Home Credit Default Risk

Business Problem Statement

Home Credit would like to provide customers with the ability to request loans despite having a lack of credit history. Without credit history, it is difficult to predict the likeliness of a given customer to pay off their debt which is a risk lenders usually avoid. Home Credit will need to create a model to determine which customers will pay and which will default in order to remain profitable and continue stable operations.

Benefit of a Solution

The benefits of an accurate probability model include:

- Increased accuracy in predicting which customers are likely to default
- Increased accuracy in predicting which customers are likely to pay off their loans
- Removing the traditional banking requirement of a stable credit history to be approved for loans.

Analytics Approach

To create a model that can accurately predict a customer's likeliness of credit default, we will need to train a model using historical customer data given by Home Credit. This will be a supervised classification model that determines the probability of the target variable: credit default.Through the combination of calculating the significance of one variable and the relationship between multiple variables, a test set can be created to provide a probability of the target variable happening.

Success Metrics

The main metric for determining a successful model is if the test model produces a number that is close to 1.0 for the area under the ROC curve. The area under the curve represents how far the predicted probability is from the observed target. The higher that number is, the more accurate the model will be in helping Home Credit determine which customers are likely to default. On the other hand, the model should also accurately predict who is eligible for a loan to ensure that there is not a significant amount of customers who are turned away due to being labeled as likely defaulters.

Scope

Deliverables

Deliverables will include a prediction model that determines likeliness of a default and a summary report of the process, insights, and results.

Out of Scope

Determining if variables for customers with credit history significantly affect the probability of default in the model used for customers with a lack of history. We will not test the accuracy of the model with customers who have credit history at this moment.

Future Enhancements

Some future enhancements can include using a different statistical approach to create the prediction model, using “not default” as the target variable to confirm consistency for predictions, and including other variables that may affect the probability of default.

Details and Timeline

Data analysts at Home Credit will begin the project on May 17, 2018. The final model will be completed on August 29, 2018. We will then meet that Friday, the 31st, to share the model and the report that will include the results of this project.

Start - May 17, 2018
Deadline - August 29, 2018
Presentation - August 31, 2018