UNIVERSITI TEKNOLOGI MARA

DEVELOPING DEBT RISK MODEL USING DATA MINING FOR IRBM

HAMISAH BINTI MOHAMED AZIDIN

IT Project submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Information Technology

Faculty of Computer and Mathematical Sciences

July 2015

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of University Technology MARA. It is original and is the result of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

In the event that my report be found to violate the conditions mentioned above, I voluntarily waive the right of conferment of my degree and degree to be subjected to the disciplinary rules and regulations of University Technology MARA.

Name of Student	:	Hamisah binti Mohamed Azidin
Student I.D. No.	:	212696034
Programme	:	Master of Science in Information Technology
Faculty	:	Computer and Mathematical Sciences
IT Project Title	:	Developing Debt Risk Model Using Data Mining for IRBM
Signature of Student	:	.....
Date	:	July 2015

ABSTRACT

This paper describes the development of a debt risk predictive model for individual taxpayer in Inland Revenue Board Malaysia (IRBM). Using data mining to predict taxpayer's compliance and non-compliance has gained attention in recent times. However, very little research has been done to predict taxpayers who have debt with tax organizations around the world. The objective of this study is to choose a suitable data mining methodology or framework and what is suitable data mining technique to build a predictive model debt risk for taxpayers who have debts with IRBM. This study also to get the behavior or pattern of data and identify important variables in predicting taxpayer with debt risk. Data individual taxpayers who have debt value until 31.12.2013 obtained from the database data warehouse and data mart IRBM. These data were analyzed using IBM SPSS version 16.0 applications. From the study of data mining framework and technique in theory, we choose CRISP-DM data mining life cycle framework include business understanding, data preparation, build models, evaluation and test the model as a framework and CHAID decision tree as technique. The results showed that the method CHAID Decision Tree in the model built to the study of type of test the value of 80% accurate. The result revealed that he is accurate model to use to predict future taxpayers which have outstanding debt with IRBM.

ACKNOWLEDGEMENT

Alhamdulillah, praise to Allah the Almighty and The Merciful for giving me the strength and health to the completion of this research. Without His blessings and consent, I might not have enough courage and determination to complete this research. All my thanks and appreciation will be lay upon Him.

My deepest gratitude is extended to Associate. Prof. Madziah Hj Hashim, for all assistance, advice, guidance, encouragement, new ideas and invaluable support given as my project supervisor for a better quality in my research. I also would like to express my gratitude and sincere appreciation to Associate. Prof. Dr.. Haryani binti Haron for her invaluable knowledge, comment and recommendation on how to enhance my research.

A special thanks to all staff of ReMS Division, Information Technology Department of IRBM, all the lecturers, friends and colleagues of Master Science (Information Technology) for their support and encouragement during the process of completing this research.

Finally, I would like to express my deepest gratitude to my family and parents for their continuous support, understanding and patience during my journey in this study. Their supports were just what I needed to give me the spirit and strength to keep on going. Without their personal sacrifices and being a constant source for encouragement, especially in the final stages, this thesis would not have been possible.

Thank You.

TABLE OF CONTENTS

	Page
AUTHOR'S DECLARATION	**i**
ABSTRACT	**ii**
ACKNOWLEDGEMENT	**iii**
TABLE OF CONTENTS	**iv**
LIST OF TABLES	**vii**
LIST OF FIGURES	**viii**
CHAPTER ONE: INTRODUCTION	1
1.1 Introduction	1
1.2 Background of Study	1
1.2.1 Inland Revenue Board Malaysia	3
1.3 Problem Statement	4
1.4 Research Questions	5
1.5 Research Objectives	5
1.6 Significance of Study	5
1.7 Scope of Study	6
1.8 Report Outline	6
CHAPTER 2: LITERATURE REVIEW	8
2.1 Definition of Data Mining	8
2.2 Data mining Framework	9
2.2.1 CRISP-DM (Crisp-DM stand for Cross Industry Standard Process)	9
2.2.2 SEMMA (Sample, Explore, Modify, Model and Assess)	13
2.2.3 KDD (Knowledge Discovery in Database)	15
2.2.4 Comparison between the Data Mining Framework	16
2.3 Data Mining Techniques	17
2.3.1 Neural Network	17