AN APPLICATION OF PREDICTING STUDENT PERFORMANCE USING KERNEL K-MEANS AND SMOOTH SUPPORT VECTOR MACHINE

SAJADIN SEMBIRING

MASTER OF SCIENCE (COMPUTER)

UNIVERSITI MALAYSIA PAHANG

AN APPLICATION OF PREDICTING STUDENT PERFORMANCE USING KERNEL K-MEANS AND SMOOTH SUPPORT VECTOR MACHINE

SAJADIN SEMBIRING

Thesis submitted in fulfillment of the requirements
For the award of the degree of
Master of Science (Computer)

Faculty of Computer Systems & Software Engineering
UNIVERSITI MALAYSIA PAHANG

AUGUST, 2012

UNIVERSITI MALAYSIA PAHANG

DECLARATION OF THESIS AND COPYRIGHT

Author’s full name : ______________________________

Date of birth : ______________________________

Title : ______________________________

Academic Session : ______________________________

I declare that this thesis is classified as :

□ **CONFIDENTIAL** (Contain confidential information under the Official Secret Act 1972)*

□ **RESTRICTED** (Contain restricted information as specified by the organization where research was done)*

□ **OPEN ACCESS** I agree that my thesis to be published as online open access (Full text)

I acknowledge that University Malaysia Pahang reserve the right as follows:

1. The Thesis is the Property of University Malaysia Pahang
2. The library of University Malaysia Pahang has the right to make copies for the purpose of research only
3. The library has the right to make copies of the thesis for academic exchange.

Certified by:

(Student’s Signature)

(Signature of Supervisor)

New IC/Passport number
Date :

Name of Supervisor
Date :

NOTES : *If the thesis is CONFIDENTIAL or RESTRCTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Master of Science (Computer).

Signature
Name of Supervisor: MOHD. AZWAN MOHAMAD@ HAMZA
Position: LECTURER
FACULTY OF COMPUTER SYSTEMS & SOFTWARE ENGINEERING
UNIVERSITI MALAYSIA PAHANG
Date: AUGUST, 2012

STUDENT'S DECLARATION

I hereby declare that the work in this thesis is my own except for quotations and summaries, which have been duly acknowledged. The thesis has not been accepted for any degree and is not concurrently submitted for award of other degree.

Signature
Name: SAJADIN SEMBIRING
ID Number: MCC 08001
Date: AUGUST, 2012

DEDICATION

Dedicated to my parents, family, friends, and my country.

Their positive contributions throughout my life have made this work possible.

ACKNOWLEDGEMENTS

I am grateful and would like to express my sincere gratitude to my supervisor Mohd. Azwan Mohamad @ Hamza and Professor. Dr. Abdullah Embong for their germinal ideas, invaluable guidance, continuous encouragement, and constant support in making this research possible. They always impressed me with their outstanding professional conduct, their strong conviction for science, and his belief that a Master program is only start of a life-long learning experience. I appreciate their consistent support from the first day I applied to graduate program to these concluding moments. I am truly grateful for their progressive vision about my training science, his tolerance of my naïve mistakes, and their commitment to my future career. I also sincerely thanks for the time spent proofreading and correcting my many mistakes.

My sincere thanks go to all my lab mates and the staff of Faculty of Computer Systems & Software Engineering, UMP, who helped me in many ways and made my stay at UMP a pleasant and unforgettable. Many special thank go to members of Data Mining & Knowledge Management research group for their co-operation, inspirations and supports during this study.

I acknowledge my sincere indebtedness and gratitude to my parents, Mr. Syarifuddin Sembiring and Mrs. Pekenasa Br Tarigan, my wife Siti Salehati, Ama.Pd for their love, dream and sacrifice throughout my life. I acknowledge the sincerity of my-parents-in-law, Sabaruddin Tarigan (Alm) and Kasmi Nurjannah Br Sembiring, who consistently encouraged me to carry on my higher studies in Malaysia. I am also grateful to my brothers; Muhammad Sidik Sembiring, BSc, Ahmad Fajar Sembiring, BA, Syahrial Sembiring, BSc and special thank to my sister Siti Rohani Br Sembiring (Almh) and her husband for their support, sacrifice, patience, and understanding that were inevitable to make this work possible. I cannot find the appropriate words that could properly describe my appreciation for their devotion, support and faith in my ability to attain my goals. Special thanks should be given to my committee members. I would like to acknowledge their comments and suggestion, which was crucial for the successful completion of this study.

ABSTRACT

This thesis presents the model of predicting student academic performances in Higher Learning Institution (HLI). The prediction of students' successful is one of the most vital issues in HLI. In the previous work, there are many methods proposed to predict the performance of students such as Scholastic Aptitude Test (SAT) or American College Test (ACT), Intelligent Test, Fuzzy Set Theory, Neural Network, Decision Tree and Naïve Bayes. However, the fact remains found in a variety of debate among educators in higher learning institution, especially those related to predictor variables that used and the resulting level of prediction accuracy. This shown that the rule model in predicting student performance is still a gap and it is urgent for educators to obtain a more accurate prediction results. The objective of this study is to create a rule model in predicting of students performance based on their psychometric factors. In this study, psychometric factors used as predictor variables, there are Interest, Study Behavior, Engaged Time, Believe, and Family Support. The rule model developed using Kernel K-means Clustering and Smooth Support Vector Machine Classification. Both of these techniques based on kernel methods and relatively new algorithms of data mining techniques, recently received increasingly popularity in machine learning community. These techniques successfully applied in processing large amounts of data, especially on high dimensional data that are non-linearly separable. The data collection from student academic databases and surveyed the psychometric factors of undergraduate student in semester 3 sessions 2007/2008 at Universiti Malaysia Pahang. The result of this study indicates a positive correlation between the proposed predictor variables and the students' performance. These predictor variables contribute significantly in increasing or decreasing student performance that is equal to 52.2% (R^2=0.522). The study also found the cluster model of students based on their performance. Each member of the clusters labeled with their performance index to describe the current condition of student performance. The prediction accuracy of predicting model proposed have the lowest accuracy 61% (R^2 = 0.61) in predicting "Good" performance index and the highest accuracy 93.67% (R^2 = 0.9367) in predicting "Poor" Performance index. This study showed that the kernel method has a capability as data mining technique on educational data mining. The results of this study are suitable to be used in monitoring the progression of students' performance semester by semester and supported the decision making process by decision maker in HLI.

ABSTRAK

Tesis ini menghasilkan model peramalan prestasi akademik pelajar bagi Institusi Pengajian Tinggi (IPT). Meramal kejayaan pelajar telah menjadi satu isu yang amat penting di IPT. Dalam kajian rintis yang dilakukan, terdapat beberapa kaedah yang dicadangkan untuk membuat ramalan prestasi pelajar seperti Scholastic Aptitude Test (SAT) atau American College Test (ACT), Intelligent Test, Fuzzy Set Theory, Neural Network, Decision Tree dan Naïve Bayes Namun begitu masih terdapat banyak fakta yang diperdebatkan di kalangan pendidik di IPT khususnya berkaitan pembolehubah ramalan yang digunakan serta tahap ramalan yang dihasilkan. Ini menunjukkan bahawa masih terdapat jurang yang menyebabkan keperluan untuk membangunkan model peraturan dalam meramal prestasi pelajar yang mendesak para pendidik untuk mendapatkan hasil ramalan yang lebih tepat. Tesis ini bertujuan untuk mencipta model peraturan dalam meramal prestasi pelajar berdasarkan faktor-faktor psikometri mereka. Di dalam kajian ini, faktor psikometri yang digunakan sebagai pembolehubah ramalan adalah Minat, Sikap Pelajar, Penggunaan Masa, Kepercayaan, dan Sokongan Keluarga. Model peraturan ini dibangunkan dengan menggunakan Kernel K-means Clustering dan Pengkelasan Smooth Support Vector Machine. Kedua-dua teknik ini adalah berdasarkan kaedah kernel yang merupakan satu algoritma baru dalam teknik perlombongan data yang kini semakin banyak digunakan dalam bidang mesin pembelajaran. Teknik-teknik ini telah berjaya dilaksanakan untuk pemprosesan data dalam jumlah yang besar, terutama bagi data berdimensi tinggi yang bersifat non-linear berasingan. Pengumpulan data adalah daripada pangkalan data akademik mahasiswa manakala kajian ke atas faktor psikometri adalah daripada pelajar sarjana muda semester 3 sesi 2007/2008 di Universiti Malaysia Pahang. Keputusan daripada kajian ini menunjukkan bahawa hubungan antara pembolehubah ramalan yang dicadangkan dengan prestasi pelajar mempunyai korelasi positif. Pembolehubah-pembolehubah ramalan memberikan sumbangan yang signifikan dalam meningkatkan atau menurunkan prestasi pelajar iaitu sebanyak 52.2% (R^2=0.522). Kajian ini juga mendapati terdapat model klaster terhadap pelajar berdasarkan prestasi mereka. Setiap ahli dari klaster telah dilabel dengan indeks prestasi mereka bagi menggambarkan keadaan semasa bagi prestasi pelajar. Ketepatan ramalan bagi model peramalan yang dicadangkan mempunyai ketepatan terendah 61% (R^2 = 0.6100) dalam membuat peramalan indeks prestasi "Baik" dan ketepatan tertinggi 93.67% (R^2 = 0.9367) dalam membuat peramalan Indeks Prestasi "Lemah". Kajian ini membuktikan bahawa kaedah kernel boleh digunakan dan sesuai sebagai teknik perlombongan data dalam bidang pendidikan. Keputusan kajian ini juga sesuai digunakan untuk memantau perkembangan prestasi mahasiswa dan juga dapat meningkatkan proses membuat keputusan oleh yang membuat keputusan di IPT.

TABLE OF CONTENTS

CHAPTER 3 METHODOLOGY

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Φ	Non-linear mapping function
δ	Value Indicator function
Θ_u	Variable denoting the cluster
λ_i	Lagrangian Multiplier
$\bar{\xi}_i$	Non-negative slack variable
α	Smoothing parameter
γ	Value of parameters
μ	Value of parameter for non linear case data.

LIST OF ABBREVIATIONS

UMP	Universiti Malaysia Pahang
ANN	Artificial Neural Network
PAMS	Performance Assessment Monitoring System
CGPA	Cumulative Grade Performance Academic
SSVM	Smooth Support Vector Machines
KDD	Knowledge Discovery Databases
DM	Data Mining
CURE	Clustering Using Representatives
SVD	Singular Value Decomposition
EM	Expectation-Maximization
PAM	Partitioning Around Medoids
CLARA	Clustering LARge Applications
SVM	Support Vector Machines
RKHS	Reproducing Kernel Hilbert Space
SRM	Structural Risk Minimization
GSVM	Generalize Support Vector Machines
SPSS	Statistical Package for Social Science
ANOVA	Analysis of Variance
KEEL	Knowledge Extraction based on Evolutionary Learning
10-CV	Tenfold Cross Validation
RMSE	Root Means Square Error
VIF	Variant Inflation Factor
RBF	Radial Basis Function

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND OF THE PROBLEM

Data mining is to mine the knowledge interested by people from a great deal of data, and this knowledge is the useful information but connotative and prior unbeknown (Han and Kamber, 2003). The data mining techniques is acquired to satisfy application in many fields, and application of data mining techniques in university can accelerate the innovation and development of the education system. The data mining technology can find useful knowledge from a great deal of data, and this knowledge provides important foundation to improve the process of decision making in management system of university (Luan, 2002).

Clustering and classification are two of the most common data mining tasks used frequently for data categorization and analysis in both industry and academia (Han and Kamber, 2006). Clustering is the process of organizing unlabeled objects into groups which members are similar in some way (Larose, 2006). Clustering is a kind of unsupervised learning algorithm. It does not use category labels when grouping objects. Classification is the procedure to assign class labels. A classifier is constructed from the labeled training data using certain classification algorithm, and then it will be used to predict the class label of the test data. Classification is a kind of supervised learning algorithm.

Unsupervised data mining in the educational data used for situation in which particular groupings or patterns are unknown. For example, not much is known about which courses are usually taken as a group, or which course types are associated with which student types.

Supervised data mining on educational data used to predict the group membership of the data instances that given works of a student, one may predicate his/her final grade. Classification rules are prediction rules to describe the future situation.

Data mining in higher education is a new emerging field, called Educational Data Mining (Romero and Ventura, 2007). Educational Data Mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in (Baker and Yacef, 2009).

Data mining, also called Knowledge Discovery in Databases (KDD), is the field of discovering novel and potentially useful information from large amounts of data (Witten and Frank, 1999). It has been proposed that educational data mining methods are often different from standard data mining methods, due to the need to explicitly account for (and the opportunities to exploit) the multi-level hierarchy and non-independence in educational data (Baker, 2008). For this reason, it is increasingly common to see the use of models drawn from the psychometrics literature in educational data mining publications (Barnes, 2005.; Desmarais and Pu, 2005.; Pavlik et al., 2008). Fuzzy set theory applications involving in educational assessment and performance regarded as efficient and effective in uncertain situations involving performance assessment (Nolan, 1998). Ma et al. (2000) applied a data mining approach based in Association Rules in order to select weak tertiary school students of Singapore for remedial classes. The input variables included demographic attributes (e.g. sex, region) and school performance over the past years and the proposed solution outperformed the traditional allocation procedure. Artificial Neural Network (ANN) is used to predict persisters and non-persisters, although the best model or typical rule for persisters and non-persisters are highly useful in understanding but they do not assist in understanding what is in the dataset (Luan, 2002). In 2003 (Minaei et al., 2003), online student grades from the Michigan State University were modeled using three classification approaches (i.e. binary: pass/fail; 3-level: low, middle, high; and 9-level: from 1 - lowest grade to 9 - highest score). The database included 227 samples with online features (e.g. number of corrected answers or tries for homework) and the best results were obtained by a classifier ensemble (e.g. Decision Tree and Neural Network) with accuracy rates of 94% (binary), 72% (3-classes) and 62% (9- classes). Waiyamai (2003) used data mining to assist in development of new curricula and to help engineering students to select an appropriate major. Kotsiantis et al. (2004) applied several data mining algorithms to predict the performance of computer science students from a university distance-learning program. For each student, several Demographic (e.g. sex, age, marital status) and performance attributes (e.g. mark in a given assignment) were used as inputs of a binary pass/fail classifier. The best solution was obtained by a Naive Bayes method with an accuracy of 74%. Also, it was found that past school grades has a much higher impact than demographic variables. Merceron and Yacef (2005) conducted a case study that used data mining to identify behavior of failing student to warn students at risk before final exam. Delavari et al. (2005) used the educational data mining methods to enhance educational process in higher educational system, which can improve their decision making process. Romero and Ventura (2007) carried out a survey on educational data mining between 1995 to 2005, they concluded that educational data mining is promising area of research and it has a specific requirements not presented in other domain. Ogor (2007) used data mining techniques to build prototype Performance Assessment Monitoring System (PAMS) to evaluate student performance. El-Halees (2008) analyzed learning behavior of students by mining student's data.

Baker and Yacef (2009) stated one of the key applications of educational data mining methods seeks empirical evidence to refine and extend educational theories and well-known educational phenomena, towards gaining deeper understanding of the key factors affecting learning, often with a view to design better learning systems. For instance, Gong et al. (2009) investigated the impact of self-discipline on learning and found that, whilst it correlated to higher incoming knowledge and fewer mistakes, the actual impact on learning was marginal. Perera et al. (2009) used the Big 5 theory for teamwork as a driving theory to search for successful patterns of interaction within student teams. Madhyastha and Tanimoto (2009) investigated the relationship between consistency and student performance with the aim to provide guidelines for scaffolding instruction.

The university as a community of scholars surrounded by pupils and auditors is no longer in tune with reality, with the need to engage in a process of change, adjustment and innovation (Cunha et al., 2000). The higher education institutions will only be able to solve the major challenges that they have been facing long by making creative use of information and communication technologies (Bresfelean et.al., 2007). Contemporary facts that knowledge is becoming a central economic dynamic might be, with the transfer from the concept of “information society” to that of “knowledge societies”. In Loing (2005) led to a re-consideration of the impact of the educational process, that universities are driven to situations of competition, among one another and with the private sector, particularly with the growth of e-learning and trans-national systems. Two main objectives can be distinguished in the data mining process integrated in the management system: a description objective consisting in establishing the eloquent variables and its influences; and a prediction objective (Rusu and Bresfelean, 2006).

Data mining technology can helps managers by finding the knowledge from the data to set up new hypothesis, in a short period of time, which was unattainable or time-consuming in the past, in view of large datasets and previous methods (Luan, 2004). The comprehension of students’ opinions, satisfactions and discontentment regarding the each component of the educational process, the prediction of their preference in certain fields of study, and the choice in continuing their education, is a factual and imperative preoccupation for every higher education institution manager (Luan, 2004).

The task of exploratory data analysis was studied for application in higher education. There is vital issue for the higher educationist to identify and analyze the relationships among different entities such as student, subjects, lecturer’s environment and organization to ensure the effectiveness of their important process (Ogor, 2007). Educational Data Mining (EDM) methods can help bridging this knowledge gaps in higher education system. Therefore, the hidden pattern, association, and anomalies, can be discovered by some data mining techniques, each of them is potential to be used in improving the effectiveness, efficiency, and the speed of the process decision making in higher education.

This study addressed the capabilities of educational data mining in improving the quality of decision making process in higher learning institution by proposing the model of student performance predictors. The data was analyzed using multiple regression and two of kernel methods: Kernel K-Means clustering and Smooth Support Vector Machine classification. Both of these techniques are relatively new algorithms in a variety of data mining techniques. These techniques currently received increasing popularity in machine learning community and were success applied in many field such as image processing, computational biology, bioinformatics, communication and medicine (Mitra and Acharya, 2003). These techniques have a good capability in processing of high dimensional data, and non-linearly separable. Through the experiments researcher used the student academic databases, and surveyed the psychometric factors (intrinsic motivation and behavioral) of undergraduate students at Universiti Malaysia Pahang using closed questionnaire. The questionnaire especially designed to investigate the relationships between psychometric factors of students and their performance. Demographic and psychometric factors variables (Interest, Study Behavior, Engage Time, Believe, and Family Support) of students during their study at university used to measure the correlation of performance predictors proposed with student performance.

1.2 PROBLEM STATEMENT

The amount of data in educational environment maintained in electronic format has seen a dramatic increase in recent time. The data can be collected from historical and operational data reside in the databases of educational institutes. The task to manage the large amount of data and determine the relationships among variables in the data is not easy to be done.

The prediction of success in tertiary institutions is one of the most vital issues in higher education (Golding and Donaldson, 2006). There is no certainty if there are any predictors that accurately determine whether a student will be an academic genius, a drop out, or an average performer. The task to develop effective predictors of academic success is a critical issue for educators (Golding and Donaldson, 2006).

The face value assessment of students at the point of entry can only be confirmed or dispelled by the dynamic follow-up monitoring of students' performance during the course of study leading to serve as an indicator of the suitability and unsuitability of students before admission and during their course of study. Performance predicate is dependent upon motivation, attitudes, peer influence, curriculum and by the continued real-time monitoring of student's performance using a simple rapid response system and as noted predicts correctly which student may need some attention or reinforcements in the course of their education (Ogor, 2007).

The research studies revealed that various factors are responsible for scholastic failure of students, such as low socio-economic background, student's cognitive abilities, school related factors, environment of the home, or the support given by the parents and other family members (Chohan and Khan, 2010).

In present day's educational system, a student's performance in any universities is determined by the combination of internal assessment and external mark. An internal assessment is carried out by the teachers upon the student's performance in various evaluation methods such as tests, assignments, and seminar, attendance, and extension activities. An external mark is the one that is scored by the student in semester examination. Each student has to get the minimum pass mark in internal and as well as in external examination.

The current educational system does not involve any prediction about pass or fail percentage based on performance. The system does not deal with dropouts. There is no efficient method to caution the students about the deficiency in attendance. It does not identify the weak student and inform the teachers.

The most likely place where data miners may initiate data mining project in higher education area is Institutional Effectiveness. Thus, some of the research questions that arise in higher education data analysis can be stated as follows:

(i) What variables or combination of variables collected can be used as predictors of students' performance final grade?

(ii) How the discovered knowledge from academic data can aid decision makers to improve decision making processes?

(iii) How to apply the kernel methods in constructing the model of student performance predictors?

