

HOME CREDIT DEFAULT RISK PREDICTION

Rajith Lakshan Wijewickrama Rathnayake

A dissertation submitted in partial fulfilment of the requirement
for Master of Science degree in Advanced Software Engineering

**Department of Computing
Informatics Institute of Technology, Sri Lanka
in collaboration with
University of Westminster, UK**

2020

Abstract

There is a risk associated with any loan because the borrower may or may not repay the loan(Zamore et al., 2018). Measuring this risk before approving the loan is a difficult task because there are so many parameters which may affect the repayment of the loan. And these parameters may vary from one applicant to another. Banks and financial institutes try to minimize this risk by requesting for guarantors and recommendations. Even though this will reduce the risk, there is no accurate method to measure the probability of risk in a loan.

Financial institutes have a large collection of past data of loan applicants and the repayment details. These can be used to predict the outcome of a new loan application(Lakhani et al., 2019). Since the amount of data is large and the data is not organised, data analysis tools needs to be used for processing the data accurately and faster. Pre-processed data were used for predicting the result of a new loan application using machine learning. Income of the client, credit amount of the loan, loan annuity, number of days before the application the person started current employment, number of days before the application did client change the identity document with which he applied for the loan.

Home Credit Default Risk Prediction System implements machine learning models with different algorithms and predict the credit default risk of a loan application. The algorithms used are Logistic regression, random forest classification and neural network. The models are trained using all the features in the dataset and then trained using the selected features with most importance to the outcome.

The user can train the system with a training dataset for the selected models and predict the outcome of a new loan application using the trained models. The results will be stored in CSV format.

Keywords: Credit Default Risk, Machine learning, Python, Neural Networks, Algorithm selection.