# Election Forecasting and Elastic-Net Model Comparisons

**Darian S. Martos**
Department of Statistics
Stanford University
Stanford, CA 15213
dtmartos@stanford.edu

## Abstract

Within literature for election forecasting, ensemble methods reign supreme for their ability to capture aggregate predictions for inference in predicting probabilities of party wins in elections. Here, we investigate some of the most frequent models of classification and ensembling to understand performance differences from classical logistic regression. Using logistic regression and various regularization penalty terms, we explore some of the most impactful features incorporated into election forecasts for 2020 and 2022 and understand how varying regularization alters predictions. We find that regularized logistic regression models focus significantly on political factors including polling data and state party alignment, while ensemble models tend to incorporate diverse features including demographic ones. Regularization has no effects in adding new variables to models, it instead shifts the importance of a set of predictors relative to standard logistic regression.

## 1 Introduction

Models in election forecasting have transformed national media for nearly decades, having become a staple within American electoral politics and the election media system. The first possible discussion of election prediction comes all the way from 1948 by Louis H. Bean [1], where models focused on past election results to help predict future results. These simplistic models, with the sole features being past presidential election results, proved successful through the 1940s and 1950s and were instrumental in developing political science concepts such as “bellweather states." Bellweather states, defined as states that where a statewide party win often correlates with a federal party win, became a prominent topic in election forecasts that emerged out of primitive models such as these past-election models from Bean.

The models of today take up more dynamic approaches, with many forecasts taking up standard classification frameworks as well as newer Bayesian approaches. Deep learning has yet to find presence in election forecasts and even tabular data at all [2]. As a result, ensemble methods and especially those incorporating Bayesian methods are most prominent within the space, as it is well established that Bayesian methods work best in spaces where we form no “ a priori" assumptions of the most useful factors that influence an election outcome. [3] Across most forecasting platforms, ensembling of various forms are the standard, with one frequent method being standard ensemble methods such as random forests or boosting. Perhaps the most common form of forecasting is the use of repeated regressions methods that run thousands of forecasts and then form a histogram of the most likely outcomes. This helps create probability distributions that aid in predictive inference. [7]

Preprint. Under review.

## 2 Background, Goals, and Tasks

For this project, we explore the different features that affect success in election forecastind models. While Bayesian methods are effective, we reserve our space of exploration to the frequentist viewpoint that arises in standard logistic regression. This is done primarily to preserve the baseline of logistic regression for classification, but we still provide forecasts from random forests and gradient boosting towards improved prediction outcomes.

As a side goal to this task, we also investigate regularization as a shrinkage method within these forecasts. With logistic regression as our base method, we explore regularization penalties for minimizing the provided loss objective, known as elastic-net penalized loss, defined below:

$$\min_{w,c} \frac{1-\alpha}{2} w^{\intercal} w + \alpha ||w||_1 + \sum_{i=1}^{n} \log\left(\exp\left(-y_i(X_i^{\intercal} w + c)\right) + 1\right)$$

Where $\alpha$ is a general mixing parameter that determines how much of a ridge penalty is applied versus the LASSO penalty. We see that $\alpha = 1$ applies the LASSO regularization penalty, while $\alpha = 0$, gives the ridge regularization penalty. We vary $\alpha$ over even values within the interval $[0, 1]$ as a way to understand further what coefficients get zeroed out or reduced, and work towards understanding how Elastic-Net alters the feature landscape for training. This regularization is especially helpful in scenarios where we have a smaller $n$ than the number of features applied. [5]

As tasks towards this goal, we (1) first “classify" the 2020 US Presidential election results as closely as possible to the real results, and (2) develop a forecast towards the upcoming 2022 US Senate elections. As covered in multiple citations below, especially within models developed in [4][7][8][9]. Much of the methodology covered by Decision Desk HQ is especially applied, and we make heavy use of many of the features (about one-fifth) from Decision Desk's Data Dictionary.

