Matthew Chapman

BUS 440 (Business Intelligence Capstone)

Capstone Project Summary

My project used marketing data from a company in a attempt to create a model that could predict whether customers would accept or decline the current marketing campaign offer. The data had information about the traits of each enrolled customer and their history with the company. Some customer traits include their age, income, education, marital status, and the number of children and teens in their household. The customer history data included information such as their outcome on past marketing campaigns, the amount they've spent on various products, and the types of purchases they make. The target variable for this project was 'Response', which indicates whether the customer has accepted an offer for the current campaign. The goal of the model was to identify customers likely to accept the offer so that marketing efforts can be focused on them. This will save time and money for the company.

I used a logistic regression model to predict the outcome variable. In order to do this I had to manipulate the data. I had to change variables involving dates such as date of birth and date of enrollment to numeric variables (age and number of days enrolled). I also had to change the categorical variables such as education and marital status to multiple binary variables with boolean values. I then had to clean the data by deleting rows with null values and outliers. When I was able to put together my first model, I noticed the metrics were not good. Recall was the most important metric for this use case and it was very low. I used techniques such as adjusting the decision threshold and SMOTE to increase the recall metric.

In the end I achieved nearly a 97% recall metric. While precision was low at 20%, it is more important to reach all customers who are likely to accept the offer. Of the 441 customers in the test set for the model, the model predicted that 134 customers would not accept the offer while maintaining a recall of 97%. In other words, the company can reduce marketing efforts by 30.39% while maintaining the nearly the same amount of accepted offers using this variation of the model.

The most challenging part of this project was deciding between recall and how much we could save on marketing efforts. While other variations of the model had higher precision, from a business perspective I recommend that the model with 97% recall is deployed. This model nearly maintains the same number of accepted offers but reduces costs by 30%. The model also exposes 246 other people to the business who then have increased awareness about the business in the future.