America's Next Top Model

Demystifying Two Methods for Election Prediction

Haley Johnson, Bella Karduck, Rohit Maramraju, Philip Menchaca

Opening the Black Box of Political Polling

The political consulting industry generates billions of dollars in business. The 2012 election saw approximately $3.6 billion dollars – more than half of the total amount spent on races—directed to consulting firms (Sheingate 2016, 1). Political polling has been a foundational part of these services, and has shaped the public's understanding of elections, since George Gallup began popularizing political polling in the 1930s and Emil Hurja used it to advise FDR's 1932 campaign (Sheingate 2016, 92). Today, in addition to contributing to campaign strategies, political polling fuels the "horse race" media coverage of elections—yet the methods used to produce election predictions receive limited coverage and are not generally understood. Our project has two goals. First, we shed light on the methods of political prediction by developing and documenting in an open way. Secondly, we develop a novel approach for election prediction in an attempt to improve on the current gold-standard statistical approach.

We use 2020 election data to develop our model, which combines a standard multilevel regression and poststratification (MRP) approach with machine learning, and apply it to a January 2024 Reuters poll to generate a state-by-state prediction for the 2024 presidential election. Our model was able to accurately predict 35 out of 50 states plus the District of Columbia in the 2020 election. Based on an evaluation of win margins by state, the classical MRP model and our novel machine learning approach performed very similarly. Our model predicts a win for Donald Trump in the 2024 election, although these results should be read with some skepticism as detailed in the limitations section.

A companion website to this project (https://www.pollrbear.com/) provides an accessible overview of polling methods as well as a quick review of the U.S. election process to inform a general audience. Additionally, our code is documented on GitHub (https://github.com/pmench/mrp-team) for those who want a more technical view of our work.

Methods

Predictive Models: Combining MRP and Machine Learning for Prediction

Political polling faces the fundamental problem of needing to make post-sampling adjustments to correct for bias in the polling data that is used to build predictions and model public opinion. The current foundational approach to making these adjustments, which forms the basis of our model, is multilevel regression with poststratification, which was introduced by Gelman and Little in 1997 (Ornstein 2023, 101; Gelman, Little, and Witter 1997). In brief, MRP first "fits a hierarchical regression model" to polling data, pooling survey respondents by, for example, demographic groups and then higher-level geographic groups. Then, it adjusts the model using population sizes to weight the sample data to be more representative of the true population (Gao et al. 2021, 3). The final model outputs a propensity score for each demographic group that indicates the probability that the group will favor a particular candidate. While initially pioneered in political polling, this method has been applied to social research more broadly (Downes et al., 2018). Poststratification scales these propensities to get expected counts based on the size of the population. For instance, say the MRP model predicts that 80% of individuals in group A will vote for candidate 1. If 10,000 people belong to group A in our population, we would expect to see 8,000 votes for candidate 1.

MRP has commonly been applied to public opinion data because it can adjust for sampling biases in non-representative surveys (e.g. for a specific state) (Kiewiet de Jonge, Langer, and Sinozich, 2018). For instance, say a poll has just a handful of respondents who are older women in Minnesota. Multilevel regression will move this group's propensity score closer to the mean to help correct for sampling bias. This produces a better, less variable estimate than just reporting the mean level of support among that group with no adjustment (Downes et al. 2018; Wang et al., 2015).

Our theory was that these scores could serve as input to a machine-learning model that could learn more complex interactions between features in the survey data. While certain demographic characteristics are predictive of partisan affiliation, voting decisions are complex and often cannot be explained by the voter's identities alone. For instance, Kim and Zilinsky (2022) find that logistic regression and tree-based machine learning models could only correctly predict approximately 60% of two-party vote choices using only demographic features. We aim to combine demographic propensities with richer information about voter's beliefs and preferences, and thus improve upon current state-of-the-art methods for election prediction.

Machine Learning Workflow

We developed our machine learning model using data from 2020 polls, and then applied it to data from the 2024 election cycle. Since results from 2024 are not yet available, prototyping our approach on 2020 allowed us to evaluate how well our model predicted key states and voting groups.

We aggregated multiple polls with varying questions regarding voter's beliefs into our machine learning pipeline to increase the size and representativeness of our dataset. For example, some of the questions we considered included questions on the following: whether then-President Donald Trump was focused on issues relevant to the voter; voter enthusiasm; candidate favorability; top household concerns; and the perceived stamina of a candidate.

We preprocessed data from each individual poll, assigning dummy variables to categorical data, and splitting the data into train and test sets, with the answer to questions such as "Who will you vote for in the election?" as the outcome variable. Note that our outcome variable was who the voter **intended to support at the time the poll was taken.** We have no way of verifying if that individual voted for their stated preferred candidate.

We performed GridSearch Cross Validation on varying parameter combinations, using both Random Forest and Gradient Boosting Classification models to find the best hyperparameter configuration for each model. The purpose of this is to train individualized machine learning models to specifically best fit different sample sizes and demographic variations that are inherently found in unique polls. Additionally, this allows us to accommodate questions that are unique to different polls and which might greatly contribute to the ability to predict a certain set of voter beliefs.

We hoped that the machine learning model would be able to learn complex interactions between voter’s beliefs and their demographic information. While partisan polarization across demographic lines has been well-documented, identity is not destiny. Our machine learning model was trained on questions from our polling dataset and the voter’s estimated MRP propensity. We hypothesized that using the MRP score as an input would give the machine learning model a “hint” about that individual’s preference without being overly deterministic.

