Tyler Gelfand

Credit Card Marketing Report

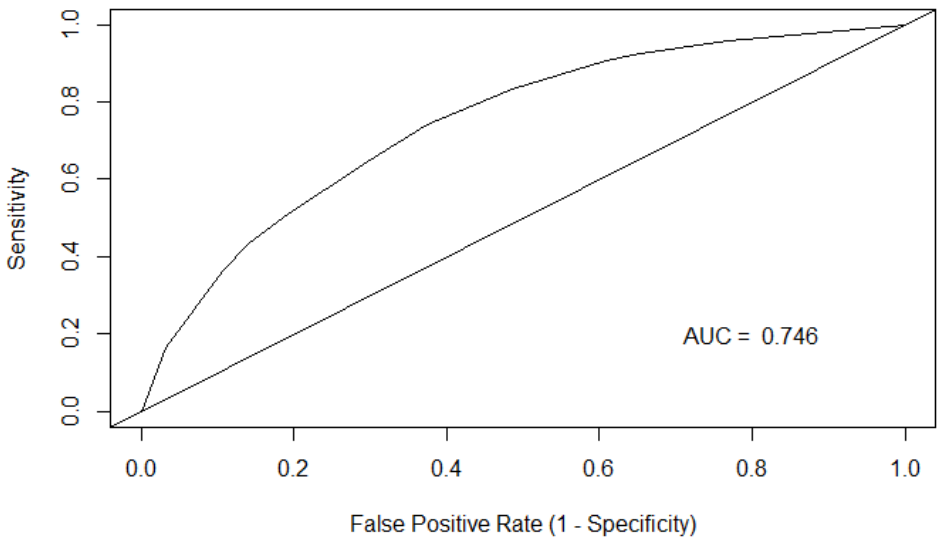
Modeling Methods:

To correctly predict whether a customer will open a new account, three separate but similar models were run. A normal classification tree model utilizing the full data set, a partitioned classification tree model which splits the data into a test and training set with a 70/30 split, and a random forest model, also utilizing the partitioned data.

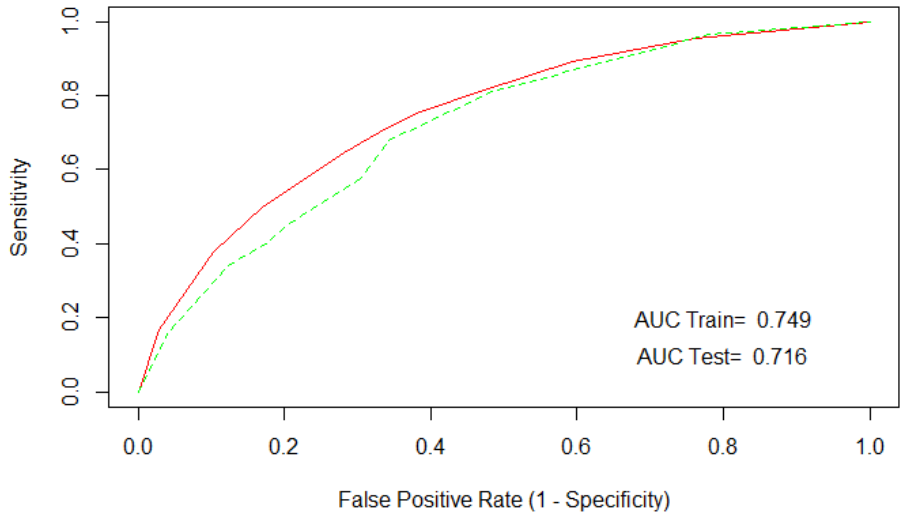
Model Performance:

One way to measure the model performance is by building an ROC Curve. As we can see below in each model the Area Under Curve was about .72-.75, meaning that if we select 2 customers, one that will open an account and one that will not, the model will be able to tell them apart anywhere from 72-75% of the time.

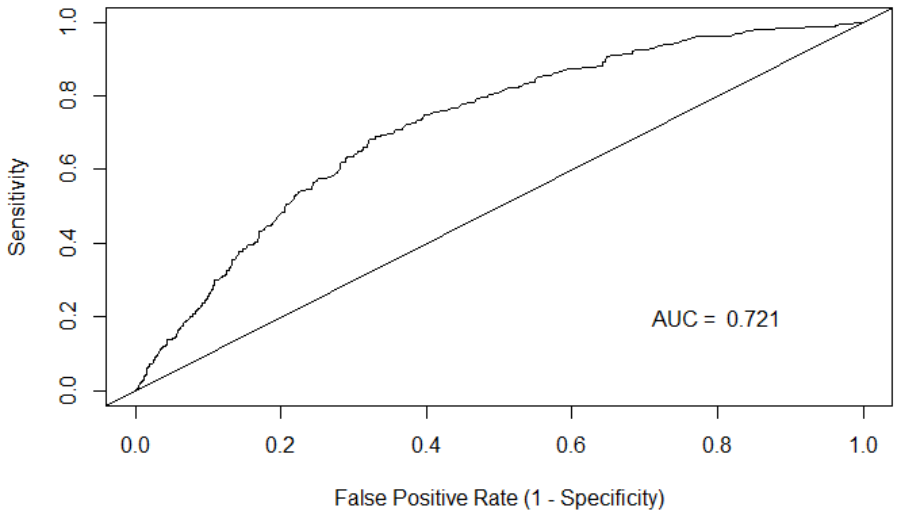
ROC Curve- Classification Tree(Non-Partitioned)



ROC Curve (Partitioned)



ROC Curve- Random Forest



Another way to measure the model performance is by looking at the confusion matrix. The below confusion matrix was created using the non-partitioned classification tree model. The matrix shows that about 81% of the customers were given the correct label. The major issue with this confusion matrix is the sensitivity score. It appears that the model only correctly predicted about 16% of customers who were going to accept the credit card offer. However the model did correctly predict about 97% of customers who would deny the offer. The positive predicted value shows that there is about a 55% success rate from people who are targeted by the marketing.

