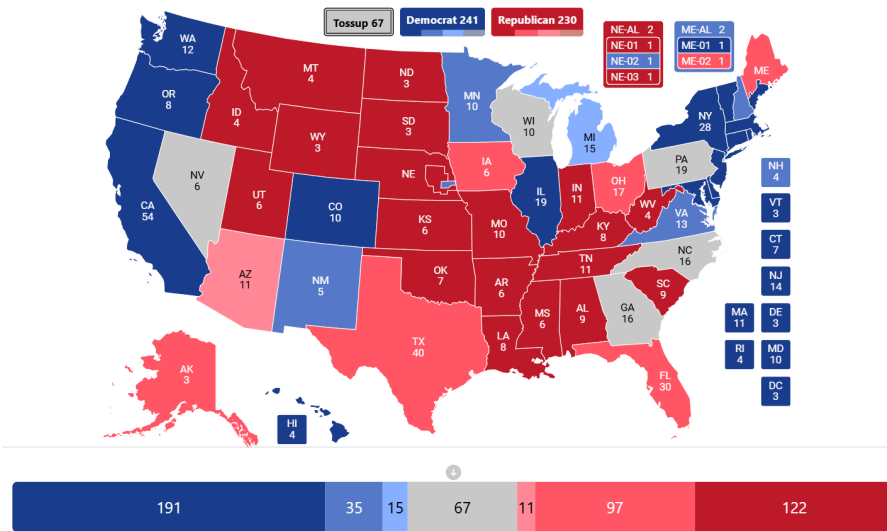


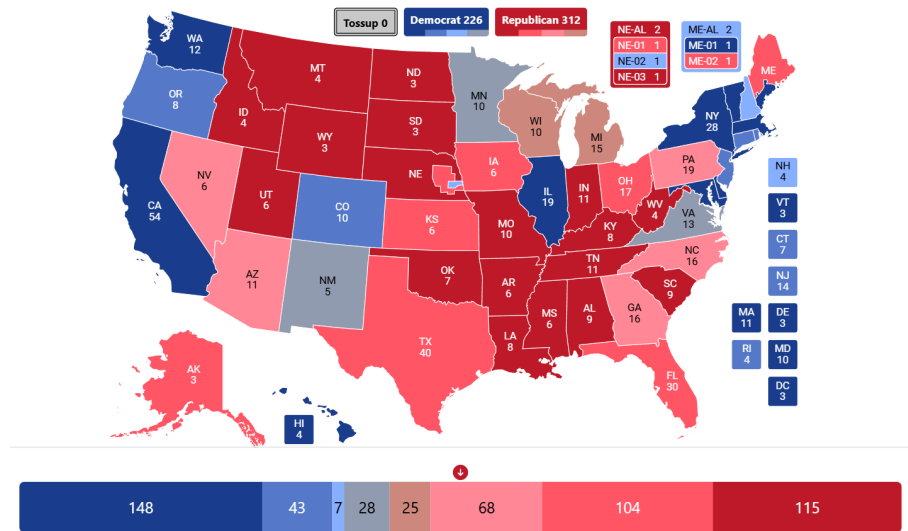
# US Election Retrospective

This note is to reflect on the results of the US election in November 2024, the bias of the polls just prior to the election date, and to evaluate this page's machine learning model's effectiveness at predicting the results using the polling data.

For sake of comparison, the maps below show the final predictions from the highly reputable political analysts at [FiveThirtyEight](#) (part of the ABC News network) on the left, against the machine learning model's prediction on the right.



Final Prediction from FiveThirtyEight



Projected results by the machine learning model (3 Nov 2024)

It should be noted that FiveThirtyEight differentiate between levels of party support by the probability of victory in each race, as calculated by their analysts, whereas the machine learning model differentiates by expected vote margin based on its calculations. FiveThirtyEight specified 5 states as being “Tossups”, meaning that they calculate that neither candidate has over a 60% chance of victory in such a race, however they additionally projected that the most likely and neutral outcome was that the Democrat would win the election with 270 Electoral College votes, with the Republican attaining 268 votes. This is equivalent to the scenario where the Democrat wins WI and PA, and the Republican wins NV, GA and NC.