After finding the best fit model, we predicted the vote of each individual in our test set for all of our individual polls and their corresponding model. Then, we aggregated the predicted votes by the voter’s respective age, gender, race, and education and developed a new set of propensity scores for all of the polls that were assessed via a classification model. Subsequently, after adjusting for null values where the polls did not contain data regarding a certain demographic, we were able to weight both the original MRP scores along with the machine learning propensity scores to predict the vote breakdown for each state based on its demographic breakdown.

Voter turnout can have significant effects on electoral outcomes. While poststratification typically scales propensities by the raw population counts, we experimented with ways to modify voter turnout. This was important because not all demographic subgroups head to the ballot box in equal proportions. Likewise, we believed incorporating voter turnout into our poststratification step could implicitly account for voter suppression and ease of voting in different states. We multiplied raw population numbers in each state with the turnout rate from the previous presidential election to account for this.

Finally, we calculated the total votes for each candidate in each state. Due to the winner-takes-all nature of the Electoral College, we were easily able to make a national prediction using the state-level projections. We also predicted the Electoral College using only the MRP model’s propensity scores to understand how the machine learning model impacted overall predictions.

Our model for 2024 was trained from polling data with fewer opinion or preference questions, which is another limitation of having little access to data and predicting this far out from the election.

Data

To build our model, we used two kinds of data: demographic data from the U.S. Census Bureau's American Community Surveys (5-Year) and opinion polling data from multiple sources. The Census data was used in the post-stratification step to adjust the regression results and correct for sampling bias in the survey data. For survey data, we used polling reports taken before the 2020 presidential election that asked respondents a variety of questions about their opinions on related political issues, including who they intended to vote for or who they had actually voted for (if they had voted early). In order to properly pool respondents, we needed national polls that reported data at the level of individual responses and included key demographics for each participant, including at a minimum their age, gender, education level, race, geographic location (state/region), and vote preference. These requirements limited the number of polls we could use for building the model. We found three sources that fit our requirements for the 2020 election: Monmouth University opinion polling reports (March, June, July and August 2020); the University of Texas at Dallas COMETrends pre-election survey (October 2020); and the Cooperative Election Study Common Content (2020) dataset. In total, the aggregated polls, before data cleaning, represented responses by 67,000 people (61,000 adults from the CES Common Content dataset, 2,500 adults from COMETrends, and 2,900 from Monmouth surveys). For our 2024 prediction, we used a January 2024 Reuters/Ipsos web-based poll with 4,677 respondents.

Poll	Date	Respondents	Method
Monmouth University Poll Reports	March 2020	851	Telephone
	June 2020	867	Telephone
	July 2020	402	Telephone
	August 2020	868	Telephone
COMETrends, University of Texas at Dallas	October 2020	2,500	Web
Cooperative Election Study Common Content	Nov 8-Dec 14, 2020	4,331	Web
Reuters/Ipsos	January 2024	4,677	Web

Table 1: Description of polls used for training and validation.

Given time constraints, we focused on a core set of features in the data. We removed respondents who indicated a third-party vote preference (or gave no response) and for the MRP model included only the basic demographic categories noted above (most consequently, this reduced the number of respondents in the Cooperative Election Study from 61,000 to 4,331).

The MRP model thus produced baseline propensity scores that were used to train the machine learning model, along with a few more nuanced survey responses, such as religious affiliation, impressions of a candidate's performance, and satisfaction with the overall direction of the country.

As expected, our polling dataset was not nationally representative. For instance, some states, such as Texas, were significantly overrepresented in the survey. Others, like Arizona and Michigan, both key swing states, were sampled approximately in proportion to their share of the national population.

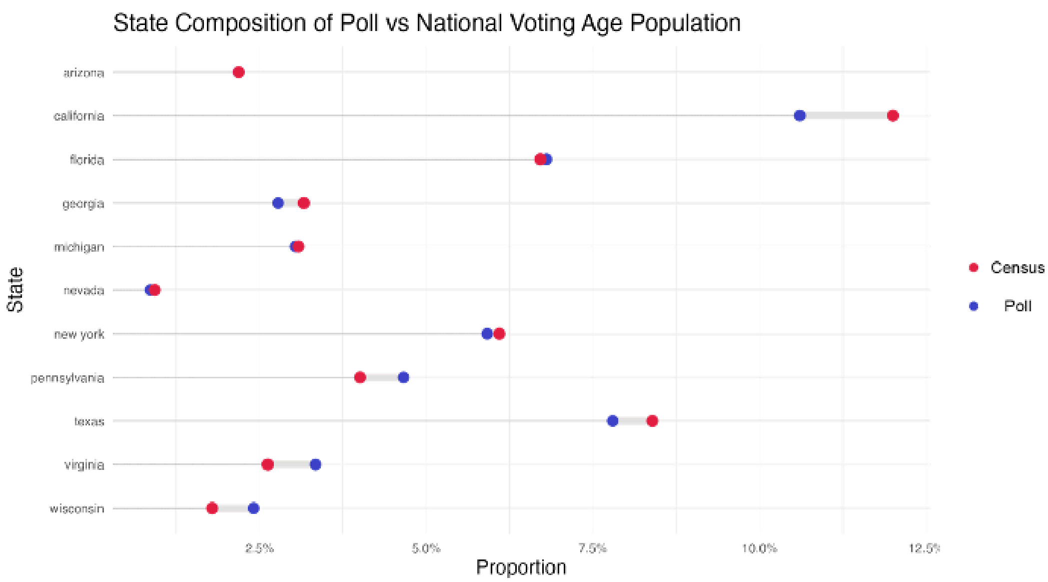


Figure 1: Proportion of U.S. population vs proportion of poll respondents for select states. Census data is represented in red and poll data in blue.

