%[2]. This outcome sparked global criticism and skepticism about the reliability of statistical models in predicting elections.

The 2024 election scenario differed from previous predictions, as polls indicated a very close race. According to *The New York Times*, the probability of a Harris victory was 49%, compared to 48% for Trump [3]. Despite early forecasts suggesting that the 7 swing states were too close to call, the actual results showed Trump winning all of them, albeit by narrow margins. This outcome secured him 312 electoral votes, far surpassing the 270 needed for victory.

It was evident that the decisive factor in the election was the outcome of these 7 battleground states: Arizona, Georgia, Michigan, Nevada, North Carolina, Pennsylvania and Wisconsin. In 2020, these same states had been won by Biden, yet in 2024, they shifted to the Republican.

Our project aims to make a forecast for predicting the winner of the election, focusing on this very important states. And most important seeks to explore the factors influencing election outcomes and provide meaningful insights into the dynamics of voter behavior.

1.2 Strategy

Classification model to predict the winner of the 2024 election.

Data: This project is organized within the specified Github Repository were we have added the datasets in [4].

- *Surveys of public interests that influenced their voting decisions:* Aggregated rankings of issues Americans prioritized in their votes. Key concerns include: economy, abortion, safety, immigration, and others.
- *State socio-economic indicators:* Possible variables include: income, immigration rates, crime rates, education levels, rural vs. urban population ratio, unemployment rates, religion,...
- *Past elections results*
- More that we find interesting and useful.

1.3 Motivation of the problem

- To better understand today's social dynamics by analyzing a compelling topic: the intricate intersection of human behavior and politics.
- To explore why forecasts often fail to predict actual outcomes, and why it is so challenging to forecast electoral results with precision.

1.4 Challenges

- Small number of observations (# states=50).
- Something new that has not been studied yet and we cannot compare results or take more ideas.
- Find all the necessary complete datasets.

1.5 Expect to Learn

- Gain a deeper understanding of voting behavior and how key issues influence voter decisions in the US.
- Evaluate the effectiveness of predictive models in forecasting election outcomes.
- Identify the challenges and limitations in predicting election results accurately.
- Explore how socio-economic factors impact voting patterns and election results.
- Understand the influence of polls and public opinion in shaping electoral predictions.

Note: This proposal is challenging a priori, but we are very interested and expect to gain valuable insights, even if on a small scale. We are still evaluating which datasets to include or exclude from our analysis. If you have any feedback on our approach or recommendations for datasets/tools that could enhance our work, we would greatly appreciate it.

2 Teacher Comments

Would'nt it be more useful to concentrate on the decisive states only? Maybe adding Ohio. As data, I would try to get and use county level analysis. These data is easy to obtain in (clickable, etc) graphical form; I do not know if it can be obtained in numerical form. To this you can add the "sure victory" predictions to get an overall prediction. Moreover, this would increment a lot the number of observations (total counties). Use models that deliver probabilities.

References

[1] Nate Silver. 2016 election forecast, 2016. Accessed: 2024-11-16.

[2] The New York Times. Presidential polls forecast, 2016. Accessed: 2024-11-16.

[3] The New York Times. 2024 presidential polls forecast, 2024. Accessed: 2024-11-16.

[4] Alicia Chimeno and Silvia Ferrer. Github Repository. `https://github.com/aliciachimeno/us-2024-election-prediction`, 2024.