

Modeling State Legislative Elections

By Eric Stone






Goal

To create an effective model for predicting state legislative races in all 50 states in order to aid in resource allocation

Classify every race as either safe Republican, Democratic, or toss-up





Methodology

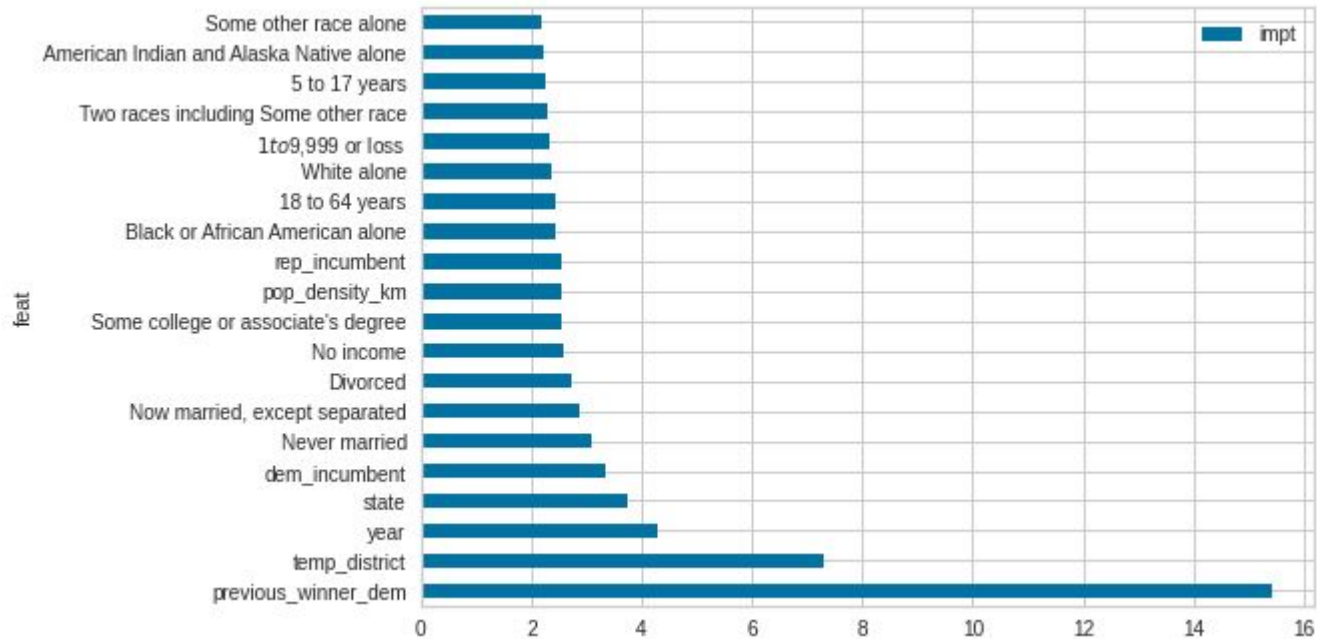
National Model

- ◎ CatBoostClassifier
 - ~80% Accuracy
 - ~42% of toss-ups properly classified
- ◎ Random Forest Classification
 - ~85% Accurate
 - ~47% of toss-ups properly classified

50 State Models

- ◎ Ran all 50 states through 6 models to determine the best models for each
- ◎ Average accuracy ~81%
- ◎ Most states performed worse
- ◎ Possible correlation between strong state models and state population
 - NY, CA, TX among best
 - RI, WV, SD among worst

Most Influential factors



According to CatBoostClassifier Model


Results

- ◎ Seat history is most important
- ◎ Democratic Incumbency stronger indicator than Republican
- ◎ Percentage of lower income people stronger indicator than higher income

A decorative network diagram in the top right corner, consisting of various sized circles (nodes) connected by thin lines (edges). Some nodes are solid grey, while others are hollow with a grey outline. The connections form a complex, branching structure.

Conclusion

Modeling electoral outcomes accurately requires a great deal of context. State legislative races are particularly challenging due to lack of general attention or polling.

A decorative network diagram in the bottom left corner, similar to the one in the top right, featuring a cluster of interconnected nodes and edges.

Further Steps

- ◎ Further data manipulation including Feature Unions
- ◎ Collecting district info from commercial sources
- ◎ Expand time period analyzed
- ◎ Analyze relationship between up ballot and down ballot races
 - Presence of state-wide and federal races
 - Effect on turnout and relation to outcome