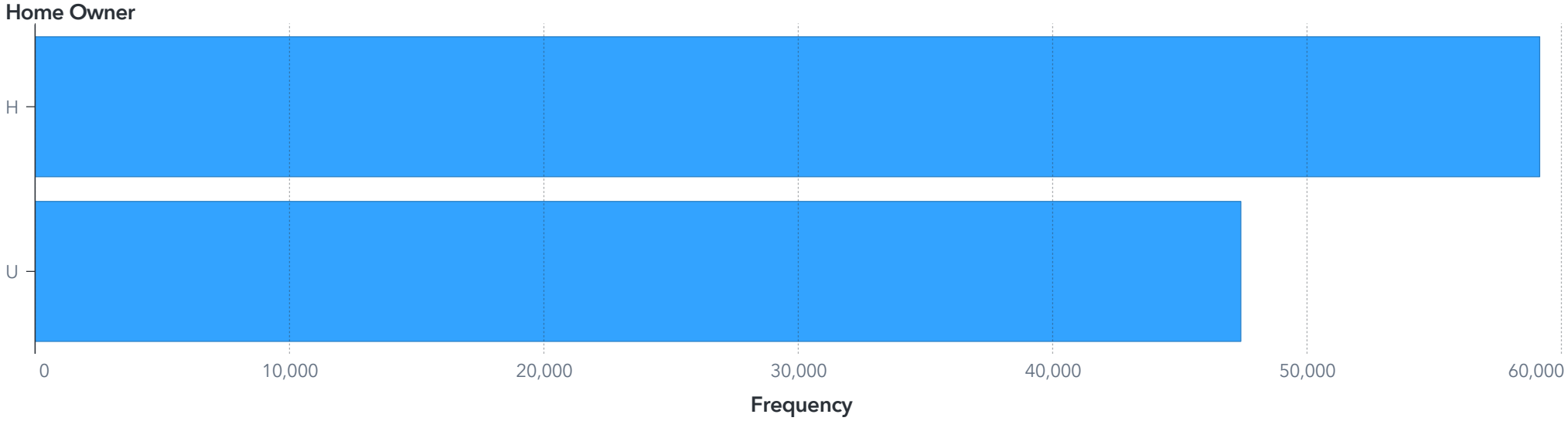
PVA

Creation Date: Sunday, 8 May 2022 17:22:17

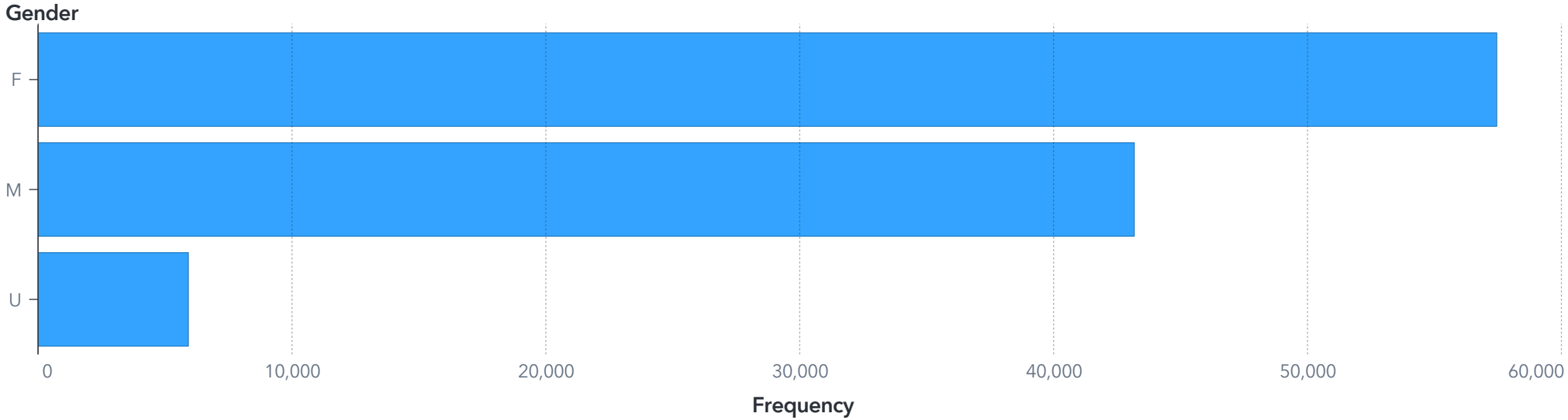
Author: joyee.banerjee@edu.dsti.institute

