Universiti Teknologi MARA

Development of a Prototype Application to Detect Bad Behaviour Pattern

Mohamad Khairul Firdhaus B. M. Dollah

Thesis submitted in fulfillment of the requirements for
Bachelor of Computer Science (Hons)
Faculty of Information Technology And
Quantitative Science

November 2008

ACKNOWLEGEMENT

In the name of Allah, Most Gracious and Most Merciful

First and foremost, I would like to praise Allah, with Allah will, I finally able to finish this project proposal although it took me quite some time. This project is fulfillment of the requirements for Bachelor of Computer Science (Hons). I am thankful for the opportunity to conduct this proposal and for the strength and the capability that Allah S.W.T gave me throughout the process.

Special thanks to my supervisor Dr. Siti Salwa Salleh, for the opportunity to take me as her supervisee and for her trust. I appreciated her time for helping, guiding, and encouraging me throughout the process and I'm sure it's not wasted.

Thanks to my family for always supporting and trusting me in whatever I do especially my parent. Thanks for the love and kindness and always be there for me, they are always in my heart.

Thanks also to my coordinator, PM. Syed Ahmad Sheikh Aljunid and also to all of CS230's lecturers. Thanks for everything they have taught me.

Lastly, I want to thank you all again for everything you have done to me. Every help for the project proposal you gave to me is not wasted at all, the completion of the project is the result of your kindness of time you spent for me. My Allah blesses you all.

Abstract

Bad behaviour can be prevented by knowing their behaviour patterns before the problem occurs. Data mining is one way that can help in predicting or identifying the patterns behavior of certain individual that is involved in terrorist activity. Data mining is a process of finding hidden information and patterns in databases. One of the tasks for data mining purposes is classification refers to assigning cases into categories based on a predictable attributes. This method can be applied to detect patterns of bad behaviour UiTM students. By using auto discretization and rules compositions based on table and decision tree, classification method of data mining can be applied to the prototype application. This project is developing a prototype application to identify patterns of bad behavior among student in Universiti Teknologi Mara (UiTM), Shah Alam. The data is obtain at Unit Tatatertib UiTM Shah Alam. There are 98 data collected from year 2006 to 2007 which consist of the student data attribute that is age, gender, faculty code, semester, year, type of offence and the case status. A test has been done to validate the rules generated in the process. The result also shows the patterns of behavior that is acquired. The result

Keyword: Data Mining, Classification Methods, Bad Behaviour, Pattern detection, Decisions Tree, Discretizations

TABLE OF CONTENTS

CONTENT	PAGE
DECLARATION	ii
APPROVAL	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
LIST OF TABLE	ix
LIST OF FIGURES	x
CHAPTER 1: INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement	2
1.3 Research Question	2
1.4 Research Purpose	3
1.5 Objective	3
1.6 Scope of Project	3
1.7 Project Aim/Contribution	4
CHAPTER 2: LITERATURE REVIEW	5
2.1 Introduction	5
2.2 Data Mining Overview	5
2.3 Basic Data Mining Tasks	10
2.3.1 Classification	10
2.3.2 Regression	10
2.3.3 Time Series Analysis	11
2.3.4 Prediction	11
2.3.5 Clustering	11

2.3.6 Summarization 12
2.3.7 Association Rule 12
2.3.8 Sequence Discovery 12
2.4 Common Technique in Data Mining 13
2.5 Discretization Overview 14
2.5.1 Discretization Technique 15
2.6 Decision Tree Overview 17
2.7 Related Works 19
2.5 Crime Pattern Detection Using Data Mining 19
2.6 Using Data Mining Technique for Detecting Terror-related Activities on the Web 21
2.7 Microsoft Time Series Algorithm 21

CHAPTER 3: METHODOLOGY 23
3.1 Introduction 23
3.2 Project Outline 26
3.3 Information Gathering 26
3.4 Identifying Algorithm 27
3.5 System Development 27
3.5.1 Indentify Student Attribute 28
3.5.1.1 Database Design 29
3.5.2 Applying Discretization 30
3.5.3 Applying Classification Algorithm 33
3.6 System Testing 35
3.7 Documentation 35
3.8 Project Planning 36
3.9 Hardware and Software Requirement 36
3.10 Gantt Chart 37
3.10 Summary 37