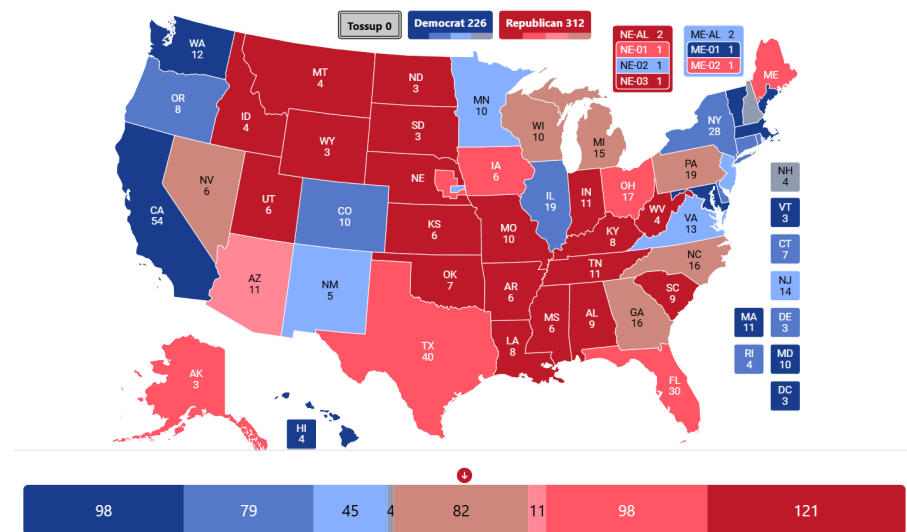
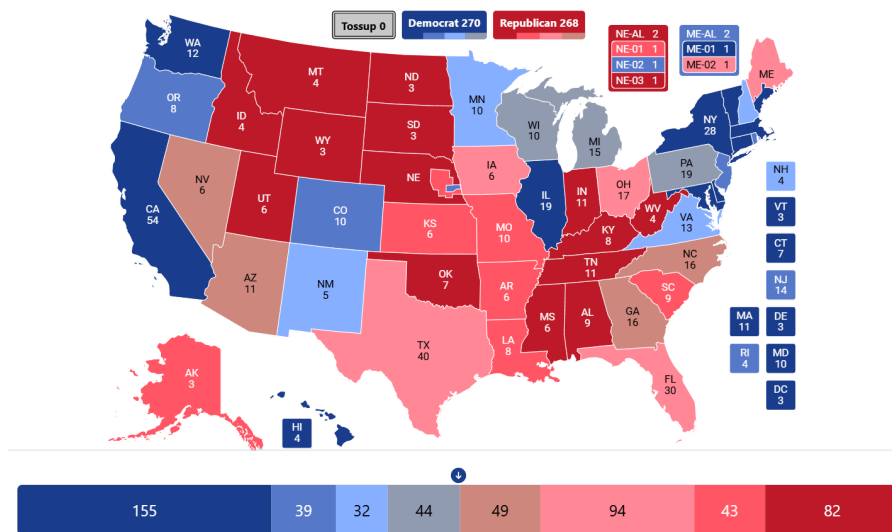
## Polling Data and Final Results

The polling data was collated as at 3 November 2024, two days prior to the election, missing a very small number of polls that were published on 4 November 2024.

In this election cycle there were no polls published for the District of Columbia, which is allocated 3 Electoral College votes, however this reliably goes to the Democrat party in elections by an overwhelming margin.

Several states are still counting votes as of writing this note, however all of the state races have been declared. The results data used in this note were collated on 10 November 2024.

The maps below show the expected election result according to the polling data and the final results.



*Expected results according to the polls (3 Nov 2024)*

*Actual results (10 Nov 2024)*

The margins for differentiating between the categories of party support used in these maps are:

- Tilt: 0%-3.83% lead
- Lean: 3.83%-7.66% lead
- Likely: 7.66%-15.33% lead
- Safe: 15.33%+ lead

These margins are calculated in accordance with the model's calculated error margin of 3.83% against the test data set, and they represent boundaries of 1, 2 and 4 error margins.

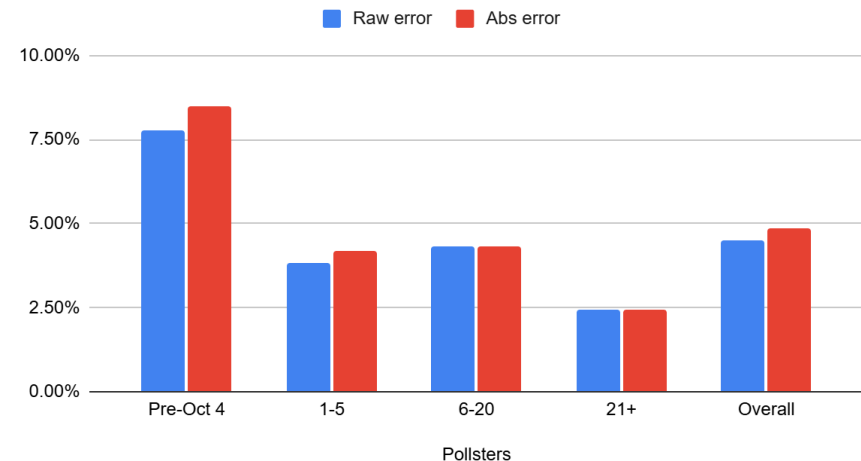
As can be seen above, the polling data overestimated the performance of the Democrat Party significantly. Out of 55 races (excluding the District of Columbia race due to the absence of polling data), the polls miscategorised 3 races (WI, MI and PA) as going to be Democrat wins, when they actually were won by the Republican party. Further, the polls overestimated the margin of Democrat wins in 8 races, and underestimated the margin of Republican wins in 11 races. No Democrat wins were underestimated to a different category and no Republican wins were overestimated.