Excercise Page 72

Frequency of Home Owner

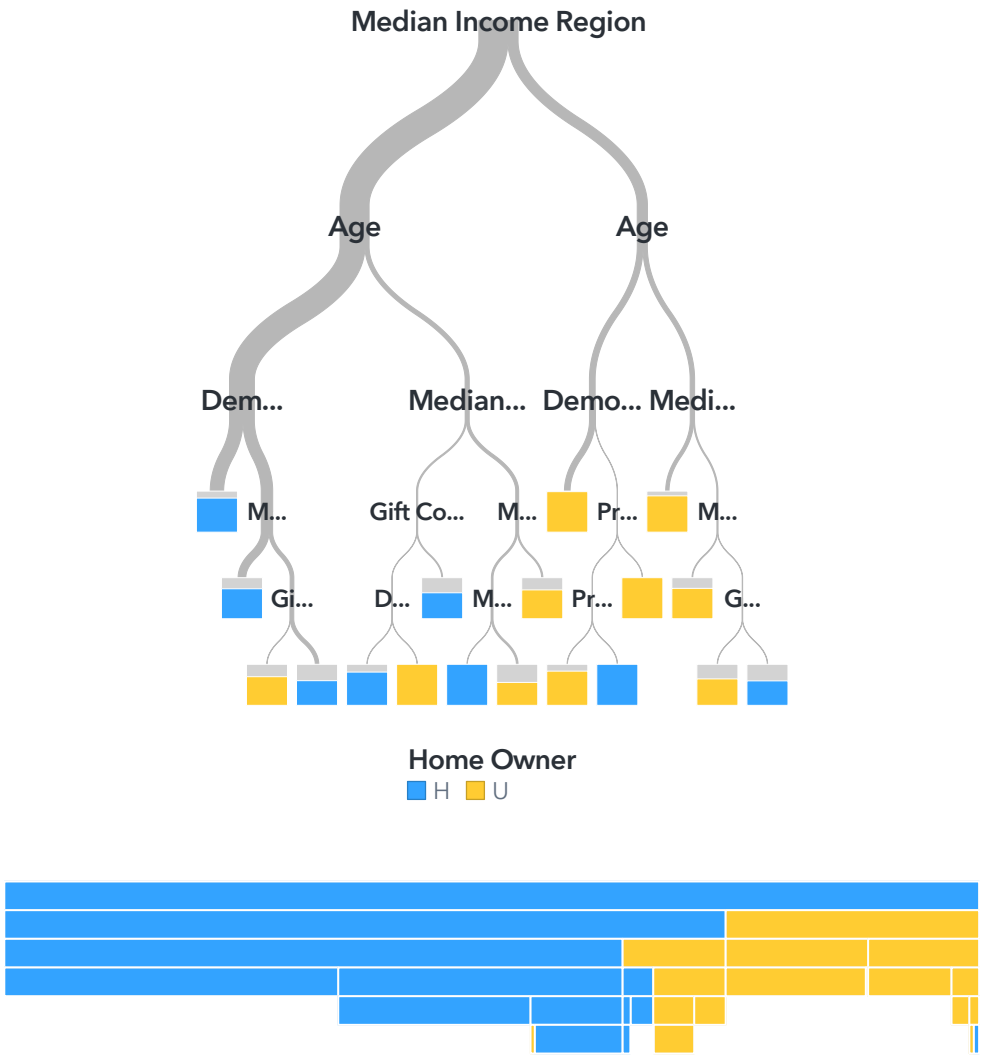


Frequency of Gender

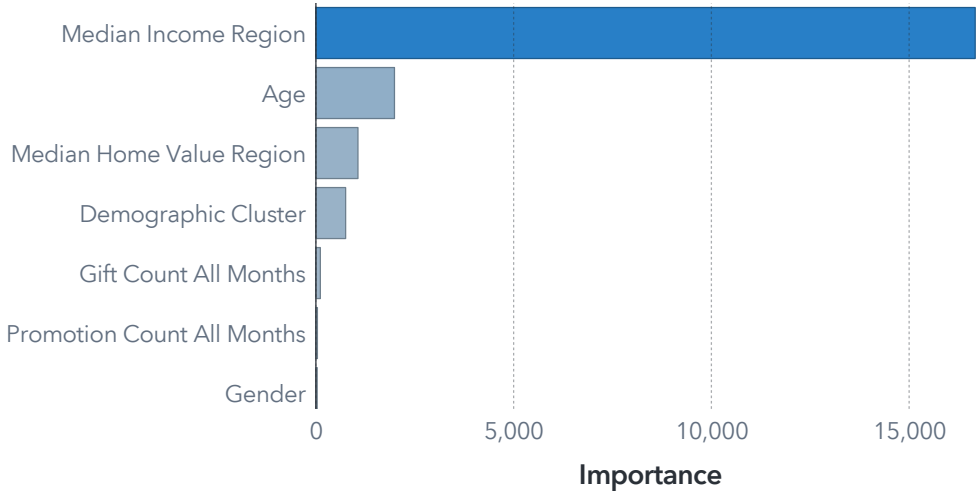


Decision Tree **Home Owner** (event=U) KS (Youden) **0.5552** Observations Used **106,546**