1.3 RESEARCH OBJECTIVES

The main objectives in this study are to apply the kernel methods as the data mining techniques in educational data; Kernel K-means Clustering and Smooth Support Vector Machine Classification techniques. Both of the techniques will be used to construct the model of student's performance predictors based on intrinsic motivation and learning behavior of students during their course of study. The detail of research objectives in this study as follows:

(i) To apply two of kernel methods: Kernel K-means clustering and Smooth Support Vector Machine classification in application of predicting student performance.

(ii) To create a model of student performance predictors in Higher Learning Institution that can be used as decision support system.

(iii) To evaluate Kernel K-means and Smooth Support Vector Machine as methodology approach in predictive Educational Data Mining modeling.

1.4 RESEARCH SCOPE

The scope of this research is to create a students' performance prediction model by using psychometric factors of students as variable predictors and study of two Kernel Methods as data mining techniques: Kernel K-means algorithm for clustering and Smooth Support Vector Machine for classification. The sample data of this research come from student academic databases and the surveyed intrinsic motivation and behavioral of undergraduate students in semester 3 session 2007/2008 at Universiti Malaysia Pahang.

1.5 RESEARCH MOTIVATION

One of the bigger challenges that higher education faces today is predicting the academic paths of student. Many higher education systems specially, community colleges are unable to guide students in selecting career paths, deciding majors, and detecting student population who are likely to drop out because of lack of information and guidance from the learning system. To better manage and serve the student population, the institutions need better assessment, analysis, and prediction tools to analyze and predict student related issues.

There are two major motivations of this study. The first is the critical issues for educators on the prediction of success in tertiary institutions. Although, various types of studies exist in predicting student performance, the prediction accuracy was considerably low with an overall accuracy around 60% (Zlatko, 2010). This fact indicates that there remains a challenging task to devise the most effective in predicting student performance (Sahay and Karun, 2010).

The second motivation of this study is related to methodology used in extracting the data on educational area and the variables employed to construct the model of student performance predictors. The key trend on Educational Data Mining is the increase in prominence of modeling frameworks from Item Response Theory, Bayes Nets, and Markov Decision Processes. The increase in the commonality of these methods is likely a reflection of the integration of researchers from the psychometrics and student modeling communities into the EDM community (Baker and Yacef, 2009). Application of kernel methods recently received increasing popularity in machine learning community and succeed applied in pattern recognition, bio informatics and image processing fields (Byun and Lee, 2003) but poor application in educational area. Currently, a majority of the research on online and off-line student persistence has focused on finding a causal relationship between one variable and its effect on persistence (Sahay and Karun, 2010). There are few research in the Educational Data Mining community that uses the intrinsic motivation and behavioral of students as variables in the prediction model of students' performance. According to Tella (2007) gave the statement that the motivation of students has a significant impact on academic achievement. The issues of motivation of students in education and the impact on academic performance are considered as an important aspect of effective learning. In fact, psychologists believe that motivation is a necessary ingredient for learning (Biehler and Snowman, 1996).

Considering the need of university development and actuality of the data management, it is necessary to set up the analytical guideline as decision support system. Information of students' final grade prediction and the pictures of student groups are important factors to improve their performances. The management needs the data of experience and information to enhance their students' performances.

1.6 THESIS CONTRIBUTIONS

This research attempts to understand the relationship among psychometric factors of student with their performance. Quantitative research was used to explore and examine among all variables.

The first contribution of this study is to evaluate the kernel methods; Kernel K-means clustering and Smooth Support Vector Machine as methodology approach in predictive educational data mining modeling. This has not been empirically tested in previous research yet.

The second contribution of this study is to create a student performance prediction model in higher learning institution by using psychometric factors of students as variables predictor. According to Baker and Yacef (2009) this is one of the key areas of application in educational data mining community.

The third contribution of this study is to provide reference and comparative study for the next researcher in application of the kernel methods in educational data mining.

1.7 THESIS ORGANIZATION

This thesis is organized into five chapters. Chapter one introduces the topic of the study, the research objectives, its research questions, and the significance of the study. Chapter two provides a review of the literature relevant to the topic as introduction to the theoretical framework and Educational Data Mining methods. While in Chapter three provides a description of research problem and the methodology used in the study, including the description of two data mining tools used; Rapid Miner communities and SSVM Ten-fold cross validation. Chapter four reports the experiment results and analysis, while Chapter five presents the conclusion of the study and the implication for future practice and research.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

The purpose of this chapter is to provide a review of the past research efforts related to prediction of student performance and discussion to the theoretical framework in Educational Data Mining methods. A review of Kernel Methods; Kernel K-Means and Smooth Support Vector Machine and other relevant research studies are also provided. The review organized chronologically to offers insight on how past research efforts laid the groundwork for subsequent studies, including the present research efforts. The reviews have been done in detail, so that; the present research can be properly tailored to add the present body of literature, as well as, the scope and direction of the present research effort.

2.2 DATA MINING: KNOWLEDGE DISCOVERY DATABASES

The digital revolution had made digitized information easy to capture, process, store, distribute, and transmit. With significant progress in computing and related technologies and their ever-expanding usage in different ways of life, huge amount of data of diverse characteristics continue to be collected and stored in databases. The rate at which such data is stored growing phenomenally. Discovery of knowledge from this huge volume of data is a challenge indeed. Data mining as an attempt to make sense of the information explosion embedded in this huge volume of data (Olson and Delen, 2008).

The advanced database management technology today enables the system to integrate different types of data such as, image, text, video, and other numeric as well as non-numeric data, in a provably single database in order to facilitate multimedia processing. As a result, traditional ad hoc mixtures of statistical techniques and data management tools are no longer adequate for analyzing these vast collections of mixed data. Data mining techniques is a possible solution (Chakrabarti et al., 2009).

One of the concepts for efficient managing codified knowledge is data mining. Although the definition of data mining is not consensual, most would agree that it refers to extracting knowledge from large amount of data. This thesis adopted the broader definition given by Han and Kamber (2001). Data mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses or in other information repositories. For interesting knowledge with mean patterns, rules or relationships between data are not easily identifiable using the cognition capabilities of a human being; that is, non-trivial, implicit and potentially useful information. Data mining intended to provide support in the rich complex data but poor of information situations. DM has recently attracted a great deal of attention both in industry and academia (Chakrabarti et al., 2009).

Data mining is an interdisciplinary field of computer science. The type of data mining techniques to use depends on the problem to be solved, and on the type of patterns and data to be mined. Moreover, depending on the specific problem, techniques from various fields can be used, including machine learning (Witten and Frank, 1999), artificial intelligence, statistics, information retrieval (Mitra and Acharya, 2003), natural language processing, pattern recognition and visualization (Chakrabarti et al., 2009). Different patterns can be mined using different data mining techniques such as concept or class description, association analysis, classification and regression, cluster analysis, trend analysis, deviation analysis and similarity analysis (Han and Kamber, 2001).

The general process of knowledge discovery using data mining techniques involves an interactive sequence of the following steps: data cleaning, and data integration (Preprocessing step), data selection, data transformation, modeling

(modeling step), pattern evaluation, and knowledge representation (Han and Kamber, 2006). During the preprocessing steps, data is analyzed in order to remove noise, missing value and inconsistencies. The resulting preprocessed data are stored in a data warehouse. During the modeling step, artificial intelligent techniques such as, Neural Network, Pattern Recognitions and others are used to extract and then evaluate data patterns. The patterns found relevant can be presented to users using visualization and representation techniques (Chakrabarti et al., 2009).

Data mining is often used to build predictive/inference models aimed to predict future trends or behaviors based on the analysis of structured data (Han and Kamber, 2001). In this context, prediction is constructing the model and used to assess the class of an unlabeled example, or to assess the value or value ranges of an attribute that a given example is likely to have. Classification and regression are the two major types of prediction techniques, where classification technique is used to predict discrete or nominal values while regression technique is used to predict continuous or ordered values (Larose, 2006). The steps of KDD process as seen on figure 2.1.

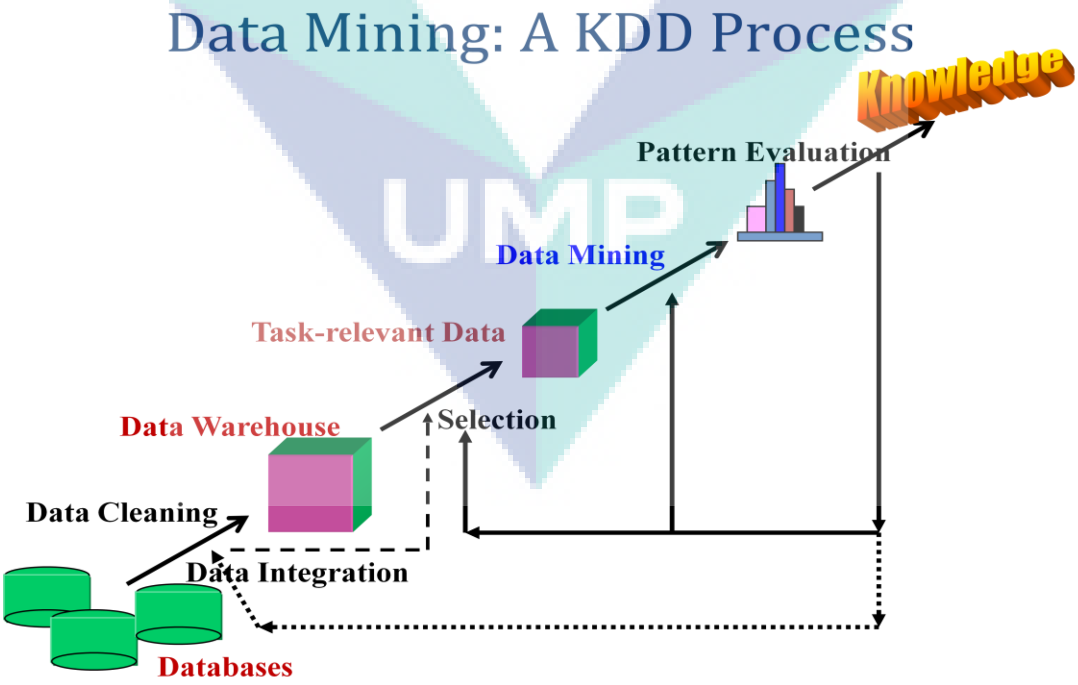


Figure 2.1: Data mining: A KDD Process

Source: Fayyad (1996); Han and Kamber (2006)

Tasks well suited to data mining include the following:

(i) Prediction- determining the value of one variable based on patterns found in others.

(ii) Classification- dividing the data into predefined categories based on their attributes.

(iii) Clustering- finding similarities and differences in a data set's attributes in order to identify a set of cluster to describe the data. The cluster may be mutually exclusive and exhaustive or consist of overlapping categories.

(iv) Description- putting a given data pattern or relationship into human interpretable form (Fayyad, 1996).

Data mining should work the same way as a human brain. It uses historical information (experience) to learn. However, in order for data mining technology to get information out of the database, the user must "tell it" what the information looks like (i.e. what is the problem that the user would like to solve). It uses the description of that information to look for similar example in database, and uses these pieces of information from the past to develop a predictive model of what will happen in the future.

The essential ingredient in building a successful predictive model is to have some information in the database that describes what has happened in the past. Data mining tools are designed to "learn" from these past success and failure (theoretically as a human being would), and then be able to predict what is going to happen next. However, one of the major advantages of a data mining tool over a human mind is that data mining tool can automatically go through a very large database quickly, and find even the smallest pattern that may help in a better prediction. Clearly, this is something humans cannot easily do, and most do not explore a very large database as mentioned above.

2.3 DATA MINING TECHNIQUES

Data mining involves the evaluation of datasets to find patterns that can be used in prediction. Stepwise logistic regression is a popular procedure for selecting variables to build a predictive model. In addition, in spite of its popularity, stepwise logistic regression has certain insurmountable problems. Stepwise regression as a tool for variable selection has also been under severe criticism. It was found that stepwise regression tends to yield conclusions that cannot be replicated because this model building approach capitalizes on sampling error (Thompson, 1995). It is also a well known fact that the results of stepwise regression are affected by the order of entry of the variables (Glymour, 2001).

Cluster analysis could be divided into hierarchical clustering and non-hierarchical clustering techniques. Examples of hierarchical techniques are single linkage, complete linkage, average linkage, median, and Ward. Non-hierarchical techniques include K-Means, Kernel K-Means, adaptive K-Means, k-Medoids, and fuzzy clustering. To determine which algorithm is good is a function of the type of data available and the particular purpose of analysis. Objectively, the stability of clusters can be investigated in simulation studies. The problem of selecting the "best" algorithm/parameter setting is a difficult one. A good clustering algorithm ideally should produce groups with distinct non-overlapping boundaries, although a perfect separation cannot typically be achieved in practice (Hinnenburg and Keim, 2003).

The concept of stability of a clustering algorithm was considered in (Rousseeuw, 1987). The K-Means algorithm provides an easy method to implement approximate solution to objective function. The reasons for the popularity of K-Means are ease and simplicity of implementation, scalability, speed of convergence and adaptability to sparse data. The K-Means algorithm can be thought of as a gradient descent procedure, which begins at starting cluster centroids, and iteratively updates these centroids to decrease the objective function. The K-Means always converge to a local minimum. The particular local minimum found depends on the starting cluster centroids. The K-Means algorithm updates cluster centroids till local minimum is found.

K-Means is an unsupervised learning algorithm that partitions the data set into a selected number of clusters under some optimization measures. For example, to minimize the sum of squares of the Euclidean distance between the samples and the centroids needed. The assumption behind this measure is the belief that the data space consists of isolated elliptical regions. Sometimes it is not sufficient for a given learning machine to work in the input space because the assumption behind the machine does not match the real pattern of the data. For example, SVM and Perceptron require the data are linearly separable, while K-Means with Euclidean distance expects the data distribute into elliptical regions. When the assumption is not held, some kind of transformation technique can be applied to the data, mapping them to a new space where the learning machine can be used. Kernel function provided the mean to define the transformation (Zhang and Rudnicky, 2006).

The extension from K-Means to Kernel K-Means is simply realized by expressing the distance in the form of kernel function (Klaus and Sebastian, 2001.; Zhang and Rudnicky, 2006). The idea behind this validation approach is that an algorithm should be rewarded for consistency. The traditional K-Means clustering algorithm and Kernel K-Means Clustering is used to analyze the student's data. The expected results of these clustering algorithms may serve as a good benchmark to monitor the progression of students' performance in higher learning institution. The result can also potentially be used to enhance the decision making process by academic planners to monitor the candidates' performance semester by semester by improving on the future academic results in the subsequent academic session.

Decision Trees; most famous technique in Data mining not only classifies, but are also able to estimate the importance of each factor and how it affects the model (Apte and Weiss, 1997). A decision tree learner is an algorithm that constructs a decision tree as classifier. A decision tree is a hierarchical structure, consisting of nodes and directed edges. Its base is the root node, which has no incoming edges and has some outgoing edges. Every outgoing edge points to an internal node or to a leaf node. An internal node has one incoming edge and two or more outgoing edges, while a leaf node has one incoming edge and no outgoing edges. The root node and the internal nodes all contain a test condition, used to separate records. There are many variations of decision tree algorithms available in the literature reported. For example, decision tree C4.5 algorithm uses the concept of information gain or entropy reduction to select the optimal split. The information gain is equal to the total entropy for an attribute if for each of the attribute values a unique classification can be made for the result attribute. In the general case the relative entropies subtracted from the total entropy are 0 (Al-Radaideh et al., 2006).

Consider a variable x whose k possible values have probabilities p1, p2, . . . , pk The smallest number of bits, on average per symbol, needed to transmit a stream of symbols representing the values of x observed called the *entropy of* x and defined as:

$$H(x) = -\sum_j p_j \; log_2(p_j) \tag{2.1}$$

For an event with probability p, the average amount of information in bits required to transmit the result is $-\log_2$ (p). For example, the result of a fair coin toss, with probability 0.5, can be transmitted using $-\log_2$ (0.5) = 1 bit, which is a zero or 1, depending on the result of the toss. For variables with several outcomes, we simply use a weighted sum of the $\log_2$ (p_j)'s, with weights equal to the outcome probabilities, resulting in the formula

$$H(x) = -\sum_j p_i \; log_2(p_i) \tag{2.2}$$

Where, p_i represents the proportion of records in subset i and then define the *information gain* to be gain(S) = H(T) − HS(T), that is, the increase in information produced by partitioning the training data T according to this candidate split S. At each decision node, C4.5 chooses the optimal split to be the split that has the greatest information gain, (Larose, 2006)

In recent years, Support Vector Machine (SVM) is relatively new algorithm base on kernel method has been successfully developed and has becoming a powerful tool for solving data mining problems such as classification problem and feature selection. In classification problem, SVM determines an optimal separating surface that classifies data point into different categories. This optimal separating surface is generated by solving an underlying optimization problem. SVM can discriminate between complex data patterns by generating highly non-linear separating surface that is implicitly defined by a nonlinear kernel map. This ability makes SVM applicable to many important applications in pattern classification, image processing, bioinformatics, and database marketing (Lee and Mangasarian, 2001). However, the computational cost (CPU time and optimization package needed) and memory storage limitation of SVM prohibit their use on massive datasets, especially when a nonlinear classifier used. Lee and Mangasarian (2001) proposed new algorithm as further development of conventional SVM called Smooth Support Vector Machine (SSVM). Newton-Armijo algorithm has the ability to solve the SSVM problem and show that this algorithm globally and quadratically converges to unique solution of the SSVM.

2.4 KERNEL METHODS

Nonlinear phenomena encountered in many engineering problems. Traditional signal processing techniques are linear, which makes them unable to extract the complex, nonlinear patterns that may lie in the data available in such scenarios. Therefore, problems concerning nonlinear data analysis have traditionally been tackled by polynomial filters (Mathews and Sicuranza, 2000) which provide straightforward extensions of many linear methods, or by neural network approaches (Haykin, 1999), which are able to learn nonlinear relationships. In Haykin (1999), learning defined as “the process by which the free parameters of a neural network are adapted through a process of stimulation by the environment in which the network is embedded”. Hence, the goal of a learning process is to make the network respond in a certain desired way to its environment. The learning process itself ensures that the performance of the network improves with experience. This definition of learning applies to a broad class of machine learning systems.

In general, machine learning is not concerns learning processes based on a set of predefined rules, but learns relations from the data itself. There exist many types of learning, including supervised learning, in which the desired responses for all training data are given, unsupervised (or blind) learning, in which a data set is provided without the corresponding desired responses, and reinforcement learning, where only indirect feedback on the performance is available. Two important concepts to keep in mind when designing a learning machine are capacity and generalization (Vapnik, 1995).

The capacity of a learning machine refers to the capability of this machine to represent complex and highly nonlinear functions. The generalization capability allows a learning machine to generalize beyond the training data to new, unseen data. Clearly, there exists a trade-off between the capacity and the generalization capability of a learning machine, as a high capacity will allow representing the patterns in the training data very accurately, but it will usually not generalize well to new data. In contrast to linear techniques, which typically offer elegant formulations and efficient algorithms, nonlinear learning machines such as neural networks require more computation and often involve nonlinear optimization problems with multiple local minima.

Both neural networks and kernel methods are universal function approximations (Hornik et al., 1990) and (Micchelli et al., 2006), which means that they can approximate a nonlinear mapping with any given accuracy. However, neural networks usually require a high number of parameters, and their optimal configuration is found by performing an iterative nonlinear optimization process, often implemented through back-propagation. For a multitude of problems, this training procedure is slow, and it does not guarantee convergence to the optimal solution but rather encounters one of the multiple local minima (Boyd and Vandenberghe, 2004).

Kernel methods, on the other hand, generally admit a more elegant solution that stems from the framework of RKHS and the convexity of the resulting optimization problem. Therefore, much kernel-based algorithms have a unique global solution that can be found by solving a convex optimization problem. As a result, although kernel methods are only a decade old, they now represent an established framework to solve machine-learning problems and they are backed by an extensive list of experimental accomplishments. Some of the best known kernel methods are support vector machines (SVM) (Vapnik, 1995), kernel principal component analysis (kernel PCA) (Schölkopf et al., 1998), kernel-based regression techniques (Schölkopf and Smola. 2002), kernel canonical correlation analysis (kernel CCA) (Bach and Jordan, 2002.; Hardoon et al., 2003), kernel Fisher discriminant analysis (KFD) (Mika et al., 1999) and spectral clustering (Ng et al., 2001). An attractive alternative framework is offered by kernel methods (Schölkopf and Smola, 2002.; Shawe and Cristianini, 2004). Kernel methods are powerful machine learning techniques that exhibit a less complex architecture and provide a straightforward approach to transform nonlinear problems into convex optimization problems. Common analysis tasks in kernel-based learning compromise classification, regression and clustering Successful applications of these algorithms have been reported in many fields, such as image processing, computational biology, bioinformatics, communications and medicine.

