Credit Risk Analysis of Small and Medium-sized Enterprise Loans

G.D.T.D. Chandrasiri
198744F

Faculty of Information Technology
University of Moratuwa
July 2022

Credit Risk Analysis of Small and Medium-sized Enterprise Loans

G.D.T.D. Chandrasiri
198744F

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the fulfillment of the requirements of Degree of Master of Science in Information Technology

July 2022

Declaration

We declare that this thesis is our own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student
G.D.T.D Chandrasiri

Signature of Student
UOM Verified Signature

Date:

Supervised by

Name of Supervisor
Dr. S.C. Premaratne

Signature of Supervisor
UOM Verified Signature

Date: 29/07/2022

Dedication

I dedicate my research work affectionately to the following:

My research supervisor, Dr. S.C Premaratne, for providing the valuable guidance on this and for dedicating his valuable time for me,

My parents, for providing me all the courage and their precious love,

My sisters and brother, for helping me as always,

Finally, to all my best friends for supporting me in many ways.

Thank you.

Acknowledgments

First, I would like to convey my sincere gratitude towards my research supervisor, Dr. Saminda Premaratne, Senior Lecturer, Faculty of Information Technology, University of Moratuwa, for his valuable guidance, motivation, and valuable time spent during this research project.

Further, I would like to thank Dr.M.F.M Firdhous who conducted Research Methodology course and Literature Review and Thesis Writing course which laid the foundation for this research and Dr. Chaman Wijesiriwardane who coordinated the research project course.

It is a pleasure to thank all my collegues of the 13th batch of MSc. in IT degree program who encouraged me always.

Importantly, I would like to extend my heartiest thank to my family who gave me their precious love through my entire life. Finally, I must be thankful for all the people whose names are not appeared here, but their diligent effort was very important to make this study a success.

Abstract

Analyzing the credit risk is important in banking systems to ensure that debtors pay the loans regularly, on schedule. The inability of managing the credit risk may lead severe losses in financial institutes. Every financial institute must predict and manage the credit risk to avoid financial crises. Hence, finding an effective method for credit risk analysis is vital. Among various types of loans, Small and Medium-sized Enterprise (SME) loans dominate since SMEs are considered the backbone of any economy. With the higher amount of SME loans, the associated risk also gets increased. SME sector is considered as risky and costly than the large enterprises. The amount of non-performing SME loans has increased at a higher pace throughout the last few quarters in Sri Lanka. Hence, this study analyzes the credit risk of SME loans by using a data set received from a financial institute in Sri Lanka. It recognizes financial and non-financial attributes affecting the credit risk of SME loans. The data mining techniques were used to analyze the data and extract knowledge. It is an emerging technology that provides significant improvements in terms of making accurate decisions. Data mining is used widely for financial analysis since it facilitates knowledge extraction from large data sets and making effective decisions. The study will help financial institutes to predict the credit risk of potential SME borrowers and avoid inefficiencies in the lending process. It identifies the credit risk of debtors in different aspects. Ultimately, it provides valuable insights into effective decision making of an economy.

Keywords: *Credit Risk, SME Loans, Data Mining*

Contents

Chapter 1 : Introduction 1
1.1 Prolegomena 1
1.2 Background of the Study 1
1.3 Problem Statement 3
1.4 Aim and Objectives 4
1.4.1 Aim 4
1.4.2 Objectives 4
1.5 Proposed Solution 4
1.6 Structure of the Thesis 5
Chapter 2 : Literature Review 6
2.1 Introduction 6
2.2 Credit Risk Analysis of SME Loans 6
2.3 Data Mining Techniques for Credit Risk Analysis 10
2.4 Summary 12
Chapter 3 : Adopted Technologies 13
3.1 Introduction 13
3.2 Data Mining 13
3.3 Data Mining Process 14
3.4 Supervised Learning 15
3.5 Classification Algorithms 15
3.5.1 Decision Tree 15
3.5.2 K- Nearest Neighbor 16
3.5.3 Naïve Bayes 16
3.5.4 Rule Induction 17
3.5.5 Random Forest 17
3.5.6 Linear regression 18