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 3991  854
         1  132  167

               Accuracy : 0.8083
                 95% CI : (0.7973, 0.819)
    No Information Rate : 0.8015
    P-Value [Acc > NIR] : 0.1136

                  Kappa : 0.1792

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.16357
            Specificity : 0.96798
         Pos Pred Value : 0.55853
         Neg Pred Value : 0.82374
             Prevalence : 0.19848
         Detection Rate : 0.03247
   Detection Prevalence : 0.05813
      Balanced Accuracy : 0.56577

       'Positive' Class : 1
```

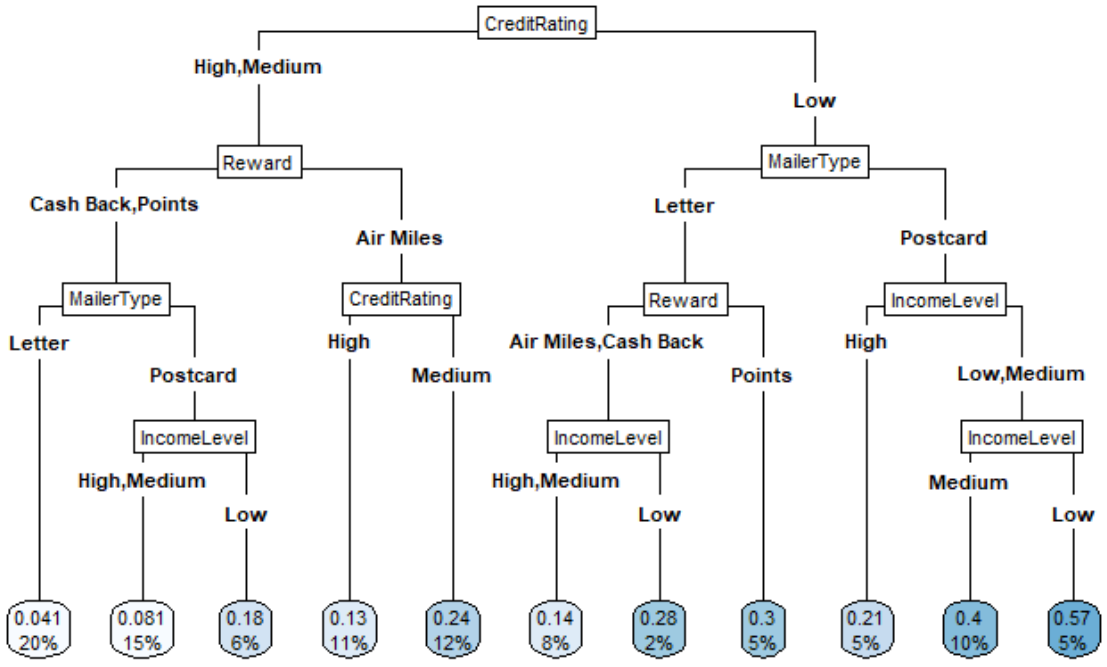
Variable Importance:

The below chart is a measure of variable importance. According to the chart, the top three most important variables are credit rating, mailer type and income level.

	No	Yes	MeanDecreaseAccuracy	MeanDecreaseGini
CreditRating	8.740461e-03	0.0419181285	1.499343e-02	78.83004
MailerType	3.204711e-03	0.0202269437	6.408410e-03	33.76586
IncomeLevel	4.116342e-03	0.0153168867	6.229785e-03	41.74478
Reward	1.114160e-03	0.0054433242	1.922828e-03	34.47423
NumCreditCardsHeld	-1.406089e-04	0.0042998140	6.946650e-04	38.20762
AverageBalance	8.872465e-04	-0.0007527921	5.737865e-04	146.10288
NumHomesOwned	6.530854e-04	-0.0001868489	4.961635e-04	21.36766
OwnYourHome	-3.270512e-05	0.0016367636	2.863870e-04	19.02911
OverdraftProtection	1.550497e-04	-0.0008859200	-4.139736e-05	14.69766
HouseholdSize	-4.097635e-04	0.0014818295	-5.350459e-05	53.74266
NumBankAccountsOpen	-6.278074e-05	-0.0018493900	-3.984911e-04	22.10071

Letters or Postcards?

The below classification tree is taken from the partitioned data. Looking at the mailer type, across both credit ratings, income levels and reward types, letters were accepted a total of 35% of the time while postcards were accepted a total of 46% when sent out. Therefore, postcards are slightly more effective.



How can this be used?

The classification tree model can be used by the credit card company to distribute and target the marketing materials more effectively. Much of the effort should be placed towards those with high or medium credit score as about 2/3 of accepted responses came from that group. Those with high- or medium-income levels also accepted most of the offers. As stated above, postcards were the most effective marketing communication method.