Group Work Handout - Credit Scoring with R

SBD2 - Data-driven Visualization and Decision Making

September 27, 2024

The main objective of the group work assignment is to produce a R markdown file on credit scoring using a data set uploaded on Moodle.

Note: Each group works with a different version of the dataset. Namely, each group should work with the dataset that corresponds to their group number in Moodle. For example: group 1 should work with the dataset "loan_sample_1".

Throughout the script, you should use comments to include descriptions/interpretation of results as well as answers to specific questions. The target variable of the analysis is the "Status" feature included in the dataset which contains the values 0 if the loan contract did not default and 1 if the loan contract defaulted. The description of the other variables included in the data set is provided in Table 1.

Exercise 1

Using the dataset 'loan_data.csv,' please conduct an exploratory data analysis. Perform quality control checks to identify and address any issues present in the data, ensuring that all steps taken are thoroughly explained and justified.

Exercise 2

Using the pre-processed dataset obtained in Exercise 1, apply a machine learning model to predict loan defaults. Select and justify the choice of model and evaluate its performance.

Exercise 3

Thinking about the preprocessing steps that you carried out before modeling and what you learned about the dataset, is there a way in which you could improve the performance of the model? How?

Exercise 4

Finally, think about putting your model into action and basing credit decisions on the prediction that it generates. Which opportunities and challenges your model brings to daily business decisions? How could companies mitigate the challenges you described?

Variable	Description
annual_inc	The self-reported annual income provided by the borrower during registration.
application_type	Indicates whether the loan is an individual or joint application
dti	Borrower's total monthly debt payments divided by monthly income.
grade	Assigned loan grade by the financial service provider
home_ownership	The home ownership status
int_rate	Interest Rate on the loan
loan_amnt	The listed amount of the loan applied for by the borrower
open_acc	Number of open trades in last 6 months
purpose	A category provided by the borrower for the loan request.
revol_bal	Total credit revolving balance
revol_util	Revolving line utilization rate.
tot_cur_bal	Total current balance of all accounts
total_acc	The total number of credit lines currently in the borrower's credit file
total_rec_int	Interest received to date
total_rev_hi_lim	Total revolving high credit/credit limit.
verification_status	Indicates if the co-borrowers' joint income was verified

Table 1: Variables' description