## 3 Methods

### 3.1 Data and Features

For our features, we use a mix of about 17 demographic features and 16 political features for our presidential forecast, and 19 demographic features and 20 political features for our Senate forecast. These features and their sources are summarized below:

- Demographics: state region, state population, GDP (GDP per capita is used in a later forecast), educational breakdown by proportions (% high school educated, % associates, % Bachelors holders, % advanced degree holders), racial categories by proportions (American Indian, Asian, Black, Hispanic/Latino, Pacific Islander, White), median household income. As a somewhat arbitrary choice, the Senate forecast also includes state employment rates and % urban population.

  **Sources** All of these demographic features, other than GDP/per capita and urban percentages are derived from the US Census profile. The latter two aforementioned features are scraped from Wikipedia.

- Political: Both forecast tasks initially incorporated four past presidential election results (2000/2004/2008/2012), state partisan alignment (by a measure known as PVI), incumbency status (whether the current candidate running was already in the Senate seat/presidency), finances between the candidates (total receipts/disbursements for their campaign), and polling averages from five most recent polls. Past Senate performances for most six recent elections (i.e. three most recent elections for both Senate seats) were used for the Senate forecast, while the Presidential forecast used the two most recent Senate elections as predictors. In a later experiment, we removed a few of these past Presidential and Senate results within both forecasts.

  **Sources** Many of these political features are scraped from Wikipedia. Campaign finance information was pulled from Federal Election Commission data and polling averages were either hand-computed, or computed from RealClearPolitics' polling averages. Notably, some states lack polling data in cases where a presidential or Senate race is expected to be noncompetitive, so for these results I used previous election results within the respective races as estimates.

Both forecasts incorporate a different form of training and test split for their respective tasks. For the presidential forecasting task, we use the 2016 Presidential election results and its polling/2010 population counts as training data for the task. The “test set" in this case is the the features for 2020, with the goal of predicting the 2020 Presidential election as closely as possible. As for the Senate forecasting task, we preserve all features as described above as from 2020-2022. Instead of using past Senate races as training (a difficult task, as only one-third of the 100 Senate seats are contested during any singular election), we use states that do not have elections during the 2022 cycle and their 2020 results, as well as noncompetitive states, as training data for the Senate forecast. Noncompetitive states are judged roughly based off whether all or all-but-one of their forecasts over the 2022 Senate Elections Wikipedia page are all “Solid" towards a party. At the time of writing of this report, the competitive states and test set based off this above criteria are: Alaska, Arizona, Colorado, Florida, Georgia, Missouri, Nevada, New Hampshire, North Carolina, Ohio, Pennsylvania, Washington, Wisconsin.

### 3.2 Models

Across the first set of experiments, we used logistic regression, LASSO-regularized logistic regression, ridge-regularized logistic regression, decision trees, random forests, and gradient boosting. The second set of experiments used most of these models, as well as the Elastic-Net regularized logistic models as described above, with $\alpha \in \{0.2, 0.4, 0.6, 0.8\}$. The second set of experiments also included the removal of decision trees due to poor learning from the first set of experiments.

### 3.3 Model Hyperparameters

Beyond the $\alpha$ parameters described above, other hyperparameters used for our models include:

- Standard scaling is applied to every logistic regression and regularized model, including the standard model with no penalty. The SAGA solver was also applied to optimize each logistic model.
- Random state = 83 for all tree-based models, and RFs and GB both have $120$ estimators.
- For boosting, the base score is $0.5$, percentage of features used per tree = $0.3$, learning rate = $0.1$, and $\alpha = 10$ for L1 regularization.

