

**Universiti Teknologi Mara**

**Prediction of Secondary Public School  
Selection Based on Parents' Preferences  
Using Naïve Bayes**

**Mohamad Hafizi Bin Masdar**

**Thesis submitted in fulfilment of the requirements  
for Bachelor of Computer Science (Hons.) Faculty  
of Computer and Mathematical Sciences.**

**July 2017**

## **STUDENT DECLARATION**

I certify that the thesis and the project to which it refers is the product of my own work and that any ideas or quotation from the work of other people, published or otherwise are fully acknowledged in accordance with the standard referring practices of the discipline.

.....  
MOHAMAD HAFIZI BIN MASDAR  
2014247684

July 24, 2017

## ABSTRACT

Choosing a school for the secondary students is not a simple task nowadays since there are many factors needed to be considered by parents. These might be an overhead to the parents to filter each school and try to sort it out based on their requirements. Therefore, a predictive model which is a quantitative research that implemented artificial intelligence (AI) strategy using Naïve Bayes (NB) technique been applied in order to predict the chosen secondary school in Gombak based on parents' preferences. Through six different independent variables (IV) namely distance of home from school (DSH), sibling enrolment in the same school (SE), highest parent's level of education (PLE), household income (HI), employment status (ES) and race, 64 prediction models were produced and tested. Among all these models, the highest accuracy detected was the combination of four IVs which are DSH, PLE, ES and race, that resulted more than 55% of accuracy. This finding then implemented into a system which is Prediction of Secondary Public School System (PSPSS) to help parents making school selection easier since parents' concerns are taken into consideration. NB provided a new way of solving a problem in determining the vital factors for making a school choice. Since this is the preliminary and exploratory study, the highlighted achievement is not the accuracy percentage of the prediction, but a new angle on the ability of how AI can solve the problem within the domain of this study. This finding can be further improved by using the same technique which is NB through modification of IV used and the number of data collected. Other AI techniques such as support vector machine (SVM), multiple linear regression (MLR) and artificial neural network (ANN) also can be used to compare with the current NB as well as an enhancement to readily available findings. This study also may be broadened to other regions or states to help more parents making school selection easier.

## TABLE OF CONTENTS

CONTENT	PAGE
<b>SUPERVISOR APPROVAL</b>	ii
<b>STUDENT DECLARATION</b>	iii
<b>ACKNOWLEDGEMENT</b>	iv
<b>ABSTRACT</b>	v
<b>TABLE OF CONTENT</b>	vi
<b>LIST OF FIGURES</b>	x
<b>LIST OF TABLES</b>	xi
<b>LIST OF ABBREVIATIONS</b>	xii
<b>CHAPTER ONE: INTRODUCTION</b>	1
1.1 Background of Study	1
1.2 Problem Statement	2
1.3 Objectives	3
1.4 Project Scope	4
1.5 Significance of Project	6
<b>CHAPTER TWO: LITERATURE REVIEW</b>	7
2.1 Introduction	7
2.1.1 School Definition	8
2.1.2 Secondary Schools in Malaysia	9
2.1.2.1 Regular National Secondary School (SMK)	9
2.1.2.2 Fully Residential Secondary School (SBP)	10
2.1.2.3 Technical and Vocational Education	10
2.1.2.4 Government-Aided Religious School (SABK) and National Religious Secondary School (SMKA)	11

	2.1.2.5 Malaysia Art School	11
	2.1.2.6 Malaysia Sports School	11
2.1.3	Parents and the School Choice	12
2.1.4	Parents' Role towards Children Education	13
2.1.5	Factors and Related Problems Regarding School Choice	13
	2.1.5.1 Academic Performance	13
	2.1.5.2 School's Location	14
	2.1.5.3 Co-curricular Activities	14
	2.1.5.4 School's Distance	15
	2.1.5.5 Family Background	15
	2.1.5.6 Social and Racial Composition	16
2.2	Previous Research	17
2.3	Artificial Intelligence (AI)	19
	2.3.1 Machine Learning	20
	2.3.1.1 Clustering	20
	2.3.1.2 Optimization	20
	2.3.1.3 Controlling	20
	2.3.1.4 Classification	21
	2.3.2 AI Techniques under Classification Type	21
	2.3.2.1 Multiple Linear Regression (MLR)	24
	2.3.2.2 Support Vector Machine (SVM)	24
	2.3.2.3 Artificial Neural Network (ANN)	25
	2.3.2.4 Naïve Bayes (NB)	26
2.4	Conclusion	28
<b>CHAPTER THREE: METHODOLOGY</b>		<b>29</b>
3.1	Introduction	29
	3.1.1 Agile Model	29
3.2	Detail of Project in Agile	32
	3.2.1 Overview of Project Framework	33
	3.2.2 Analysis Phase	34
	3.2.3 Planning Phase	35
	3.2.4 Design and Development Phase	35
	3.2.5 Testing Phase	36
	3.2.6 Documentation Phase	36