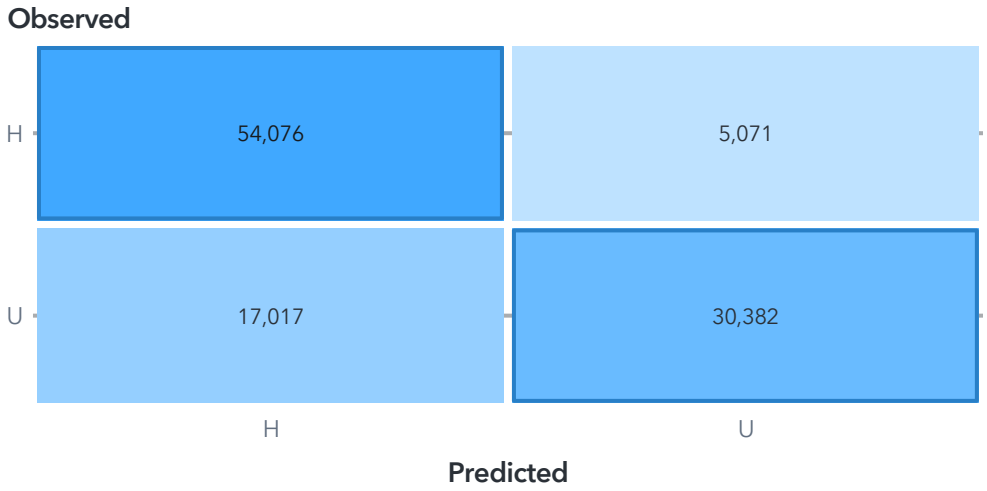
Tree



Variable Importance

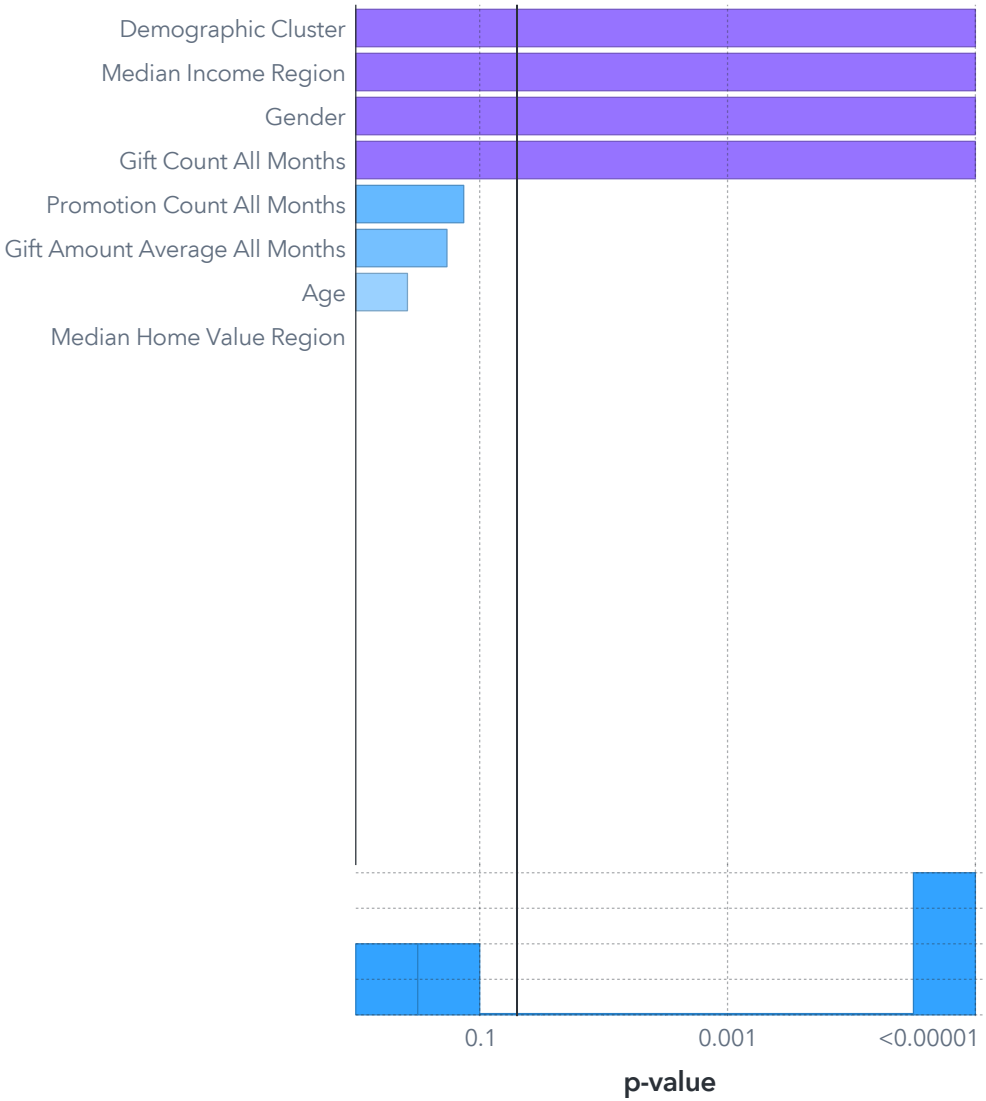


Confusion Matrix

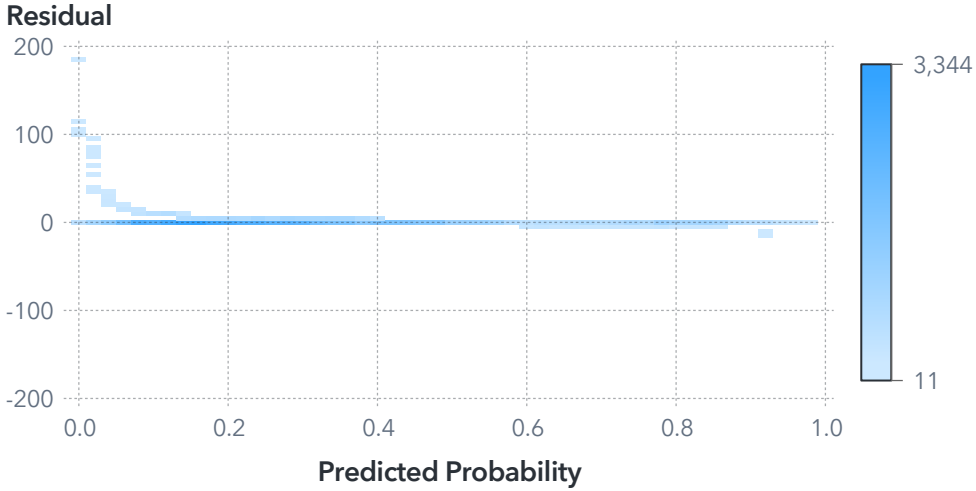


Logistic Regression **Home Owner** (event=U) KS (Youden) **0.4076** Observations Used **80,069** Unused **26,477**