Sampling biases were much more pronounced with respect to demographic categories. White Americans and older Americans are overrepresented in our dataset, while younger Americans and People of Color are underrepresented.

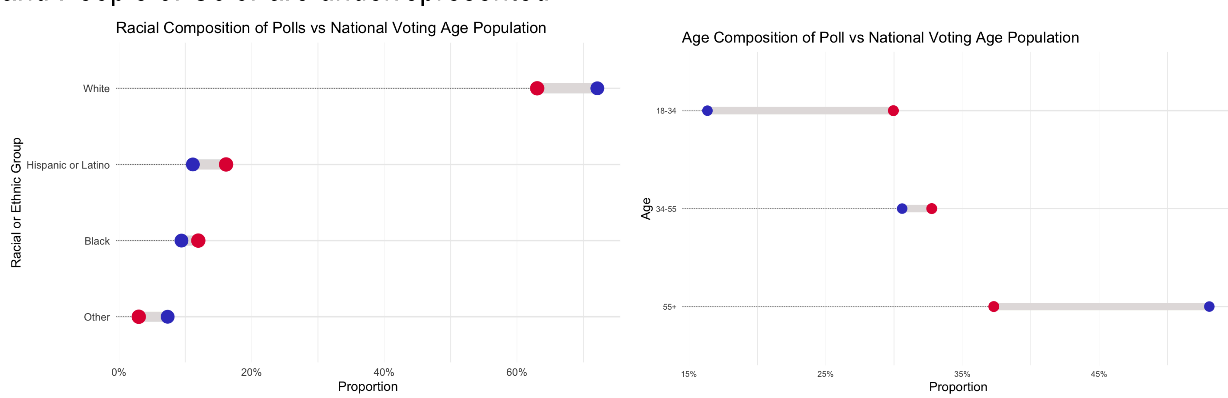


Figure 2: Age and racial composition of poll respondents. Census data is represented in red and poll data in blue.

Feature	Description
Is Trump focused on important issues?	Asks if respondent believed then-President Trump was focused on important issues
Identifies as Republican	Respondent's party affiliation
Identifies as Democrat	Respondent's party affiliation
Favorable of Trump	Respondent's favorability towards sitting President Trump
Leans Liberal	Respondent's ideology affiliation
MrP Propensity	Propensity scores imported from MRP workflow
Favorable of Biden	Respondent's favorability towards candidate Biden
Optimistic about election	Asks if respondent feels optimistic about the upcoming election
Leans Conservative	Respondent's ideology affiliation
Leans Other	Respondent's ideology affiliation

Table 2: Description of features used in machine learning model.

The Prediction Pipeline

We used a two-stage process to generate predictions. First, we predicted the probability of supporting each candidate in a two-party vote scenario based on demographic information using MRP. These propensity scores, alongside additional data, became the input for our machine learning model. The machine learning model made a binary prediction for whether each individual would support "candidate 0" or "candidate 1." We used these binary decisions to calculate new propensity scores, which were applied to Census data in a final post-stratification step.

We developed our approach using polling data from 2020 and used polling data from 2024 to generate our final predictions. Limited polling disaggregated data for this election cycle is currently available; after creating train-test splits that reduced data available for prediction, our final predictions were based on information from just over 1,000 respondents.

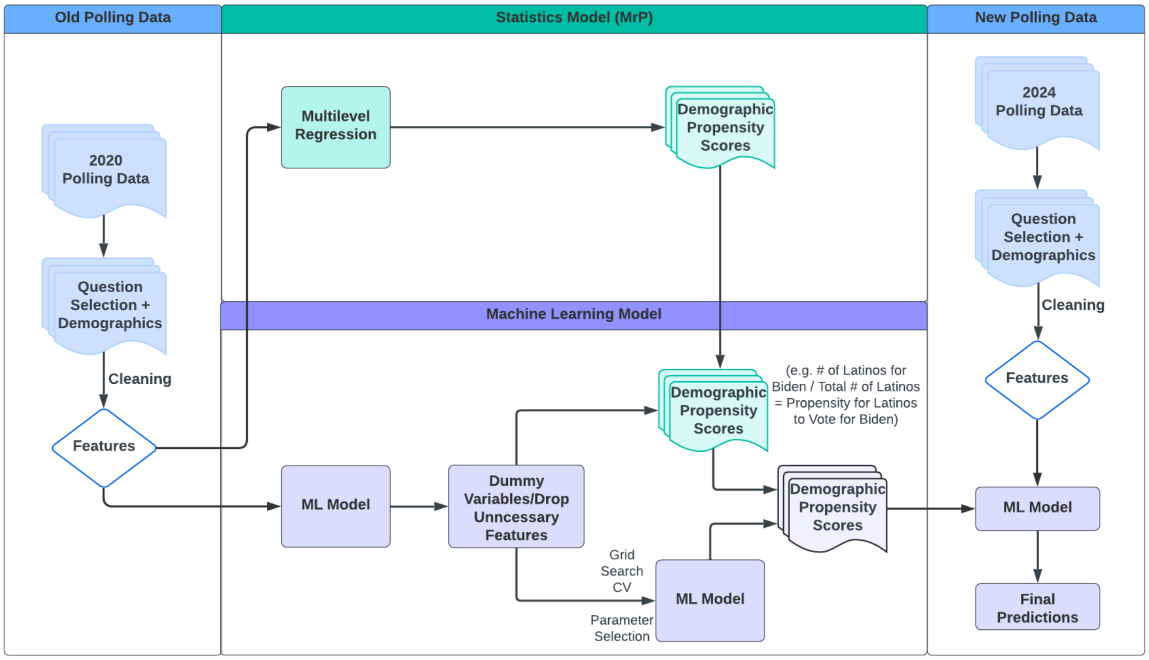


Figure 3: Full workflow for MRP and ML Model.