Based on the presented literature, kernel methods has been successfully applied in many fields, but is extremely lacking in the educational field and becomes the main motivation on this study. The study applies kernel methods: Kernel K-Means Clustering and Smooth Support Vector Machine (SSVM) Classification to explore the data from the education system.

2.4.1 Kernel Function

Sometimes it is not sufficient for a given learning machine to work in the input space because the assumption behind the machine does not match the real pattern of the data. For example, SVM (Support Vector Machine) and Perceptron require the data to be linearly separable, while K-Means with Euclidean distance expects the data to be distributed into elliptical regions. When the assumption is not held, some kind of transformation method can be applied to the data, mapping them to new space where the learning machine can be used.

If given a set of samples $\mathbf{x_1, x_2, x_3,\ldots, x_N}$, where $\mathbf{x_i}$ $\boldsymbol{\varepsilon}$ $\boldsymbol{R^D}$, and a mapping function $\boldsymbol{\Phi}$ that maps $\mathbf{x_i}$ from the input space $\boldsymbol{R^D}$ to a new space $\mathbf{Q}$. The kernel function is defined as the dot product in the new space $\mathbf{Q}$ (Zang and Rudnicky, 2006):

$$\mathbf{H\,(x_i\,,\,x_j) = \Phi(x_i)\,.\,\Phi\,(x_j)} \tag{2.3}$$

An important fact about kernel function is that it is constructed without knowing the concrete form of Φ, namely, the transformation is defined implicitly. Three commonly used kernel function are listed below (Girolami, 2002):

$$\textbf{Polynomial} \quad \mathbf{H\,(x_i\,,x_j\,)\;=(x_i\,.\,x_j+1\,)^d} \tag{2.4}$$

$$\textbf{Radial} \quad \mathbf{H\,(x_i\,,x_j\,)} = \boldsymbol{exp}(-\boldsymbol{r}\ \|\ X_i\ -\ X_j\ \|^2\ ) \tag{2.5}$$

$$\textbf{Neural} \quad \mathbf{H\,(x_i\,,x_j\,)} = \boldsymbol{tanh}\ \mathbf{(\ ax_i\,.\,x_j\ +b)} \tag{2.6}$$

The main weaknesses of kernel function include: First, some properties of the new space are lost, e.g. its dimensionality and value range, due, the lack of the explicit form for Φ. Second, the determination of the appropriate kernel form for a given data set has to be realized through experiments. In addition, the computation and storage cost are increased by wide margin.

2.4.2 Kernel Trick

The dot product is often regarded as a measure of similarity between two input vectors. The dot product **Φ(x$_i$) . Φ(x$_j$)** can be regarded as a measure of similarity between two instances x$_i$ and x$_j$, in the transformed space (Girolami, 2002).

The kernel trick is a method for computing similarity in the transformed space using original attribute set. Consider the mapping function Φ given in equation (2.7).

$$\Phi\ =(\ x_1,\ x_2\ )\rightarrow (x_1^2\ , x_2^2\ , \sqrt{2x_1}\ , \sqrt{2x_2}\ , 1 \tag{2.7}$$

The dot product between two input vectors **u** and **v** in the transformed space can be written as follows:

$$\begin{aligned}\Phi(u)\ .\ \Phi(v) &= (u_1^2\ , u_2^2\ , \sqrt{2u_1}\ ,\ \sqrt{2u_2}\ , 1)\ .\ \ (v_1^2\ , v_2^2 , \sqrt{2v_1}\ ,\ \sqrt{2v_2}\ , 1\ )\\ &= u_1^2\ v_1^2 + u_2^2\ v_2^2 + 2\ u_1\ v_1 + 2\ u_2\ v_2 + 1\\ &= (\mathbf{u}\ .\ \mathbf{v} + 1\ )^2\end{aligned} \tag{2.8}$$

This analysis shows that the dot product in the transformed space can be expressed in terms of similarity function in the original space (Zang and Rudnicky, 2006):

$$K(\mathbf{u},\mathbf{v}) = \Phi(\mathrm{u}) \, . \, \Phi(\mathrm{v}) = (\mathbf{u} \, . \, \mathbf{v} + 1)^2 \qquad (2.9)$$

The similarity function, *K,* which is computed in the original attribute space, is known as the *kernel function*. The kernel trick helps to address some of concerns about how to implement non-linear Support Vector Machine (SVM). Firstly, we do not have to know the exact form of the mapping function Φ because the kernel functions used in nonlinear SVM must satisfy the mathematical principle known as *Mercer's Theorem*. This principle ensures that the kernel function can always be expressed as the dot product between two input vectors in some high-dimensional space. Figure 2.2 described the process of kernel map by using kernel function. The transformed space of the SVM kernels is called a Reproducing Kernel Hilbert Space (RKHS). Secondly, computing the dot products using kernel function is considerably cheaper that using the transformed attribute set Φ (x). Thirdly, since the computations are performed in the original space, issues associated with the curse of dimensionality problem can be avoided (Girolami, 2002).

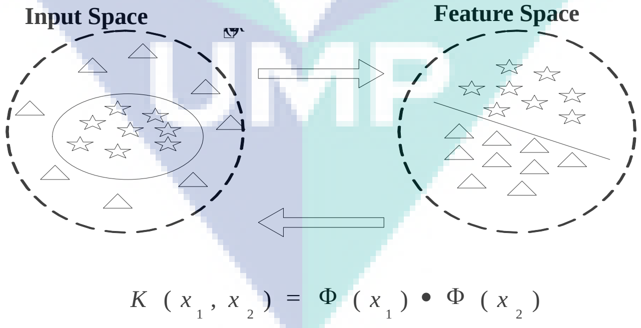


Figure 2.2: Kernel Mapping Process

Source: Schölkopf and Smola (2002)

2.5 REPRESENTATIVE ALGORITHMS

Some of the most powerful kernel-based algorithms can be found in the areas of classification, regression and clustering. This study focused on kernel methods in classification and clustering.

Discussion on this section started with explanations of detail representative algorithms that are used in experiment. The brief overview of the basic concept of Kernel K-Means clustering algorithm is provided and the details study of a new formulation Smooth Support Vector Machine (SSVM) has been described, which is a further development of a Support Vector Machine (SVM) that is the best known in kernel methods.

2.5.1 Kernel K-Means Clustering Algorithm

The first and perhaps most prominent kernel method is the support vector machine (SVM), which optimizes a maximum margin criterion in the kernel feature space. The K-Means algorithm has perhaps been the most popular clustering technique since it was introduced in the late 1960’s. It maximizes the squared Euclidean distance between the cluster centers. However, it is well known that it is only optimal for (*linearly separable*) Gaussian distributed clusters. Different methods for executing this algorithm in the kernel space instead i.e. Kernel K-Means have been derived. In (Zang and Rudnicky, 2006) a stochastic optimization technique was developed using *kernel trick*, while in Girolami (2002) the actual data mapping was approximated by the eigenvectors of so-called kernel matrix.

Experimentally, these works on Kernel K-Means demonstrated that the limitation of ordinary K-Means was overcome, and good results were achieved also for data sets having non-linear cluster boundaries. The motivation for wanting to execute K-Means in the kernel feature space was expressed rather loosely as “the problem of non-linear separability classes can be circumvented by the mapping the observed data to a higher dimensional data space in non-linear manner so that each cluster for each class unfolds into a simple form. However, it is not obvious how Kernel K-Means relates to an operation on the *input* space data set. It is also not obvious how to connect the kernel width to properties of the input data set. Some thoughts alluding to these points have made in (Girolami, 2002.; Shawe and Cristianini, 2004).

Usually the extension from K-Means to Kernel K-Means is simply realized by expressing the distance in form of kernel function (Girolami, 2002.; Klaus and Sebastian, 2003). However, such implementation suffers serious problems such as the high clustering cost due to repeated calculation of kernel values, or insufficient memory to store the kernel matrix, that make it unsuitable for large amount of data.

Given the data set has N samples x_1, x_2,… x_N. K-Means algorithm aims to partition the N samples into *K* clusters, C_1, C_2, …, C_K, and then returns the centre of each cluster, m_1, m_2, …., m_K as the representatives of the data set. Thus, an N-point data set is compressed to a *K*-point "code book". The batch mode K-Means clustering algorithm using Euclidean distance works as follow (Zang and Rudnicky, 2006):

Algorithm 2.1

Step 1 Select *K* initial centers: m_1, m_2, …., m_K

Step 2 Assign each sample x_i ($1 \le i \le N$) to the closet center, forming *K* clusters.

Namely, compute the value of indicator function δ (x_i, C_k), ($1 \le k \le K$).

$$\boldsymbol{\delta}(\boldsymbol{x_i}\,,\boldsymbol{C_k}\,) = \begin{cases} \mathbf{1},\ andD(\boldsymbol{x_i}\,,\boldsymbol{m_k}) < D(\boldsymbol{x_i}\,,\boldsymbol{m_j}\,)\boldsymbol{for\ all\ j}\ \neq \boldsymbol{k} \\ \qquad\qquad \mathbf{0},\ and\ otherwise \end{cases}$$

Step 3 Compute the new centre $\mathbf{m_k}$ for each cluster $\mathbf{C_k}$

$$\boldsymbol{m_k} = \frac{\mathbf{1}}{|\boldsymbol{C_k}|}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{\delta}(\boldsymbol{x_i}\,,\boldsymbol{C_k}\,)\,\boldsymbol{x_i}$$

Where $|C_k|$ is the number of samples in C_k

$$|\mathbf{C_k}| = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{\delta}(\boldsymbol{x_i}\,,\boldsymbol{C_k}\,)$$

Step 4 Repeat step 2 and 3 until converge,

Step 5 Return m_k ($1 \le k \le K$)

In Step 2, $D(x_i\,,m_k)$ is the Euclidean distance satisfying

$$\mathbf{D^2\ (x_i\,,\,m_k\,) = \|\,x_i - m_k\,\|^2}$$

The key issue extending traditional K-Means to Kernel K-Means is the computation of distance in the new space. Let $\boldsymbol{u_i} = \boldsymbol{\Phi}(\mathbf{x_i})$ denoting x_i's transformation. The Euclidean distance between $\boldsymbol{u_i}$ and $\boldsymbol{u_j}$ is written as:

$$\begin{aligned} D^2(\boldsymbol{u_i}, \boldsymbol{u_j}) &= \left\| \boldsymbol{\Phi}(\mathbf{x_i}) - \boldsymbol{\Phi}(\mathbf{x_j}) \right\|^2 \\ &= \boldsymbol{\Phi}^2(\mathbf{x_i}) \mathbf{-2}\, \boldsymbol{\Phi}(\mathbf{x_i})\, \boldsymbol{\Phi}(\mathbf{x_j}) + \boldsymbol{\Phi}^2(\mathbf{x_j}) \\ &= \boldsymbol{H}(x_i, x_i) - 2\, \boldsymbol{H}(x_i, x_i) + \boldsymbol{H}(x_j, x_j) \end{aligned} \qquad (2.10)$$

Let $\mathbf{z_k}$ be the cluster centre in transformed space that,

$$\boldsymbol{z_k} = \frac{\mathbf{1}}{|\boldsymbol{C_k}|} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{\delta}(\boldsymbol{u_i}, \boldsymbol{C_k})\, \boldsymbol{u_i} \qquad (2.11)$$

Where $\boldsymbol{\delta}(\boldsymbol{u_i}, \boldsymbol{C_k})$ **,** is the indicator function. The distance between $\boldsymbol{u_i}$ and $\boldsymbol{z_k}$ is expressed as:

$$\begin{aligned} D^2(\boldsymbol{u_i}, \boldsymbol{z_k}) &= \left\| \boldsymbol{u_i} - \frac{\mathbf{1}}{|C_k|} \sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} \boldsymbol{\delta}(\boldsymbol{u_j}, \boldsymbol{C_k})\, \boldsymbol{u_j} \right\|^2 \\ &= \boldsymbol{H}(x_i, x_i) + f(x_i, C_k) + g(C_k) \end{aligned} \qquad (2.12)$$

Where,

$$f(x_i, C_k) = \frac{\mathbf{2}}{|\boldsymbol{C_k}|} \sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} \boldsymbol{\delta}(\boldsymbol{u_j}, \boldsymbol{C_k})\boldsymbol{H}(x_i, \boldsymbol{x_j}) \qquad (2.13)$$

$$\mathbf{g}(C_k) = \frac{\mathbf{1}}{|C_k|^2} \sum_{\boldsymbol{j=1}}^{\boldsymbol{N}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{\delta}(\boldsymbol{u_j}, \boldsymbol{C_k})\boldsymbol{\delta}(\boldsymbol{u_i}, \boldsymbol{C_k})\boldsymbol{H}(x_j, \boldsymbol{x_i}) \qquad (2.14)$$

The main different between Kernel K-Means and its traditional version exist in step 5, in Kernel K-Means algorithm. Since the cluster in the transformed space cannot be expressed explicitly, a pseudo centre should be choosing instead. By applying (2.10) to the traditional K-Means, the kernel based K-Means algorithm obtained as follows:

Algorithm 2.2

Step 1 *Assign* $\boldsymbol{\delta}(\boldsymbol{x_i}, \boldsymbol{C_k})$ *$(1 \leq i \leq N,\ 1 \leq k \leq K)$ with initial value, forming K initial cluster $C_1, C_2, \ldots, C_K$.*

Step 2 *For each cluster C_k, compute $|C_k|$ and $g(C_k)$.*

Step 3 *For each training sample* x_i *and cluster* C_k *, compute* $f(x_i\ ,C_k)$. *And then assign* x_i *to the closest cluster.*

$$\boldsymbol{\delta}(\boldsymbol{x_i}, \boldsymbol{C_k}) = \begin{cases} \mathbf{1}, & and f(x_i\ ,C_k) + g\ (C_k) < f(x_i\ ,C_j) + g\ (C_j) \\ & \boldsymbol{for\ all\ j\ \neq k} \\ \mathbf{0}, & and\ otherwise \end{cases}$$

Step 4 *Repeat step 2 and 3 until converge.*

Step 5 *For each cluster* $C_{k,}$ *, select the sample that is closest to the centre as the representative of* C_k*.,* $m_k = Arg\ min\ D(\Phi\ (x_i), z_k)$. X_i, *that* $\boldsymbol{\delta(X_i, C_k) = 1}$

The factor H (x_i , x_i) is ignored because it does not contribute to determine the closest cluster.

Kernel K-Means has been extended to efficient and effective large scale clustering (Zang and Rudnicky, 2006) , since the original Kernel K-Means had serious problems, such as the high clustering cost due to the repeated calculations of kernel values, or insufficient memory to store the kernel matrix, that make it unsuitable for large amount of data.

Zang and Rudnicky (2006) solved the problem and proposed the new clustering scheme for large scale data. The first step in this algorithm is to change the clustering order from the sequence of sample to the sequences of kernel that enable us to take an efficient way handling the kernel matrix *H*. The second step is used the disk space to make up the insufficient of memory. Split the kernel matrix H into blocks which size is determined according to the I/O capability and the affordable memory. For example, assuming H has N^2 kernels and the memory can store S^2 of them, H is then split into ½N ^{2} /S ^{2} blocks. These blocks are moved into memory successively and processed there. Note that the block is moved as a whole, so the number of I/O operations to move the entire H is equivalent to the block number which is the minimum value (Zang and Rudnicky, 2006).

2.5.2 Smooth Support Vector Machine

Smoothing methods have been extensively used for solving important mathematical programming problems (Chunhui and Mangasarian, 1996). Lee and Mangasarian, (2001) has proposed a new formulation of support vector machines with linear and nonlinear kernels for pattern classification using smoothing methods. It is called Smooth Support Vectors Machines (SSVM).

The basic idea of SSVM is converting SVM primal formulation to a non smooth unconstrained minimization problem. Since the objective function of this unconstrained optimization problem is not twice differentiable, smoothing techniques can be applied to smooth this unconstrained problem.

The problem of classifying m points in the n-dimensional real space R^n,' represented by m x n matrix A, according to membership of each point A_i in the classes 1 or -1 as specified by a given m x m diagonal matrix D with ones or minus ones along its diagonal. For this problem the standard support vector machine with a linear kernel is given by the following for some v>0 (Lee et al, 2001):

$$\underset{(w,\gamma,y)\epsilon R^{n+1+m}}{min} \quad ve'y + \frac{1}{2}\, w'w \qquad (2.15)$$

$$\text{S t . } D(Aw - e\gamma) + y \geq e \;.$$

Here w is the normal to the bounding planes:

$$x'w - \gamma = +1$$
$$x'w - \gamma = -1 \qquad 2.16)$$

And γ, determine their location relative to the origin. The first plane above bounds the class 1 points and the second plane bounds the class -1 points when the two classes are strictly linearly separable, that is when the slack variable y = 0. The linear separating surface is the plane.

$$x'w = \gamma \qquad (2.17)$$

Midway between the bounding planes (2.16), if the classes are linearly inseparable then two planes bound the two classes with a *soft margin* determined by a nonnegative slack variable *y*, that is:

$$\begin{aligned} x'w - \gamma + y &\geq +1, \quad for\ x' = A_i and\ D_{ii} = +1 \\ x'w - \gamma + y &\geq -1, \quad for\ x' = A_i and\ D_{ii} = -1 \end{aligned} \tag{2.18}$$

These constraints can be written as a single matrix equation as follows:

$$D(Aw - e\gamma) + y \geq e \tag{2.19}$$

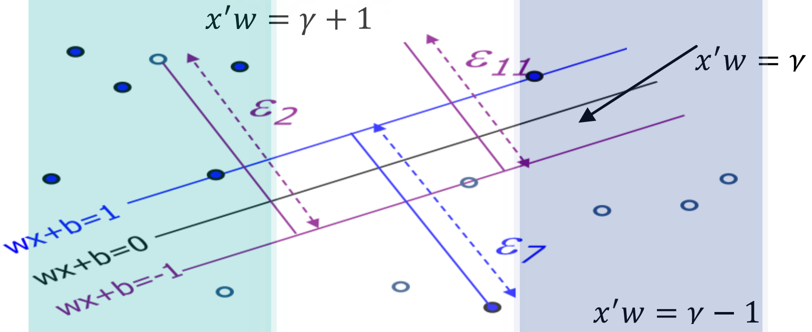


Figure 2.9: *Soft Margin* in linearly inseparable

Source: Nugroho et.al. (2003)

In the SSVM approach, the square of 2-norm of the slack variable y is minimized with weight $\frac{v}{2}$ instead of the 1-norm of y as in (2.15). In addition the distance between the planes (2) is measured in the (n+1)-dimensional space of $(w,\gamma)\epsilon R^{n+1}$, that is $\frac{2}{\|w,\gamma\|_2}$. Thus using twice the reciprocal squared of the margin instead, yields the modified SVM problem as follows:

$$\begin{aligned} &\min_{(w,\gamma,y)\epsilon R^{n+1+m}} \frac{v}{2}\ y'y + \frac{1}{2}\ (w'w + \gamma^2 \\ &\qquad \text{s.t. } D(Aw - e\gamma) + y \geq e \\ &\qquad\qquad y \geq e \end{aligned} \tag{2.20}$$

at a solution of problem (2.49) y is given by:

$$y = (e - D(Aw - e\gamma)) \qquad (2.21)$$

Thus, The replace y in (2.20) by $(e - D(Aw - e\gamma))$ and convert the SVM problem (2.20) into an equivalent SVM which is an unconstrained optimization problem as follow:

$$\min_{(w,\gamma,y)} \frac{v}{2} \ \|(e - D(Aw - e\gamma))\|^2_{2+} \frac{1}{2}\ (w'w + \gamma^2) \qquad (2.22)$$

This problem is a strongly convex minimization problem without any constraints. It is easy to show that it has a unique solution. However, the objective function in (2.22) is not twice differentiable. The smoothing techniques apply and replace x_+ by a very accurate smooth approximation that is given by *p(x, α)*. This *p* function with a smoothing parameter α is used here to replace the plus function in (2.22) to obtain a smooth support vector machines (SSVM) for linearly separable case.

$$\min_{(w,\gamma,y)\epsilon R^{n+1}} \Phi_\alpha(w,\gamma) = \min_{(w,\gamma,y)} \frac{v}{2} \|p(e - D(Aw - e\gamma), \alpha)\|^2_2 + \frac{1}{2}\ (w'w + \gamma^2) \qquad (2.23)$$

The generalized support vector machine (GSVM) generates a non linear separating surface by using a completely arbitrary kernel (Mangasarian, 2000). The GSVM solves the following mathematical program for a general kernel $K\,(A, A^{'})$:

$$\min_{(w,\gamma,y)\epsilon R^{n+1+m}} \quad ve'y + f(u)$$
$$\text{s.t. } D(K(A, A')Du - ey) + y \geq e \qquad (2.24)$$
$$y \geq 0$$

Here $f(u)$ is some convex function on R^m which suppresses the parameter $\boldsymbol{u}$ and $\boldsymbol{v}$ is some positive number that weights the classification error $\boldsymbol{e'y}$ versus the suppression of ***u*.** A solution for this mathematical program for u and γ leads to the nonlinear separating surface.

$$K(A, A')Du = \gamma \tag{2.25}$$

The linear formulation (2.23) is obtained if $K(A, A') = AA'$, $w = A'Du$, and $f(u) = \frac{1}{2}\, u'DAA'Du$. The different classification objective now used that is not only suppresses the parameter $\boldsymbol{u}$ but also suppresses γ in nonlinear formulation:

$$\begin{aligned} \underset{(w,\gamma,y)}{min} \quad & \frac{v}{2} y'y + \frac{1}{2}\,(u'u + \gamma^2) \\ \text{s.t. } & D(K(A, A')Du - ey) + y \geq e \\ & y \geq 0 \end{aligned} \tag{2.26}$$

The same arguments with SSVM for linearly separable case to obtain the SSVM with a nonlinear kernel *K (A, A*′*):*

$$\underset{(w,\gamma,y)}{min} \frac{v}{2} \|p(e - D(K(A, A')Du - ey), \alpha)\|_2^2 + \frac{1}{2}\,(u'u + \gamma^2) \tag{2.27}$$

The smooth support vector machines have important mathematical properties such as strong convexity and infinitely often differentiability. Based on these properties, it can be proved that when the smoothing parameter α in the SSVM approaches infinity, the unique solution of the SSVM converges globally to the unique solution of the original optimization problem (Lee and Mangasarian, 2001). The SSVM can be solved by using a fast Newton-Armijo Algorithm, (Chunhui and Mangasarian, 1996).

