Title: Prediction of clients’ repayment capabilities in an international sales finance company

Abstract

In credit markets, people with inadequate or deficient credit history are less likely to secure a loan, though this population might be of lower income groups and perhaps in a more severe need of getting a loan. Home Credit, a Eurasian consumer finance provider, seeks to provide lending opportunities to this underserved population with less or no credit history. One main challenge is to identify clients who need loans but also are capable of repayment. In this project, we aim to predict the repayment capabilities of Home Credit clients, with a prediction diagnostic ability calculated using ROC curve. We will use the Home Credit Default Risk (HCDR) dataset that contains numerous categorical and numeric features regarding clients’ loan application history, monthly credits balance, history of monthly balance of loans and credit card loans in Home Credit, and payment history. We will make pipelines to transform and standardize the variables, and impute the missing values with mean, median, and multivariate feature imputation strategies. After preprocessing the dataset, we will use different classifiers, including but not limited to logistic regression, different ensembles of decision trees, Support Vector Machine, and a voting classifier. We will report the test and training data set accuracy of the classifiers.