Results

Our project contributes to the field of opinion estimation by integrating classical statistics with machine learning techniques. By comparing these two methodologies, we discovered that their combination does not surpass the current state-of-the-art methods in accuracy. As expected, the MRP model helped "moderate" extreme estimates towards the center. In Figure 4, the y-axis represents the percentage of the population that we estimate will vote for Donald Trump. The MRP predictions tended to be more moderate and have smaller standard errors than the ones obtained from polling data alone. Our model predicted a Trump win in the 2020 election and a Trump win in the 2024 election. The machine learning enhanced model accurately predicted 35 out of 51 elections (50 states plus the District of Columbia) in the 2020 election, incorrectly predicting outcomes in the following states: Arizona, Colorado, Louisiana, Maine, Michigan, Minnesota, Mississippi, Nevada, New Hampshire, North Carolina, Oregon, Pennsylvania, Texas, Vermont, Virginia, and Wisconsin. The classical MRP model alone accurately predicted 39 out of 51 elections in the 2020 election. It made incorrect predictions for: Arizona, Colorado, Maine, Michigan, Minnesota, Mississippi, Nevada, New Hampshire, Oregon, Pennsylvania, Virginia, and Wisconsin.

	Machine Learning Model	**MRP Model**
Accuracy	68.627%	76.471%
Correct Predictions	35	39

Table 3: Comparison of model accuracy.

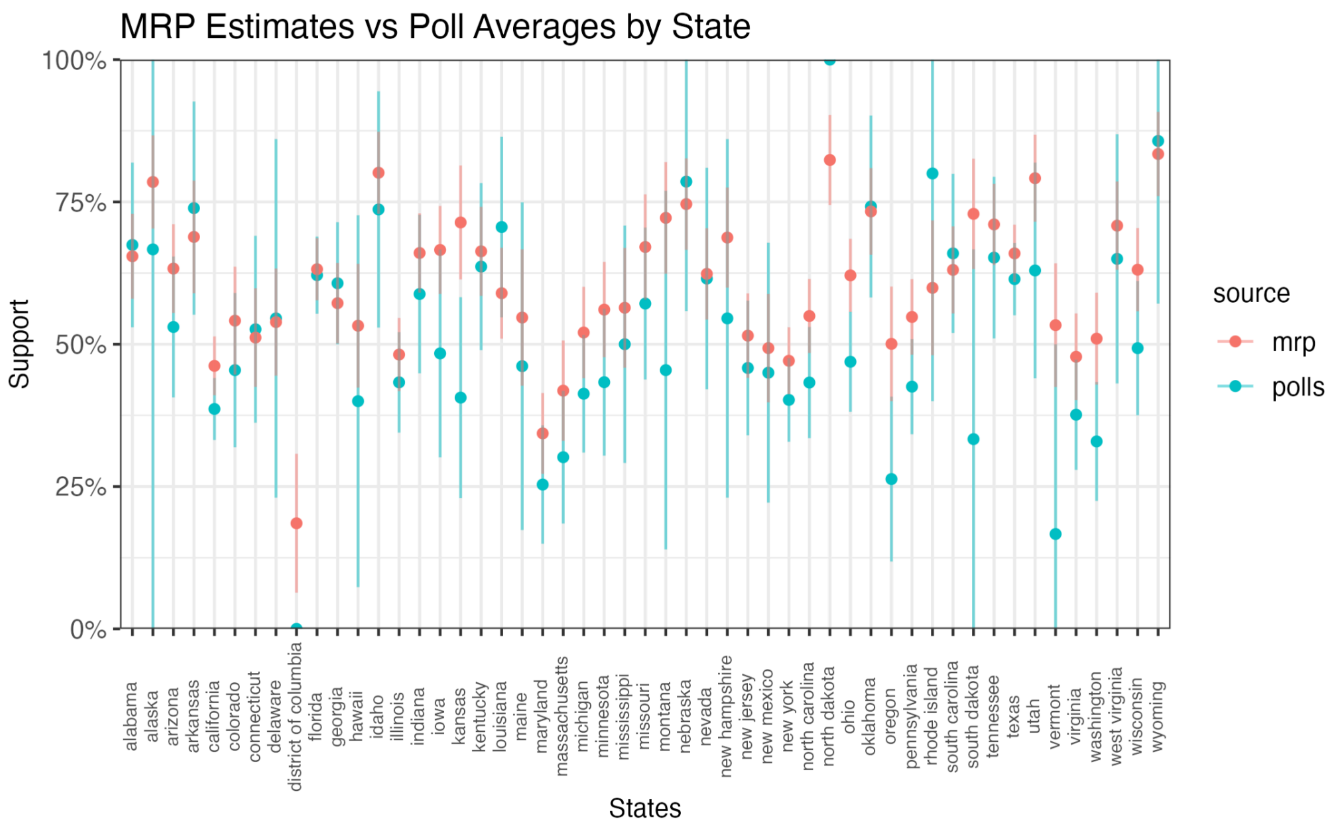


Figure 4: MRP estimates vs polling average.

We anticipated that applying our hybrid approach to the 2024 polling data would improve the accuracy of our forecast but the results show a decrease in accuracy when machine learning is applied. At a more granular level, we found that the ML model and the MRP model performed quite similarly, although the MRP model still outperformed the ML model. We calculated the predicted win margin for each state (% of votes for Biden - % of votes for Trump) and compared it to the actual win margin for each state, using Mean Squared Error as our evaluation metric.

