Capstone Project Proposal

Predicting Peer-to-Peer Loan Payback Amount

Peer-to-peer lenders risk a financial loss if a loan defaults and is not paid back in full. One way to try to minimize the loss is to try try to predict whether a loan will be paid back or not (classification). While predicting if a loan will default or not is valuable, it doesn't take everything into account. For example, there could be a situation where a high-interest loan defaults after being nearly paid off, that is actually pays back the lender more than a low-interest rate loan that is paid in full. Therefore, this project will take a different approach and predict the amount that will be paid back.

My client is an investor that wants to maximize their return on their lending. When an investor obtains a list of available loans they can invest in, they could run the model and generate a prediction of how much will be paid back from their investment, and choose the loans that will maximize their return.

The data for this project comes from lendingclub.com. The data is available for download in csv format.

The approach to solving this problem will be to first inspect the data and perform data wrangling where necessary. Next I will perform exploratory data analysis to better understand the data. Finally, I will do a linear regression, random forest, and xgboost model to predict how much will be repaid and do a comparison to find the best model.

The deliverables for this project include the code written python using Jupyter Notebook, a paper, and a slide deck.