

Model Release Notes

Model: Activist

Type: Survey-Response

Geography: National

Refresh Date: August 2019



Model Description

An ensemble method classifier model was created to target and identify political activists. The model was constructed using responses to ten questions from an online and telephone survey conducted in August 2017. The model scores here are expressed on a 0-100 scale and represent the probability that a person is an activist. Higher scores represent a higher probability that a person is an activist. The model was used to score over 250 million voting age persons.

Process Overview

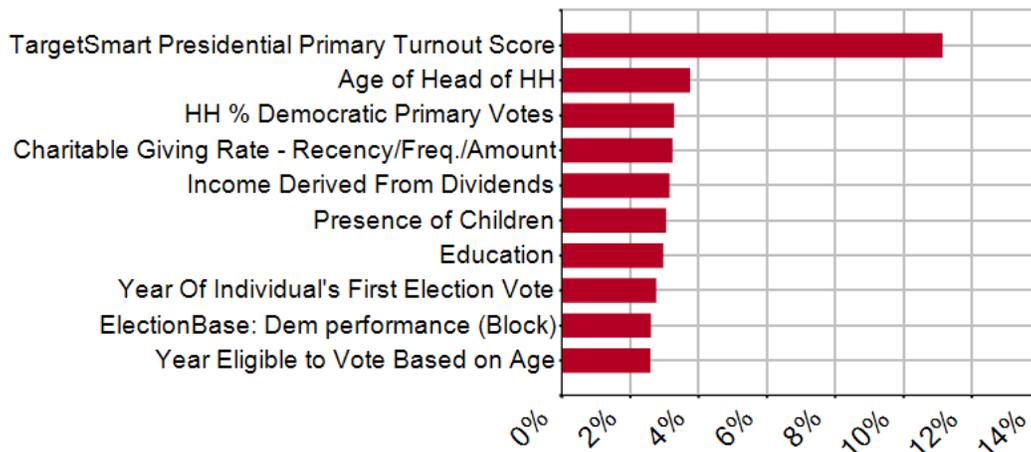
The model was trained on survey respondents who were divided into two groups based on their responses to a series of ten questions about ways to get involved in politics or causes:

- (1) **Targets:** Respondents who have performed multiple political acts in the past, such as volunteering time on a campaign, demonstrating, going to town halls, reaching out to elected officials, or donating money to campaigns.
- (2) **Non-Targets:** Respondents who have performed none of the ten selected political acts in the past.

Several sub-models were built using a rules-based classifier on a variety of consumer, political, credit, and demographic variables. In total, over 1,000 variables were considered during model creation. The model building process determined the most appropriate variables in identifying the differences between target respondents and non-target respondents. The final score was generated using a boosted decision tree classifier. To validate the model, a randomly selected group of survey records were held out from the model building process. These holdout records were then scored with the model and analyzed for accuracy.

Key Variables

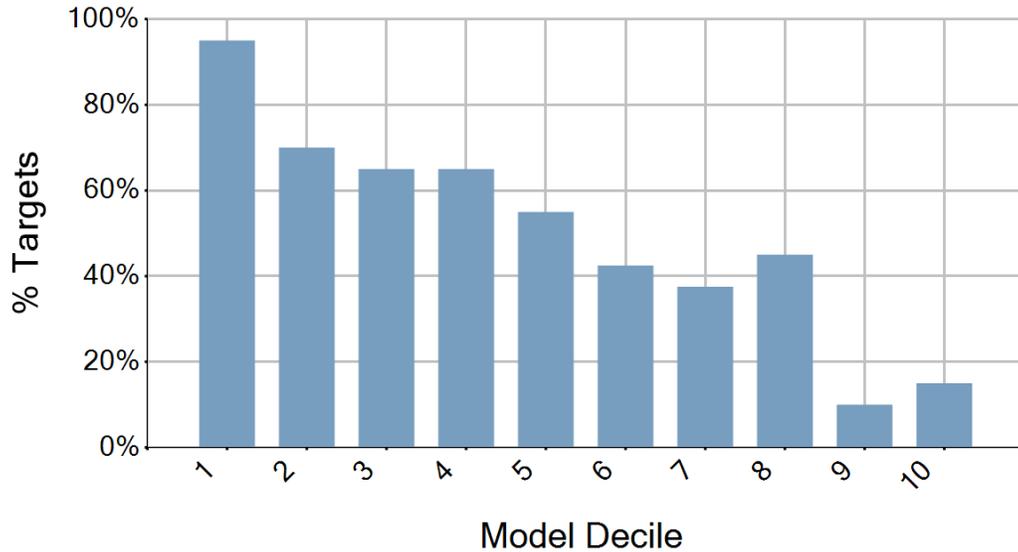
The key variables and relative weights used in the model include:



Validation

The model was validated by scoring a holdout sample of survey respondents. 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 voting age persons nationwide. The scores range from 0-100. Higher scores indicate a higher likelihood of being an activist.