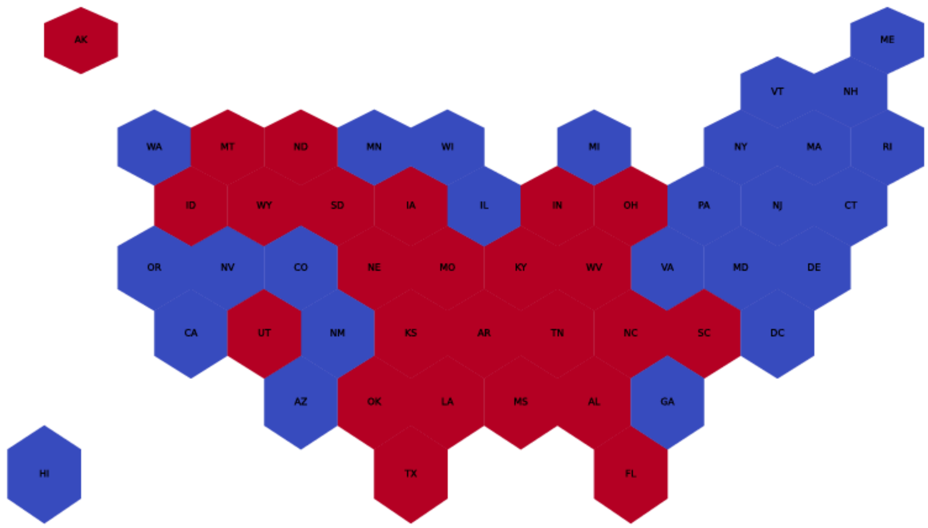
Machine Learning Model MSE	**MRP Model MSE**
0.020997	0.020035

Table 4: Comparison of model performance on predicting margins of victory by state.

We found that the most substantial gains in accuracy stemmed from expanding the quantity of data used in predictions, though (as expected) multilevel regression did help address issues arising from non-representative survey data, enabling us to derive subnational estimates from national polls. However, we identified high-quality survey data availability as a primary obstacle in modeling public opinion effectively. While post-stratification techniques begin to encapsulate the electorate, experimenting with voter turnout modeling noticeably impacts election predictions. With MRP, more nuanced factors such as voter enthusiasm or access to voting remained unmodeled, however. Thus, our strategy to incorporate non-demographic features was achieved through combining machine learning and our propensity scores from MRP. While the accuracy of the machine learning technique was lower than just the MRP, their MSE scores were very close. This is promising, as it seems that the machine learning technique picks up trends that may not be found in the MRP model. The key to determining the true effectiveness of this strategy is further testing with larger polling sets. It's important to note that our ability to model is confined to the demographics provided by census data, leaving certain groups, such as Middle Eastern and North African individuals, inadequately represented by common demographic categorizations. This limitation underscores the need for broader demographic considerations in future modeling efforts.

The deviations we see between the actual and the model electoral colleges are incorrect Democratic wins in Louisiana, Mississippi, North Carolina, and Texas. These are contrasted by incorrect Republican pickups in Arizona, Colorado, Maine, Michigan, Minnesota, Nevada, New Hampshire, Oregon, Pennsylvania, Vermont, Virginia, and Wisconsin. Although some of these states are swing states that routinely swap parties over each election, some of the larger discrepancies in states such as Texas, Louisiana and Mississippi voting Democratic need to be highlighted. These are primarily driven by our model underestimating the propensity of people of color voting Republican in a two-party system.

2020 Actual Election Results



2020 ML Predicted Election Results

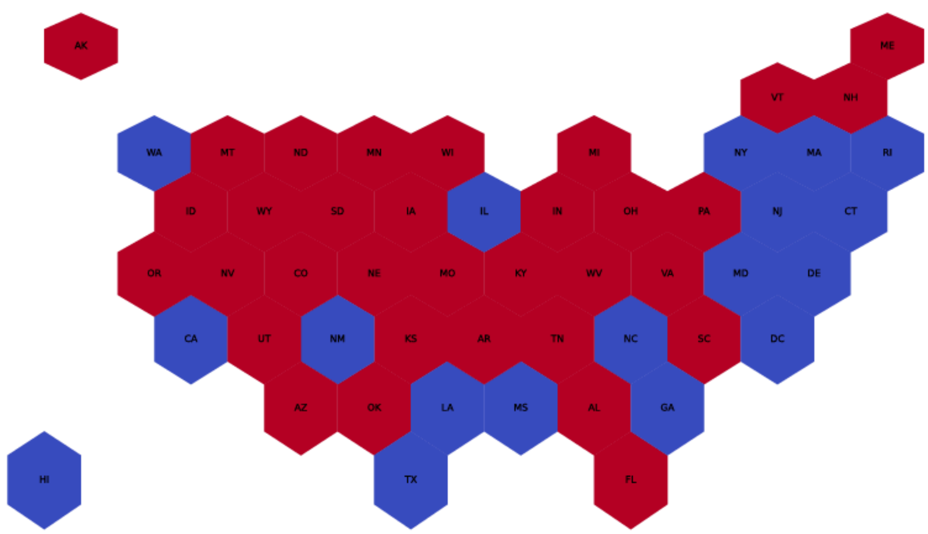


Figure 5: Actual vs. Predicted 2020 Electoral College Results. The actual total for this election had candidate Joe Biden winning 306 -232 over incumbent Donald Trump. Our model predicts a Trump win by 278 to 260.

