

Interest Rate and Loan Default Prediction

Instructor: Prof. Daniel Acuna

Team Members: Kshitij Sankesara, Nitin Nagpal, Siddharth Srivastava, Tavleen Allagh

Problem & Objectives:

Problem 1. We aim to predict the Interest rate based on the financial history of loan borrowers.

Problem 2. Many loans aren't completely paid off on time resulting in borrowers defaulting on their loan. So, we are also developing an algorithm that predicts if the loan will be paid off on time or not.

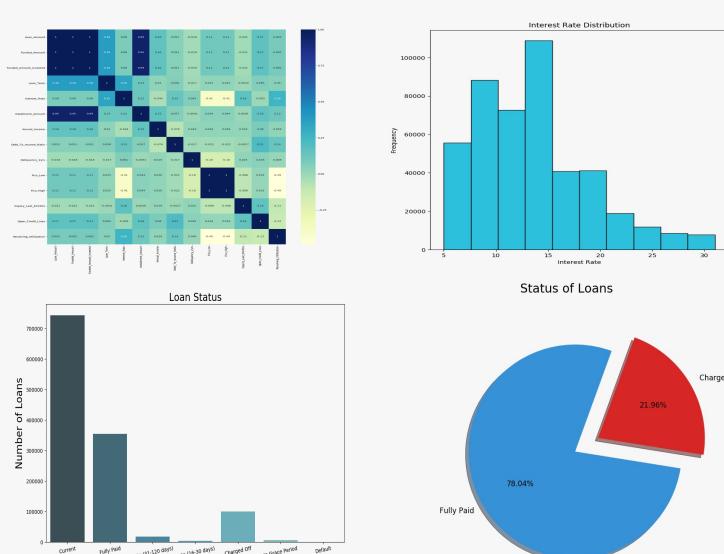
This will help Lending Club in modeling borrower's credit risk and their investors in knowing the chances of getting a return on investments.

Data Description:

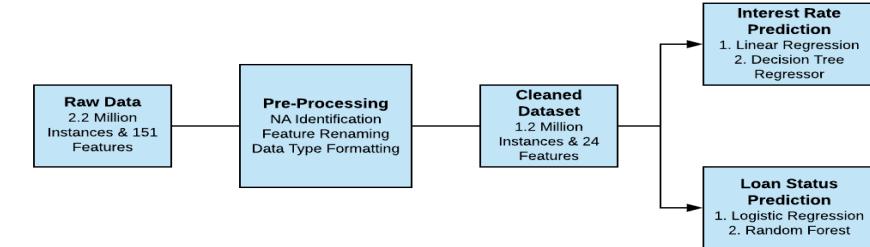
Lending Club is a marketplace for loans that match the loan borrowers with the investors. We have obtained the data from Lending Club's website for our analysis.

The dataset consists of **151** features and about **2.2** million records ranging from 2007 – 2018.

Exploratory Analysis:



Data Flow:



Class Imbalance Problem in Classification:

The data has **22%** defaulters and **78%** non-defaulters. This has created a Class Imbalance problem. To solve this issue, we have used

1. Class weight parameters in Logistic Regression
2. Stratified Sample by Majority down sampling in Random Forest

Model Comparison Metrics:

Interest Rate Prediction - Regression

Model	Hyperparameter Tuning	Evaluation Metric
Linear Regression	maxIter = [10 - 15]	RMSE = 1.239; R^2 = 0.942
Decision Tree Regressor	maxDepth = [10 - 15]; maxBins = [32 - 64]	RMSE = 1.0003

Loan Default Prediction – Classification

Model	Hyperparameter Tuning with CV = 3	Validation AUC
Logistic Regression	RegParam = [0.01, 0.2]; elasticNetParam = [0.4, 0.6]	1
Random Forest	maxDepth = [1, 3, 6]; numTrees = [60, 40, 20]	1

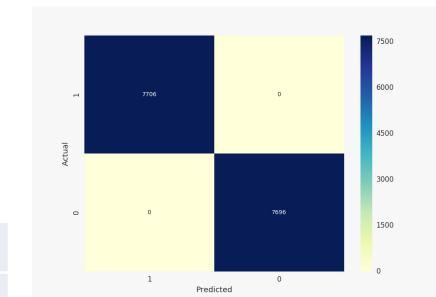
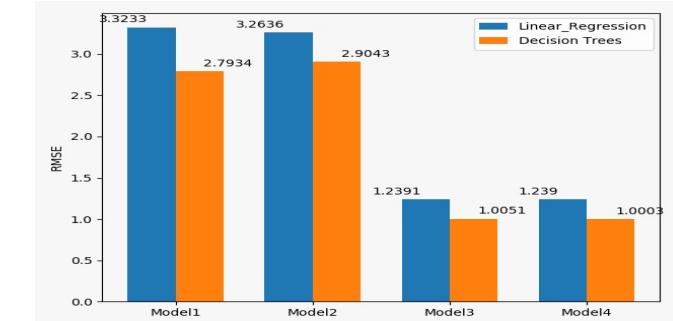
Model Comparison Measures:

Training data: **60** percent, Validation data: **30** percent, Test data: **10** percent
Generalized Performance Measure:

Linear Regression: **RMSE & R²** Decision Tree: **RMSE**

Logistic Regression: **AUC** Random Forest: **AUC**

Evaluation Metrics:



Results & Inference:

For Interest rate prediction, the best model chosen is Decision Tree Regressor with **RMSE of 1.0003**.

For Loan Default prediction, both Logistic Regression and Random Forest are showing **AUC of 1**.

Debt_To_Income_Ratio is the most significant feature that predicts the Loan defaults.

Conclusion:

Our analysis can be used by Lending Club to set different Interest rates for different individuals based on their financial history.

We can also predict if the loan will be paid off on time or not.

Therefore, Investors can also leverage the model to know the chances of getting a return on their Investments.