

Assignment 13

Objective

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 (Equated Monthly Instalments) on the due date.

Dataset

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 minimizing the default rates.

Data: <https://www.kaggle.com/avikpaul4u/vehicle-loan-default-prediction>

Possible Work to be done

1. Write a Data Science Proposal for achieving the objective mentioned.
2. Perform exploratory analysis on the data.
3. Perform data wrangling / pre-processing.
4. Apply any 2 features engineering technique.
5. Plot top 10 features.
6. Identification of the performance parameters to be improved, for the given problem statement.
7. Design Machine learning models – Logistic regression and Decision tree to predict.
8. Compare the performance of selected feature engineering techniques.
9. Compare the performance of the 2 classifiers – Logistic regression and Decision tree to predict.
10. Present the conclusions/results in the format shared.