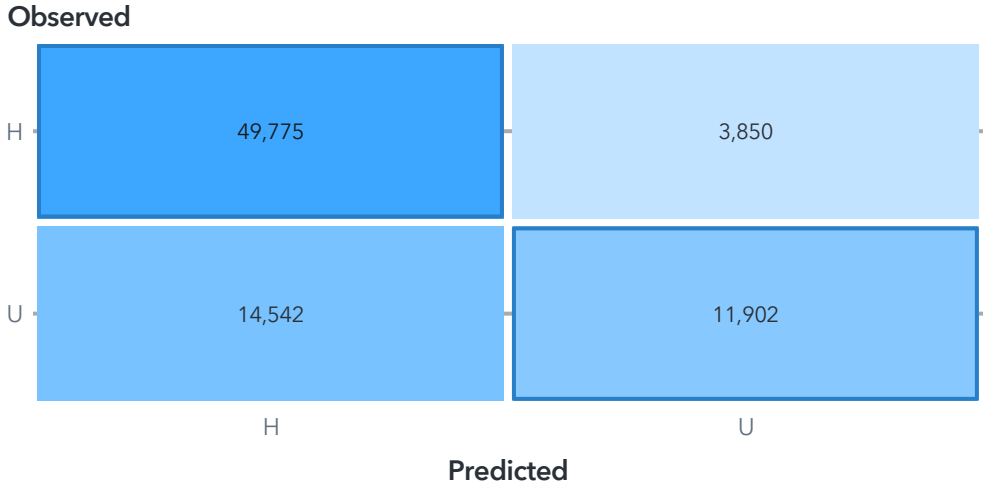
Fit Summary



Residual Plot

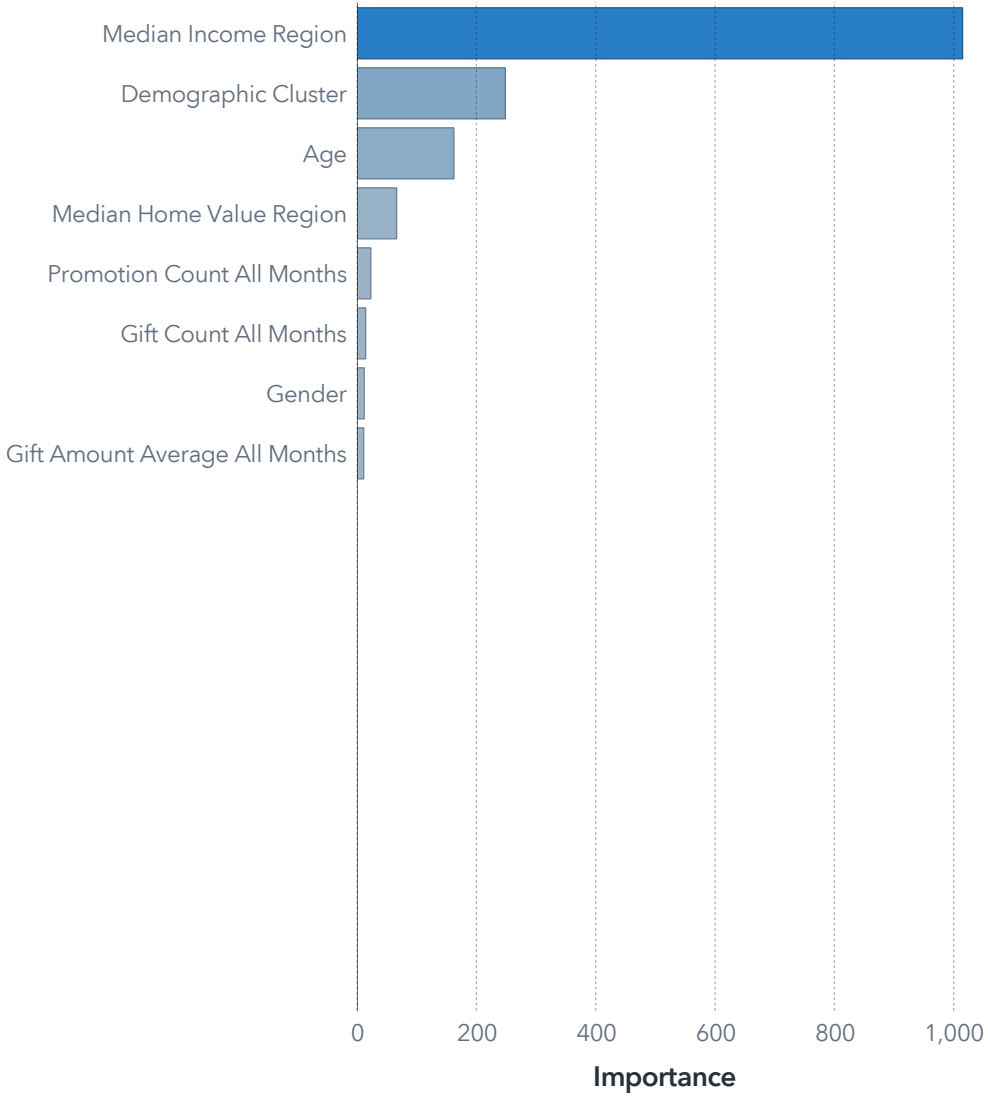


Confusion Matrix

