SCHOOL OF INFORMATION
UNIVERSITY OF MICHIGAN

America's Next Top Model: Demystifying Two Methods for Election Prediction

Bella Karduck, Haley Johnson, Rohit Maramraju, Philip Menchaca

Project Statement

If knowledge is power then when it comes to election predictions, the public is in the dark. Media reports are filled with opinion polling data and pundits expound on which candidate will win, but how are these predictions made?

We test two methods of election prediction—a classical statistical approach and a machine learning method—and make them understandable to a general audience. We aim to accurately predict the 2024 election and identify key issues that will determine election results.

Background

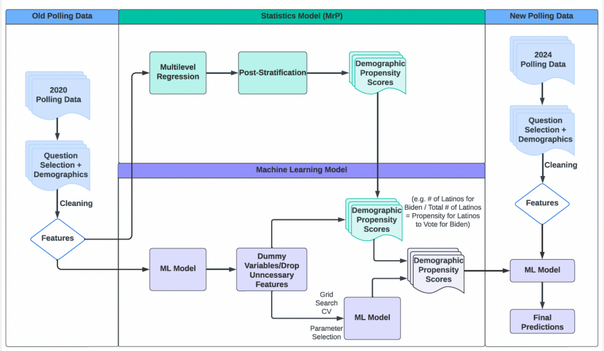
In an era where campaigns invest considerable financial resources in polling to gauge public opinion **the transparency and methodologies behind data aggregation remain shrouded in mystery.** Our project confronts this challenge head-on by offering a transparent deep dive of polling practices and election prediction models.

We have designed two predictive models, both open-source, to serve as practical examples and educational tools. These models invite continuous iteration and improvement from the community, fostering a collaborative and transparent approach to understanding public opinion.

Methodology

Our research aims to model public opinion and election outcomes by leveraging both classical statistical techniques and advanced machine learning algorithms. Recognized as the current state-of-the-art, **Multi-Level Regression and Post-Stratification (MrP) serves as our foundational model**.

In parallel to refining our MrP prototype, we augment this classical model with machine learning (ML). **Recognizing the potential limitations of classical approaches in capturing the complexity of voting patterns, we use machine learning to unveil deeper insights**.

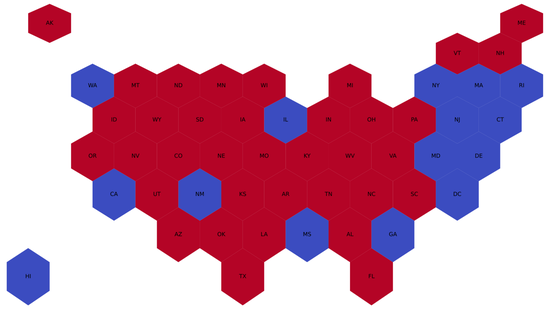


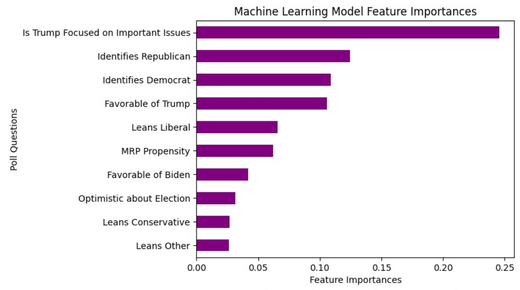
Who will win the 2024 election?

