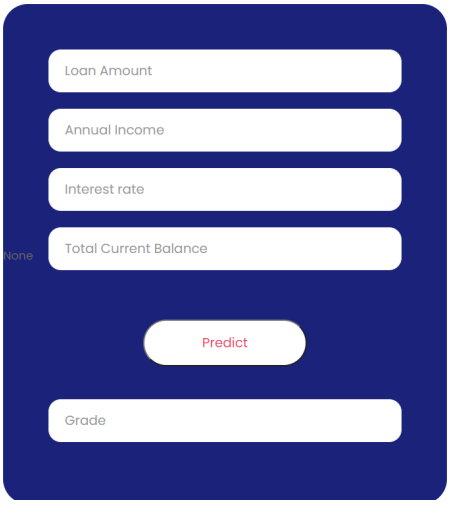
Optimizing Lending Club's Financial



Abstract

Traditionally, loan-level risk is measured as credit risk—the probability of default to measure then expected loss. Using machine learning techniques, we modeled credit risk and expected payoff maximization on the ROC, to help LendingClub optimize their risk.

Introduction

We analyzed LendingClub's dataset of roughly 2.2M loans between 2008–18. We chose to only analyze loans that were paid off in full, charged off or defaulted in this case. There are over 400 borrower characteristics at time of application and loan characteristics at time of issuance. Considering the parameters, we will predict if a person can pay back his loan or not.

Overview

- Examine and overview the dataset
- Building and training the model
- Deployment on Flask

Dateset

The dataset consists of 22 lakh rows and 150 columns about the details of different customers and their history of loan payment. The dataset is in the form of comma-separated values form.

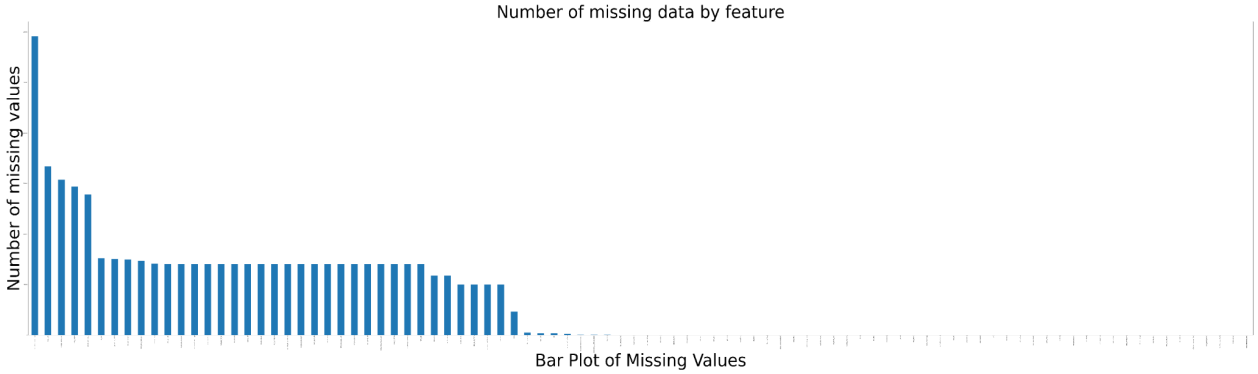
Examine and Visualize the data

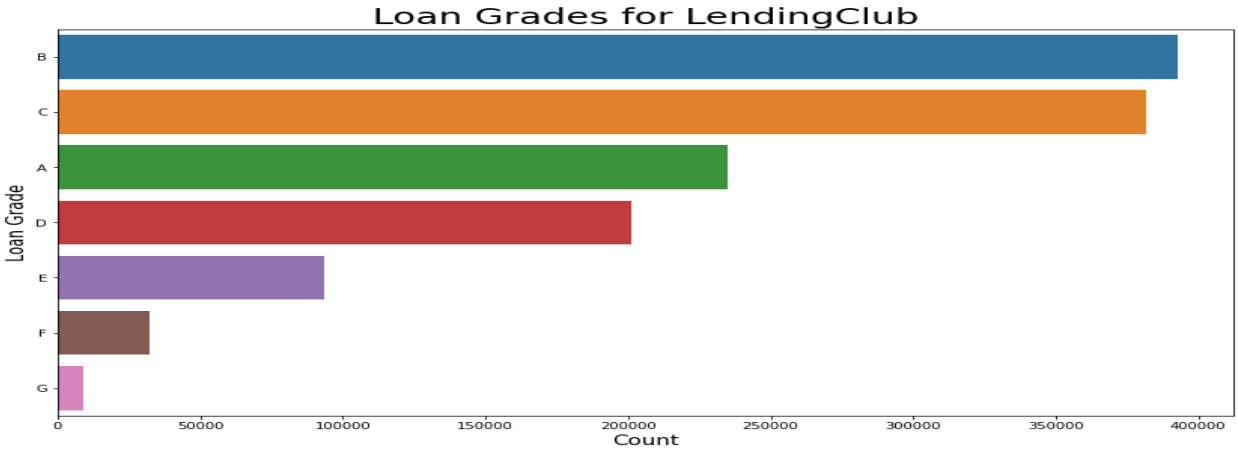
Firstly, we have removed the unnecessary columns like the columns which are not required for prediction.

