Capstone Project Proposal: Loan Default Prediction

Problem

Create a model to predict the probability of loan-default using Lending Club historical loan data.

Lending Club enables borrowers to create unsecured personal loans and investors can search and browse the loans. Investors select loans based on the borrower's information, loan amount, loan grade, loan purpose and make money from the interest.

The risk to the investor is if the borrower misses the payments and the loan defaults. If the risk is predicted and provided to the investor during the selection of the loans then it helps the investor to make a decision whether to fund the loan.

Data Set

https://www.lendingclub.com/info/download-data.action

The Lending Club dataset includes detailed information for every loan issued by Lending Club from 2007 to 2017. The dataset contains a comprehensive list of features that we can employ to train our model for prediction.

Approach

Data Collection: Extract datasets from Lending club publicly available datasets

Data Preprocessing: Apply data wrangling techniques to clean and organize the data. Ex: handling missing data

Exploratory Data Analysis: Explore data by applying visual and numerical techniques to

- Find insights in the data
- Extract important features
- Detect outliers and anomalies

Modeling: Build different models by applying machine learning algorithms on the training dataset.

Evaluation: Evaluate the performance of each model using test data.

Deliverables

- Code: iPynb on Github
- Report: Methodology write-up
- Slide Deck