2.6 DATA MINING APPLICATION IN HIGHER EDUCATION SYSTEM

Nowadays, many higher education systems generate mountains of administrative data about students, courses, and staff including lecturers, organizational personnel, and managerial systems and so on. This data is a strategic resource for higher educational institutions. Making the most use of these strategic resources will lead to the main objective of higher education system, which is improving the quality of processes.

To retain qualification in educational domain, a deep understanding of the knowledge hidden among the data is required. Data mining techniques can be used to extract unknown pattern from the set of data and discover useful knowledge, which would assist decision makers to improve the decision-making and policy-making procedures. The results can be used in extracting greater value from the raw data set, and making use of strategic resources efficiently and effectively.

In the next time soon to be high on the agenda for researchers and educators in higher learning institution is the adoption of data mining techniques. Higher learning institution will find larger and wider applications of data mining than its counterpart in the business sector, because higher learning institutions carry three (3) duties that are suitable for data mining intensive research: scientific research that relates to the creation of knowledge, teaching that concerns with the transmission of knowledge, and institutional research that pertains to the use of knowledge for decision making (Luan, 2001). Data mining can be best explained as the process of extracting useful knowledge and information including, patterns, associations, changes, anomalies and significant structures from a great deal of data stored in databases, data warehouses, or other information repositories (Han and Kamber, 2001). Prior to the great usage, that this technology brings into many application areas such as biomedical and DNA analysis (Mitra and Acharya, 2003), retail industry and marketing, telecommunications (Chang and Lee, 2000), web Mining (Madhyastha and Tanimoto, 2009) computer auditing, banking (Han and Kamber, 2001), fraud detection (Chang and Lee, 2000), financial industry (Han and Kamber, 2001) and medicine (Baylis, 1999) it recently has also been an interesting area of research in educational domain (Luan, 2001.; Wayamai, 2003.; Bersfelean, 2007.; Baker and Yacef, 2009.; Sahay and Karun, 2010).

Nowadays, the important challenge that higher education institutions face is reaching a stage to facilitate the universities in having more efficient, effective and accurate educational processes (Bodea et al., 2006). As discussed before, lack of deep and implicit knowledge in higher education system may prevent the management to achieve the quality objectives (Luan, 2001). Data mining technology can help in bridging this knowledge gaps in higher education system. Therefore the hidden patterns, association and anomalies, which are discovered by some data mining techniques (Luan, 2004). As a result, this improvement may bring a lot of advantages to the higher education system such as maximizing education system efficiency, decreasing student's drop-out rate, increasing student's promotion rate, increasing student's retention rate, increasing student's transition rate, increasing education improvement ratio, increasing students' success, increasing students' learning outcome, and reducing the cost of system processes. In order to achieve the above quality improvement, data mining technology is needed to provide the knowledge and insights for the decision makers in the higher education system.

2.6.1 Educational Data Mining (EDM)

Data mining in higher education is a new emerging field, called Educational Data Mining (Romero and Ventura, 2007). The Educational Data Mining community defines educational data mining as follows: "Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in"(Baker and Yacef, 2009).

In Witten and Frank (1999) data mining also called Knowledge Discovery in Databases (KDD) is the field of discovering novel and potentially useful information from large amounts of data. Baker (2008) has been proposed that educational data mining methods are often different from standard data mining methods, due to the need to explicitly account for the multi-level hierarchy and non-independence in educational data. Based on this reason, it is increasingly common to see the use of models drawn from the psychometrics literature in educational data mining publications (Barnes, 2005.; Desmarais and Pu, 2005.; Pavlik et al., 2008.; Baker and Yacef, 2009).

2.6.2 Educational Data Mining Methods

Educational data mining methods are drawn from a variety of literatures, including data mining and machine learning, psychometrics and other areas of statistics, information visualization, and computational modeling. Romero and Ventura (2007) categorize work in educational data mining into the following categories:

i. Statistics and visualization
ii. Web mining
   a. Clustering, classification, and outlier detection
   b. Association rule mining and sequential pattern mining
   c. Text mining

The first viewpoint is focused on applications of educational data mining to web data, a perspective that accords with the history of the research area. To a large degree, educational data mining emerged from the analysis of logs of student-computer interaction. This is perhaps most clearly shown by the name of an early Educational Data Mining workshop (according to the EDM community website, the third workshop in the history of the community – the workshop at AIED 2005 on Usage Analysis in Learning Systems (Choque et al., 2005). The methods listed by Romero and Ventura (2007) as web mining methods are quite prominent in EDM today, both in mining of web data and in mining other forms of educational data.

A second viewpoint on educational data mining is given by Baker (2008) which classifies work in educational data mining as follows:

(i) Prediction
   a. Classification
   b. Regression
   c. Density estimation
(ii) Clustering
(iii) Relationship mining
   a. Association rule mining

b. Correlation mining
c. Sequential pattern mining
d. Causal data mining

(iv) Distillation of data for human judgment
(v) Discovery with models

The first three categories of Baker's taxonomy of educational data mining methods would look familiar to most researchers in data mining (the first set of sub-categories are directly drawn from Moore's categorization of data mining methods) (Moore, 2006).

The fourth category, though not necessarily universally seen as data mining, accords with Romero and Ventura's category of statistics and visualization, and has had a prominent place both in published EDM research (Kay et al., 2006) and in theoretical discussions of educational data mining (Tanimoto, 2007).

The fifth category of Baker's EDM taxonomy is perhaps the most unusual category, from a classical data mining perspective. In discovery with models, a model of a phenomenon is developed through any process that can be validated in some fashion (most commonly, prediction or knowledge engineering), and this model is then used as a component in another analysis, such as prediction or relationship mining. Discovery with models has become an increasingly popular method in EDM research, supporting sophisticated analyses such as which learning material sub-categories of students will most benefit from (Beck and Mostow, 2008) how different types of student behavior impact students' learning in different ways (Cocea and Weibelzahl, 2007), and how variations in intelligent tutor design impact students' behavior over time (Jeong and Biswas, 2008).

Historically relationship mining methods of various types have been the most prominent category in EDM research. In Romero and Ventura's survey of EDM research from 1995 to 2005, 60 papers were reported that utilized EDM methods to answer research questions of applied interest (according to a post-hoc analysis conducted for the current article). 26 of those papers (43%) involved relationship mining methods. 17 more papers (28%) involved prediction methods of various types. Other methods were less common. The full distribution of methods across papers is shown in Figure 2.3.

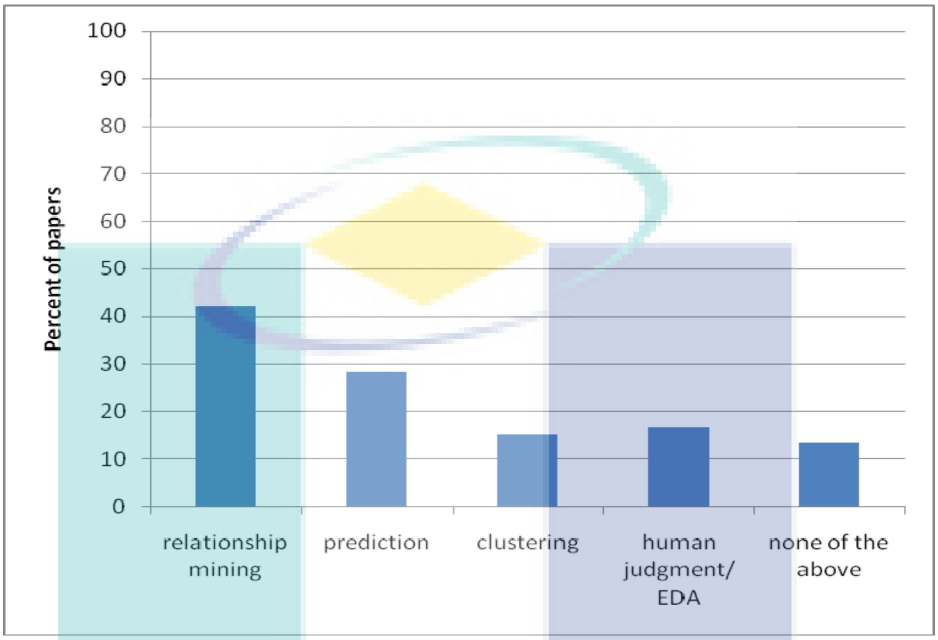


Figure 2.3: The proportion of papers involving each type of EDM method

Source: Baker and Yacef (2009)

A very different pattern is seen in the papers from the first two years of the Educational Data Mining conference (Baker, 2008.; Barnes et al., 2009) as shown in Figure 2.4. Whereas relationship mining was dominant between 1995 and 2005, in 2008-2009 it slipped to fifth place, with only 9% of papers involving relationship mining. Prediction was in second place between 1995 and 2005, moved to the dominant position in 2008-2009, representing 42% of EDM 2008 papers. Human judgment/exploratory data analysis and clustering remain in approximately the same position in 2008-2009 as 1995-2005, with (respectively) 12% and 15% of papers.

A new method, significantly more prominent in 2008-2009 than in earlier years, is discovery with models. Whereas no papers in Romero and Ventura’s survey involved discovery with models, by 2008-2009 it has become the second most common category of EDM research, representing 19% of papers

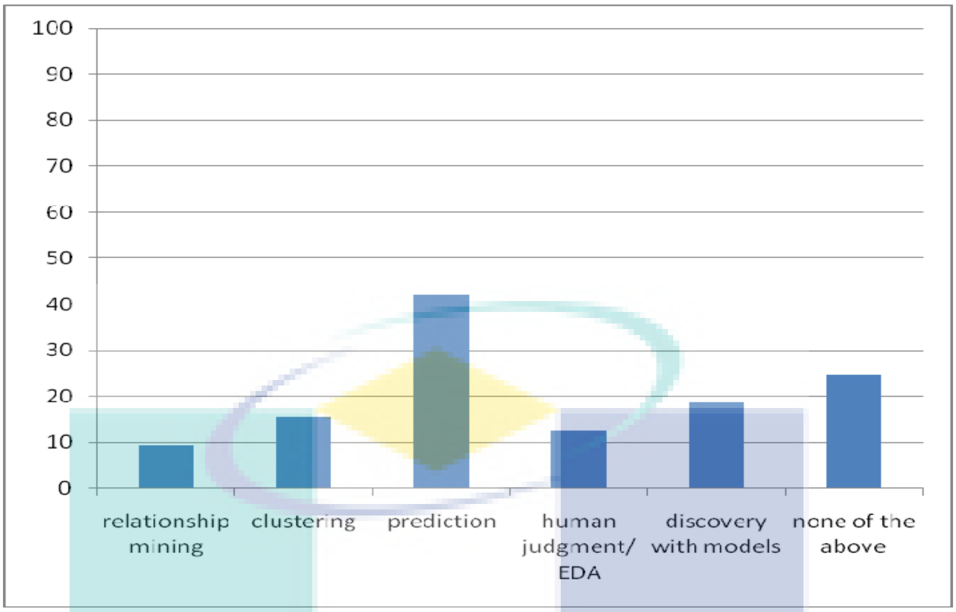


Figure 2.4: The proportion of papers involving each type of EDM method, In the proceedings of Educational Data Mining 2008 and 2009

Source: Baker and Yacef (2009)

2.6.3 Application Area of Educational Data Mining

Educational Data Mining researchers study a variety of areas, including individual learning from educational software, computer supported collaborative learning, computer-adaptive testing (and testing more broadly) and the factors that are associated with student failure or non-retention in courses.

Baker and Yacef (2009) report in the review of Educational Data Mining 2008-2009 was stated in five key area application of Educational Data Mining Methods. The first key area of application has been in the improvement of student models. Student models represent information about a student's characteristics or state, such as the student's current knowledge, motivation, meta-cognition, and attitudes. Modeling student individual differences in these areas enables software to respond to those individual differences, significantly improving student learning (Corbett, 2001).

Educational data mining methods have enabled researchers to model a broader range of potentially relevant student attributes in real-time, including higher-level constructs than were previously possible. For instance, in recent years, researchers have used EDM methods to infer whether a student is gaming the system (Baker et al., 2004) experiencing poor self-efficacy (Mcquiggan et al., 2008) off-task (Baker, 2007) or even if a student is bored or frustrated (D'Mello et al., 2008). Researchers have also been able to extend student modeling even beyond educational software, towards figuring out what factors are predictive of student failure or non-retention in college courses or in college altogether (Dekker et al., 2009.; Romero et al., 2008.; Superby et al, 2006).

The second key area of application of EDM methods has been in discovering or improving models of a domain’s knowledge structure. Through the combination of psychometric modeling frameworks with space-searching algorithms from the machine learning literature, a number of researchers have been able to develop automated approaches that can discover accurate domain structure models, directly from data. For instance, Barnes (2005) has developed algorithms which can automatically discover a QMatrix from the data, Desmarais and Pu (2005), Pavlik et al (2009) have developed algorithms for finding partial order knowledge structure (POKS) models that explain the interrelationships of knowledge in a domain (Pavlik et al., 2009).

The third key area of application of EDM methods is in studying pedagogical support (both in learning software, and in other domains, such as collaborative learning behaviors), towards discovering which types of pedagogical support are most effective, either overall or for different groups of students or in different situations (Beck and Mostow, 2008). One popular method for studying pedagogical support is learning decomposition (Beck and Mostow, 2008). Learning decomposition fits exponential learning curves to performance data, relating a student’s later success to the amount of each type of pedagogical support the student received up to that point. The relative weights for each type of pedagogical support, in the best-fit model, can be used to infer the relative effectiveness of each type of support for promoting learning.

The fourth key area of application of EDM methods is in looking for empirical evidence to refine and extend educational theories and well-known educational phenomena, towards gaining deeper understanding of the key factors impacting learning, often with a view to design better learning systems. For instance Gong et al. (2009) investigated the impact of self-discipline on learning and found that, whilst it correlated to higher incoming knowledge and fewer mistakes, the actual impact on learning was marginal. Perera et al. (2009) used the Big 5 theory for teamwork as a driving theory to search for successful patterns of interaction within student teams. Madhyastha and Tanimoto (2009) investigated the relationship between consistency and student performance with the aim to provide guidelines for scaffolding instruction, basing their work on prior theory on the implications of consistency in student behavior (Abelson, 1968).

The fifth key area of application is the trend to increase in prominence of modeling frameworks from Item Response Theory, Bayes Nets and Markov Decision Processes. These methods were rare at the very beginning of educational data mining, began to become more prominent around 2005 (appearing, for instance, in (Barnes 2005) and (Desmarais and Pu, 2005) and were found in 28% of the papers in EDM2008 and EDM2009. The increase in the commonality of these methods is likely a reflection of the integration of researchers from the psychometrics and student modeling communities into the EDM community.

2.7 STUDENT ACADEMIC PERFORMANCE

Almost every public university in Malaysia such as UMP has the infrastructure and facilities sufficient to support the process of learning and teaching in order for students to take advantage of these conditions to achieve high performance. In addition, students are also expected to obtain a certain degree of academic excellence and maybe superior to other institutions and even in some private colleges. In educational establishments, the academic excellence is used as a yardstick to gauge the success of students (Mohd. Noor and Mohd. Afifi, 2005).

The academic performance is one criterion becomes more important to potential employers later, as it is always used as an indicator of work performance. It may be misleading in some ways, but that is rather the closest readily available and more trusted instrument for assessment. In the general context of human capital development, academic performance is also measured on par with the capability of students on vocational skill, interpersonal skills, leadership and social skill or personality traits. However, since personality traits constitute intangible parameters, this thesis only concern to the academic performance of students in higher learning institution.

Academic Performance of students is the prime importances as the university is geared and more focused to teaching and research. Various measures are planned and implemented at faculty level to ensure that academic performance is placed as prime priority. Aggressive motivational and practice to enhance interpersonal skill programmed, such as orientations, talks, camps, workshops, discussion sessions, are conducted by and for students to achieve academic excellence. Some faculties even approved fund to conduct these motivational programmed (Mohd. Noor and Mohd. Afifi, 2005)

In higher learning institution, the faculty members are mainly responsible for their students' performance. Generally, a committee would monitor and look into problems in this area. It also plans strategies for improvement measures. The activities include academic discourses, camps, and workshops, such as lateral thinking, study techniques, optimal learning, job orientation, academic help lines, etc. (Mohd. Noor et al., 2004).

A faculty is playing a proactive role in putting in and monitoring the continual improvement measures for an even better output. All these initiatives are aimed at obtaining students' academic excellence, including the following;

(i) To strive for zero-failure, similar to zero-defect

(ii) To increase counseling and training sessions for all academic advisors

(iii) To reduce the number of weak students, through a series of motivational activities

(iv) To increase the number of good students, through a series of motivational activities

(v) To reduce the number of weak students through the 'change of study status' scheme

The public university in Malaysia is usually implements a 2-semester academic session. It allows a maximum duration of study of (n + 2) sessions for all students, where n is the program duration and 2 years or 4 normal semesters is the maximum extension time. Students should not go beyond the maximum duration, or else they are considered terminated. Therefore, a student undergoing a 4-year program will have the allowable maximum study duration of 6 years or sessions.

The program conducted by each faculty will declare its total number of credits for graduation. This is usually found in the faculty guide to enable students to schedule their study. Total credits would depend on the curriculum, usually in a range of 125 – 135, for a 4-year or 150 – 165 for a 5-year Engineering degree program.

The academic status of individual student in the university is indicated by the GPA and CPA. Both of these parameters are numerical figures that have maximum scores of 4.0. GPA, or Grade Point Average, defines the 'performance' of the student in one particular semester, whereas the CPA, or Cumulative Point Average, is the cumulative average performance of the student throughout their duration of study thus far.