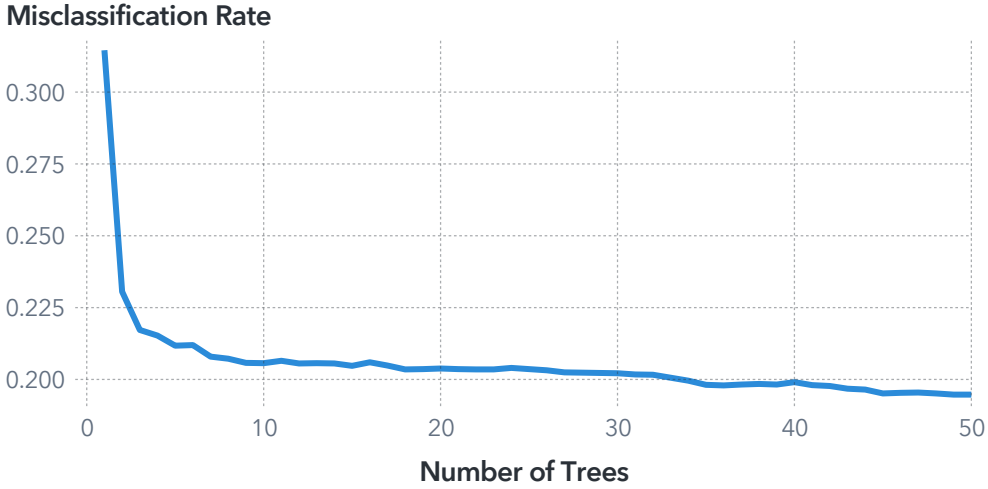


Gradient Boosting **Home Owner** (event=U) KS (Youden) **0.5911** Observations Used **106,546**

