Informatics Institute of Technology
In collaboration with
University of Westminster

Credit scoring with macroeconomic factor analysis using machine learning techniques

A dissertation by

Mr. Jithendra Anthony Dilrukshan Thenuwara

Supervised by

Mr Sameer Hamza

Submitted in partial fulfilment of the requirements for the BEng (Hons)

Software Engineering Degree Department of Computing

BEng/BEng. (Hons) in Software Engineering

May 2019

Abstract

Non-performing loans (NPLs) are loans which are several months overdue. An increase in NPLs can cause a huge impact to the economy and has even caused financial crisis in the past. Moreover, macroeconomic conditions such as GDP, Inflation and unemployment rate have an impact on the increase or decrease of NPL ratios. Credit scoring is the method of evaluating good clients from bad clients, therefore improving the accuracy of the credit scoring process would in turn reduce the NPL ratios. Many advanced machine learning models have been developed to assist lenders to identify the good clients from the bad clients, however there are only a few models which consider the transparency of the machine learning results and furthermore no models which consider the use of macroeconomic conditions in the credit scoring. This research project aims to fill this gap by selecting a suitable machine learning algorithm to provide transparency in the results and to incorporate macroeconomic conditions in to machine learning models to improve the prediction accuracy.

The results of the implementation suggested that the boosted decision tree algorithm produces the greatest accuracy from the selected machine learning algorithms and all machine learning models which were trained by including macroeconomic conditioned performed better that the machine learning models which were trained without including macroeconomic conditions. The final conclusion derived from this research project is that macroeconomic conditions have an impact on NPLs and the inclusion of these macroeconomic conditions in the machine learning models produces higher accuracy in prediction of defaults.

Key Words:

Credit Scoring, Machine Learning, Non-performing loans, Classification Algorithms, Boosted Decision Trees