Early detection of an emerging or ongoing academic problem in a student is important to academic advisors. It is from here that they can assist their students during the critical times by giving appropriate advice. Students' behavioral changes are not easily detected unless advisors are in constant contact with them, but slacking results can be easily traceable at the end of each semester.

2.8 PSYCHOMETRIC FACTORS AND ACADEMIC PERFORMANCE

In the effort to improve students cognition and affective outcomes in mathematics and/or school learning, educational psychologists and mathematics educators, have continued to search for variables (personal and environmental) that could be manipulated in favor of academic gains. Of all the personal and psychological variables that have attracted researchers in this area of educational achievement, ‘motivation’ seems to be gaining more popularity and leading other variables. (Tella and Tella, 2003). Chohan and Khan (2010) carried out a study to examine the impact of educational support given by the parents on the academic achievement and on the self concept of students. By using statistical analysis, the overall parental support variable related significantly to academic performance and self concept of the students.

Ones et al. (2004) suggesting that the psychological support from significant others have substantial impacts on adolescents’ academic self esteem, interest in academic work; overall academic performance, and perhaps also retention in school imply the importance of a holistic approach in maintaining and enhancing adolescents’ expectancy of success in achieving educational goals. The approach would need to involve the wider school community and to gain support from various sources that may have significance to the individual students. Because the positive impacts found in the present study were based on the students’ perceived psychological support from significant others, the findings also suggest that it is not only important to provide support to school students, but also essential that the support is explicitly known to the individual. It seems to be particularly important for the teacher to explicitly exhibit support for students in their academic pursuit because of their relatively greater influence on the students.

Wang and Walberg (1997) analyzed the content of 179 handbook chapters and reviews and 91 research syntheses and surveyed 61 educational researchers in an effort to achieve some consensus regarding the most significant influences on learning. They examined 28 categories of influences. Among the top 11 categories that affected learning, 8 involved *social–emotional influences: classroom management, family support, student–teacher social interactions, social–behavioral attributes, motivational–affective attributes, the peer group, school culture, and classroom climate*. Other influences, such as state, district, or school policies, organizational features such as site-based management, curriculum and instruction, and student and district demographics, had the least influence on learning. Wang and Walberg (1997) concluded that "direct intervention in the psychological determinants of learning promise the most effective avenues of reform".

The theoretical framework for conceptualizing student motivation is an adaptation of a general expectancy-value model of motivation (Pintrich, 1989). There are three motivational components that may be linked to the three different components of self regulated learning as follows:

(a) An expectancy component, which includes students' beliefs about their ability to perform a task,
(b) A value component, which includes students' goals and beliefs about the importance and interest of the task, and
(c) An affective component, which includes students' emotional reactions to the task.

The expectancy component of student motivation has been conceptualized in a variety of ways in the motivational literature (e.g., perceived competence, self efficacy, attribution style, and control beliefs), but the basic construct involves students' beliefs that they are able to perform the task and that they are responsible for their own performance. In this sense, the expectancy component involves students' answers to the question, "Can I do this task?" Different aspects of the expectancy component have been linked to students' meta-cognition, their use of cognitive strategies, and their effort management. In general, the research suggests that students who believe they are capable engage in more meta-cognition, use more cognitive strategies, and more likely to persist at a task than students who do not believe they can perform the task (Pintrich and Groot, 1990).

In scientific literature there are a number of studies based on students' evaluations, exams, behavior: experiments with the purpose of learning the factor which contributes to the high learning performance and identifying the students with high probability risk to fail exams (Bodea et al., 2006) using Oracle Data Mining decision tree algorithms; quantitative methodologies adopted to mine data from Learning Management Systems (LMS) to establish usage patterns and online learning designs within the various organizational levels operating in the university (Heathcote and Dawson, 2005); data mining models based on clustering techniques used to detect cheats in online student assessments (Burlak et al., 2006); surveys and analysis intended for emphasizing the connections between the university and master degree studies and continuing education through students behavior (Bresfelean et.al., 2007), which used questionnaires or data provided by database backed LMS.

2.9 STUDENT PERFORMANCE PREDICTION

Traditionally, American universities have adopted standardized test scores such as the Scholastic Aptitude Test (SAT) or American College Test (ACT) as their major admission criterion. Intelligence tests have achieved wide acceptance as tools to predict future success (Golding and Donaldson, 2006). However**,** these types of instruments have been widely criticized as barriers to the enrollment of non-traditional students and insufficient to address the persistence and success of all students (Golding and Donaldson, 2006).

SCIT has taken the traditional approach in the form of an aptitude test coupled with passes in CXC or GCE exam. The traditional approach emphasizes that students' ability to perform is based heavily on cognitive ability (Ones et al., 2004). Evidence of the relationship between cognitive ability and performance is highlighted in a study of a group of profoundly gifted students who obtained exceptional SAT scores before the age of 13 and achieved 10 years later, a long list of impressive accomplishments including numerous scientific publications (Lubinski et.al., 2001). These results support the argument that one's cognitive ability is directly related to performance. However according to (Fer, 2004), although IQ tests may assess analytical and verbal aptitude well, they are not an accurate test of creativity, of practical knowledge, and other skills involved in problem solving.

A number of theoretical models have been developed in the literature to explain what keeps students on a course. Based on an extensive literature review of dropout in e-learning environment, where Jun (2005) was identified variables that may affect and have been developed the theoretical models of dropout. He classified them into five variables factors such as individual background, motivation, academic integration, social integration and technological support. The background characteristics such as academic and socio-demographic variables (age, sex, ethnic origin, marital status, and financial aid) have been identified in retention literature as potential predictor variables of dropout (Zlatko, 2009). Pascarella et al. (1983) stated that the students' characteristics are a factor of equal if not greater importance when deciding to stay or discontinue the study, more than the actual experience once enrolled. The conceptual model of non-traditional student attrition a set of background characteristics stated by Bean and Metzner's (1985) is causally linked to the effect that academic performance and environmental variables have on the outcome of persistence or dropout. As Tharp (1998) stated after an extensive literature review, background information taken alone as predictors of dropout have not performed well in the case of regular and full-time students. However, the background information of students was significant in the case of distance students or open education, where the social integration and institutional commitment are not central in the student experience.

Kotsiantis et al. (2004) at the Hellenic Open University used the demographic variables and assignment marks in the supervised machine learning algorithms such as decision tree, artificial neural networks, naïve Bayes classifier, instance-based learning, logistic regression and support vector machines in predicting student's performance. When only the demographic variables were used the prediction accuracy varied from 58.84% (when using neural network) to 64.47% (when using support vector machines). However, when other variables beside demographic variables were included, the naïve Bayes classifier algorithm was found to be the most accurate technique for predicting students' performance.

In others study the decision trees, neural networks and linear discriminant analysis techniques was used by Vandamme et al. (2007) for the early identification of three categories of students: low, medium and high-risk students. Some of the background information such as demographics characteristics and academic history of the first-year students in Belgian French-speaking universities were significantly related to academic performance. Those were: previous education, number of hours of mathematics, financial independence, and age, while gender, parent's education and occupation, and marital status were not significantly related to the academic success. However, all three methods used to predict academic success did not perform well. Overall, the correct classification rate was 40.63% using decision trees, 51.88% using neural networks and the best result was obtained with discriminant analysis with overall classification accuracy of 57.35%.

When Yu et al. (2008) at Arizona State University used a data mining approach to differentiate the predictors of retention among freshmen enrolled. They used the classification tree technique based on an entropy tree-splitting criterion; they concluded that 'cumulated earned hours' was the most important factor contributing to retention. There are some of demographic variables were not identified as significant such as gender and ethnic origin.

Al-Radaideh et al. (2006) stated the high school grade is the most contributed to the separation of students in different clusters. They used the classification trees to predict the students' final grade of the Information Technology and Computer Science Faculty at Yarmouk University in Jordan. Among background variables used gender (both students and lecturers), residence, and funding were used to construct the classification tree.

Decision trees, random forests, neural networks and support vector machines, were used to predict the secondary student grades of two core classes by Cortez and Silva (2008). They used the past school grades, demographics, social and other school related data to obtain the results. When the past grades were included they achieved the high level of predictive accuracy. In some other cases, their models also use included the school related features, demographics (student's age, parent's job and education) and social variables.

The binomial probity model used by Boero et al. (2005) and founded that gender is one of the principal determinants of the probability of dropping out. The variable was entered in a quadratic form to allow the effect of age to have diminishing effect on the dropout probability. Males have a higher probability of dropping out relative to the reference group of females. They also found that age has a significant positive effect. With regard to pre- university educational qualifications, the type of school attended had a significant effect on the probability of dropping out.

Herrera (2006) stated the variables that affect persistence at one academic level will not necessarily affect persistence at a different academic level. He concluded that many variables vary in their success at predicting persistence, depending on the academic level. This means that different models that differentiate between dropout and persistent student should be constructed for each academic level. Herrera (2006) also stated the same results could be expected at the same course levels. That would mean that we would get different probabilities of leaving or staying on the course even for the same student depending upon the course. Educational resilience, also discussed by Herrera (2006) which refers to at-risk students who completed the course / diploma /degree in a timely manner despite the risk factors such as biological or psychosocial factors that increase negative outcomes. She also points to the paradigm shift where the focus is now on success rather than on failure. Identifying factors that contribute to the success of an at-risk student might help educational institutions increase students' persistence. In other data mining studies based on enrolment data, the following factors were found to be significant: faculty and nationality (Siraj and Abdoulha, 2009) and the secondary school science mark (Dekker et al., 2009).

In summary, there are mixed evidence on whether the contribution of background information prediction of student success is significant or not. It depends on the list of variables included, students population and classification methods used. Even when the background information was significantly related to the academic performance, the prediction accuracy was considerably low with an overall accuracy around 60% (Zlatko, 2010).

Despite many proposed data mining method in predicting student performance mining method in predicting student performance, however, few researchers use psychometric as predictor variables. Most of the method used in predicting student performance with applied decision tree and neural network. Both of these techniques have limitations in processing high dimensional data. Plumpness of the techniques has good prediction accuracy on demographic variables of students. Bayes Classifier and support vector machine has a good prediction accuracy when other variables beside demographic included.

In this study, Kernel K-means Clustering and Smooth Support Vector Machine is used to develop the model in predicting student performance based on psychometric factors of students. Many researchers reported in the literature that psychometric factors have a strong correlation with performance of students (Pintrich et al., 1990.; Wang et al., 1997.; Fer, 2004.; Jun, 2005.; Herrera, 2006.; Chohan and Khan, 2010).

It is important as educators to devise the most effective standards in predicting academic performance. With this in mind, this research tries the effectiveness of psychometrics parameter: intrinsic motivation and behavioral of students in order to ascertain more accurate factors of prediction by using two algorithms of Kernel methods: Kernel K-means and Smooth Support Vector Machine

CHAPTER 3

METHODOLOGY

3.1 INTRODUCTION

This chapter presents a discussion of the methodology used in this research. The framework of this research described and the rationale of constructing the questionnaire as research instrument to measure the relationship between psychometric factors of student with their academic performance in higher learning institution. In the following, researcher describes the essential steps that were conducted to fulfill the purpose of the research including identification of research variables, population and sample, design of the research, design for data collection, instrumentation and strategies of data analysis.

3.2 RESEARCH DESIGN

The research was conducted using a case study approach that was supported by the survey. The case study is a detailed study of a specific object within a certain time is done with sufficient depth and comprehensive, including the environment and conditions of the past. In this study, the information of psychometric factors of student that can affect the students' academic achievement gathered. The survey is a study that takes a sample of the population using a questionnaire as a tool to collect the data and in general using the Statistical methods.

This study uses primary data through surveys, the data collected is student interest, engage time, Belief, Family Support, and Study Behavior given by students who are still in active learning.

Four types of data mining approaches were conducted in this study. The first approach is descriptive analysis, that is concerned with the nature of the dataset such as the frequency table and the relationship between the attributes obtained using cross tabulation analysis (contingency tables). In addition, feature selection is conducted to determine the importance of the prediction variables to model the student performance. This experiment shows the significance of the correlation between the predictor variables proposed with student performance and which is the strongest variable.

The second approach is to classify the students using Kernel K-Means clustering with mixture dataset. Evaluation of clustering results is again an application dependent problem. When the goal is to characterize the groups obtained, a strategy sometimes employed consists in defining a set of external characterization features that are not used in the learning processes directly. For example, in customer segmentation for marketing purpose is commonly used to detect the groups according to behavioral and demographic information and then complete the profile of these groups using business value characteristics such as profitability. In this experiment, a similar scheme was followed using an external feature that indicates the students' condition. The numerical and categorical data are mixed in one dataset file. As a result, a profile of each cluster was obtained and described both in terms of the input features in our notion profitability (Student Performance).

The third approach is the classification tree used to examine the variables related to student performance. It is important to note, that the data mining is focused on pattern recognition, hence no probabilistic inference is involved. Also unlike regression, that returns a subset of variables, classification trees can rank order the factors that affect student performance. From this approach, the best rule model to predict student performance was identified.

Finally, the rule model is from the decision tree that will be used to predict students' performance. SSVM classification used, to evaluate the variables related to the performance of the students. The experiments used 10 fold cross validation which entails dividing dataset into 10 equal sets, then using 9 sets to train and 1 set to test. It loops through this process 10 times until all sets were used as the test set. The performance of training and testing accuracy of each performance predicate presented. Figure 3.1 present the research framework of the study.

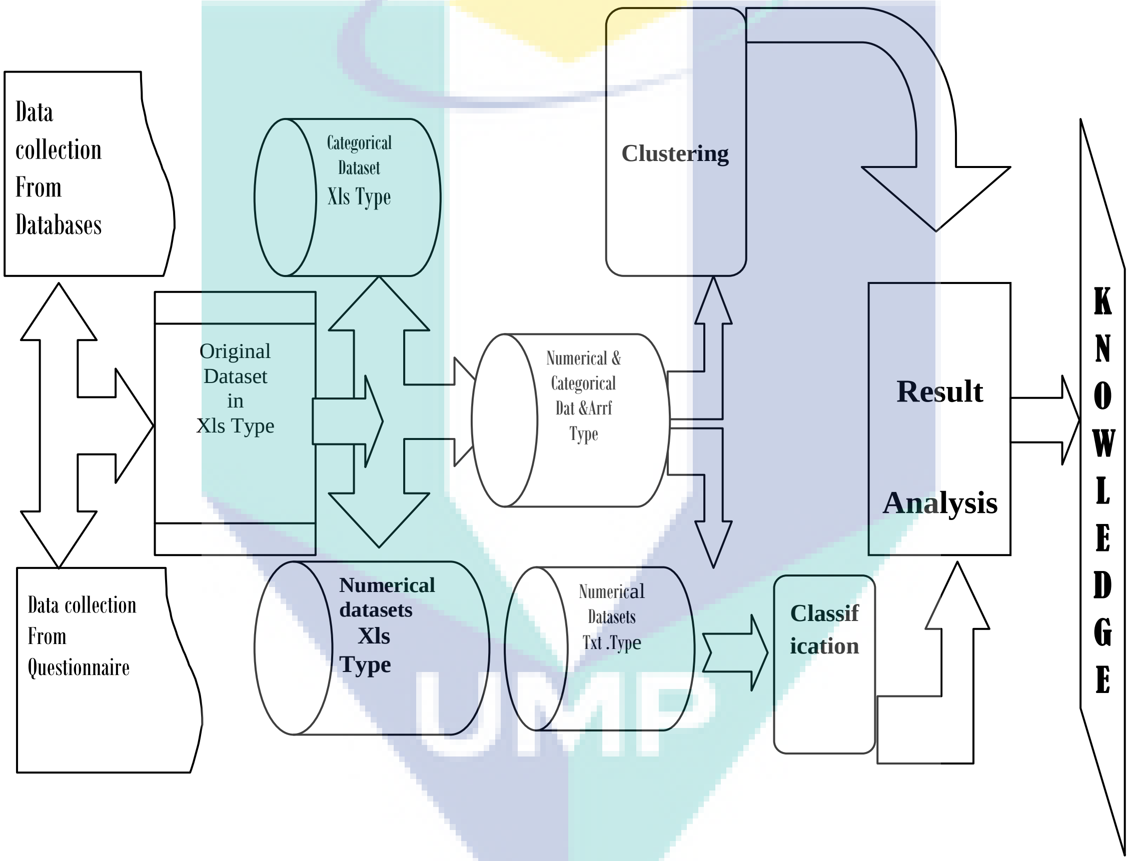

Data collection From Databases
Categorical Dataset Xls Type
Clustering
Original Dataset in Xls Type
Numerical & Categorical Dat &Arrf Type
Result
Analysis
K N O W L E D G E
Data collection From Questionnaire
Numerical datasets Xls Type
Numerical Datasets Txt .Type
Classif ication

Figure 3.1: Research Framework.

3.3 RESEARCH INSTRUMENT

3.3.1 Measurement Scale

Malhotra (1999) stated that "measurement means assigning numbers of others symbols to characteristics of objects according to certain prespecified rules". While the scale is defined as the generation of a continuum on which measured objects are located (Malhotra, 1999). According to Masson et al. (2003) there are four scales of measurement are nominal, ordinal, interval, and ratio. To measure these variables, the measurement instrument is a scale that will be used as a reference for determining the long / short or large / small interval, to produce quantitative data. By using this measure, the value of the variable to be measured with certain instruments can be expressed in numerical form, so it would be more accurate, efficient and communicative.

In the present study, in order to measure attitude, opinion and one's perception of a phenomenon and the measurement of socioeconomic status, Likert scale used. By using a Likert scale, the variables to be measured are translated into indicator variables. Then the indicator is used as a starting point to construct the items in the form of instruments questions or inquiries. Furthermore, for purposes of quantitative analysis, then the answer sheet was given a score ranging from 4 to 1 (Strongly agree to strongly disagree).

3.3.2 The Grid Research Instruments

This study conducted uses several variables in understanding the association of independent variable toward dependent variable. The research variables used in this study are described below:

(1) Dependent variable	: Students' Performance	(Y)
(2) Independent Variable	: Students' interest	(X1)
	: Students' Belief	(X2)
	: Students' Family Support	(X3)
	: Students' Study Behavior	(X4)
	: Students' Engage Time	(X5)

Figure 3.2 presents the proposed model of student performance predictors. It hypothesizes that interest, engage time, belief, family support, and study behavior of students can directly influence students' performance.

