[A2: National City Bank](https://www.canva.com/design/DAFd2difK2w/Jlig20x02S3cvezEvD6TsQ/view?utm_content=DAFd2difK2w&utm_campaign=designshare&utm_medium=link&utm_source=recording_view)

Dalton Lee Everett

1. Introduction

A picture containing text, hitting, tennis

Description automatically generated

1. Table of Contents

Graphical user interface

Description automatically generated with medium confidence

1. Problem



1. Data-Historic

Chart, bar chart

Description automatically generated

1. Data-Prospective

The data between our prospective customers and historic customers is nearly the same. There is no major difference between them, and that is why it is important to create an accurate model that will give us the top 100 customers that are most likely to accept.

Chart, bar chart

Description automatically generated

1. Methods

This section [Data Preparation], is one of the most important steps and one of the first. Without the proper setup, our code will not work throughout our modeling. This 'automated variable processing' is essential to our data manipulation as not only does this create the new binary variables, but it also fixes any nulls a single column might have. This saves us tremendous time as we do not have to go through every single column to fix the data and create dummies to get to a binary dataset. Creating the 'plan' we utilize the data we want to use, [informativeFeatures] are the columns we want to engineer into binary’s, our target variable (Y\_AcceptedOffer), and [successClass] (Accept) that we will use in our vtreat to create better and more accurate variables through binary creation. Vtreat is how we can bring our 'plan' into our training data that we use for our model. After vtreat, our treated data consist of many more variables because it has turned all those we want to select [informativeFeatures] to binary.

Graphical user interface, text, application, chat or text message

Description automatically generated

1. Model Explanation

Going through the various models helps us tremendously to see which is the most useful. The decision tree was an extremely useful one as you can visually see the different steps that the model takes in determining the outcome. Random forest is also great as it uses different decision trees across the the dataset while avoiding a problem of overfitting. Logistic regression is the 'ole trusty' model that we have been using a majority of. the time throughout our program. Using vtreat to create binarys make the logistic regression more accurate, higher precision, and better F-1 score than all the other models.

Timeline

Description automatically generated with low confidence

1. Model Exploration-Logistic Regression

Logistic regression was the most successful out of all the models, and that is why we will use it in our prediction. The confusion matrix shows us the amount of times we have been successful and unsuccessful. The most detrimental Is a false negative. Luckily, our true positive and tru negative total results in a higher number than false positive and false negative giving us a accuracy of 73%, higher than any of the other models.

Chart

Description automatically generated with medium confidence

1. Top 100 Prospective Clients

Since we have already created our Logistic Regression model than, all we must do is input our prospective client list into the model to give us an accurate reading. Creating a cutoff, we are making a 'label point' to where everything above is accepted and below is not accepted. From here we can do further work to find the top 100 prospective. To do that, we are having to create a prediction probability column that shwos the probability of the clients saying yes. To do this, we subset the data that has only accepted. From there, we sort by decsending, and then head our top 100 to view the top 100 clients. To print this our, we create a variable from the head(. . . , n=100) that will save the top 100 into a variable that we then can export as a csv file.

Graphical user interface, text, application, email

Description automatically generated

1. Insights

Chart

Description automatically generated

1. Closing

Graphical user interface

Description automatically generated