

LOAN DEFAULT PREDICTION IN ABC FINANCE COMPANY- A MACHINE LEARNING CASE STUDY

MSc Business Analytics

Academic Year – 2021 / 2022

Module: CMM 799 – MSc Project - BA

IIT ID: 20191244	RGU ID: 1915097	STUDENT NAME: Thamali Palihadaru
-------------------------	------------------------	---

DECLARATION

I certify that the attached material is my original work. I have not used another person's work or ideas without acknowledgment. I have not submitted the contents of this assignment for examination / assessment in any other course or unit at this or any other institution unless specifically stated.

Date	05th May 2023
-------------	----------------------

Abstract

Finance companies equally play a vital role in the Financial System in Sri Lanka. Loans and Advances represent the majority of the asset class of the financial position which accounted for 76.8% of the total assets in 2021. While financing through credit being a driving force of the economy, it has become the biggest risk for any financial institution. Non- Performing Loans (NPL) of the Non- Banking Finance sector have also been on the rise, reaching 13.9% and 11% in 2020 and 2021. As a solution to the problem statement, we have conducted a case study on employing machine learning models for credit risk modeling in a well reputed Finance Company in Sri Lanka. Five machine learning models namely, Logistic Regression, Decision trees, Random Forest, XGboost and Adaboost were used and evaluated the performance of each model. It was revealed that XGBoost model outperforms the other models with the highest model performance. (Accuracy (0.78), Weighted Avg Precision (0.87), Weighted Avg Recall (0.77), Weighted Avg F1-score (0.82) and AUC-ROC (0.60)). It was evident that feature selection and hyperparameter optimization will impact the performance of the model. Correlation coefficient heatmap, Chi- Squared test of independence, Select K best, Recursive Feature Elimination and Random Forest Feature importance were used as feature selection. The highest performances were shown with Random Forest feature importances. It was observed that the importance of features namely, No of Rental (Term), Effective Rate, Rental, Age, Income, LTV and vehicle age are found to be significantly high. Addition to that, we discuss model interpretability using LIME method.