This gives a total of 22 races out of 55 where the polls overestimated Democrat performance enough to change the category of party support, with 0 races where the polls overestimated Republican performance to the same extent.

Looking at the underlying data in more detail, we find that the polls produced significant errors consistently, mostly in favour of the Democrats. The table below outlines the average raw error of the polls, the average absolute error, and the bias metric obtained by dividing the raw error by the absolute error. A chart displaying the raw and absolute errors by number of pollsters for each race is also below.

Number of Pollsters	Raw error	Absolute error	Bias metric
Pre-Oct 4 polls only	7.75%	8.50%	91.18%
1-5	3.84%	4.18%	91.86%
6-20	4.31%	4.31%	100%
21+	2.46%	2.46%	100%
<b>Overall</b>	<b>4.52%</b>	<b>4.85%</b>	<b>93.33%</b>

Polling errors by number of pollsters



A positive raw error means a pro-Democrat bias in the polls, with a negative raw error meaning a pro-Republican bias. A bias metric close to 0% means neutrality, whereas a bias metric close to 100%/-100% means a consistent bias in favour of the Democrats/Republicans respectively.

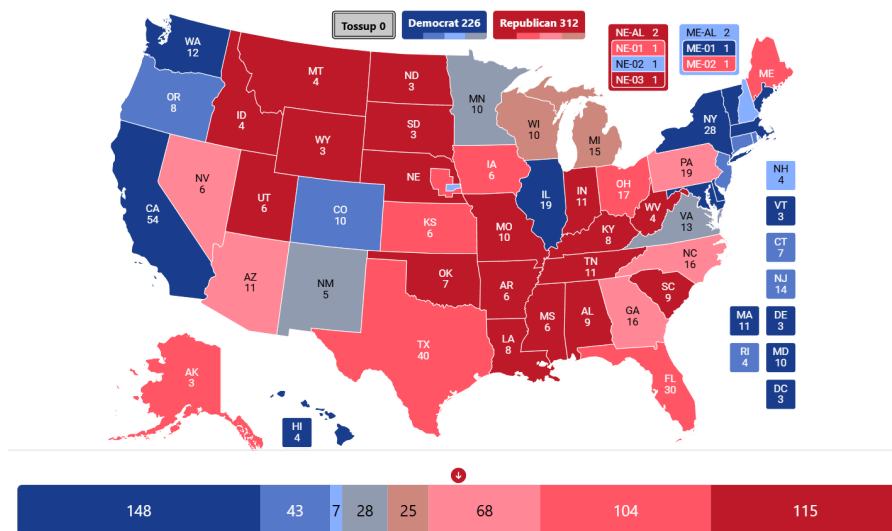
## Model Final Prediction and Evaluated Performance

The machine learning model compares previous election results to the final polling averages in each race, to analyse the biases found in polling en aggregate and then make predictions using current polling data in light of the analysis. For the election cycle in 2024, the model learned using the 2016 and 2020 election cycles' data. From its analysis it estimated the following polling biases:

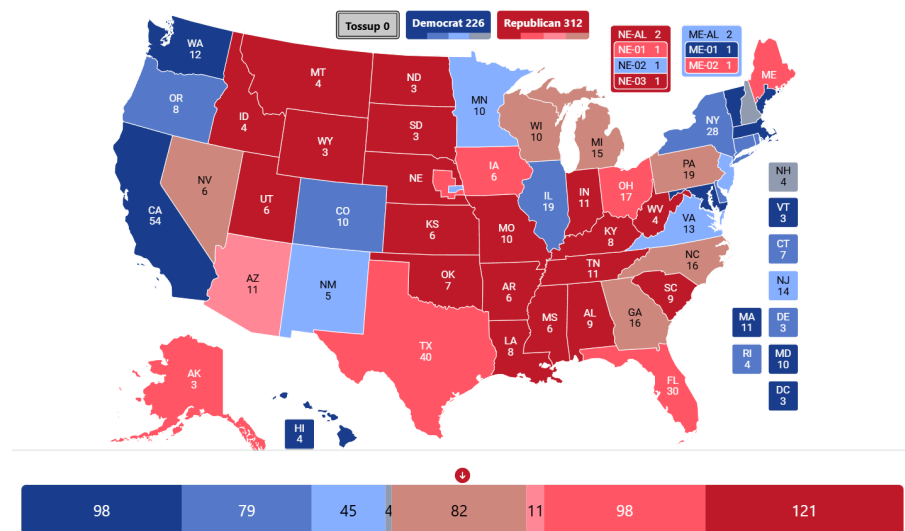
- Inherent bias in favour of the Democrats of 3.80%
- Underestimation of stronghold support equivalent to 17.84% of the published polling margin
- Overestimation of support for other candidates, resulting in a net swing bias in favour of Democrats equivalent to 11.93% of the published polling other support