## 4 Experiments

As alluded to above, the first set of experiments use all the features and models as described within the Models section above. For the second set of experiments, we apply similar features but make the following modifications:

- As stated above, removed the decision tree model and added ElasticNet-regularized logistic models.
- We converted the state GDP feature into a GDP per capita predictor, this was done largely to assess whether a more nuanced factor for states would be applied over the models. GDP per capita captures more of a state characteristic than overall GDP for a state, which does not capture as much nuance in population dynamics.
- For the Senate forecast, we removed four older past Senate results and also 2000/2004/2008 presidential election results as features. In the Presidential forecast we removed the past Senate results entirely. We do this to understand other features that may be more related to forecasting, especially in the case of the Senate.

The training and test split and modeling was unchanged through the experiments.

Table 1: First set of experiments for presidential forecasting, most important features in prediction are reported. Coefficient estimates for logistic regression models and feature importances for tree-based models are given.

| Model | 1st Important Feature | 2nd Imp. Feat. | 3rd Imp. Feat. | 4th Imp. Feat. |
|---|---|---|---|---|
| Logistic Regression | State Party Align. $(\hat{\beta}_1) = 1.960474$ | Lead Polling Party $(\hat{\beta}_2) = 0.771840$ | 3rd Recent Sen. Result $(\hat{\beta}_3) = 0.689105$ | Median Age $(\hat{\beta}_4) = 0.553478$ |
| l1-Penalized Log. Reg. | State Party Align. $(\hat{\beta}_1) = 2.422856$ | Lead Polling Party $(\hat{\beta}_2) = 0.655385$ | 3rd Recent Sen. Result $(\hat{\beta}_3) = 0.248870$ | 2012 Pres. Result $(\hat{\beta}_4) = 0.217693$ |
| l2-Penalized Log. Reg. | State Party Align. $(\hat{\beta}_1) = 1.239582$ | Lead Polling Party $(\hat{\beta}_2) = 0.576008$ | 3rd Recent Sen. Result $(\hat{\beta}_3) = 0.4128377$ | 2012 Pres. Result $(\hat{\beta}_4) = 0.409187$ |
| Decision Tree | State Party Alignment $(\mathrm{FI}_1) = 1$ | N/A $(\mathrm{FI}_2) = 0$ | N/A $(\mathrm{FI}_3) = 0$ | N/A $(\mathrm{FI}_4) = 0$ |
| Random Forests | State Party Alignment $(\mathrm{FI}_1) = 0.200451$ | Lead Polling Party $(\mathrm{FI}_2) = 0.148936$ | Median HH Income $(\mathrm{FI}_3) = 0.080751$ | % Adv. Degrees $(\mathrm{FI}_4) = 0.069273$ |
| Gradient Boosting | State Party Alignment $(\mathrm{FI}_1) = 0.267325$ | 2012 Pres. Result $(\mathrm{FI}_2) = 0.170682$ | % Adv. Degrees $(\mathrm{FI}_3) = 0.096194$ | Median HH Income $(\mathrm{FI}_4) = 0.095062$ |

Table 2: Same table as table 1 but for Senate forecasting. **Note:** The project poster for this project errroneously reported feature importances were the same for most of the models. This is not true, this was due to a coding error and the accurate, correct feature importances for the Senate forecast model are provided below.