3.5.7 Logistic Regression 18
3.5.8 Artificial Neural Networks (ANN) 18
3.5.9 Support Vector Machines (SVM) 19
3.6 Data Mining Tools 19
3.7 Summary 20
Chapter 4 : Methodology for Analyzing the Credit Risk of SME Loans 21
4.1 Introduction 21
4.2 Hypothesis 21
4.3 Input 21
4.4 Output 21
4.5 Process 22
4.6 Users 22
4.7 Features 22
4.8 Data Selection 23
4.9 Data Preparation 24
4.10 Data Pre-Processing 25
4.11 Data Mining Scheme 25
4.12 Summary 26
Chapter 5 : Analysis and Design 27
5.1 Introduction 27
5.2 Research Design 27
5.3 Detailed Research Design 27
5.3.1 Research Question One 28
5.3.2 Research Question Two 28
5.3.3 Research Question Three 28
5.4 Analysis 29
5.5 Summary 29

Chapter 6 : Implementation 30
6.1 Introduction 30
6.2 Solution for Research Question One 30
6.3 Solutions for Research Question Two 30
6.4 Solutions for Research Question Three 30
6.5 Data Preprocessing 31
6.6 Data Mining Algorithms 31
6.7 Parameter Tuning 32
6.8 Cross Validation 32
6.9 Split Validation 33
6.10 Summary 34
Chapter 7 : Evaluation 35
7.1 Introduction 35
7.2 Model Evaluation 35
7.3 Evaluation Parameters 35
7.4 Confusion Matrix 36
7.5 Prediction 01 – Status of the payments 37
7.6 Prediction 02 – Payment Percentage of Non-performing Loans 38
7.7 Prediction 03 – Exit Method 39
7.8 Prediction 04 – Exact Exit Point 40
7.9 Prediction 05 – Number of Exit Points 40
7.10 Summary 41
Chapter 8 : Conclusion and Future Works 42
8.1 Introduction 42
8.2 Conclusion 42
8.3 Limitations 42
8.4 Future Development 43

8.5 Summary 43
References 44
Chapter 9 : Appendix 51
9.1 Statistical Visualizations of the Data Set 51
9.2 Evaluation Results of the Selected Algorithm in Prediction 01 53
9.3 Evaluation Results of the Selected Algorithm in Prediction 02 54
9.4 Evaluation Results of the Selected Algorithm in Prediction 03 55
9.5 Evaluation Results of the Selected Algorithm in Prediction 04 56
9.6 Evaluation Results of the Selected Algorithm in Prediction 05 57

List of Figures

Figure 3.1: Stages in Data Mining Process 14
Figure 3.2: Scalable Architecture of RapidMiner Studio 20
Figure 5.1: The High Level Design of the Proposed Model 28
Figure 6.1: Data Mining Process in RapidMiner 33
Figure 7.1: Confusion Matrix 36
Figure 7.2: Performance Vector of Prediction 01 38
Figure 9.1: Statistical Details of SME Loan Data - 1 51
Figure 9.2: Statistical Details of SME Loan Data - 2 51
Figure 9.3: Gender-wise Representation of Loan Amount 52
Figure 9.4: Business Type and Gender-wise Representation of Loan Amount 52
Figure 9.5: Accuracy of Random Forest Algorithm 53
Figure 9.6: Results of the Prediction 01 53
Figure 9.7: Confusion Matrix of Prediction 01 53
Figure 9.8: Performance Vector of Prediction 02 54
Figure 9.9: Results of the Prediction 02 54
Figure 9.10: Results of the Prediction 03 55
Figure 9.11: Performance Vector of Prediction 03 55
Figure 9.12: Performance Vector of Prediction 04 56
Figure 9.13: Results of the Prediction 04 56
Figure 9.14: Performance Vector of Prediction 05 57
Figure 9.15: Results of the Prediction 05 57

List of Tables

Table 2.1: Summary of machine learning techniques used in predicting credit risk ... 12
Table 4.1: The Definitions of Variables 23
Table 7.1: Prediction 01- Performance Metrics of Different Algorithms 37
Table 7.2: Prediction 02- Performance Metrics of Different Algorithms 39
Table 7.3: Prediction 03- Performance Metrics of Different Algorithms 39
Table 7.4: Prediction 04- Performance Metrics of Different Algorithms 40
Table 7.5: Prediction 05- Performance Metrics of Different Algorithms 41