PREDICTION OF REPAYMENT STATUS OF INDIVIDUAL BORROWERS IN THE SRI LANKAN BANKING INDUSTRY USING SUPERVISED MACHINE LEARNING ALGORITHMS

Ravi Wijewardane Rathnayake

A dissertation submitted in partial fulfilment of the requirement for Master of Science degree in Business Analytics

Department of Computing

Informatics Institute of Technology, Sri Lanka
in collaboration with
Robert Gordon University, Scotland

2021

ABSTRACT

The banking sector has faced numerous challenges in recent years, including economic downturns, increasing non-performing advances, and unexpected shocks such as the Easter attack and the COVID-19 pandemic. As a result, banks have had to implement various strategies to manage risks prudently and navigate through these turbulent times successfully. One crucial aspect of this is identifying probable overdue borrowers and predicting their repayment status in advance.

This study aimed to identify the factors that impact individual borrowers' repayment status and explore the feasibility of identifying probable customers to be categorized into different repayment statuses. To achieve the objectives, seven machine learning models that incorporated ensemble and boosting techniques were utilized: Decision Tree, Random Forest, K-Nearest Neighbours, XG Boost, Ada Boost, Gradient Boosting Machine and Recurrent Neural Network Multi-Layer Perception. Out of these, the XG Boost classifier was able to accurately classify borrowers into different repayment statuses, achieving overall accuracy, macro average precision/recall/F1 scores of 99%. Accordingly, the study found the most significant factors influencing an individual borrower's repayment status such as risk rating, principal paid to date, interest paid to date, repayment period, number of personal guarantors etc.

In this research, we discuss the challenges associated with predicting the repayment status of borrowers such as data preprocessing, handling outliers and missing values and propose solutions to overcome these challenges. We focus on using machine learning models, including ensemble and boosting techniques, highlighting the importance of feature selection, imbalanced data and hyperparameter tuning. We also provide empirical evidence to support our proposed solutions by evaluating the accuracy of the models on real-world data from the banking sector. By addressing these challenges and improving the accuracy of predicting the repayment status of borrowers, banks can manage risks prudently and navigate through turbulent times successfully.