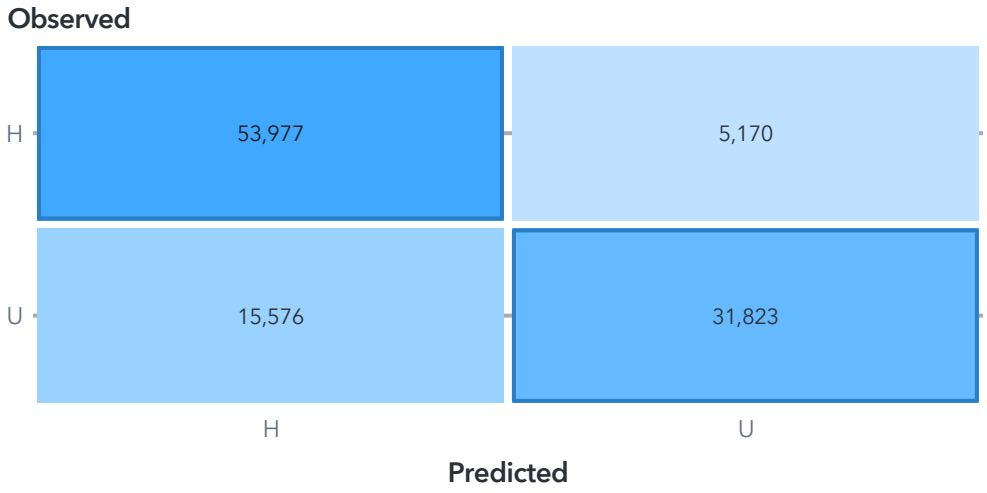
Variable Importance



Iteration Plot

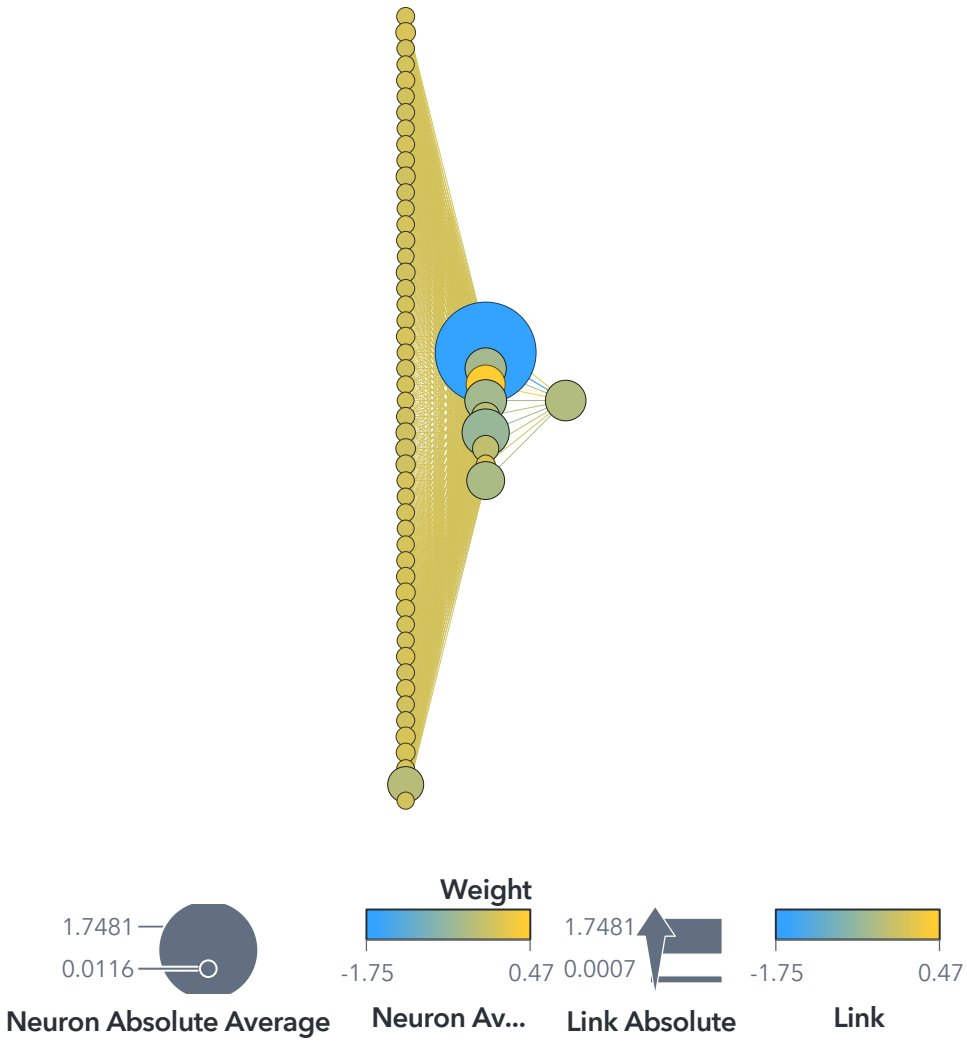


Confusion Matrix



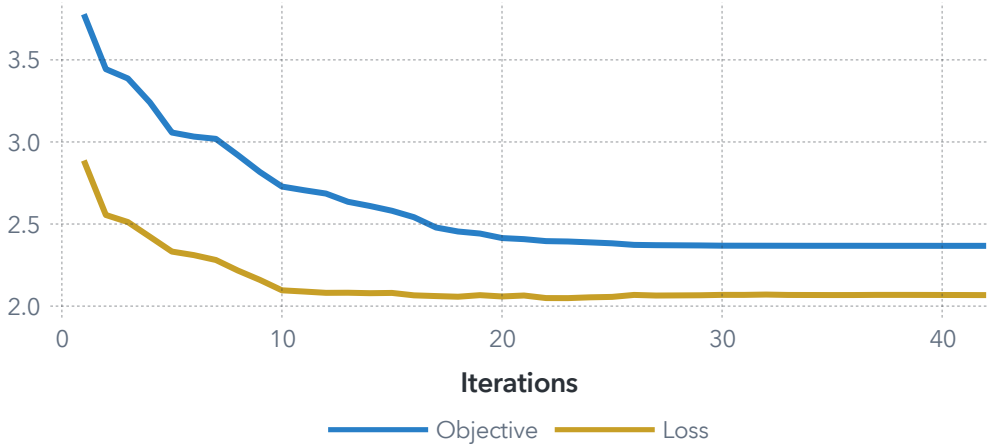
Neural Network **Home Owner** (event=U) KS (Youden) **0.4033** Observations Used **80,069** Unused **26,477**