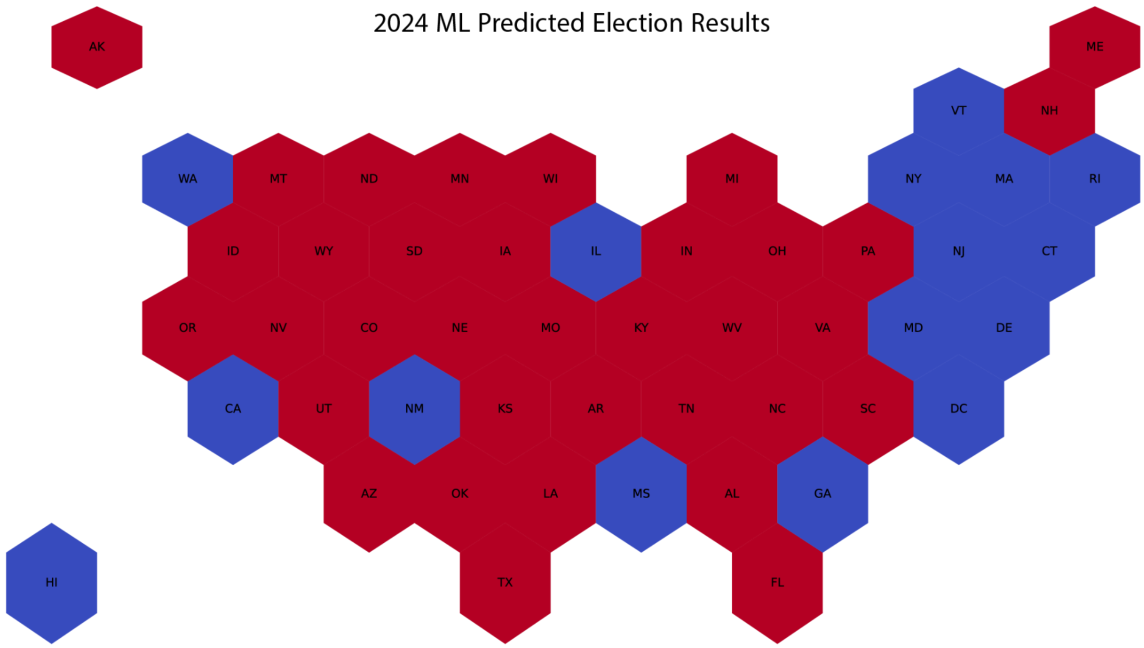


Figure 6: Model prediction of 2024 Electoral College results. Our model predicts a win for candidate Donald Trump by a margin of 339-199.

Predictive Factors

In 2020, we see that features such as candidate favorability, election optimism, and opinions regarding the candidate’s current performance are key indicators of what guides our predictive modeling. Given differences in surveys, our models for 2024 did not include the same type of opinion/preference questions as the 2020 models leading to differences in feature importance. The features that were most important for those models were the voter’s party identification. From this, it can be concluded that any polls fed into the ML dataset need to contain questions assessing features beyond voter demographics.

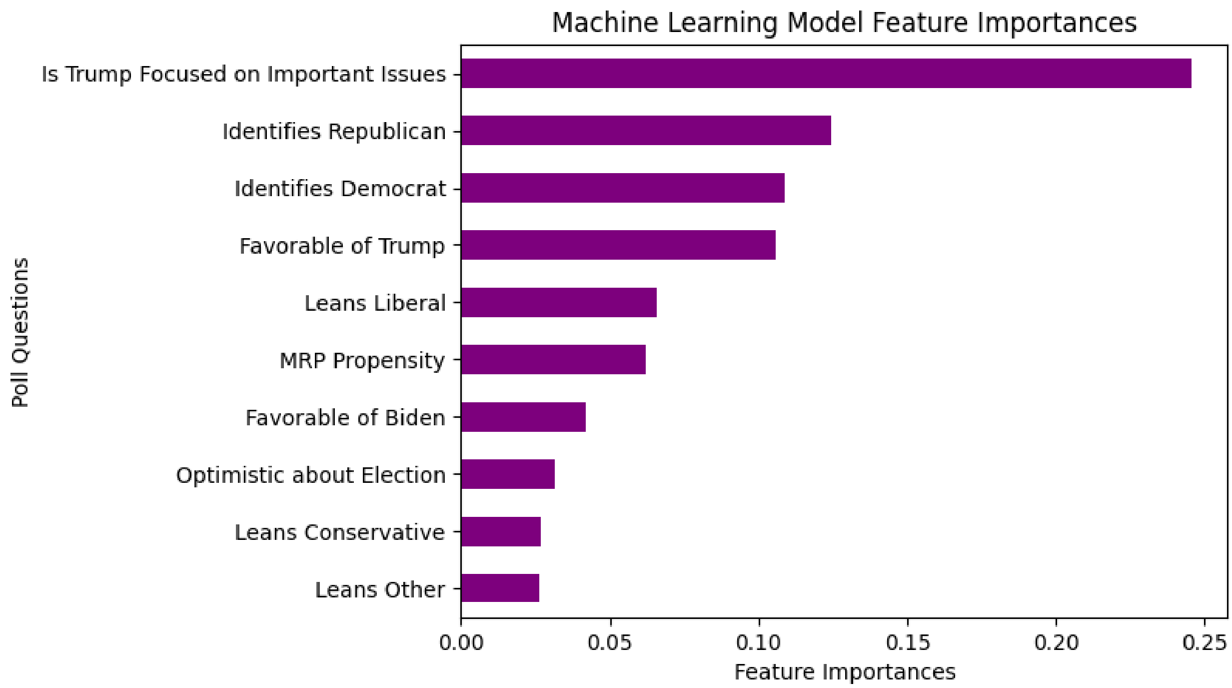


Figure 7: Random Forest Model Features from March 2020 ordered by importance.

