Predicting the 2024 U.S. Presidential Election

By: Jack Caruso

The 2024 election has been by the polls as one of the closest elections we have ever seen. Almost every swing state (Michigan, Wisconsin, Pennsylvania, North Carolina, Georgia, Arizona, and Nevada) have consistently polled within 2 points within the aggregate, making this election being seen as a toss-up to most people. Because of this, it has led to people on either side to either feel massive anxiety that their candidate is going to lose, or overconfidence that they will easily win this election. This wasn’t always the case for this election cycle though. There was a point that when Joe Biden was still the nominee, it seemed that it was nearly impossible for Donald Trump to lose the election. With historic unfavorables, Biden really showing his age, concerns about inflation and foreign affairs in Russia-Ukraine and the Middle East, Biden was projected to not only lose every swing state, but potentially use non-swing states. After his horrendous debate back in June, there was a poll with him losing in New Jersey and Trump within 10 in New York. If these numbers held, Biden was going to lose in a landslide. Now that he has dropped out and Kamala Harris has become the nominee, they have given themselves a chance to win this election.

In this project, I am going to try to predict the electoral college for the 2024 election. In the folder, I have three main datasets I used for this model, polling data going back 2016, economic data since 2000, and election data since 2008. For the polling data, I wanted my model to see how good the polling was compared to the actual election season in 2016 and 2020 to train it to pick up any trends potentially. Now, something to factor in this is that in the last two elections, Democrats have underperformed the polling. In 2016 they did not break down education demographics accurately leading them to overestimate Clinton but in 2020, theories are more about the pandemic and the fact that Democrats were more isolated due to fear of COVID-19 where Republicans were more out and about, leading to more Democratic voters to respond. Another question I have to ask about the polling data is that there was not any ratings on the 2024 polling data I got from FiveThirtyEight, so there are some internal polling that can be skewing the data a little and also just some bad polling industries. I did want to keep a large sample size so I decided to keep the data in. Lastly, I had to take out Biden’s 2024 polling data due to it skewing the data against him mainly in swing states. I noticed in non-swing states there was very few polls so I decided to just leave his numbers in. This could possibly lead to some of Harris’s numbers in non-swing states to be not as much in her favor but historical election data should make those states Democratic anyway. This would make the 2024 polling data the smallest sample of the three elections I took but there is not a better alternative.

For election data, I wanted the model to understand how states basically vote and also to pick up on any possible trends state-by-state. I wanted again to have a relatively big sample size so I went all the way back to 2008. By doing this, it should stabilize any states that we all know should go red or blue like California, Alabama, etc. I did not want to go back much more than 2008 because by doing so I would not be factoring enough possible demographics changes that likely happened before then.

Lastly, I decided I wanted some economic indicators in my model to have it understand how the economy is doing at the time. I wanted economic factors going back to 2008 so that the model could see the correlations to a strong economy and a candidate winning. I did not go crazy with the economic indicators so I used Inflation Rate, Unemployment Rate, University of Michigan Consumer Sentiment Index, and Wage Growth Rate. This will tell if people have been seeing more money in their jobs, if they have been able to get a job, how expensive things have become, and how confident consumers are. I decided to have each indicator be monthly and take the average for the entire year to give the model and understanding of how people view the economy throughout the entire election year. There is one problem with this though. The model does not know which party to help or punish if the economy is doing well or bad. What I decided to do is tell the model which party was seen as the incumbent party and if the economy was considered a recession. The incumbent party is which party has the president right now, and since Joe Biden is the incumbent, Democrats are the incumbent party currently. The last two recessions we had were in 2008 and 2020 and the incumbent party lost.

Now that all the data has been cleaned and merged I can now do the analysis. I imputed each of the features into the model and ran a Random Forest Regressor to train and test the model. The result was Democrats win 303-235, which is the exact same as they won in 2020! Some might say that could be problematic but this is a reasonable outcome to expect in this election. This is what the model believes will happen. I also put in the projected vote shares in the GitHub folder. A map of the united states

Description automatically generated

It had six of the seven states as tilt Dem or GOP, which makes sense based off the polling. Michigan was the only state with a margin bigger than two (3.41), making it lean Dem. Florida was also within three (2.9) which could be something, but personally I think this will not happen because in 2022 Florida was one of the reddest states that election. Demographics have shifted red a lot the last decade.

Something I noticed overall, was that Democrats actually overperformed the polling. This is likely due to an overall good economy and since they are the incumbent party, it likely helped them. Even though some people may not feel the economy is as good as it is. Inflation has gone down since its peak in 2022 and dropped to below 2.5 going into the election. Unemployment is also around 4% which is very strong. Consumer confidence is not strong which is fair to acknowledge.

Lastly, I want to list some factors that I did not put in my model that could alter the election. First thing is foreign policy. Will Arab voters in Michigan come out to vote like they did in 2020 even with what is happening in the Middle East? There have been polls showing Trump closing his lead over Harris where in 2020 Biden won easily. In the 2024 primary, many voters voted uncommitted due to being unsatisfied with how Biden has handled that region. Another factor is favorability ratings. Harris is relatively even popularity wise and has a lot of excitement from her democratic base that potentially has not been seen since Obama. Trump on the other hand is unpopular like he has always been. JD Vance also is unpopular while Tim Walz is seen as the most popular of the four. We also need to factor in that how much will Roe V Wade being overturned factor into this election. In 2022 and special elections, Democrats have overperformed polling due to women turning out more than usual, leading to a smaller red wave than expected despite Biden being unpopular and inflation peaking in 2022. If this is true, then even my model is underestimating Harris. Is it also possible that polls just underestimate Trump again? A political party has never seen a polling error in their favor three times in a row, but it could happen. If that is true than Trump likely wins close to what he won in 2016. What about down the ballot races like Mark Robinson in North Carolina being down 15-20 points or abortion being on the ballot in several states? Could this help Harris, hard really to tell. Also hard to truly measure how good either campaign has been to the median voter which matters as well.

Overall, my model predicts a Harris victory by the same margin Biden won four years ago. I see factors where Harris does even better than the model thinks if women come out in droves and polling hasn’t factored that in enough, and scenarios Trump still wins and potentially 312 electoral votes if he can get young men to come out which does not happen often. This is an election that there are many scenarios I can personally see, but if you are a democratic voter, this model should give optimism considering where they potentially would have been if they kept Biden in the race.