SDAIA Academy T5C04 Bootcamps: Data Science

Classification Module Project Proposal

Classification: Predicting the risk of loan default

Date: October 19, 2021

Prepared for: Nawras Boufaied

Prepared by: Nada Rambu

Objective

The loan is one of the most important schemes of banks. Usually, the banks are willing to give loans to the customers based on their credit histories. However, there is always a risk associated with the loan such as late payments and loan defaults. As a result, banks are making a great effort to manage this risk. They primarily seek to assess their customers' repayment abilities well in advance before deciding on approval or rejection of loans.

The goal of this project is to build a machine learning classification model to predict whether a customer's loan would be approved or denied depending on his or her risk of defaulting. There are two distinct classes to group the datapoints by: a loan approval or a loan denial.

Data Description

The data for this project will be read into a CSV file (Find the dataset on the following link: https://www.kaggle.com/panamby/bank-loan-status-dataset/data). The obtained dataset consists of over 111,000 loan records with 18 features, which are listed below:

- **Loan ID -** A unique Identifier for the loan information.
- **Customer ID -** A unique identifier for the customer. Customers may have more than one loan.
- **Loan Status -** A categorical variable indicating if the loan was paid back or defaulted. This will be the target (dependent) variable.
- **Current Loan Amount -** This is the loan amount that was either completely paid off, or the amount that was defaulted.
- **Term -** A categorical variable indicating if it is a short term or long term loan.

- **Credit Score -** A value between 0 and 800 indicating the riskiness of the borrower's credit history.
- **Years in current job -** A categorical variable indicating how many years the customer has been in their current job.
- **Home Ownership -** Categorical variable indicating home ownership. Values are "Rent", "Home Mortgage", and "Own". If the value is OWN, then the customer is a home owner with no mortgage.
- **Annual Income -** The customer's annual income
- **Purpose -** A description of the purpose of the loan.
- **Monthly Debt -** The customer's monthly payment for their existing loans
- **Years of Credit History -** The years since the first entry in the customer's credit history
- **Months since last delinquent -** Months since the last loan delinquent payment
- **Number of Open Accounts -** The total number of open credit cards
- **Number of Credit Problems -** The number of credit problems in the customer records.
- **Current Credit Balance -** The current total debt for the customer
- **Maximum Open Credit -** The maximum credit limit for all credit sources.
- **Bankruptcies -** The number of bankruptcies
- **Tax Liens -** The number of tax liens.

Tools

The following tools will be used to carry out the project:

1. **Pandas** library will be used to create data frames for easier data manipulation.
2. **Scikit-learn** library will be to implement the classification models.
3. **Matplotlib** library will be used to visualize and discuss the results of the analysis.