The machine learning models trained for 2024 developed an overreliance on party identification, which seemed to skew the predictions away from the direction of current election polling trends. We believe this inability to capture historical trends either demonstrates the need for more nuance within the polling questions used as features, or alternatively an unprecedented shift in the electorate (which is highly unlikely). However, for our 2024 predictions, it is essential to note the challenge of validating the predictive models for an election that is over 6 months away. There is a lack of recent individual poll data that is accessible to the general public and the current political climate continues to shift in unprecedented ways.

Discussion and Limitations

Our results demonstrate that machine learning could be a useful tool for enhancing the accuracy of survey-based political prediction models but that additional work is needed to demonstrate that it can outperform an MRP model alone. Given that both models performed similarly based on a MSE evaluation of their predictive ability for state-by-state win margins, additional investigation is warranted. The lower accuracy of the ML model compared to the classical MRP model could be explained both by the limited number of additional features included in our ML model and by the closeness of results in some states in the 2020 election. Future work that compares the performance of a more complex MRP model to the performance of a more complex MRP-ML pipeline would help provide additional insight into whether an ML-enhanced model could capture more nuanced voter preferences not accounted for by the MRP's method of creating hierarchical and demographic-enhanced groupings.

Our work also highlights a variety of limitations to political prediction models. In addition to the previously mentioned difficulty of obtaining suitable data, a variety of assumptions must be made about the data that can introduce errors. At the most basic level, polls generally assume that respondents are honest in their responses (or that dishonesty is rare enough that a sufficiently large survey will compensate for it). However, it is possible that polls are better at capturing voter sentiment rather than voter intention; just because a voter does not like a candidate does not mean they will not vote for the candidate. Additionally, political polling models that aggregate multiple polls must assume independence among respondents. While this would appear to be a generally valid assumption, it is possible that certain people are more likely to be surveyed and respond to surveys. Techniques such as MRP can help compensate for unrepresentative data at group levels but cannot adjust for the possibility that some significant number of people are represented across multiple polls. We discuss a few more limitations specific to our project in the following sections.

Demographic Categories

To aggregate across multiple polls and apply the final poststratification step, we needed to standardize the demographic we used. This typically meant reverting to the lowest-level of granularity found in our datasets. For instance, Monmouth Polls reports age as one of three categories: 18-34, 35-54, and 55+. Although the Comet and Harvard datasets included the respondent's raw age, we needed to discretize it into the same bins Monmouth uses.

Similarly, there are known issues with the way the U.S. Census Bureau categorizes race. Over 3.5 million Americans of Middle Eastern and North African (MENA) descent are labeled as "white" by the census, even though they are culturally and ethnically distinct (Zraick et al., 2024). The Biden Administration is currently reviewing a proposal to add MENA as a distinct racial category to the census (Zraick et al., 2024). Harvard's poll included detailed information about the respondents' ethnicity that could have been used to construct a MENA racial category, but we could not use this information because we needed to match the census's categories in order to perform post-stratification.

Subnational Predictions

There were some states the model consistently struggled to predict. In particular, multiple iterations of our model, including the final one, predicted that northeastern states such as Maine and Vermont would vote for Donald Trump and that southern states such as Louisiana, Mississippi, and Texas would vote for Joe Biden.

We believe this is because our model struggled to understand voting behavior in these regions that run contrary to national trends. Our propensity scores from the machine learning models are not computed on a state basis, due to the lack of thorough data encompassing even the current demographics. Thus, the regional factor that contributes to differences in voting patterns is lost when the machine learning and MRP propensity scores are mapped to one another. In particular, we hypothesize that our model could not capture the preferences of highly educated white voters in the northeast and conservative Black and Hispanic voters in the south. Future implementations of this method should consider a way to weight the machine learning propensity scores based on state or region, requiring the aggregation of larger polling samples. Alternatively, voter suppression in the south may also contribute to these errors; in other words, the predictions for states like Louisiana, Mississippi, and Texas reflect what electoral outcomes might be if all racial groups in those states had comparable turnout rates. An improved way of understanding and modeling interactions between voter turnout and gender, age, education, race, and region is essential for unlocking the key to balancing MRP propensity scores with machine learning predictions.

The Third-Party Problem

Our model uses a single two-party vote choice scenario. In other words, we ignore the existence of third-party candidates. This approach greatly simplifies our model, but can lead to inaccurate predictions in states where the margin between candidates is close. For example, in the 2016 presidential election, Donald Trump won Michigan by just 11,000 votes (*The New York Times,* 2017). Over 51,000 voters cast ballots for the Green Party—if even just 1/4 of these voters would have supported Clinton, the state would've flipped. Future work may extend our methodology to account for third-party candidates and potential "spoiler effects."

Conclusion

Models are only as good as the data they're trained on. Political polling has many epistemological issues that limit our ability to make accurate predictions—people don't pick up the phone, opinions change, an "October surprise" can unexpectedly alter the electoral landscape. It's unclear to what extent polls can actually capture the phenomenon we're trying to model. Throughout our project we experimented with several methods to improve the accuracy of our predictions, but repeatedly found that the biggest performance gains came from simply aggregating more high-quality polling data. Given the strategic and economic value of political data, a key limitation in terms of developing an open-source polling method is that individual-level polling data is quite inaccessible and inconsistent. Campaigns, political consultants, and polling firms tend to limit the information they publish concerning their sources and methods. And even with additional data, there is still some irreducible error that represents the stochastic nature of politics.

