Springboard--DSC
Capstone Project 2
Project Proposal
Cassie Kluse
September 2020

Problem Statement

The goal of this project is to determine whether or not a customer will be approved for a credit card.

Context

Banks get millions of applications for credit cards each year. When making this decision, banks need to consider many characteristics of the applicant including credit score, credit history, current debt, current income, etc. However, choosing who to approve and who to reject can be a confusing and time consuming process. Banks need to make this process automatic to save time and money.

Criteria for Success

Predicting credit card approvals is a binary classification problem, where the results can only be one of two options: ‘approved’ or ‘rejected’. There are multiple models that work well in binary classification problems. Therefore, multiple models will be studied and evaluated for this problem. The best model will have a high performance, as well as account for bias in the data. Furthermore, these models will be used to study connections between the features in the dataset with the approval decision.

Scope of Solution Space

The model will only focus on features in the dataset. Although, new features may be engineered from existing features. The features included in the dataset are: gender, age, marriage status, education level, ethnicity, years employed, prior defaults, current employment status, current credit score, drivers license, citizenship, zip code, income, and whether or not the application was approved.

Constraints

The main constraint is that the dataset is anonymized, which makes it a bit harder to decipher some of the features. This presents itself as an opportunity to be exposed to cases where the dataset is anonymized when mandated.

Stakeholders

The primary stakeholders are the employees who work in the Credit Operations department. This department is in charge of the entire credit card application process, including approving applications and collecting data needed to create this model.

Data Source and Deliverables

The dataset is from the UCI Machine Learning Repository. It contains records of 690 previous credit card applicants and various features including gender,age, credit score, zip code, etc. The data has been anonymized to protect the privacy of customers in the dataset.

All deliverables will be uploaded to my Capstone 2 repository on Github. My deliverables will include separate Jupyter notebooks for each step, as well as a final slide deck and project report.