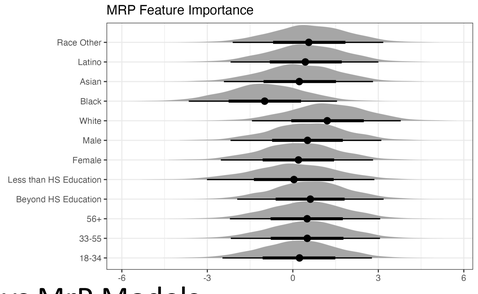
Donald Trump will win the 2024 Election

Predicted 2024
Electoral College Results

339-199

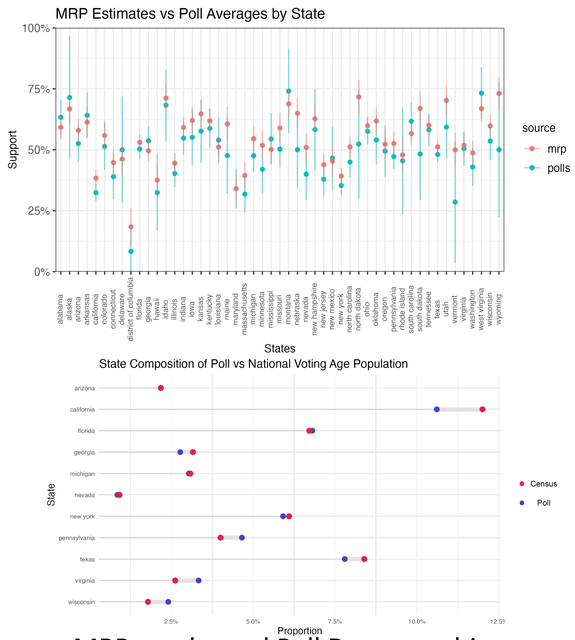




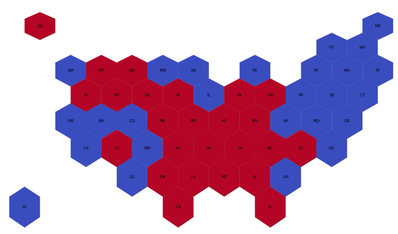


Feature Importance ML vs MrP Models

How well did our model do with the data that we had?

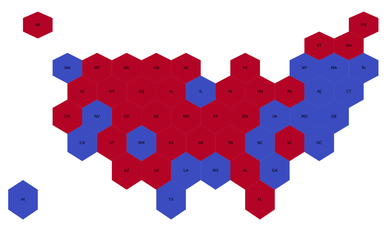


MRP results and Poll Demographics



Actual
2020
Results

2020 Electoral College (Actual): 306-232
2020 Electoral College (Model): 279-259



2020
Model
Results

2020 Hexbin Actual vs Model Data

Insights and Conclusion

Our project makes three main contributions:

1. **We compare classical statistics and machine learning** to estimate opinion. We find that **combining these two approaches outperforms current state-of-the-art methods**.
2. We apply our method to polling data from 2024 to forecast the outcome of the election.
3. We recommend **changes that could improve modeling techniques and better capture voter opinions.**

Major Takeaways:

1. The **availability of high-quality survey data is the main barrier to modeling public opinion.** While more sophisticated modeling techniques improved our accuracy, the biggest performance gains came from increasing the amount of data we were predicting on.
2. **Common modeling methods cannot account for individuals who are not well-categorized by common demographic schemes** (e.g. MENA demographic).
3. We hypothesize that **the machine learning model outperformed classical statistics because it was able to learn more complex interactions** between demographics, ideology, and partisanship.

Further Directions

- To enhance our models, we need to seek out and incorporate more sophisticated polls that provide disaggregated, anonymized voter data
- We also need polls which accurately weight demographics to capture the diversity of the voting population.
- Handling data for 3rd parties and from non-voters is a complex challenge to handle. We need to develop a way to accurately model the influence of these potential "spoilers".
- The polling industry is ripe for disruption by using innovative methods like data science and AI/ML.

Sources

1. Monmouth University Polling Institute. 2021. "Monmouth University National Poll, Number 216." UNC Dataverse. https://doi.org/10.15139/S3/36MVPN
2. Reuters. 2024. "Reuters/Ipsos Issues Survey - January 2024." https://doi.org/10.25940/ROPER-31120717
3. U.S. Census Bureau. American Community Survey 5-Year Data (2009-2022). Retrieved from https://www.census.gov/data/developers/data-sets/acs-5year.2020.html#list-tab-1806015614
4. Schaffner, Brian; Ansolabehere, Stephen; Luks, Sam, 2021, "Cooperative Election Study Common Content, 2020", https://doi.org/10.7910/DVN/E9N6PH
5. COMETrends. November 2020 Election Survey Data. Retrieved from https://cometrends.utdallas.edu/data-and-questionnaires/
6. Hill, Andrew. "U.S. States Hexgrid." CARTO. Accessed April 7, 2024. https://team.carto.com/u/andrew/tables/andrew.us_states_hexgrid/public/map