| Model | 1st Important Feature | 2nd Imp. Feat. | 3rd Imp. Feat. | 4th Imp. Feat. |
|---|---|---|---|---|
| Logistic Regression | Most Recent Sen. Result $(\hat{\beta}_1) = 0.763433$ | 2nd Recent Sen. $(\hat{\beta}_2) = 0.727866$ | 3rd Recent Sen. $(\hat{\beta}_3) = 0.637550$ | % Urban Population $(\hat{\beta}_4) = 0.509023$ |
| l1-Penalized Log. Reg. | Most Recent Sen. Result $(\hat{\beta}_1) = 1.14716$ | 2nd Recent Sen. $(\hat{\beta}_2) = 0.76192$ | 3rd Recent Sen. $(\hat{\beta}_3) = 0.358157$ | Lead Polling Party $(\hat{\beta}_4) = 0.358157$ |
| l2-Penalized Log. Reg. | Most Recent Sen. Result $(\hat{\beta}_1) = 0.499051$ | 2nd Recent Sen. $(\hat{\beta}_2) = 0.470376$ | 3rd Recent Sen. $(\hat{\beta}_3) = 0.380422$ | Senate Incumbent Party $(\hat{\beta}_4) = 0.354815$ |
| Decision Tree | Most Recent Sen. Result $(\mathrm{FI}_1) = 1$ | N/A $(\mathrm{FI}_2) = 0$ | N/A $(\mathrm{FI}_3) = 0$ | N/A $(\mathrm{FI}_4) = 0$ |
| Random Forests | Most Recent Sen. Result $(\mathrm{FI}_1) = 0.187645$ | 2nd Recent Sen. $(\mathrm{FI}_2) = 0.111745$ | Lead Polling Party $(\mathrm{FI}_3) = 0.0875037$ | 2020 Pres. Result $(\mathrm{FI}_4) = 0.079881$ |
| Gradient Boosting | Lead Polling Party $(\mathrm{FI}_1) = 0.238209$ | 2020 Pres. Result $(\mathrm{FI}_2) = 0.164538$ | % Adv. Degrees $(\mathrm{FI}_3) = 0.127514$ | Median HH Income $(\mathrm{FI}_4) = 0.125685$ |

## 5 Results

**Experiment 1** For the first set of experiments, we can see the results summarized within Table 1 and Table 2 for the most important predictors across each of the models. We see that for the presidential election forecast that State Party Alignment (based off the PVI measure) is one of the biggest predictors for how a state votes. Other important indicators include the party that leads in polling, past Senate results, and even some unique demographic indicators like the percentage of the population with post-Bachelors degrees. This is in line with most presidential forecasts, especially with respect to polling and party alignment.

For the Senate forecast, it's very apparent that past Senate election results were the biggest indicator of the 2022 forecasts. This is a bit skewed away from what is expected in standard forecasting literature, so the second experiment accounts for this skew in feature importance.

For prediction, each presidential election was classified the exact same, with all states from 2020 classified the same except for Arizona, Georgia, Michigan, Pennsylvania, and Wisconsin. Each of these states had flipped parties between the 2016 and 2020 elections. As for Senate forecasts, some "competitive states" had been classified the same party while most had a range in their predictions. These results are summarized below in Figure 1, along with Experiment 2 Senate forecasts.

Table 3: Second set of experiments for Senate forecasting, most important features in prediction are reported. Coefficient estimates and feature importances are not provided but can be found within the associated code notebooks in the project GitHub repository.

| Model | 1st Important Feature | 2nd Imp. Feat. | 3rd Imp. Feat. | 4th Imp. Feat. |
|---|---|---|---|---|
| Logistic Regression | Recent Sen. Result | Incumb. Sen. Party | 2020 Pres. Result | Lead Polling Party |
| l1-Penalized Log. Reg. | Recent Sen. Result | 2020 Pres. Result | Lead Polling Party | Incumb. Sen. Party |
| E.Net LR ($\alpha = 0.8$) | Recent Sen. Result | Incumb. Sen. Party | 2020 Pres. Result | Lead Polling Party |
| E.Net LR ($\alpha = 0.6$) | Recent Sen. Result | Incumb. Sen. Party | 2020 Pres. Result | Lead Polling Party |
| E.Net LR ($\alpha = 0.4$) | Recent Sen. Result | 2020 Pres. Result | Lead Polling Party | Incumb. Sen. Party |
| E.Net LR ($\alpha = 0.2$) | Recent Sen. Result | Incumb. Sen. Party | 2020 Pres. Result | Lead Polling Party |
| l2-Penalized Log. Reg. | Recent Sen. Result | Incumb. Sen. Party | Lead Polling Party | 2020 Pres. Result |
| Random Forests | Recent Sen. Result | % Adv. Degrees | 2012 Pres. Result | Lead Polling Party |
| Gradient Boosting | Rec. Sen. Res. (Sen. 2) | Party Alignment | % Lead Polling Party | Recent Sen. Result |

**Experiment 2** For presidential forecasts, most models featured some permutation of the party leading in polling, state party alignment, 2012 and 2008 presidential election results as political features that predicted forecasts. The percentage of advanced degree holders and median household income were the top demographic predictors. Every model had leading party in polls and party alignment as the top features for each model prediction. Due to similar results as in the previous dataset (including with prediction), we omit further discussion of these set of forecasts over the new dataset.