Network



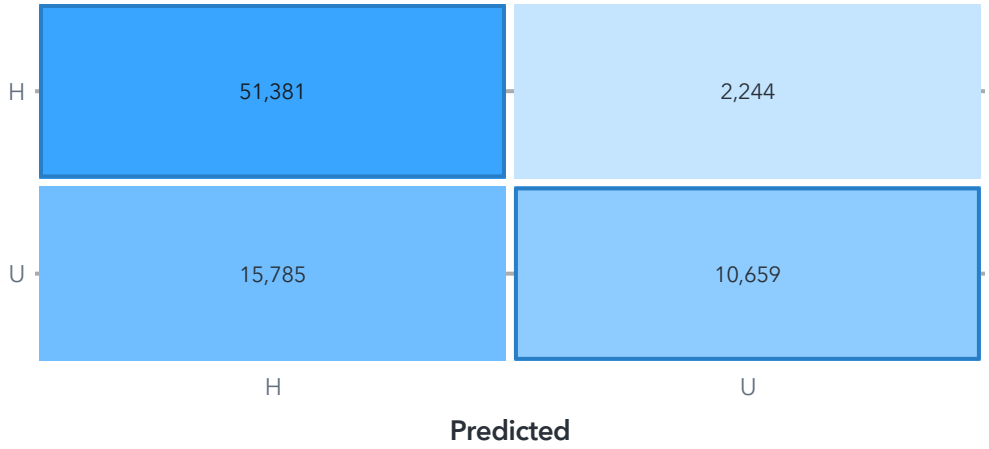
Iteration Plot

Objective / Loss



Confusion Matrix

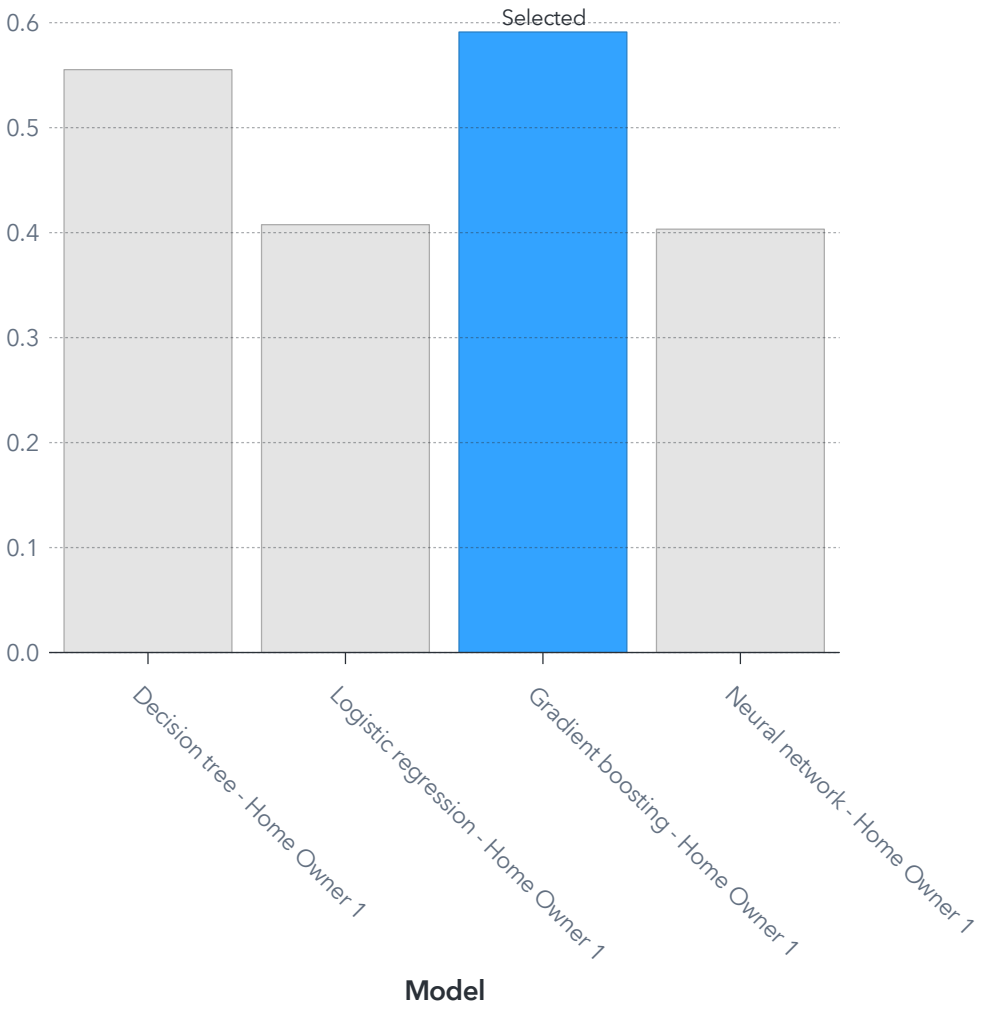
Observed



Model Comparison **Home Owner** (event=U)

