Model Release Notes

Model: Progressive Tax
Type: Survey-Response
Geography: United States
Date: November 27, 2019

targetsmart

Model Description

An ensemble method classifier model was created to target and identify supporters of progressive taxation in the United States. The model was constructed using results from an online survey conducted in October 2019. The model scores here are expressed on a 0-100 scale and represent the probability that a person would support raising taxes on the wealthy. The model was used to score over 269 million registered voters nationwide.

Process Overview

The model was trained on survey respondents who were divided into two groups based on their response to a question about taxation:

(1) **Targets**: Respondents who favor raising taxes on the wealthy because it will help fund government services and balance the budget
(2) **Non-Targets**: Respondents who oppose raising taxes on anyone, including the wealthy, because it will lead to job losses, hurt the economy, and take money out of their pockets

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:

targetsmart
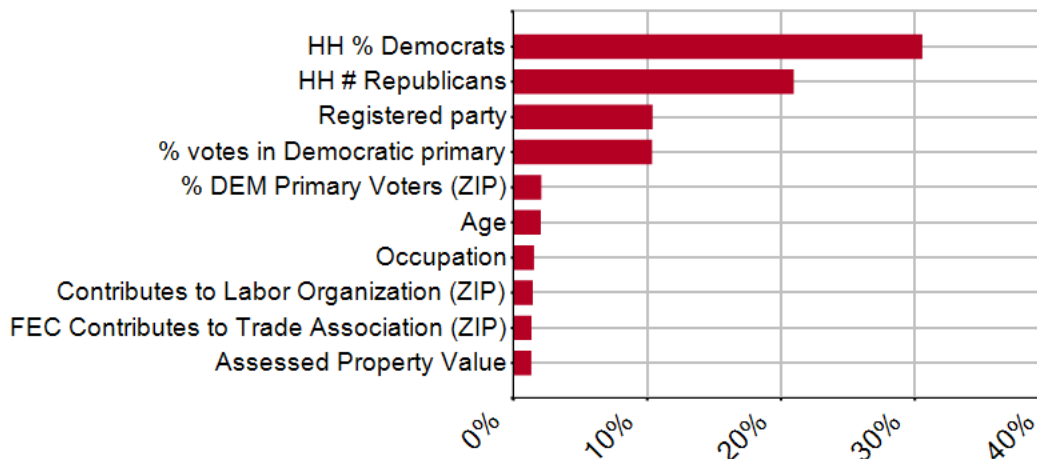
Validation

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.
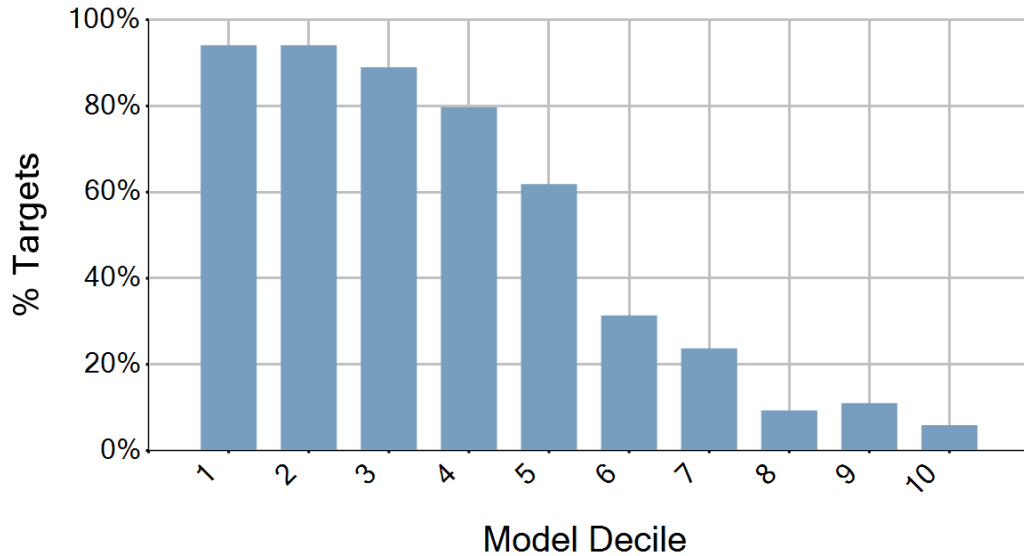
Score Distribution

The following chart shows the distribution of model scores for registered voters nationwide. The scores range from 0-100, with higher scores indicate a higher likelihood of supporting raising taxes on the wealthy.