For the Senate forecasts, since so many of the past Senate elections were removed from the set of features, we observed whether there were any changes in the types of features were included within the most important features of the model. We find that some permutation of party leading in polling, the incumbency status of the current sitting Senator (i.e. the party of the Senator currently holding the seat), 2020 Presidential Election results, and the most recent senate result and the key predictors in the models. Table 3 summarizes these findings, clearly the logistic regression models all hold similar features, with the regularization notably affecting the order in which certain features take importance.

As a quick note of sparsity, we observe that LASSO and Elastic-Net models with $\alpha > 0$ led to a general shrinkage of features across the board. Specific to LASSO, LASSO zeroed out 23 features altogether, with the four features described above and also 2008 Presidential Election results as the only overall predictors for the model. This trend of variable shrinking can be seen in the notebooks in the Github repository provided below.

For each model, except for gradient boosting, the most recent sen. result of the Class 2/3 senator within each state was the key feature for predicting the results. Notably, boosting used no presidential election results and used both of the state's Senator's most recent elections as features for prediction. We also observe that demographic features are largely absent in these models for prediction, with political features reigning supreme. The only model to ascribe any importance to a state's demographic characteristics was the random forests model, which considered the percentage of advanced degree holders as the second most important characteristic in modeling.

As for prediction, presidential results were nearly the same as in the first experiment, which makes sense as feature importances did not differ as drastically within the new dataset. For Senate forecasts, we take an average of all the predictions across the eight models and find a more different landscape than the one found in the first experiment. A summary of these results is provided in Figure 1.

## 6 Conclusion and Future Work

From the results, we observe that the models were fairly consistent with using political features that were most aligned with modern-day forecasts. Party alignment for a state, incumbency status, and polling are some of the strongest indicators in election forecasting and from the feature analyses we have above we see that these features were strongly emphasized even after modifications to our training features. Demographic features were largely absent, yet some features such as advanced degree holder percentages and median household income were strong predictors for party wins. This is consistent with longtime political alignment trends in the United States, more wealthy and educated

