The Neptune Project

2024 Swing State Predictions

Executive Summary

This report outlines the approach, methodology, and results for a machine learning project forecasting voting margins in key swing states for the 2024 U.S. Presidential Election. Using aggregate polling data, historical polling trends, and actual election results, we modeled potential voting margins in Arizona, Georgia, Michigan, Nevada, North Carolina, Pennsylvania, and Wisconsin. Data from the 538 polling aggregator and historical results from the Federal Election Commission (2000-2020) were used to train and evaluate each model.

Results and Variability

Our analysis predicts a narrow electoral college victory for Vice President Kamala Harris in the 2024 United States Presidential Election. We predict that Democrats will win this election on the back of the traditional "blue wall" states of Michigan, Wisconsin, and Pennsylvania.

Our electoral vote prediction (276-262 in favor of Harris) is based both on our internal model for the seven swing states and the current polling consensus for non-swing states. The table below shows the projected 2024 vote margin for Democrats in the selected swing states and the mean absolute error (MAE) associated with the prediction.

State	Predicted Vote Margin (+D)	Mean Absolute Error (MAE)
Michigan	+1.73%	2.69
Nevada	+0.75%	2.08
Pennsylvania	+0.31%	2.03
Wisconsin	+0.31%	2.46
Georgia	-1.15%	0.77
North Carolina	-1.23%	0.92
Arizona	-1.49%	1.25

Notably, the MAE is higher than the predicted vote margin in all four states predicted to split for Democrats, while the MAE is lower than the predicted vote margin for the three swing states projected to fall into the Republican camp. This may reflect additional volatility induced in recent historical data for those blue wall states due to President Trump's 2016 win.

What's in a Model?

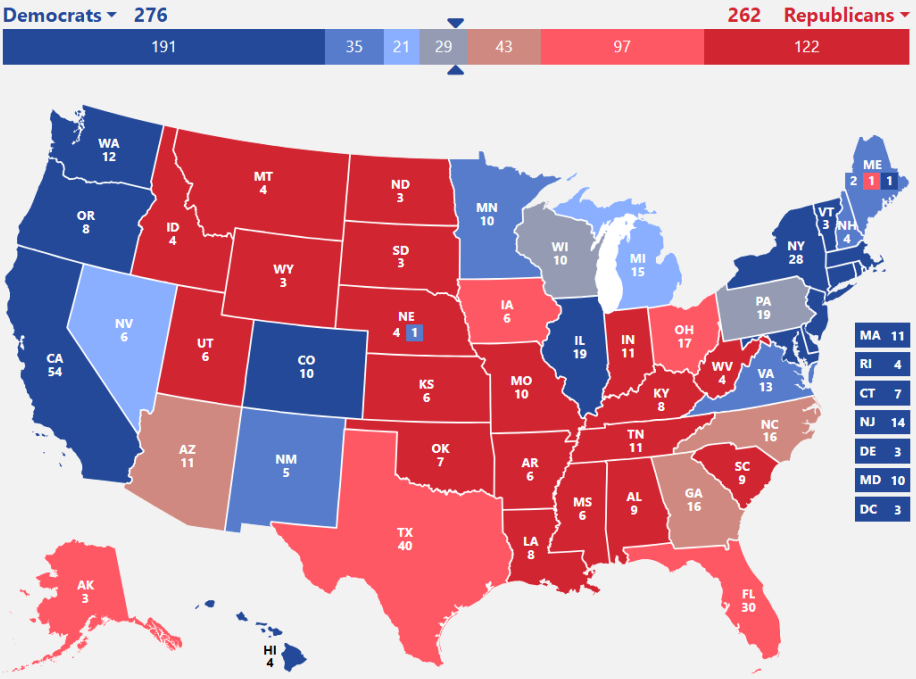
The diversity in model selection highlights state-specific variations in polling behavior and historical voting patterns:

- **RandomForest and KNeighbors** models performed well in states with high variability in voter turnout, such as Arizona and Nevada, where non-linear relationships between polling and voting outcomes may be stronger.
- **ElasticNet** regression emerged as the best model for Georgia, indicating that a mix of linear and regularized approaches can effectively capture the competitive dynamics in traditionally conservative states undergoing demographic shifts.
- In states with more stable voting patterns, such as Michigan and Pennsylvania, **LinearRegression** and **KNeighbors** provided robust results, due to their ability to generalize well from relatively stable historical data.

