ROBERT GORDON
UNIVERSITY ABERDEEN
SCHOOL OF COMPUTING SCIENCE AND DIGITAL MEDIA

MSc Business Analytics

Student Name: G.G.T.Lakshan Galanga	Matriculation Number: 2119002
Supervisor: Dr. Alroy Mascrenghe, PhD	Second Marker:
Project Title: Machine learning approach for Loan default Prediction for Sri Lankan Banking Industry	
MSc Business Analytics	Start Date:
	Submission Date: 5th May 2023

CONSENT

I agree ■
I do not agree ☐

That the University shall be entitled to use any results, materials or other outcomes arising from my project work for the purposes of non-commercial teaching and research, including collaboration.

DECLARATION

I confirm:

- **That the work contained in this document has been composed solely by myself and that I have not made use of any unauthorised assistance.**

- **That the work has not been accepted in any previous application for a degree.**

- **All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.**

Student Signature: Lakshan Galanga	Date Signed: 25/05/2023

Abstract

With the economic, socialist breakdown & covid 19 outbreak Sri Lankan banking industry is having their hardest years of the history. The citizens of Sri Lanka have encountered these factors, which have significantly contributed to the rise of non-performing loans in the banking industry. Carried literature review proved that the independent variables such as interest rate, loan amount, installment amount, tenor & income have a significant impact on the decision variable loan default.

The current report showcases a loan default prediction model that employs seven distinct algorithms. The goal of this study is to develop & identify a model that can accurately predict loan defaults based on borrower characteristics and loan attributes. The dataset used for this study consists of historical loan data which is downloaded from Kaggle website. There is 0.68 positive correlation between the loan amount and the property values of the dataset & 0.46 positive correlation between the loan amount and the income. The author has found a negative correlation between LTV and the property value which is -0.41 and a negative correlation between the income & the debt to ration value which is -0.25.

Various data preprocessing techniques were used to clean and transform the dataset before training and testing the targeted models. The results show that the XGBoost model achieved a high accuracy of 84.6% and F1-score of 90.37%. The precision and recall of the model were 96.09 and 85.26% respectively. The findings of this research contribute to the field of loan default prediction by providing valuable insights into the efficacy of machine learning models in accurately identifying borrowers at risk of default. These insights can assist financial institutions in making informed decisions regarding loan approvals, setting interest rates, and implementing risk mitigation strategies.

Future work may involve exploring additional data sources and refining the model to further improve its predictive performance.