Case Summary

The goal of this project was to develop a model that will effectively predict churn of customers within a given time period. Churn rate is the percentage of customers or subscribers that stop doing business with a company over a specific period of time. The industry looked at in this project was banking. The growth rate of theoretical vs. actual was examined, for those in stock portfolio management, as an example of customer churn on the revenue from that department. Other products such as loans and fees that are missed from churned customers were then also assessed. The proposed solution for churning customers is to provide an effective model to target those at high risk of churning. This in turn will optimize efforts of retention by offering lower rates and incentives to customers at high risk of churning, while keeping legacy customers at their higher rates, thus optimizing profits. Shallow ANN model is the model of choice as it provides a high accuracy with the best precision and recall. This means that we were able to successfully identify high churn risk customers with the least amount of false positives and false negatives. With further tuning of models, it is possible to bring the company's growth curve closer to the theoretical curve, from the current actual, which would scale exponentially.

Work Summary

The goal of this project was to develop a model that will effectively predict churn of customers within a given time period. The only cleaning and feature engineering involved included converting the Gender column into binary output (Male = 1, Female = 0), and creating dummy columns for regions categorized under the Geography column. No NaN’s were present in the initial dataset. In order to counteract the imbalance of “churned” and “not churned” in our target variable, undersampling was conducted for the initial sampling, which improved the precision and recall of the results. The theoretical and actual growth rates were explored using the current data set to show the impact of churn rate on revenue. Logistic regression, Shallow ANN, and ANN models were explored as possible predictors for predicting customer churn. The Shallow ANN model showed decent accuracy while providing much better precision and recall. Precision and recall are the two most important metrics, with recall being the most important metric, for this case, proving it to be the best model for the use case. Stochastic Gradient Descent was used in combination with all models on this project for tuning.