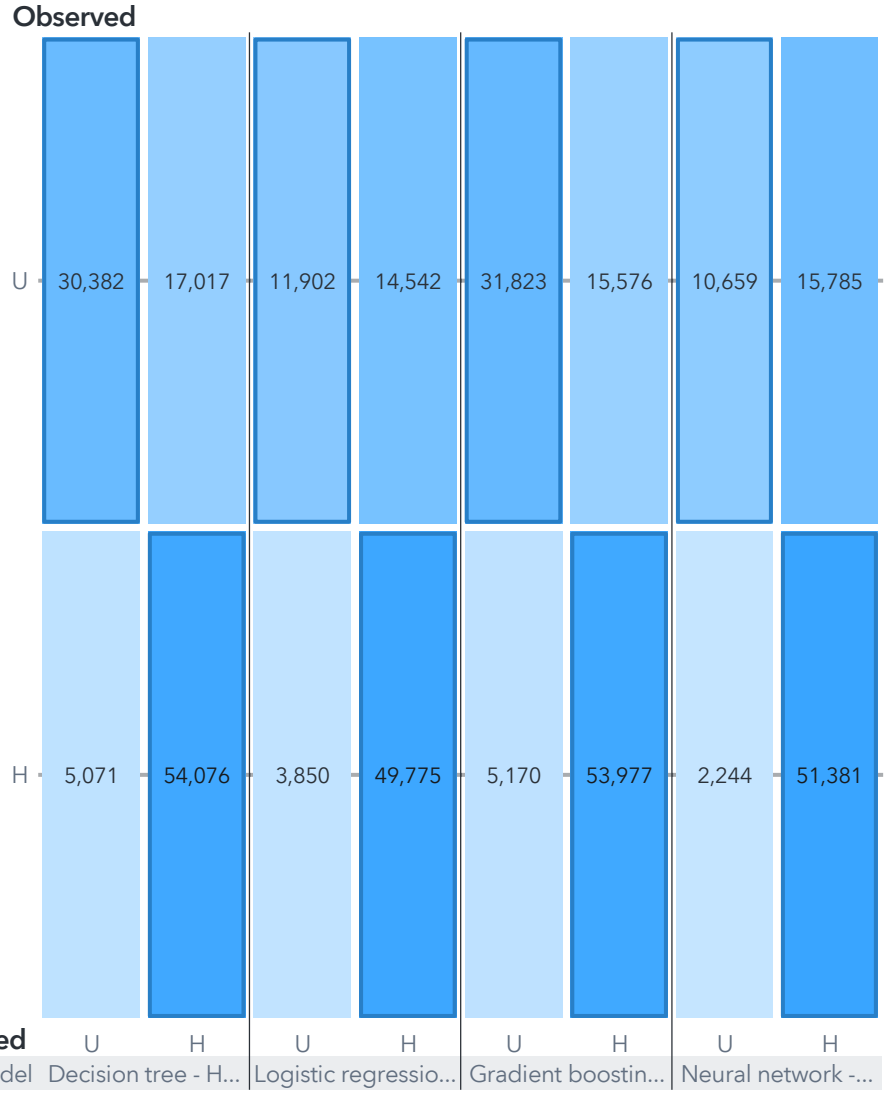
Fit Statistic

KS (Youden)



A1.1

Confusion Matrix



Executive Summary

In this analysis the target variable chosen is ‘DemHomeOwner’ which has the description ‘Home Owner’ and the Role as ‘Predictor’. This target variable was chosen so that it could process several models to be analyzed.

In the first page the variable Home Owner is analyzed against the Gender and it can be seen that in proportion females own more houses than their male counterparts. In the second page the decision tree model is analyzed to understand the segmentation of the dataset better. In the classification tree, the categorical variable Home Owner was chosen and the variables Median Income Region, Age, Gender, Demographic Cluster, Median Home Value Region, Gift Amount Average All Months, Gift Count All Months, Promotion Count All Months were chosen as predictors. Then the Logistic Regression, Gradient Boosting and Neural Network models were created. In the Model Comparison it is observed that the best model is the Gradient Boosting according to the KS (Youden) statistic which is 0.5911 and the lift curve.

Following the model analysis, to calculate the ROI the formula =100000*A2/100*(-2+3*F2*45/100) is used. There are more than 1 00 000 donors in the PVA table. The default base rate in this case is 45% and the average cost of sending a gift is 2$ and the margin is 3$. Here, selected is the 5% who’ve got the highest probability to make a donation, the default return rate from 45% will be multiplied by the lift from the model (2.24 at 5%) to give an ROI that is 5175$. Then the formula is applied to all the columns. It is seen that the maximum of the ROI is at 30%.

Thus, when selected is the 5% who have the highest probability to make a donation, we must get 5175$ back. And on the other hand, when selected is the 30% who have the highest probability to make a donation, we must get 22810$ back.

Appendix

A1.1 Fit Statistic

Warnings: Neural network - Home Owner 1: One or more of the neuron layers has too many neurons to display and has been truncated.
Number of observations for all models do not match.