Springboard: Data Science Career Track Program

Loan Prediction

Capstone Project 3 Project Proposal

Paul Kim

April 2022

Problem Description

There are many people applying for loans today from banks and third parties for many different reasons. However, not all of the people who are applying for loans are getting approval–as it is expected. More importantly, the detailed reasons for rejections are not disclosed for clients sometimes. Therefore, loan eligibility can be modeled using machine learning, no only to build models that estimate the probability of approval for clients, but also to potentially extract suggestions for actionable interventions that maximize the probability of a loan being accepted.s.

Prospective Stakeholders

The stakeholders for this project are clients who are applying for loans, and companies with clients seeking for loans. The clients can use the developed model to check the eligibility with their status before applying for loans. Also in case they are not eligible, the clients can be counseled on the reasons for denial and interventions that might lead to increasing or maximizing the probability of approval.

Also another client for the problem can be third-party loan providers. Once they have client information, they can go through the list to see if there are clients seeking loans but not eligible for approval from the financing institution, so they can reach out to them and offer effective counseling that might lead to future loan approval.

Dataset

The dataset to be used in this project is from Kaggle[1]. The dataset contains thirteen columns including all the different features for the clients, which include finance-related features and non-financial related features, like gender and marital status. The dataset contains 252,000 samples of clients indicating whether they were approved for loans or not.

---

[1]https://www.kaggle.com/datasets/subhamjain/loan-prediction-based-on-customer-behavior?select=Test+Data.csv

Anticipated Data Science Approach

This project will be developed according to the classical Data Science Project Development process: data wrangling, exploratory data analysis, baseline modeling, advanced modeling, and interpretability.

Several models will be developed and compared, keeping in mind the appropriate performance metrics for the business case, and the anticipated imbalance of the dataset.

Several model-based interpretability analyses will be attempted, and in particular counterfactual analyses will be conducted to show how to generate actionable interventions that would lead to the maximization of the likelihood of a loan being approved.

Lastly, the deliverables for this project will be a final report, slide deck and Jupyter notebooks developed.