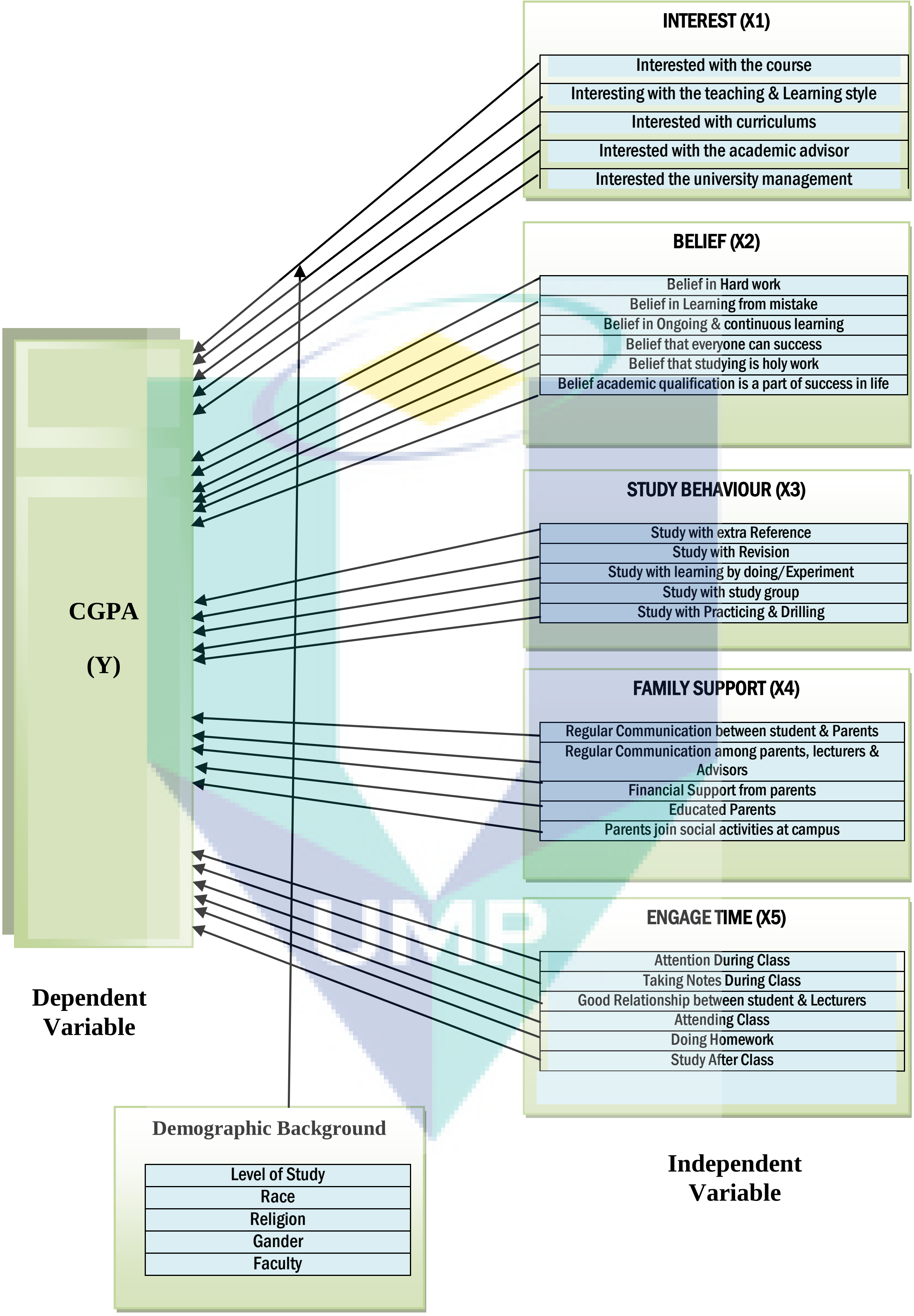

INTEREST (X1)
Interested with the course
Interesting with the teaching & Learning style
Interested with curriculums
Interested with the academic advisor
Interested the university management
BELIEF (X2)
Belief in Hard work
Belief in Learning from mistake
Belief in Ongoing & continuous learning
Belief that everyone can success
Belief that studying is holy work
Belief academic qualification is a part of success in life
STUDY BEHAVIOUR (X3)
Study with extra Reference
Study with Revision
Study with learning by doing/Experiment
Study with study group
Study with Practicing & Drilling
FAMILY SUPPORT (X4)
Regular Communication between student & Parents
Regular Communication among parents, lecturers & Advisors
Financial Support from parents
Educated Parents
Parents join social activities at campus
ENGAGE TIME (X5)
Attention During Class
Taking Notes During Class
Good Relationship between student & Lecturers
Attending Class
Doing Homework
Study After Class
CGPA
(Y)
Dependent Variable
Independent Variable
Demographic Background
Level of Study
Race
Religion
Gander
Faculty

Figure 3.2: The proposed model of student Performance Predictors

In this phase, the data are put into a form suitable for the modeling phase. If required, some selected variables are combined, transformed or used to create new variables. Variables definition and their domains are presented in Table 3.1

Table 3.1: **Description of variables and their domains**

Variable	Description (Domain)
Student Demographics	
Gender	Student gender (Binary : Male and Female)
Race	Students ethnic (Nominal: Melayu, India, China)
Religion	Students Religion (Nominal: Islam, Kristen, Catholic, Hindu, Buddha and others)
Faculty ID	Students Faculty (Nominal: Mechanical, Technology Management, Chemical, Civil Engineering, Computer)
Semester	Semester in which a course is offered (Semester 1, Semester 2 , Semester 3, Semester 4 ,or Semester 5)
Psychometric Factors	
Interest	Students Interest on the course programme (Ordinal: High, Medium and Low)
Belief	Students confidence level that he/she could succeed (Ordinal: High, Medium, and Low)
Family Support	The Support from Family (Parent's) of student (Ordinal: High, Medium and Low)
Study Behavior	The behavioral learning of student (Ordinal: High, Medium and Low)
Engaged Time	Engagement time in learning of student (Ordinal: High, Medium and Low)
Dependent Variable	
Student Performance	Students CGPA (Ordinal: Excellent, Very Good, Good, Average and Poor)

To increase interpretation and comprehensibility of the data, the numerical attributes of student performance variable (CGPA) are discretized to categorical value using normal distribution method. The performance grades are grouped into five categories, which are *excellent*, *very good*, *good*, *average* and *poor*. The race or ethnicity varied between 1 and 5, the variable Religion between 1 and 4, Faculty ID varied between 1 and 5 and the gender convert to 1 and 2. While the numerical value obtained of variables from questionnaire were categorized into *High*, *Medium* and *Low*. All of variables discretized are presented in Table 3.2.

Table 3.2: Dicretized of Variables

Student's Demographics		
Variables	Domain	Domain Discrete
Gender	Binary: Male/Female	1= Male, 2= Female
Race	Nominal: Melayu, China, India, Bumi Putera Sabah, Bumi Putera Serawak.	1= Melayu, 2=China, 3= India, 4=BP Serawak , 5= BP Sabah
Religion	Nominal: Muslim, Kristen, Hindu, Buddha,	4= Muslim, 3=Christian 2= Buddha 1= Hindu
Faculty ID	Nominal: Mechanical, Technology Management , Chemical, Civil Engineering, Computer	1=FKM, 2=FKEE 3=FSKKP, 4=FKSA 5=FKASA

Psychometric Factors				
Variable	Domain	Mean	Std.Dev	Numerical Interval Value

Interest	L=Low,
	M=Medium
	H= High
		53.36	13.66	x < 39.7=L, 39.70 <= x < 67.02 = M, x >= 67.02 = H

Belief	L=Low,
	M=Medium
	H= High
		52.21	14.63	x < 37.58 = L, 37.58 <= x <66.84 = M, x >= 66.84= H

Family Support	L=Low,
	M=Medium
	H= High
		53.56	13.23	x < 40.33 = L, 40.33 <= x < 66.80 = M, x > 66.8 = H

<table>
<tr><td>Study Behavior</td><td>L=Low,
M=Medium
H= High</td><td>39.66</td><td>9.07</td><td>x &lt; 30.59 = L, 30.59 <= x < 48.73 =M, x >= 48.73 = H</td></tr>
<tr><td>Engaged Time</td><td>L=Low,
M=Medium
H= High</td><td>53.72</td><td>12.86</td><td>x &lt; 40.87 = L, 40.87 <= x < 66.59 = M, x > 66.59 = H</td></tr>
<tr><td colspan="5">Student Performance (CGPA)</td></tr>
<tr><td colspan="3">Numerical Interval Value (4 scale)</td><td colspan="2">Performance Predicate</td></tr>
<tr><td colspan="3">CGPA &lt;= 2.25</td><td colspan="2">Poor</td></tr>
<tr><td colspan="3">2.25 &lt; CGPA &lt;= 2.64</td><td colspan="2">Average</td></tr>
<tr><td colspan="3">2.64 &lt; CGPA &lt;= 3.04</td><td colspan="2">Good</td></tr>
<tr><td colspan="3">3.04 &lt; CGPA &lt;= 4.42</td><td colspan="2">Very Good</td></tr>
<tr><td colspan="3">CGPA &gt; 3.42</td><td colspan="2">Excellent</td></tr>
</table>

3.4 DATA COLLECTION AND PREPROCESSING

3.4.1 Data Collection

The scope of this research in terms of data is limited to the data available in the student academic databases Universiti Malaysia Pahang. The questionnaire used for collecting data was administered to undergraduate students in semester 3 session 2007/2008 at Universiti Malaysia Pahang to investigate the relationships between psychometric factors of student and their academic performance. It is important to have a full understanding of the nature of the data and how it was collected and entered before proceeding further. In this phase, an initial data exploration using pivot tables was also conducted to get some insight in the data.

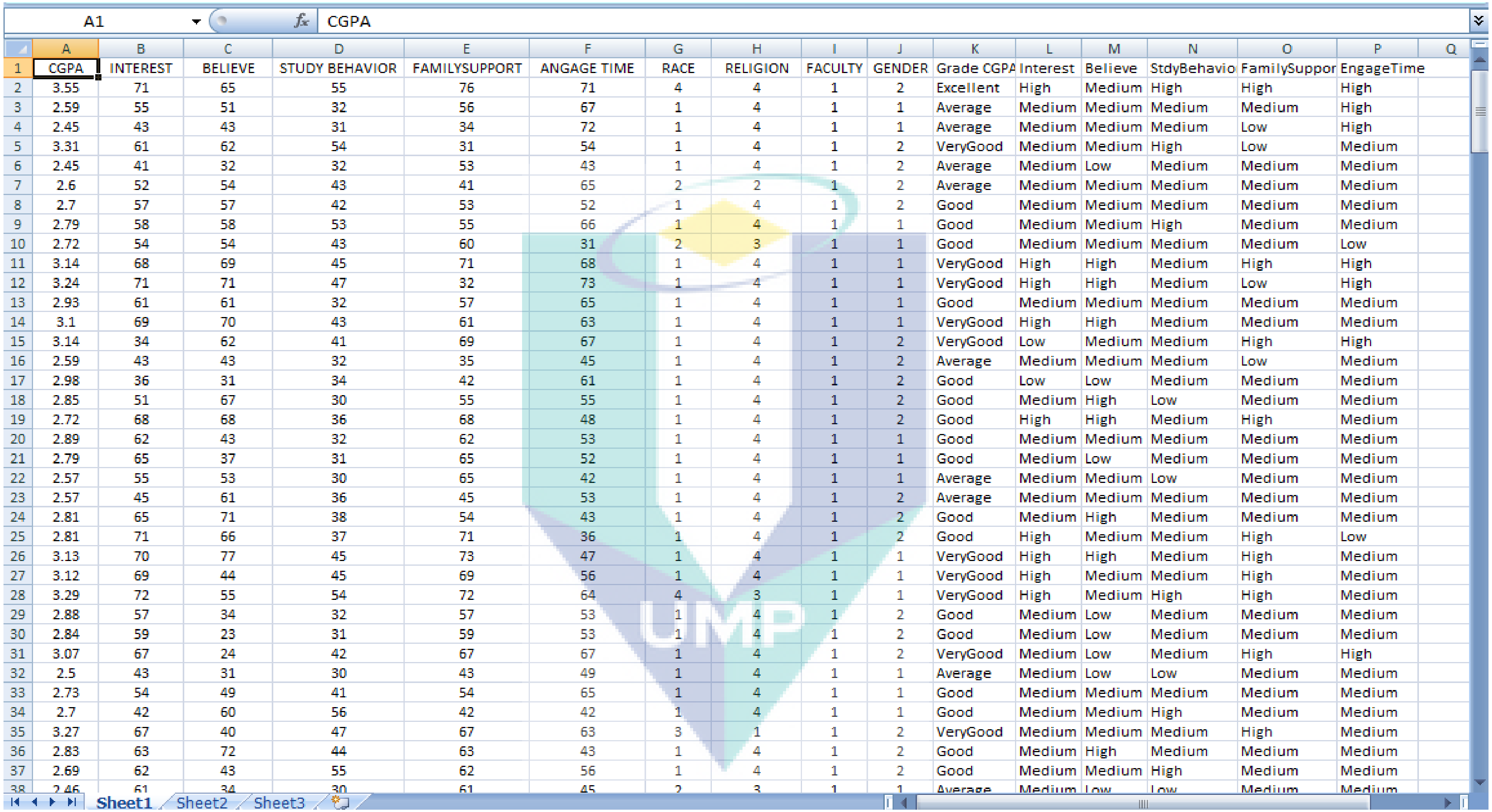
The questionnaire forms required the students to enter the following information: demographic (gender, age, ethnicity, religion, and faculty), current academic performance (CGPA), Psychometric factors (Motivation and Behavioral) (Budoff and Corman, 1974). Where the value components of student motivational are *interest* of the task, *belief* and *family support* (Brophy, 1983). While the value components of student behavioral are *engaged time* and *study behavior* of students. (Budoff and Corman, 1973).

The questionnaire was generated by the researcher and was verified and evaluated by a psychologist from Indonesia Mr. Triantoro Safaria, M.Si.Psi, where he is a Ph.D student from Department of Technology Management of Manufacturing and Technology Management Faculty at Universiti Malaysia Pahang. The questionnaire was used to investigate the relationships between psychometric factors of the students and their performance. The remaining survey items consisted of 87 items to which participant ranked their responses on a 4-point linker-type scale that ranged from "strongly agree to "strongly disagree". All 87 items questions in the survey were divided into six parts. Part one consists of 7 items to obtain the demographic of the respondents. The remaining of 80 items were related to students' Interest, Study Behavior, Engaged Time, Belief and Family Support, which are will be used to predict student performance.

3.4.2 Data Preprocessing

This study used two datasets, the first datasets collected from database management system course held at the Universiti Malaysia Pahang in semester 3 sessions 2007/2008. The value components of the data from database were: personal records, and academic record of student. The second dataset is the results obtained from questionnaire. The number of student was 1000 and we get the sample data from 350 students.

After the target population was defined, the next step was to prepare an appropriate data set. Certainty, data fields (or variables) had to be combined or transformed, and only selected fields were mined. All of the data integrated to single table. The sample was 350 students and we found 46 respondents with incomplete parameters. Hence, the remaining 304 samples were used for analysis. Figure 3.4 presented the screen shot of the mixture data sets model.



A1 CGPA

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	CGPA	INTEREST	BELIEVE	STUDY BEHAVIOR	FAMILYSUPPORT	ANGAGE TIME	RACE	RELIGION	FACULTY	GENDER	Grade CGPA	Interest	Believe	StdyBehavio	FamilySuppor	EngageTime	
2	3.55	71	65	55	76	71	4	4	1	2	Excellent	High	Medium	High	High	High	
3	2.59	55	51	32	56	67	1	4	1	1	Average	Medium	Medium	Medium	Medium	High	
4	2.45	43	43	31	34	72	1	4	1	1	Average	Medium	Medium	Medium	Low	High	
5	3.31	61	62	54	31	54	1	4	1	2	VeryGood	Medium	Medium	High	Low	Medium	
6	2.45	41	32	32	53	43	1	4	1	2	Average	Medium	Low	Medium	Medium	Medium	
7	2.6	52	54	43	41	65	2	2	1	2	Average	Medium	Medium	Medium	Medium	Medium	
8	2.7	57	57	42	53	52	1	4	1	2	Good	Medium	Medium	Medium	Medium	Medium	
9	2.79	58	58	53	55	66	1	4	1	1	Good	Medium	Medium	High	Medium	Medium	
10	2.72	54	54	43	60	31	2	3	1	1	Good	Medium	Medium	Medium	Medium	Low	
11	3.14	68	69	45	71	68	1	4	1	1	VeryGood	High	High	Medium	High	High	
12	3.24	71	71	47	32	73	1	4	1	1	VeryGood	High	High	Medium	Low	High	
13	2.93	61	61	32	57	65	1	4	1	1	Good	Medium	Medium	Medium	Medium	Medium	
14	3.1	69	70	43	61	63	1	4	1	1	VeryGood	High	High	Medium	Medium	Medium	
15	3.14	34	62	41	69	67	1	4	1	2	VeryGood	Low	Medium	Medium	High	High	
16	2.59	43	43	32	35	45	1	4	1	2	Average	Medium	Medium	Medium	Low	Medium	
17	2.98	36	31	34	42	61	1	4	1	2	Good	Low	Low	Medium	Medium	Medium	
18	2.85	51	67	30	55	55	1	4	1	2	Good	Medium	High	Low	Medium	Medium	
19	2.72	68	68	36	68	48	1	4	1	2	Good	High	High	Medium	High	Medium	
20	2.89	62	43	32	62	53	1	4	1	1	Good	Medium	Medium	Medium	Medium	Medium	
21	2.79	65	37	31	65	52	1	4	1	1	Good	Medium	Low	Medium	Medium	Medium	
22	2.57	55	53	30	65	42	1	4	1	1	Average	Medium	Medium	Low	Medium	Medium	
23	2.57	45	61	36	45	53	1	4	1	2	Average	Medium	Medium	Medium	Medium	Medium	
24	2.81	65	71	38	54	43	1	4	1	2	Good	Medium	High	Medium	Medium	Medium	
25	2.81	71	66	37	71	36	1	4	1	2	Good	High	Medium	Medium	High	Low	
26	3.13	70	77	45	73	47	1	4	1	1	VeryGood	High	High	Medium	High	Medium	
27	3.12	69	44	45	69	56	1	4	1	1	VeryGood	High	Medium	Medium	High	Medium	
28	3.29	72	55	54	72	64	4	3	1	1	VeryGood	High	Medium	High	High	Medium	
29	2.88	57	34	32	57	53	1	4	1	2	Good	Medium	Low	Medium	Medium	Medium	
30	2.84	59	23	31	59	53	1	4	1	2	Good	Medium	Low	Medium	Medium	Medium	
31	3.07	67	24	42	67	67	1	4	1	2	VeryGood	Medium	Low	Medium	High	High	
32	2.5	43	31	30	43	49	1	4	1	1	Average	Medium	Low	Low	Medium	Medium	
33	2.73	54	49	41	54	65	1	4	1	1	Good	Medium	Medium	Medium	Medium	Medium	
34	2.7	42	60	56	42	42	1	4	1	1	Good	Medium	Medium	High	Medium	Medium	
35	3.27	67	40	47	67	63	3	1	1	2	VeryGood	Medium	Medium	Medium	High	Medium	
36	2.83	63	72	44	63	43	1	4	1	2	Good	Medium	High	Medium	Medium	Medium	
37	2.69	62	43	55	62	56	1	4	1	2	Good	Medium	Medium	High	Medium	Medium	
38	2.46	61	34	30	61	45	2	3	1	1	Average	Medium	Low	Low	Medium	Medium	

Figure 3.3: Mixture Data Sets Model

3.5 DESIGN EXPERIMENTS

Experimental design is the process of planning a study to meet specified objectives. Planning an experiment properly is very important in order to ensure that the right type of data and a sufficient sample size and power are available to answer the research questions of interest as clearly and efficiently as possible.

Three experimental units will be conducting in this study to achieve the research objectives. The firstly is to examine the relationship between performance predictors' variable with the variable of students' performance. On this step, the strategy for data analysis is using multi regressions analysis technique.

The secondly is to find the rule model in prediction of students' performance. The data will be test and analyzed by using Kernel K-Means clustering and decision tree technique.

The thirdly is to examine the rule model finding and will be used to predict students' performance. The strategy used on this step is using Smooth Support Vector Machine (SSVM) classification technique. The block diagram of this research is, described on Figure 3.3.

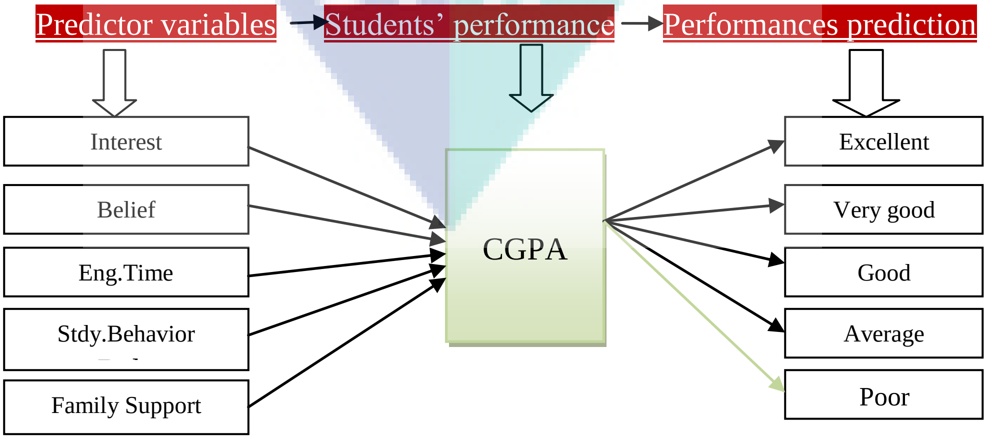


Figure 3.4: Block Diagram of the study.

3.5.1 Student Segmentation Using Kernel K-Means Clustering

The first step in this algorithm is change the clustering order from the sequence of sample to the sequence of kernel that enable to take an efficient way on handling the kernel matrix H.

