

MACHINE LEARNING 30412 – Project 2024 DATA

Banks use credit scoring algorithms to estimate the probability of default when deciding whether a loan should be granted.

In the project, your task will be to build a credit scoring algorithm predicting the probability that somebody will experience financial distress in the next two years. The goal of this project is to build a model that loan providers can use to help borrowers make the best financial decisions.

Files

You can download from Blackboard the following csv files:

- *train.csv*: training data that you can use to train your algorithms;
- *test.csv*: test data for which you have to make predictions;
- *submission.csv*: an example of the structure of the final csv file that you'll have to upload as part of your final submission.

Description of data

- **SeriousDlqin2yrs** Person experienced 90 days past due delinquency or worse (Y/N) (**this is your target!**)
- **RevolvingUtilizationOfUnsecuredLines** Total balance on credit cards and personal lines of credit except real estate and no installment debt like car loans divided by the sum of credit limits (**percentage**)
- **Age** Age of borrower in years (**integer**)
- **NumberOfTime3059DaysPastDueNotWorse** Number of times borrower has been 30-59 days past due but no worse in the last 2 years (**integer**)
- **DebtRatio** Monthly debt payments, alimony, living costs divided by monthly gross income (**percentage**)
- **MonthlyIncome** Monthly income (**real**)
- **NumberOfOpenCreditLinesAndLoans** Number of open loans (installment like car loan or mortgage) and lines of credit (e.g. credit cards) (**integer**)
- **NumberOfTimes90DaysLate** Number of times borrower has been 90 days or more past due (**integer**)
- **NumberRealEstateLoansOrLines** Number of mortgage and real estate loans including home equity lines of credit (**integer**)
- **NumberOfTime60-89DaysPastDueNotWorse** Number of times borrower has been 60-89 days past due but no worse in the last 2 years. **integer**
- **NumberOfDependents** Number of dependents in family excluding themselves (spouse, children etc.) (**integer**)

Metric

For each test input you will have to predict the probability of financial distress in the next two years.

The metric with which we will evaluate your predictions is AUC (see, for instance, [here](#) for details or ISL page 155).