

Model Release Notes

Model: Presidential Election Turnout (2020)

Source Data: Past Vote History

Geography: United States

Date: May 23, 2017



Model Description

An ensemble method classifier model was created to predict the likelihood that an individual will vote in the 2020 general election. The model was constructed using a random sample of TargetSmart's national Voterbase file. The model scores are expressed from 0-100, with the score representing the probability that person will vote in the 2020 general election. The model was used to score over 250 million voting age persons nationwide.

Process Overview

Two distinct models were built: a model that utilizes an individual's previous vote history and a model that does not utilize previous vote history. The model that does not utilize previous vote history is necessary to score recently registered voters who have no vote history. Each registered individual receives only 1 model score, based upon whether they have vote history.

For the With Vote History model (WVH), individuals were divided into the following two groups:

- (1) Targets:** Individuals who voted in the 2016 general election and who have voted in previous elections.
- (2) Non-Targets:** Individuals who did not vote in the 2016 general election, but had voted previously.

For the No Vote History model (NVH), individuals were divided into the following two groups:

- (1) Targets:** Individuals who voted in the 2016 general election, but had never voted previously.
- (2) Non-Targets:** Individuals who have no vote history, were registered, but did not vote in the 2016 general election.

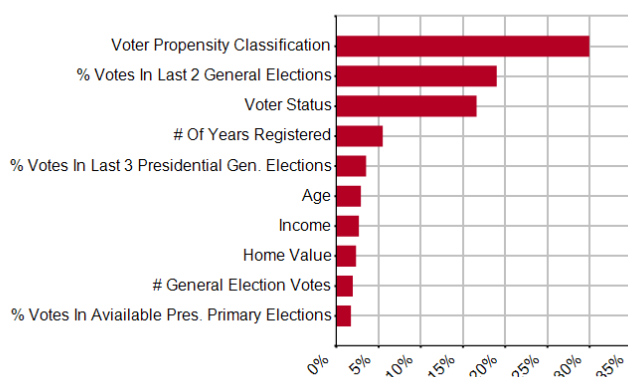
Several sub-models were built using a rules-based classifier on a variety of consumer, political, credit, and demographic variables. In total, over 100 variables were considered during model creation. The model building process determined the most appropriate variables in identifying the differences between those that will vote in the 2020 general election and those that will not vote. The final score was generated using a boosted decision tree classifier. A higher score predicts a higher likelihood that the individual will vote in the 2020 general election.

To validate the model, a random selected group of registered voters were held out from the model building process. These holdout records were scored with the models and analyzed for accuracy.

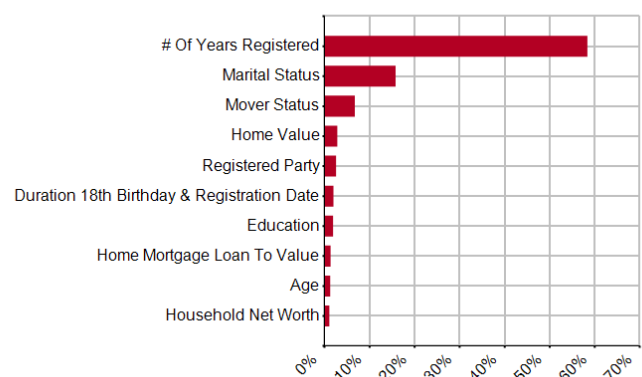
Key Variables

The key variables and relative weights used in the model include:

With Vote History



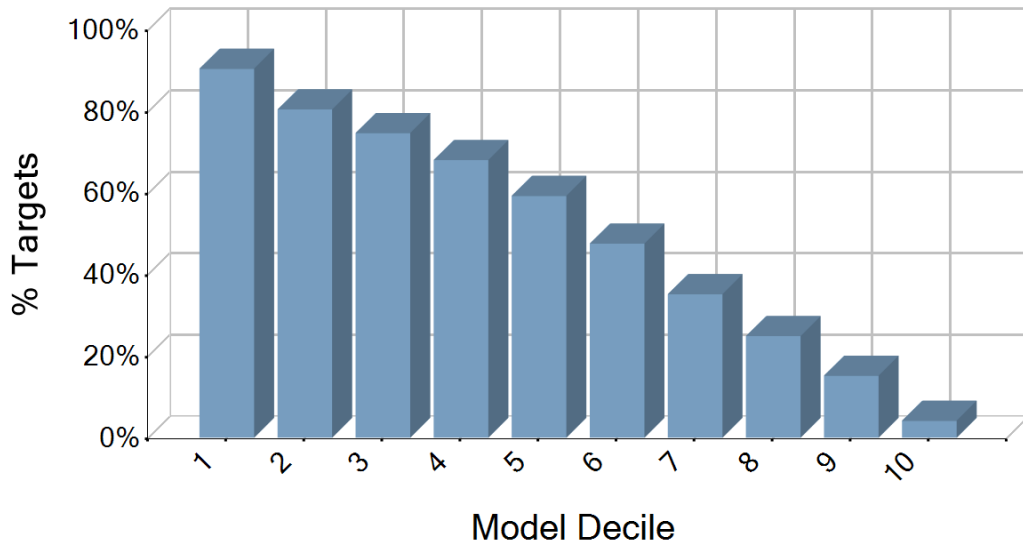
No Vote History



Validation

The model was validated by scoring a holdout sample of registered voters. The holdout records were then ranked by model score, separated into ten deciles, and evaluated. The highest scored records are found in decile 1, while the lowest scored records are found in decile 10.

Successful rank-ordering occurs when a higher percentage of target records are correctly scored higher than non-target records. We expect a valid model to show a stepwise decline from decile 1 through decile 10. Ideal or perfect models show a steep downward slope, with values near 100% in decile 1 and near 0% in decile 10.



Score Distribution

The following chart shows the distribution of model scores for registered voters in the United States. The scores range from 0-100. Higher scores indicate a higher likelihood of voting in the 2020 general election.

