

Credit Analysis using Data Mining Techniques in Banking Sector

N.A.U.H.Jayarathna

179465N

Master of Science in Information Technology

Faculty of Information Technology

University of Moratuwa

2020

Credit Analysis using Data Mining Techniques in Banking Sector

N.A.U.H.Jayarathna

179465N

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

June 2020

DECLARATION

We declare that 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 Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.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

Signature of Student

N.A.U.H..Jayarathna

.....

Date: 09th of June 2020

Supervised by

Name of Supervisor

Signature of Supervisor

S.C. Premarathne

.....

Date:

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest sincere gratitude towards my supervisor, Mr. Saminda Premarathne, Senior Lecturer, Faculty of Information Technology, University of Moratuwa for his guidance, supervision, advices and sparing valuable time thorough the research project.

I offer my obeisance to Dr.M.F.M Firdhous who taught Research Methodology and Literature Review and thesis writing subjects which were the basis for this research work.

I would also like to thank all the batch mates of the M.Sc. in IT degree program who gave their valuable feedbacks to improve the results of the research and my family for the support they provided me through my entire life and in particular.

Finally I would like to offer my heartiest gratitude to all the people whose names are not appeared but their untiring effort was very much crucial to make this study success throughout this work.

ABSTRACT

In the financial market, banking sector is one of the major sectors. The main objective of a bank is to maximize their shareholders returns. While maximizing the shareholders returns, they have to bear number of risks. Credit risk is one of their major risks. Credit risk is the risk that the bankers have to bear when they give loan facilities to the customers. Deciding whether the borrower is suitable to get the loan is such a long process. Currently this process is a manual process in the banks and the final decision is based on the credit officers' opinions.

This study has focused on to analyze the credit analysis of businesses using data mining techniques. Basic aim of this study is to sought and to analyze the best data mining techniques which can be used to credit analysis and appraisals of businesses in banking sector in order to get the accurate decisions by minimizing human errors. .

In this study it is empirically evaluated current techniques which are using for credit appraisals and the best data mining techniques which can be used to minimize the human errors in the banking sector. The sample consisted of 1500 records taken from a private bank in Sri Lanka which gives loan facilities to Small and Medium Scale Enterprises.

Keywords: Credit Analysis, Data mining, Banking Sector

Table of Contents

DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT.....	v
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER ONE	1
INTRODUCTION	1
1.1 Chapter Introduction	1
1.2 Background of the Study.....	1
1.3 Research Problem.....	2
1.4 Scope of the research.....	3
1.5 Aim of the research	3
1.6 Research Questions and Objectives	3
1.7 Proposed Solution of the study	4
1.8 Organization of the Dissertation	4
2 CHAPTER TWO	5
LITERATURE REVIEW	5
2.1 Chapter Introduction	5
2.2 Credit risk in Banking Sector	5
2.3 Issues in credit analysis and appraisal process.....	7
2.4 The need of using data mining techniques in banking sector	7
2.5 Data Mining Techniques	8
2.6 Use of data mining techniques for credit analysis and appraisal process	9
2.7 Chapter Summary.....	10
3 CHAPTER THREE	11
TECHNOLOGY ADAPTED.....	11
3.1 Chapter Introduction	11
3.2 Data Mining Techniques	11
3.3 Rapid Miner.....	14
3.4 Chapter Summary.....	14
4 CHAPTER FOUR.....	15
METHODOLOGY	15

4.1	Chapter Introduction	15
4.2	Profile of the Sample.....	15
4.3	Hypotheses	15
4.4	Input	16
4.5	Output.....	16
4.6	Process.....	16
4.7	Data Preprocessing.....	16
4.8	Classification Algorithms.....	17
4.9	Chapter Summary.....	18
5	CHAPTER FIVE	19
	RESEARCH DESIGN AND ANALYSIS.....	19
5.1	Chapter Introduction	19
5.2	Research Design.....	19
5.3	Detailed Research Design	20
5.3.1	Sub Research Question One.....	20
5.3.2	Sub Research Question Two.....	20
5.4	Chapter Summary.....	21
6	CHAPTER SIX.....	22
	IMPLEMENTATION.....	22
6.1	Chapter Introduction	22
6.2	Solution for Research Question One.....	22
6.3	Solutions for Research Question Two.....	22
6.3.1	Data Preprocessing.....	23
6.3.2	Attribute Selection	23
6.3.3	Correlation Matrix	24
6.3.4	Model Creation	25
6.3.5	Cross Validation.....	25
6.3.6	Split Validation	26
6.4	Chapter Summary.....	26
7	CHAPTER SEVEN	27
	EVALUATION.....	27
7.1	Chapter Introduction	27
7.2	Evaluation for Classification.....	27

7.3	Summary of the Accuracy of Classification Algorithm.....	28
7.4	External factors affected to the credit decision	30
7.5	Chapter Summary.....	30
8	CHAPTER EIGHT	31
	DISCUSSION AND FUTURE WORK	31
8.1	Chapter Introduction	31
8.2	Overview of the research.....	31
8.3	Limitations	32
8.4	Future work of the project.....	32
8.5	Chapter Summary.....	32
	REFERENCES	33
9	APPENDICES	35
9.1	Appendix A -Data Preprocessing -Discretization	35
9.2	Appendix B – Correlation Matrix	36
9.3	Appendix C – Results of Cross Validation and Split Validation	37

LIST OF TABLES

Table 1-1 Research Questions and Objectives.....	03
Table 6-1 Discretization.....	16
Table 7-1 Confusion Matrix.....	20
Table 7-2 Accuracy of Classification Algorithms.....	21
Table 7-3 Cross Validation Results.....	21
Table 7-4 Split Validation Results.....	22

LIST OF FIGURES

Figure 3-1 – Data Mining Techniques	13
Figure 7-1 – Summary of Accuracy Levels	13
Figure 9-1 - Discretization	35
Figure 9-2 - Attribute Selection	36
Figure 9-3 - Correlation Matrix I.....	36
Figure 9-4 - Correlation Matrix II.....	37
Figure 9-5- Cross Validation - Decision Tree.....	37
Figure 9-6 - Cross Validation - KNN	38
Figure 9-7 - Cross Validations - Naive Bayes	38
Figure 9-8 - Cross Validation - Random Forest.....	39
Figure 9-9 - Split Validation - Decision Tree	39
Figure 9-10 Split Validation - KNN	40
Figure 9-11 - Split Validation - Naive Bayes	40
Figure 9-12-Split Validation - Random Forest	41