The second steps, by used the disk space to make up the insufficient of memory. The matrix H will be split kernel into blocks which size is determined according to the I/O capability and affordable memory. The blocks are move into memory successively and processed there. The algorithm of Kernel K Means Clustering as followed:

Algorithm 3.1: Kernel K-Means Clustering

```
Step 1  Compute kernel matrix H and store every necessary block B to the disk

Step 2  Assign $\delta(x_i, C_k)$ $(1 \leq i \leq N, 1 \leq k \leq K)$ with initial value, forming K
        initial cluster $C_1, C_2, \ldots, C_K$.
Step 3  For each cluster $C_k$, compute $|C_k|$ and let $g(C_k)=0$. For each training
        sample $x_i$ and cluster
        $C_k$, let $f(X_1, C_k)=0$
Step 4  Read the next block B from disk into memory
Step 5  For every kernel $h_{u,v}$ $(1 \leq u, v \leq N)$ in B :
          • Check the cluster that $X_u$ and $X_v$ belong to
            $\Theta_u = k_1$ if $\delta(x_u, C_{k1})=1$, $\Theta_v = k_2$ if $\delta(x_v, C_{k2})=1$
            Where $\Theta_u$ and $\Theta_v$ are the variables denoting the clusters
          • If u<v, ignore this kernel. ( Only the kernel below or on the diagonal
            of H need be processed given the symmetry of H)
          • If u = v ($h_{u,v}$ is on the diagonal of H, so $x_u = x_v$ and $\Theta_u = \Theta_v$ then
            $f(x_u, \Theta_v) = f(x_u, \Theta_v) - 2 * h_{u,v} / |C_{\Theta v}|$
            $g(\Theta_u) = g(\Theta_u) + h_{u,v} / |C_{\Theta v}|^2$
          • If u > v then
            $f(x_u, \Theta_v) = f(x_u, \Theta_v) - 2 * h_{u,v} / |C_{\Theta v}|$
            $f(x_v, \Theta_u) = f(x_v, \Theta_u) - 2 * h_{u,v} / |C_{\Theta u}|$
            and if $\Theta_u = \Theta_v$
            $g(\Theta_u) = g(\Theta_u) + h_{u,v} / |C_{\Theta v}|^2$

Step 6  Repeat steps 4 and 5 until every block been processed
Step 7  for each training sample $x_i$ and cluster $C_k$
```

$$\delta(x_i, C_k) = \begin{cases} 1, & and f(x_i, C_k) + g(C_k) < f(x_i, C_j) + g(C_j) \\ & \quad for\ all\ j \neq k \\ 0, & and\ otherwise \end{cases}$$

Repeat steps 3 to 7 until converge.

3.5.2 The Rule Model for Prediction Using Decision Tree

Decision tree learners are algorithms that construct a decision tree as classifier. A decision tree learner is an algorithm that constructs a decision tree as classifier. A decision tree is a hierarchical structure, consisting of nodes and directed edges. Its base is the root node, which has no incoming edges and has some outgoing edges. Every outgoing edge points to an internal node or to a leaf node. An internal node has one incoming edge and two or more outgoing edges, while a leaf node has one incoming edge and no outgoing edges. The root node and the internal nodes all contain a test condition, used to separate records. This study used J48 Decision tree Rule learner to build the rule model for prediction of student performance. The algorithm of decision tree J48 used in this study as followed:

Algorithm 3.2: J48 Decision Tree

```
S - Training Set
A - Input Feature Set
y - Target Feature

Create a new tree T with a single root node.
IF   One of the Stopping Criteria is fulfilled THEN
     Mark the root node in T as a leaf with the most common value
     of y in S as a label.
ELSE
     Find a discrete function f(A) of the input attributes values such that splitting
     S according to f(A)'s outcomes (v1,...,vn ) gains the best splitting metric
.    IF   best splitting metric > treshold THEN
           Label t with f(A)
          FOR each outcome v i of f(A):
               Set Subtreei= TreeGrowing (σ f(A)=vi S,A,y).
               Connect the root node of tT to Subtreei with an edge
               that is labelled as vi
          END FOR
        ELSE
               Mark the root node in T as a leaf with the most
               common value of y in S as a label.
        END IF
END IF
RETURN T

TreePruning (S,T,y)

Where:
```

```
S - Training Set
y - Target Feature
T - The tree to be pruned
DO
    Select a node t in T such that pruning it
     maximally improve some evaluation criteria
    IF t≠Ø THEN T=pruned(T,t)

UNTIL t=Ø
RETURN T
```

3.5.3 Smooth Support Vector Machine To Predict Student Performance

The result of the previous section will be used and taking advantage of the twice differentiability of the objective function of SSVM algorithm, a quadratically convergent Newton algorithm with an Armijo step size prescribed (Armijo,1966. ; Dennis and Schnabel, 1983. ; Bertsekas, 1999) that makes the algorithm globally convergent. The Newton-Armijo algorithm used in C style and run on Linux Operating System as followed:

Algorithm 3.3: Newton-Armijo Algorithm

Start with any (w^0,γ^0) Є R^{n+1}.

Having (w^i,γ^i), stop

IF the gradient objectives function

$\min\limits_{w,\gamma} \quad \frac{v}{2}\|(e-D(Aw-e\gamma))_+\|_2^2+\frac{1}{2}(w'w+\gamma^2)$ is zero. That is

$\nabla\Phi_\alpha\left(w^i,\gamma^i\right)=0$.

ELSE compute $\left(w^{i+1},\gamma^{i+1}\right)$as follows:

Step 1 Newton Direction:

Determine direction $d^i \in R^{n+1}$ by setting equal to zero the linearization of $\nabla\Phi_\alpha\left(w\ ,\gamma\ \right)$ around $\left(w^i,\gamma^i\right)$ which gives n + 1 linear equations in n + 1 Variable *:*

$$\nabla^2\Phi_\alpha\left(w^i,\gamma^i\right)d^i=-\nabla\Phi_\alpha\left(w^i,\gamma^i\right)' \qquad \dots\dots\dots\dots\dots\dots\dots\dots \quad \textit{Eq:1}$$

Step 2 ***Armijo Stepsize :***

Choose a step size $\lambda_i \in R$ *such that :*

$$\left(w^{i+1},\gamma^{i+1}\right)=\left(w^{i},\gamma^{i}\right)+\lambda_i d^{i} \qquad \text{………………………….}Eq{:}2$$

Where $\lambda_i = \max\left\{1,\frac{1}{2},\frac{1}{4},...\right\}$

such that:

$$\Phi_\alpha\left(w^{i},\gamma^{i}\right)-\Phi_\alpha\left(\left(w^{i},\gamma^{i}\right)+\lambda_i d^{i}\right)\geq-\delta\lambda_i\nabla\Phi_\alpha\left(w^{i},\gamma^{i}\right)d^{i} \quad \text{….}Eq{:}3$$

Where, $\delta \in \left(0,\frac{1}{2}\right)$.

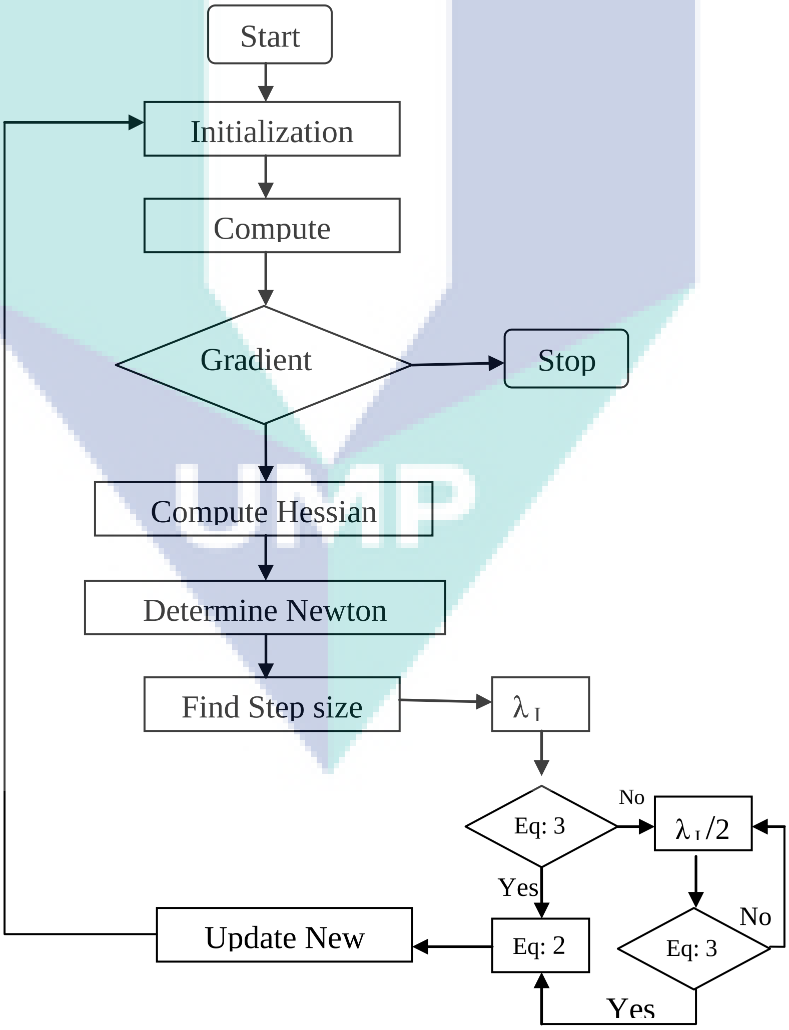


Figure 3.4: Newton-Armijo Flowchart

3.6 DATA MINING TOOLS

3.6.1 Statistical Package for Social Science

According to Wu (1999), Statistical Package for Social Science (SPSS) is considered an old timer in the field of data mining. It was originally designed for use by social scientists to analyze data from survey. SPSS allows user to pull in data and perform sophisticated statistical-analysis operation, such as computing regression and displaying a graphical presentation of data. It also uses sophisticated inferential and multivariate statistical procedures, such as Analysis of Variance (ANOVA), factor analysis, cluster analysis, and categorical data analysis. SPSS particularly well suited to survey research. SPSS 15.0 was used in this study to perform a multiple regression analysis on the data sets. Both a step-wise and full model regression was run to determine the best model to fit the data.

In this study, SPSS tools used to analyze validity and reliability of research instrumentation, to show the graphical representation and descriptive statistics of data and to find the significances of the variables performance predictors proposed.

3.6.2 Rapid Miner Community

RapidMiner (*formerly YALE*) is a free open- source environment for KDD and ML that provides a rich variety of methods that allow the prototyping of new applications. RapidMiner and its plugging provide more than 400 operators for all aspects of Data Mining. Meta operators automatically optimize the experiment designs and users no longer need to tune single steps or parameters any longer. A huge amount of visualization techniques and the possibility to place breakpoints after each operator give insight into the success of your design - even online for running experiments. (http://www.rapidminer.com). RapidMiner 4.3 was used in this study to perform the clusters of student on the data sets and displaying graphical representation of clusters. K-Means, Kernel K-Means and decision tree J48 were tools that are available in the Rapid Miner's library.

3.6.3 Smooth Support Vector Machines (SSVM) Tenfold Software

The SSVM ten fold's software is the further development of SSVM software with addition performance evaluation using tenfold cross validation and data normalization package. **SSVM Software** for classification problem was developed by Santi et al. (2008) based on Newton-Armijo algorithm to classify breast cancer. The software was built in C++ style and running on windows operating system. The SSVM tenfold was developed in C language and run on Linux Operating System and was successful applied on face recognition (Furqan et al., 2010). This software is applicable to many important applications that are needed to explore large amount of data, especially for nonlinear classifier used.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 INTRODUCTION

This chapter presents the results of the study, according to the research questions proposed at the outset. The summary of data statistics is presented to increase our understanding about the data used in this study. The experiment was conducted on two model data sets. For the first, numerical datasets and categorical datasets are mixed into single table and the second dataset is numerical data which is transformed from xls format into txt format.

The first datasets used as input data for Kernel K-Means clustering and decision tree Rule Learner. The second dataset in xls type used as input data K-Means clustering and the txt type used for SSVM. Performance accuracy training and testing data is provided. The results was evaluated using tenfold cross validation (10-CV) to guarantee that the results are valid and it can be generalized for making prediction regarding new data.

4.2 SUMMARY OF DATA STATISTICS

4.2.1 Validity and Reliability

Instrument reliability in this research was important. Reliability refers to whether an instrument consistently measures an event over time and populations (Gall et al. 1996). The survey in this study was examined in terms of internal reliability measures an instrument's degree of interrelation among test items (Brown and Alexander, 1991). This is done to assure that an instrument accurately measures, what it is intended to measure. Cronbach's alpha was administered surveys to measure internal consistency. According to Mitchell and Jolley (1999), a Cronbach's alpha at or above 0.60 is accepted as evidence of internal reliability. After the testing run to measure the validity and reliability of instrument, there are 5 items found invalid data and removed. Finally, 75 items were used in the questionnaire survey.

The overall value of alpha coefficient for each variable as described in Table 4.1 and descriptive statistics of the data presented in Table 4.2. The results of all variables in measuring validity and reliability of instrument can be seen on Appendix C.

Table 4.1: Instrument's Reliability

No	Variable's scale	Cronbach's alpha	N items
1	Interest	.926	19
2	Belief	.772	13
3	Study Behavior	.936	13
4	Family Support	.916	15
5	Engage Time	.803	15

Based on the result on Table 4.1, the minimum alpha coefficient is 0. 772 for "Belief" variable and maximum alpha coefficient is 0.936 for "Study Behavior". It means all of the variables used in this survey valid and reliable.

Table 4.2: Descriptive Statistic of Data

		CGPA	Interest	Believe	Study behavior	Familysupport	Engage time	Race	Religion	Faculty	Sex
N	Valid	304	304	304	304	304	304	304	304	304	304
	Missing	0	0	0	0	0	0	0	0	0	0
Mean		2.8382	53.3882	52.2467	39.7730	53.6842	53.7862	1.2961	3.7632	3.0395	1.4276
Std. Deviation		.38589	13.68748	14.66867	9.33027	13.19662	12.90165	.80296	.66240	1.42529	.49555
Skewness		.329	-.426	-.222	.396	-.506	-.474	2.994	-2.852	-.029	.294
Std. Error of Skewness		.140	.140	.140	.140	.140	.140	.140	.140	.140	.140
Kurtosis		-.090	-.915	-.908	.266	-.741	-.770	8.505	7.250	-1.315	-1.926
Std. Error of Kurtosis		.279	.279	.279	.279	.279	.279	.279	.279	.279	.279
Range		2.03	57.00	65.00	54.00	56.00	54.00	4.00	3.00	4.00	1.00
Minimum		1.86	20.00	20.00	20.00	20.00	20.00	1.00	1.00	1.00	1.00
Maximum		3.89	77.00	85.00	74.00	76.00	74.00	5.00	4.00	5.00	2.00

4.2.2 Significance of Correlation and Multicollinearity

The experiment tested the correlation among variables predictor of student performance and dependent variable (CGPA) using multiple regression analysis methods. The results, shown on Table 4.3, all variables give 75.2% (R^2 = .752) contributions to student performance.

Table 4.3: Significance Correlation of all variables performance predictors

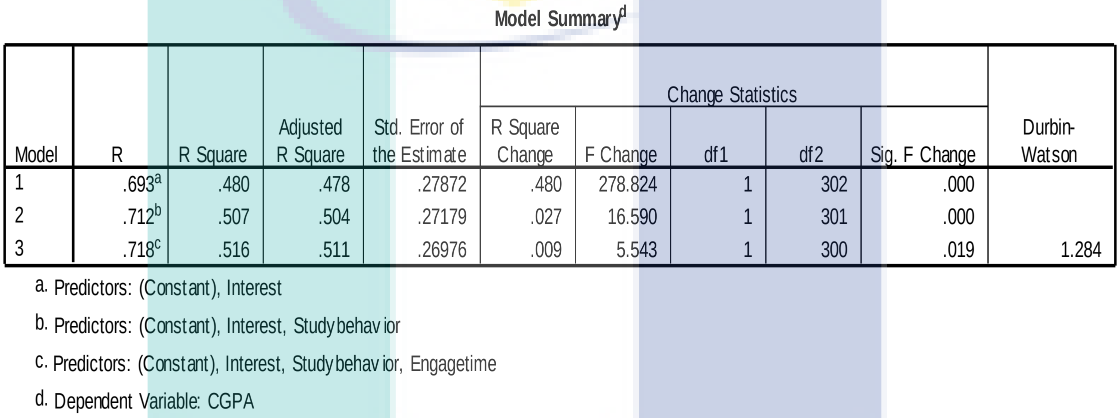
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
	R Square Change	F Change	df1	df2	Sig. F Change	R Square Change	F Change	df1	df2	Sig. F Change
1	.752(a)	.565	.552	.25827	.565	42.494	9	294	.000	1.354

a Predictors: (Constant), Sex, Race, Interest, Faculty, Studybehavior, Religion, Believe, Engagetime, Familysupport

b Dependent Variable: CGPA

Other experiments evaluate the correlation among the five predictor variables of student performance by using multiple regression analysis methods with stepwise model to find out which variables is most contributed. The results obtained showed that there are three of five variables mentioned on Table 4.3 have a significant correlation to student performance; *Interest, Study Behavior and Engage Time*. All three predictor variables above give 51.1% contributions to student performance. The experiment results shown on Table 4.4:

Table 4.4: Significant Correlation of three variables performance predictors



Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R Square Change	F Change	df1	df2	Sig. F Change	
1	.693[a]	.480	.478	.27872	.480	278.824	1	302	.000	
2	.712[b]	.507	.504	.27179	.027	16.590	1	301	.000	
3	.718[c]	.516	.511	.26976	.009	5.543	1	300	.019	1.284

a. Predictors: (Constant), Interest

b. Predictors: (Constant), Interest, Studybehavior

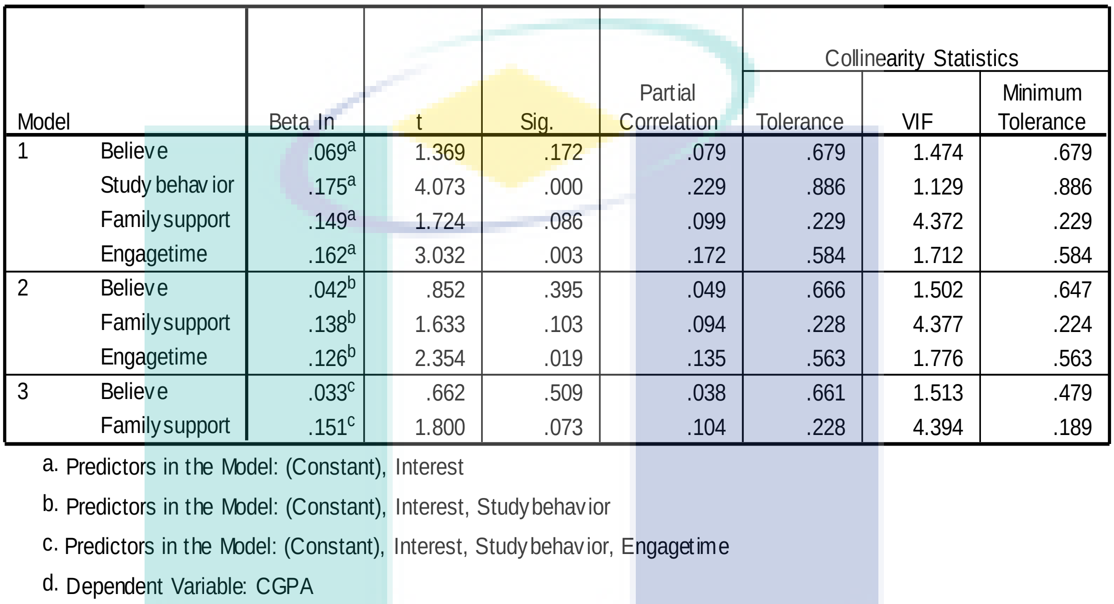
c. Predictors: (Constant), Interest, Studybehavior, Engagetime

d. Dependent Variable: CGPA

Based on the results on Table 4.4, the variable *Interest* is a superior variable predictor (see on R Square Change), contributing 48%. In other experiments using multiple regression analysis with enter model. The researcher find out the contribution of the others two variables are "Belief" and Family Support contributes only 1.1%. Therefore, the five variables provides a very significant contribution in the amount of 52, 2% of student performance (R^2=.522). It can be concluded that the five variables as mentioned above are very good to serve as a predictor for student's performance.