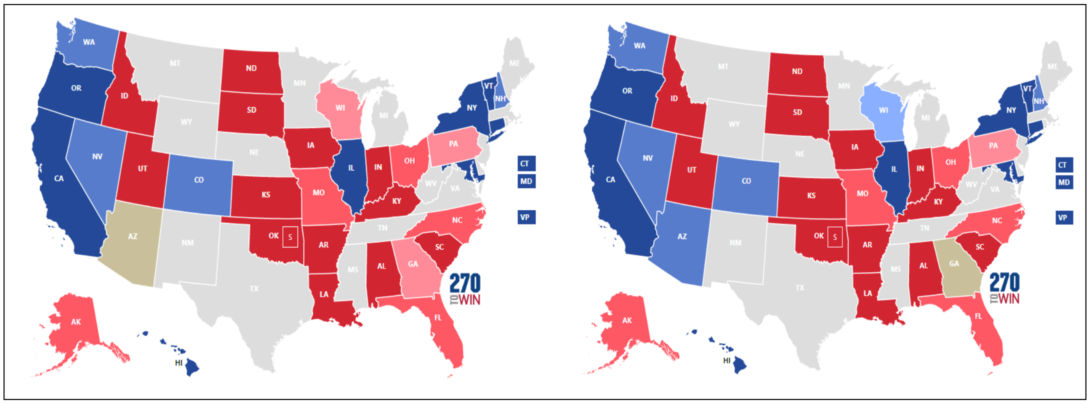


Figure 1: Left: Senate forecast from the first experiment dataset. Right: Senate forecast using the second experiment's dataset. For both visuals, the color indicates how strongly a state was predicted towards a given party over the 6 or 8 models in each respective experiment. Solid red/solid blue states are noncompetitive states that were used as part of training. Brown is a "pure tossup" - half the predictions gave a Dem win for the state while half gave it to Republicans. Varying shades of pink and blue reflects how strongly certain predictions from the sets of models gave it to a certain party. For the darkest shade of light blue or pink (that aren't solid blue or red), these were states where every model predicted only one party winning across all eight models. Notable state differences in forecasts include Arizona, Georgia, and Wisconsin.

states have been trending leftward while right-leaning states tend to be less college-educated and often deal with greater economic hardship.

In terms of the effects of regularization, we observe that within the second set of experiments that regularization did not bring in any new features. Instead, varying the ElasticNet $\alpha$ changed the feature importance of certain features rather than bringing in new features. This is somewhat expected with literature around regularization, coefficients will be altered but the feature structure overall will remain similar. For the more distinct feature differences, ensemble methods proved to provide insights into other features with predictive power, including demographic features that were largely absent across Senate forecasts. Gradient boosting on its own, across the Senate models, proved to bring new features across both Senate forecasts.

Given more time on this project, there are a couple of key changes I would consider for the models. First, I'd explore Bayesian modeling in greater depth and understand how modern ensemble Bayesian approaches are able to perform prediction without the frequentist training and test set-split that is prominent in machine learning. Beyond this, I aim to continue to explore feature optimizations that could help overall prediction. I would include further demographic features such as state density, healthcare access statistics and life expectancy, and more detailed polling and campaign finance data. Many of these metrics are prominent within modern-day models, but they often require hand-picked analysis and tuning. Lastly, I'd consider the alternative approach of predicting *when* election forecasts change for states rather than trying to predict the states' results themselves, as the task of trying to predict flips for states for Presidential elections proved to be challenging, and would often overfit to previous election results.[1]

## Broader Impact

As a disclaimer, the author acknowledges some impacts for their work. The author recognizes that this work examines demographic and political categories for the purposes of election forecasting, with various features and their interpretation as potentially sensitive topics that could suggest a correlation towards a certain political view. The author recognizes this potential issue, and affirms the potential for bias in data within census data results (especially with respect to US noncitizen populations that are known as a frequently undercounted population).

[1]Code and associated data can be found in the project Github repo: https://github.com/anti-software-club/2022-2024-elections.

## References


[1] Bean, Louis Hyman. How to Predict Elections. Greenwood Press, 1972.

[2] Schwartz-Ziv, Ravid. Tabular Data: Deep Learning Is Not All You Need. https://arxiv.org/pdf/2106.03253.pdf

[3] Montgomery, Jacob M.. Improving Predictions Using Ensemble Bayesian Model. Political Analysis, 2012-03-22. https://www.jstor.org/stable/23260318

[4] DDHQ Modeling Team. Methodology. https://forecast.decisiondeskhq.com/methodology

[5] Hastie, Trevor, et al. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2017.

[6] Lewis-Beck, Michael S.. Congressional Election Forecasting: Structure-X Models for 2014. University of Iowa.

[7] natesilver538. "A User's Guide to Fivethirtyeight's 2016 General Election Forecast." FiveThirtyEight, FiveThirtyEight, 29 June 2016.

[8] natesilver538. "How Fivethirtyeight's 2020 Presidential Forecast Works - and What's Different Because of Covid-19." FiveThirtyEight, FiveThirtyEight, 12 Aug. 2020.

[9] Lauderdale, Benjamin E. Under-performing, over-performing, or just performing? The limitations of fundamentals-based presidential election forecasting. International Journal of Forecasting.