An example of this analysis being used to make a prediction is given below, which is the calculated prediction of the race in CT (the actual result being D+14.62):

- Polling data average: D+16.00 (D 53%, R 37%, Other 10%)
- Adjustment for underestimation of stronghold support: D+2.85 (17.84% of D+16)
- Adjustment for overestimation of support for other candidates: R+1.19 (11.93% of other support at 10% of the poll)
- Adjustment for inherent bias of polls: R+3.80
- Model prediction: D+13.86



Projected results by the machine learning model (3 Nov 2024)



Actual results (10 Nov 2024)

The map above shows the final prediction made by the model after considering the polls as at 3 November 2024, with the actual results for comparison.

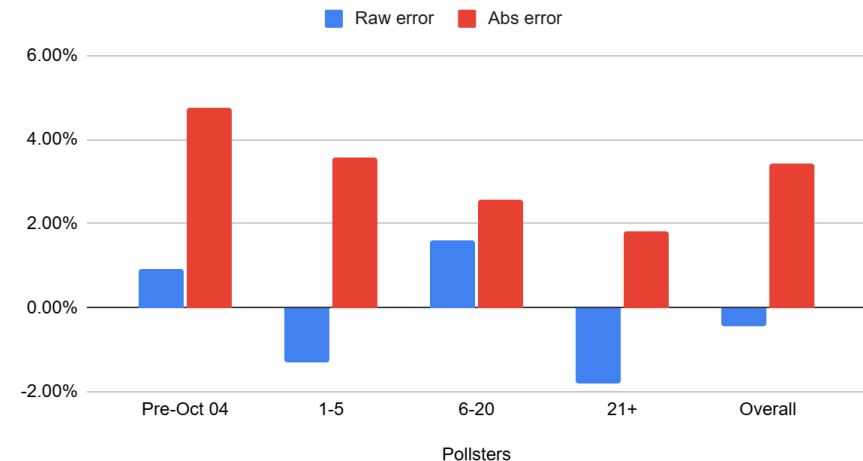
As can be seen, the machine learning model performed much better than the polling data, predicting the outcome of every race correctly. Out of 55 races, the model overestimated the margin of Democrat wins in 5 races and underestimated in 1 Republican win. Additionally, the model underestimated the margin in 3 Democrat wins and overestimated Republican wins in 4 races.

This gives a total of 6 races out of 55 where the model overestimated Democrat performance enough to change the category of party support, with 7 races where the model overestimated Republican performance to the same extent.

Examining the underlying data in the same fashion as done for the polls above, we find that the model produced far less biased predictions, with a relatively small bias in favour of the Republicans. This bias was most accentuated in races where there was a lot of pollster attention, i.e. the swing states. This may suggest that an improved approach would be to segment the learning process so that a separate machine learning model is used where there is a lot of pollster attention, or to incorporate the number of pollsters as an additional feature for the model to consider in its analysis.

Number of Pollsters	Raw error	Absolute error	Bias metric
Pre-Oct 4 polls only	0.91%	4.76%	19.12%
1-5	-1.30%	3.58%	-36.31%
6-20	1.60%	2.58%	62.02%
21+	-1.82%	1.82%	-100%
Overall	-0.45%	3.43%	-13.12%

Model errors by number of pollsters



## Revised Polling Bias Analysis

The machine learning model can now incorporate the data relating to the 2024 election cycle to its training data, to update its analysis of the polling biases. These updates are summarised in the table below.

	Inherent Bias	Stronghold Underestimation	Net Swing Bias from Other Candidates
Pre Election	3.80%	17.84%	11.93%
Post Election	4.19%	15.15%	2.96%

While the estimate of the inherent bias in favour of the Democrats has increased, the estimate of the overestimation of other support bias in favour of Democrats has decreased. Considering that the polls at 3 November 2024 averaged the support of other candidates at 5.73%, the aggregate pro-Democrat bias (combining the inherent bias and the other support bias) was analysed to be approximately 4.48% before the election, which has been reduced to a level of approximately 4.36% in light of the new election data.

Overall, the levels of bias in polling for US elections has been broadly unchanged from 2016 and 2020 levels, with the inherent bias in favour of the Democrats being beyond an error margin.