

# Model Release Notes

**Model:** Recession Sensitivity

**Type:** Survey-Response

**Geography:** United States

**Date:** August 26, 2020



## Model Description

An ensemble method classifier model was created to target and identify individuals who felt most at-risk of financial hardship during the onset of the COVID-19 outbreak in the United States. The model was constructed using results from a survey conducted in May 2020. The model scores here are expressed on a 0-100 scale and represent the probability that a person felt they were most at-risk of financial hardships. The model was used to score over 251 million voting age persons nationwide.

## Process Overview

The model was trained on survey respondents who were divided into two groups based on their survey responses to the following two questions:

- Q1) How worried are you personally about missing a rent or mortgage payment in the next three months?
- Q2) And thinking about your personal financial situation over the next few months, do you feel confident or uneasy?

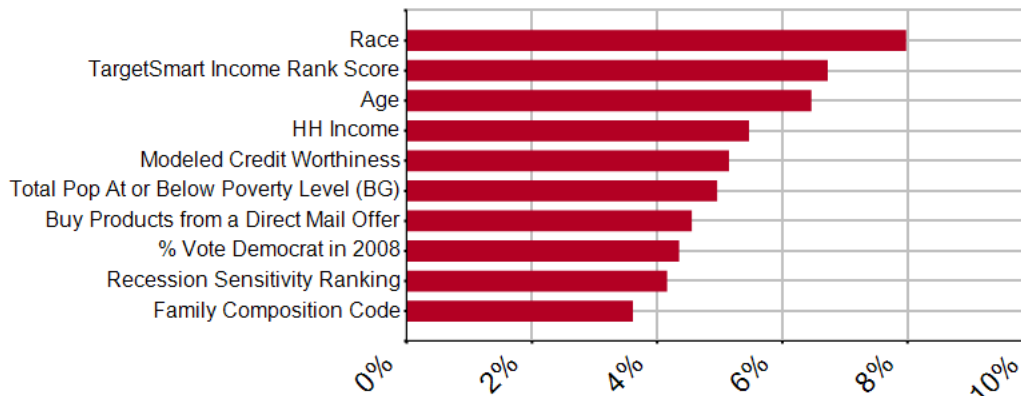
(1) **Targets:** Respondents who said they were somewhat or very worried about missing rent and felt uneasy about their personal financial situation.

(2) **Non-Targets:** Respondents who said they were not worried at all about missing rent and felt somewhat or very confident about their personal financial situation.

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

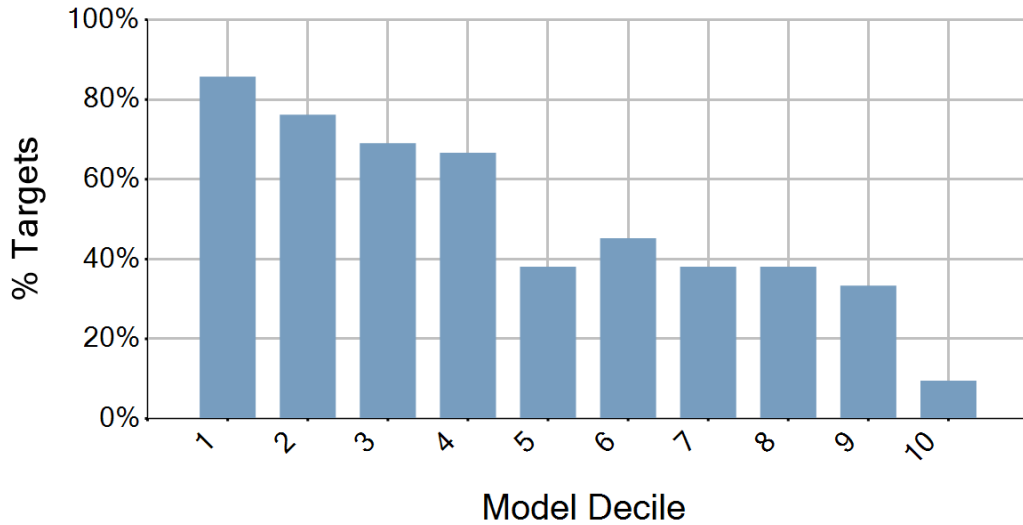
The key variables and relative weights used in the model include:



## 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 registered voters nationwide. The scores range from 0-100, with higher scores indicate individuals who felt most at-risk of financial hardship during the onset of the COVID-19 outbreak in the United States

