Loan Default Prediction

Background:

You are provided a dataset for vehicle loan default prediction where you need to predict the customer who will default for paying the EMI (Equated Monthly Instalments).

Financial institutions incur significant losses due to the default of vehicle loans. This has led to the tightening up of vehicle loan underwriting and increased vehicle loan rejection rates. The need for a better credit risk scoring model is also raised by these institutions. This warrants a study to estimate the determinants of vehicle loan default. A financial institution has hired you to accurately predict the probability of loanee/borrower defaulting on a vehicle loan in the first EMI on the due date. Following Information regarding the loan and loanee are provided in the datasets:

Loanee Information (Demographic data like age, Identity proof etc.)

Loan Information (Disbursal details, loan to value ratio etc.)

Bureau data & history (Bureau score, number of active accounts, the status of other loans, credit history etc.)

Doing so will ensure that clients capable of repayment are not rejected and important determinants can be identified which can be further used for minimising the default rates.

Tasks:

1. Split the training data into training and validation sets.
2. Using the training and validation data sets to select classification models to predict loan default. You are expected to use at least **three** methods that you learned in our class.
3. Select the best model from all your models. Predict the loan default status in the test set using the best model.
4. Report the performance (accuracy, F1-score for each class, AUC) for each of your model using different methods.

Report:

1. Submit a file of your Python notebook or R script which contains all codes and model results. Please explain what model you choose to use and what result each model produces. Please also add comments to your codes.
2. Submit a csv containing the test data and your predicted default status as the last column.
3. Submit these two files separately. Do not Zip them together.

Evaluation:

The submission will be evaluated on methodology, report (together 60%), and the F1-score of the default class in the test set (40%).

Academic Integrity

You can use the codes provided in our course. Please use the appropriate referencing. This is an **individual project**. You can discuss with each other about the project. But you need to write your own code. Your submission will be compared with online material and each other to check for similarity using plagiarism software.

In the end, the goal for this project is to sharpen your own skills!