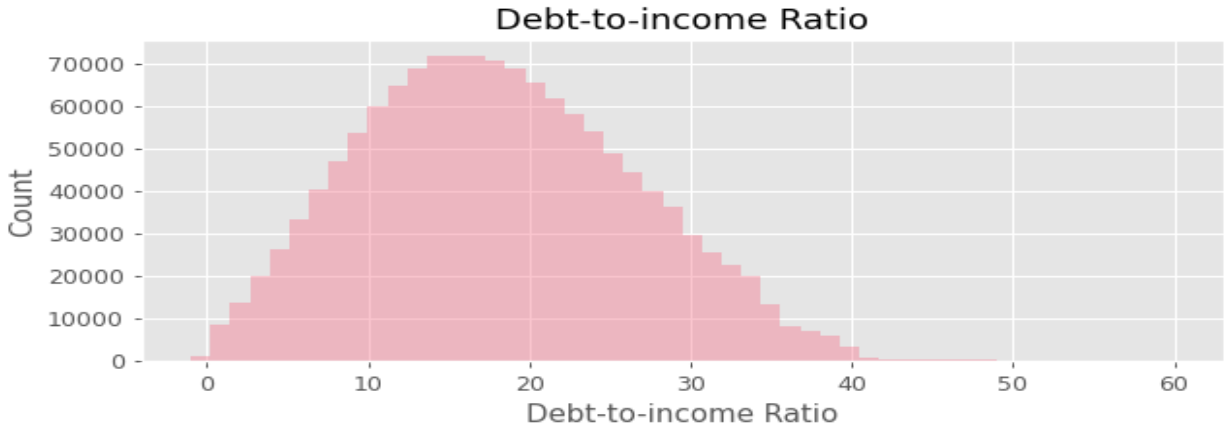
We then found out the columns having missing values and replaced them with the mean of the column.

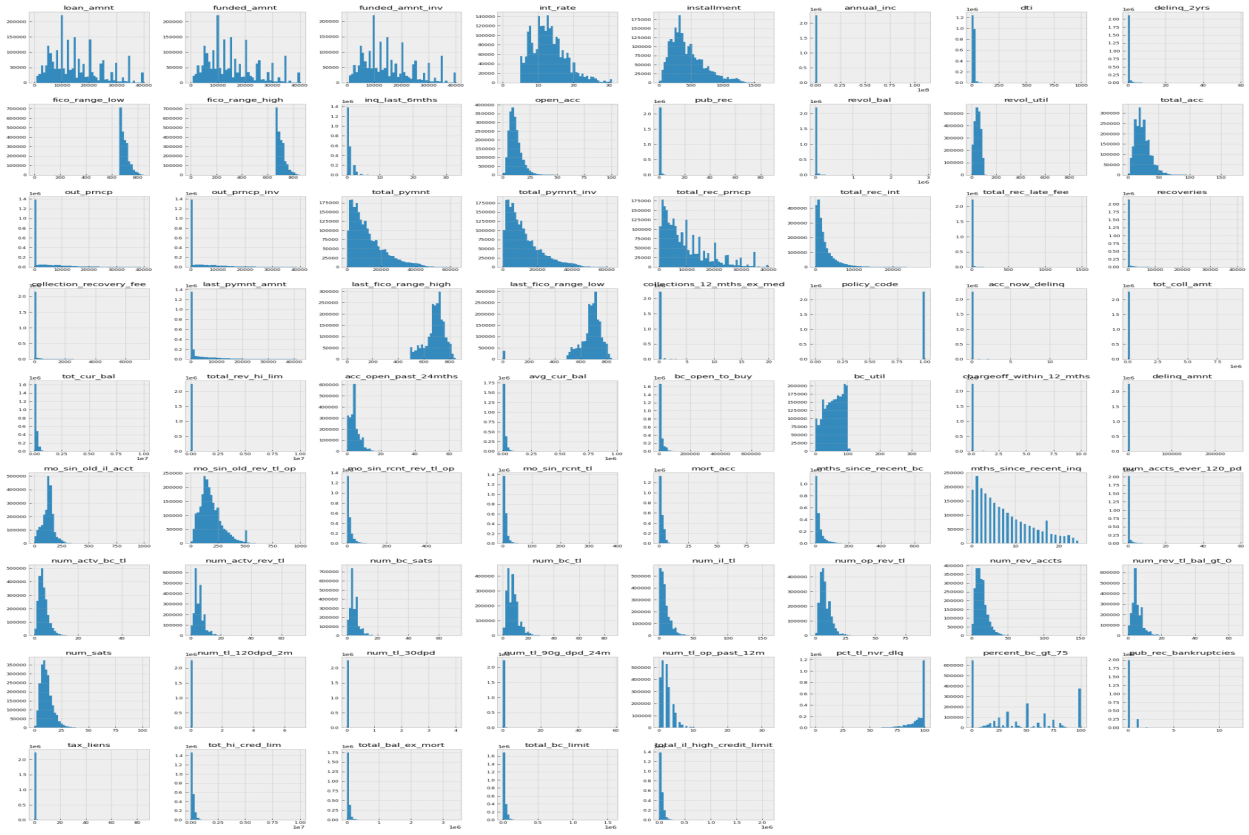
We performed scaling on the columns having non-numerical values to make all the values into one type.

Later, we performed data visualisations to the data like the bar plot, histogram, dist plot









Build and Train Model

- We used Random Forest to predict the loan payment ability. Random Forest is a classifier which contains decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.
- Initially, we split the data into training data and the test data. Train dataset is 80% of the complete data set and the remaining is the test data.
- We then took loan amount, interest rate, total current balance and annual income as variables in X to predict Y.
- After training the data set, we got an accuracy of 97.98%.
- We then converted the file into a pickle file, so that the trained model can be used for deployment.

```
[ ]  from sklearn.metrics import r2_score
```

```
[ ]
     r2_score(y_test,y_pred)

     0.9798399042064594
```

Deployment

- We created a website for lending club using HTML and CSS. Later,we integrated this website with flask to build a web app.
- Then we deployed the web application on heroku

  Website link: https://loancred.herokuapp.com/