MCIT 591 – Final Project: Loan default predictor

Ian Park (ianspark@seas.upenn.edu) / Antonius Santoso (asantoso@seas.upenn.edu)

20th Apr 2019 - Update

Our final project is a ***program to predict loan default-rate based on machine learning algorithms***. The program is using Lending Club dataset from Kaggle (https://www.kaggle.com/wendykan/lending-club-loan-data) and the machine learning used is Logistic Regression.

The program has 2 parts:

1. Machine Learning program (back-end)

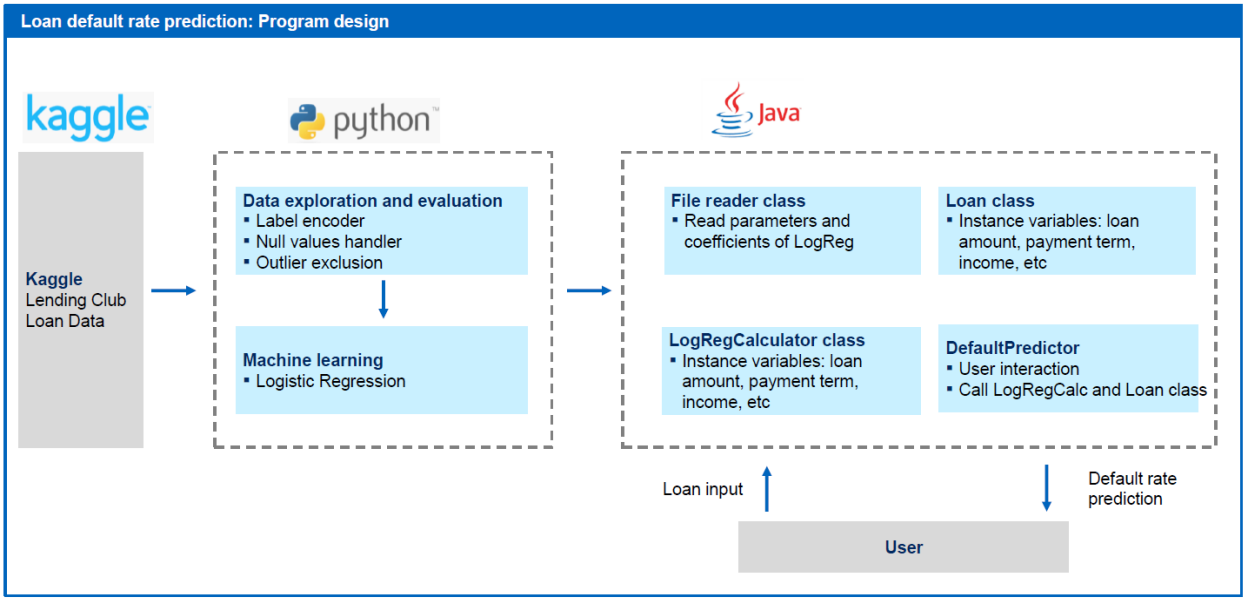
Machine Learning program is done in Python because it's easier to perform data analyses and feature engineering as well as conduct machine learning operation in Python. This Python program will load the Lending Club data and use it to train and test ML model using Logistic Regression technique. The output of this program will be regression parameters and coefficients.

Milestones achieved: Performed data exploration and analysis (Feature Engineering), Built ML model with Logistic regression

2. Loan Predictor program (front-end)

The loan predictor program is done in JAVA that will serve as main program for loan prediction. This program will receive Logistic regression parameters and coefficients from ML program and perform prediction of loan default based on user inputs.

Milestones achieved: Developed JAVA classes (DefaultPredictor, FileReader, Loan, LogRegCalculator), Built JUnit test cases



Next milestones: Refining ML model, figuring out as how to communicate between Python and JAVA, Building JAVA GUI for user interaction