Further combination of the strengths of individual models through ensemble methods may yield overall lower prediction errors and should be explored. More powerful evidence for the statistically significant strengths and weaknesses of each model must be gathered to continue.

Our Prediction

Our model, combined with current consensus polling for non-swing states, implies the adjacent electoral map for the presidential election on November 5th. It's important to note that even polls in non-swing states yield significant error; a good night for Vice President Harris may flip Republican strongholds such as Texas or Iowa, while a strong showing for President Trump could see Minnesota or Virginia flipping red. However, a mix of historical data and current polling trends indicates this map as the most likely outcome.



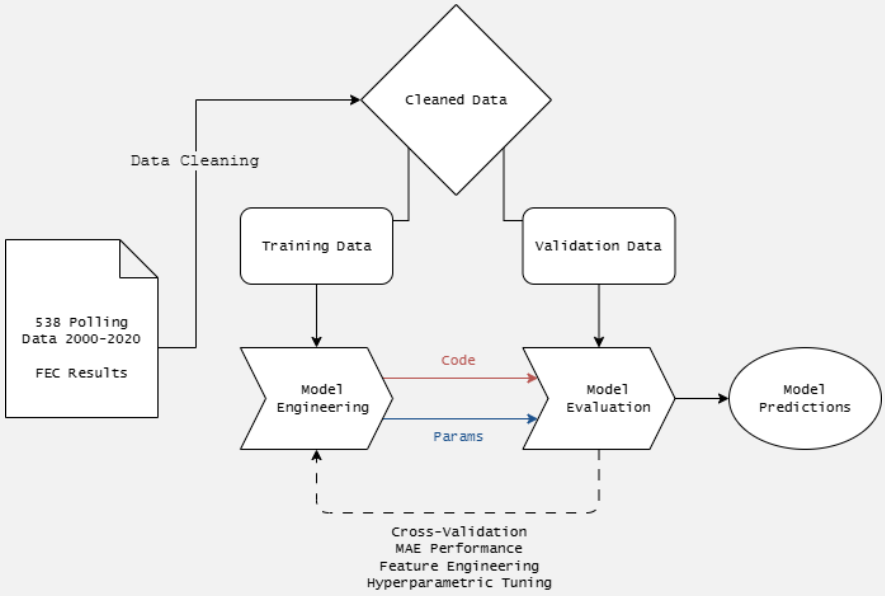
Methodology and Data

Data Sources:

- *Polling Data*: Unweighted aggregate polling data for 2024 and historical polling data from 538, including state-level trends and voter sentiment.
- *Election Results*: Official vote margin data from the FEC for past presidential elections (2000-2020), serving as the ground truth for model training and evaluation.

Model Selection: Eight machine learning models were tested per state to determine the best predictor of vote margin based on past performance and cross-validation accuracy. The models were:

- *LinearRegression*
- *XGBRegressor* (Extreme Gradient Boosting)
- *CatBoostRegressor* (Categorical Boosting)
- *RandomForestRegressor*
- *SVR* (Support Vector Regressor)
- *KNeighborsRegressor*
- *MLPRegressor* (Multi-Layer Perceptron)
- *ElasticNet*



Feature Engineering:

- *Polling Averages*: Polling data was aggregated to calculate rolling averages for each election cycle.
- *Historical Trends*: To capture longer-term political shifts, we included data from the previous five election cycles.
- *Data Cleaning*: To avoid potential confounding variables and increase the accuracy of the fitted models, all third-party votes and candidates were cleansed from the data.

Model Evaluation:

- Each model was trained and tested using cross-validation on historical data from the 2000-2020 election cycles.
- The most accurate model for each state was selected based on mean absolute error (MAE) performance.
- Predicted 2024 margins were obtained from the highest-performing model for each state.