We hope that our project will shed light on the power and limitations of political predictions. It's ultimately a reminder of why each vote is crucial; elections are indisputably uncertain and variable, regardless of what the polls say. While some pundits say we should ignore polls, they are still the primary way campaigns and the media understand public opinion, and it is unclear whether a better alternative exists (Creel 2023; Pfeiffer 2020; Rubin 2023). Crucially, though, polls do not simply capture something about public opinion—they also have the power to influence public opinion. Thus, it is imperative that we open the black box, understand what we can actually say about election predictions, and increase public understanding of the limitations and business of political predictions.

Bibliography

Bureau, US Census. 2020. "American Community Survey 5-Year Data (2009-2022)." https://www.census.gov/data/developers/data-sets/acs-5year.html.

Clarke, Harold, and Marianne Stewart. 2020. "COMETrends November 2020 Election Survey Data." https://cometrends.utdallas.edu/data-and-questionnaires/.

Creel, Nicholas. 2023. "Ignore the Polls and Pundits—Biden Will Win Reelection." *Newsweek*, November 16, 2023, sec. Opinion. https://www.newsweek.com/ignore-polls-punditsbiden-will-win-reelection-opinion-1844144.

Downes, Marnie, Lyle C Gurrin, Dallas R English, Jane Pirkis, Dianne Currier, Matthew J Spittal, and John B Carlin. 2018. "Multilevel Regression and Poststratification: A Modeling Approach to Estimating Population Quantities from Highly Selected Survey Samples." American Journal of Epidemiology 187 (8): 1780–90.

Gao, Yuxiang, Lauren Kennedy, Daniel Simpson, and Andrew Gelman. 2021. "Improving Multilevel Regression and Poststratification with Structured Priors." *Bayesian Analysis* 16 (3): 719–44. https://doi.org/10.1214/20-ba1223.

Gelman, Andrew, Thomas C Little, and Morgan Stanley Dean Witter. 1997. "Poststratification Into Many Categories Using Hierarchical Logistic Regression," September.

Institute, Monmouth University Polling. 2021a. "Monmouth University National Poll, Number 213." UNC Dataverse. https://doi.org/10.15139/S3/6TEXFU.

———. 2021b. "Monmouth University National Poll, Number 218." UNC Dataverse. https://doi.org/10.15139/S3/GCQAAI.

———. 2021c. "Monmouth University Georgia Poll, Number 220." UNC Dataverse. https://doi.org/10.15139/S3/QPC3PL.

———. 2021d. "Monmouth University National Poll, Number 222." UNC Dataverse. https://doi.org/10.15139/S3/IVDJPS.

Kim, Seo-young Silvia, and Jan Zilinsky. 2022. "Division Does Not Imply Predictability: Demographics Continue to Reveal Little About Voting and Partisanship." Political Behavior (46): 67-87. https://doi.org/10.1007/s11109-022-09816-z.

Kiewiet de Jonge, Chad P, Gary Langer, and Sofi Sinozich. 2018. "Predicting State Presidential Election Results Using National Tracking Polls and Multilevel Regression with Poststratification (MRP)." Public Opinion Quarterly 82 (3): 419–46.

Ornstein, Joseph T. 2023. "Getting the Most Out of Surveys: Multilevel Regression and Poststratification." In *Causality in Policy Studies: A Pluralist Toolbox*, edited by Alessia Damonte and Fedra Negri, 99–122. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-12982-7_5.

Pfeiffer, Dan. 2020. "Why You Should Ignore the Polls." November 4, 2020. https://www.messageboxnews.com/p/why-you-should-ignore-the-polls.

Reuters. 2024. "Reuters/Ipsos Large Sample Survey 1: January 2024." https://doi.org/10.25940/ROPER-31120717.

Rubin, Jennifer. 2023. "Opinion | I Don't Write about Polls. You Shouldn't Bother with Them, Either." *Washington Post*, September 8, 2023. https://www.washingtonpost.com/opinions/2023/09/10/pollings-unrealistic-coverage/.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rogers, and Megan Schouweiler. 2024. IPUMS USA: Version 15.0 [American Community Survey 5-Year]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V15.0.

Schaffner, Brian, Stephen Ansolabehere, and Sam Luks. 2021. "Cooperative Election Study Common Content, 2020." Harvard Dataverse. https://doi.org/10.7910/DVN/E9N6PH.

Sheingate, Adam D. 2016. *Building a Business of Politics: The Rise of Political Consulting and the Transformation of American Democracy*. Oxford University Press.

The New York Times. 2017. "Michigan Election Results 2016," August 1, 2017, sec. U.S. https://www.nytimes.com/elections/2016/results/michigan.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. "Forecasting Elections with Non-Representative Polls." International Journal of Forecasting 31 (3): 980–91.

Zraick, Karen, Allison McCann, Sarah, Almukhtar, Yuliya, Parshina-Kottas, and Robert Gebeloff. 2024. "No Box to Check: When the Census Doesn't Reflect You." *The New York Times.* https://www.nytimes.com/interactive/2024/02/25/us/census-race-ethnicity-middle-east-north-africa.html

Statement of Work

Haley Johnson was responsible for the bulk of the MRP model development. Bella Karduck was responsible for the bulk of the ML model development. Rohit Maramraju was responsible for various deliverables on the project as well as the final poster presentation. Philip Menchaca was responsible for contributions to the MRP model as well as the creation of the website (www.pollrbear.com). All group members collected, cleaned and aggregated polling data.

Philip was our wordsmith. Rohit was our fearless PM. Bella was our senior machine learning engineer. Haley was our culture hire. Darjeeling Cat (see appendix) was our vibe checker.

Appendix

Our team mascot: Ms. Darjeeling Cat.