Based on statistical assumption, the multicollinearity should be examined among variables. Multicollinearity is a common problem in many correlation analyses. This happens when the variables are redundant and can interfere with proper interpretation of the multiple regressions results. The simple way to identify collinearity is Tolerance and VIF (Variant Inflation Factor). Tolerance is the amount of variability of the selected independent variables. Stevens (2002) and Meyers et al. (2006) stated that tolerance value less than 0.1 or 0.01 and VIF greater than 10 as indicative of multicollinearity. Table 4.5 depicts that all variables did not have tolerance value less then 0.1 or 0.01, and VIF values greater than 10. The results indicates that there are no multicollinearity detection among variables in this study.

Table 4.5: **Multicollinearity Diagnostic**



Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	Believe	.069[a]	1.369	.172	.079	.679	1.474	.679
	Study behavior	.175[a]	4.073	.000	.229	.886	1.129	.886
	Family support	.149[a]	1.724	.086	.099	.229	4.372	.229
	Engagetime	.162[a]	3.032	.003	.172	.584	1.712	.584
2	Believe	.042[b]	.852	.395	.049	.666	1.502	.647
	Family support	.138[b]	1.633	.103	.094	.228	4.377	.224
	Engagetime	.126[b]	2.354	.019	.135	.563	1.776	.563
3	Believe	.033[c]	.662	.509	.038	.661	1.513	.479
	Family support	.151[c]	1.800	.073	.104	.228	4.394	.189

a. Predictors in the Model: (Constant), Interest

b. Predictors in the Model: (Constant), Interest, Study behavior

c. Predictors in the Model: (Constant), Interest, Study behavior, Engagetime

d. Dependent Variable: CGPA

4.3 STUDENT PERFORMANCE SEGMENTATION

To achieve more understanding on the data, the group of student based on their performance is presented. The performance index of students can see on Table 4.6.

Table 4.6: Students Performance Index

Performance Index	
CGPA <= 2.25	Poor
2.25 < CGPA <= 2.64	Average
2.64 < CGPA <= 3.04	Good
3.04 < CGPA <= 3.42	Very Good
CGPA > 3.42	Excellent

In Figure 4.1 shown the distribution of students' performance on the data and find out the lower CGPA is 1.86 and the higher is 3.89. Based on the results of frequency test, there are 57. 2% Male and 42.8% Female.

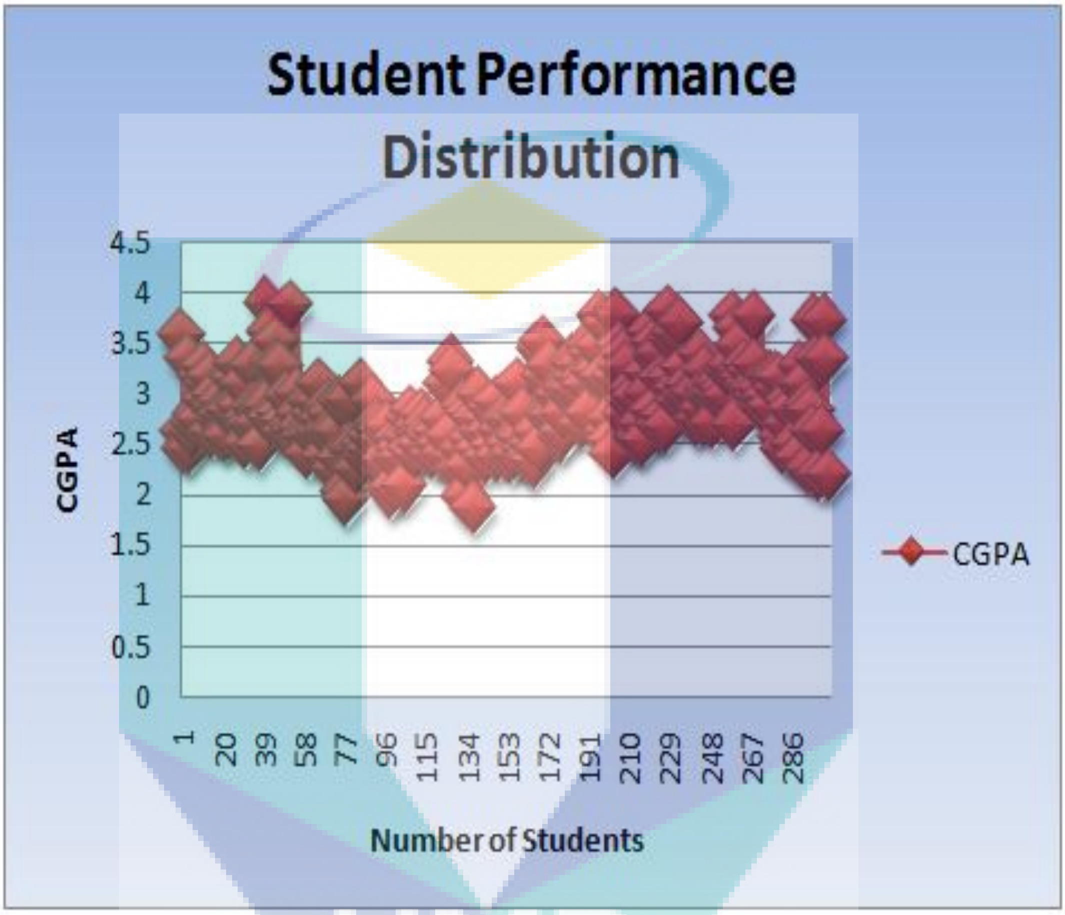


Figure 4.1: Student Performance Distribution

The students were grouped based on their performance. Based on the performance index as shown on Figure 4.2, the performance index "poor" 4.61% (14 students), performance index "good" is 35.20% (107 students), while for performance index "very good" is 21.71% (66 students) , performance index "average" is 30.59% (93 students) and performance index excellent is 7.89% (24 students).

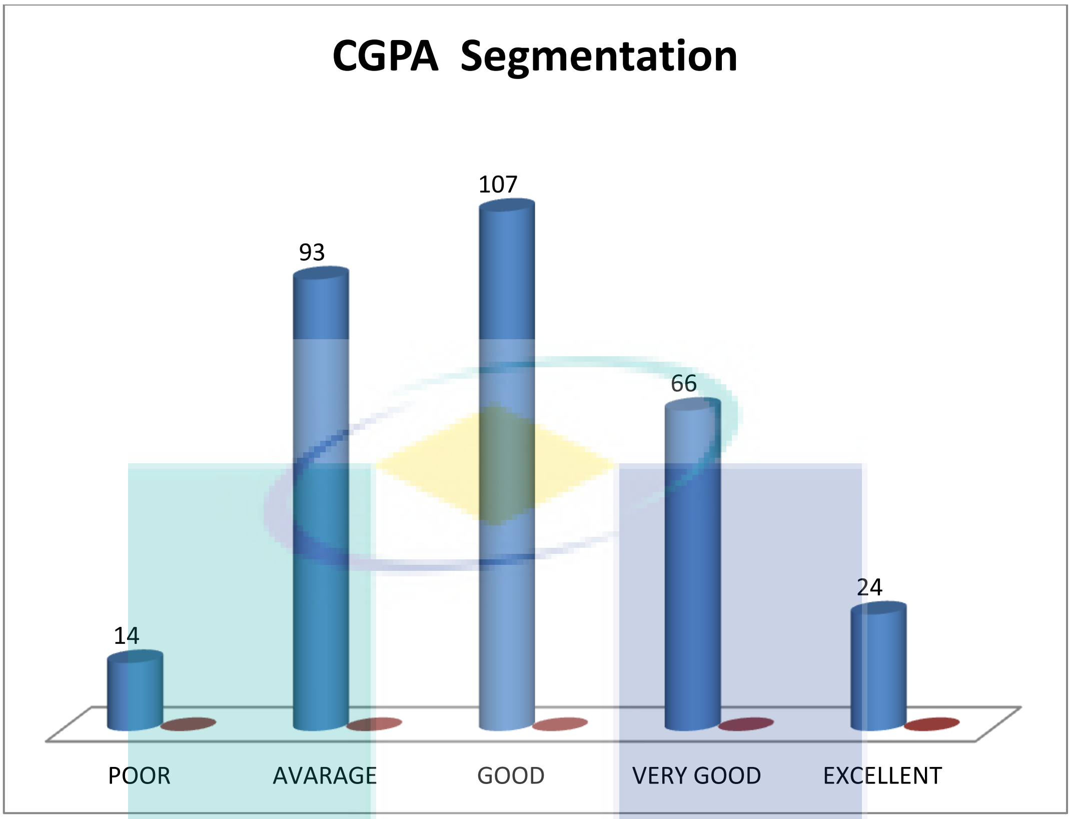


Figure 4.2: Students Performance Segmentation with Performance Index

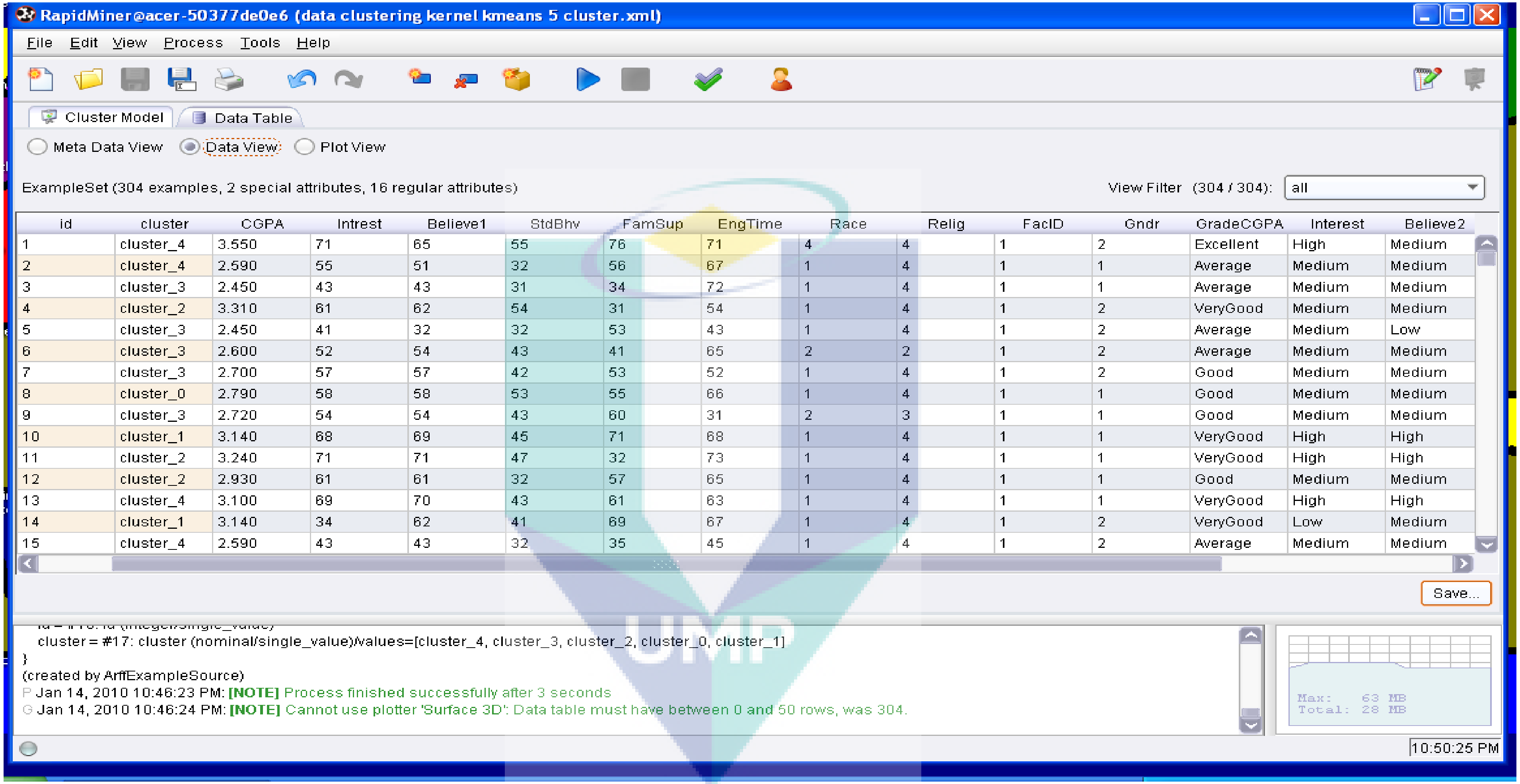
4.3.1 Student's Cluster by Kernel K-Means

As mentioned in the previous discussion, Kernel K-Means clustering works on finding the cluster centers by trying to minimize a cost function $\boldsymbol{\delta}(\boldsymbol{x_i}\ ,\boldsymbol{C_k})$. It alternates between updating the membership matrix and updating cluster centers, respectively, until no further improvement in the cost function is noticed. Since the algorithm initializes the clusters randomly, its performance is affected by those initial cluster centers.

Table 4.7: **Performance Results of Kernel K-Means Clustering**

Performance Measure	Test Runs								
	1	2	3	4	5	6	8	9	10
Number of iterations	22	20	23	20	16	19	20	24	22
Root Means Square Error (RMSE)	0.369	0.469	0.357	0.459	0.442	0.463	0.357	0.357	0.369
Accuracy	85.0 %	87.0 %	85.0 %	85.0 %	83.0 %	84.0 %	87.0 %	86.0 %	87.0 %
Time Process (sec)	3.98	3.96	4.10	3.89	3.29	3.57	3.90	4.06	4.00

As seen from the results, the best case achieved 87% accuracy and RMSE of 0.357 at test run 8. This relatively moderate performance is related to the high dimensionality of the problem; having too much dimensions tend to disrupt the coupling of data and introduced overlapping in some of the dimension that reduces the accuracy of clustering. It is also noticed that the cost function converges rapidly to a minimum value as seen from the number of iteration in each test run. However, this has no effect on the accuracy measure.



RapidMiner@acer-50377de0e6 (data clustering kernel kmeans 5 cluster.xml)
File Edit View Process Tools Help
Cluster Model
Data Table
○ Meta Data View ● Data View ○ Plot View
Save...
Max: 63 MB
Total: 28 MB
10:50:25 PM

ExampleSet (304 examples, 2 special attributes, 16 regular attributes)

View Filter (304 / 304): all

id	cluster	CGPA	Intrest	Believe1	StdBhv	FamSup	EngTime	Race	Relig	FacID	Gndr	GradeCGPA	Interest	Believe2
1	cluster_4	3.550	71	65	55	76	71	4	4	1	2	Excellent	High	Medium
2	cluster_4	2.590	55	51	32	56	67	1	4	1	1	Average	Medium	Medium
3	cluster_3	2.450	43	43	31	34	72	1	4	1	1	Average	Medium	Medium
4	cluster_2	3.310	61	62	54	31	54	1	4	1	2	VeryGood	Medium	Medium
5	cluster_3	2.450	41	32	32	53	43	1	4	1	2	Average	Medium	Low
6	cluster_3	2.600	52	54	43	41	65	2	2	1	2	Average	Medium	Medium
7	cluster_3	2.700	57	57	42	53	52	1	4	1	2	Good	Medium	Medium
8	cluster_0	2.790	58	58	53	55	66	1	4	1	1	Good	Medium	Medium
9	cluster_3	2.720	54	54	43	60	31	2	3	1	1	Good	Medium	Medium
10	cluster_1	3.140	68	69	45	71	68	1	4	1	1	VeryGood	High	High
11	cluster_2	3.240	71	71	47	32	73	1	4	1	1	VeryGood	High	High
12	cluster_2	2.930	61	61	32	57	65	1	4	1	1	Good	Medium	Medium
13	cluster_4	3.100	69	70	43	61	63	1	4	1	1	VeryGood	High	High
14	cluster_1	3.140	34	62	41	69	67	1	4	1	2	VeryGood	Low	Medium
15	cluster_4	2.590	43	43	32	35	45	1	4	1	2	Average	Medium	Medium

```
  id = #16: id (integer/single_value)
  cluster = #17: cluster (nominal/single_value)/values=[cluster_4, cluster_3, cluster_2, cluster_0, cluster_1]
}
(created by ArffExampleSource)
P Jan 14, 2010 10:46:23 PM: [NOTE] Process finished successfully after 3 seconds
G Jan 14, 2010 10:46:24 PM: [NOTE] Cannot use plotter 'Surface 3D': Data table must have between 0 and 50 rows, was 304.
```

Figure 4.3: Screen Shot Mixed Datasets after clustering process

The cluster model of students' performance as shown in figure 4.3 is to help both the instructors and management to interpret the data. The cluster model of students' has five clusters of student performance as shown on figure 4.4.

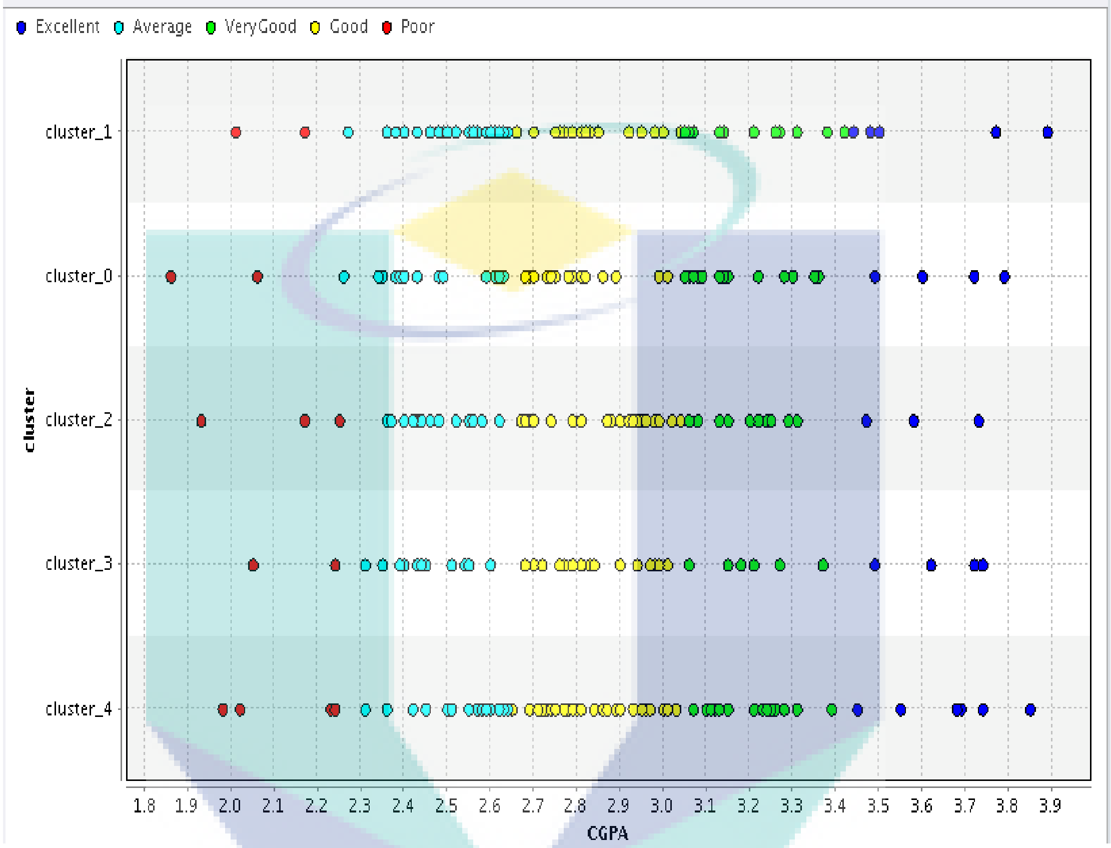


Figure 4.4: Student Cluster Model

Figure 4.4 shows the clusters model of students based on their performance which are the member of each clusters labeled with their performance index. The Cluster-0, has 65 members (21.38%), while the cluster -1, has 63 members (20.172%).Cluster-2, havs 57 members (18.75%) and cluster-3 has 59 members (18%) and the cluster-4 has 60 members (19.74%). The memberships of each cluster, could be seen on Table 4.8.

Table 4.8: Membership of Clusters

Clusters	Cluster Size	Percentage
0	65	21.38%
1	63	20.72%
2	57	18.75%
3	59	18.41%
4	60	19.74%

The cluster models of students based on their performance and related to performance predictors is presented on Figure 4.5.

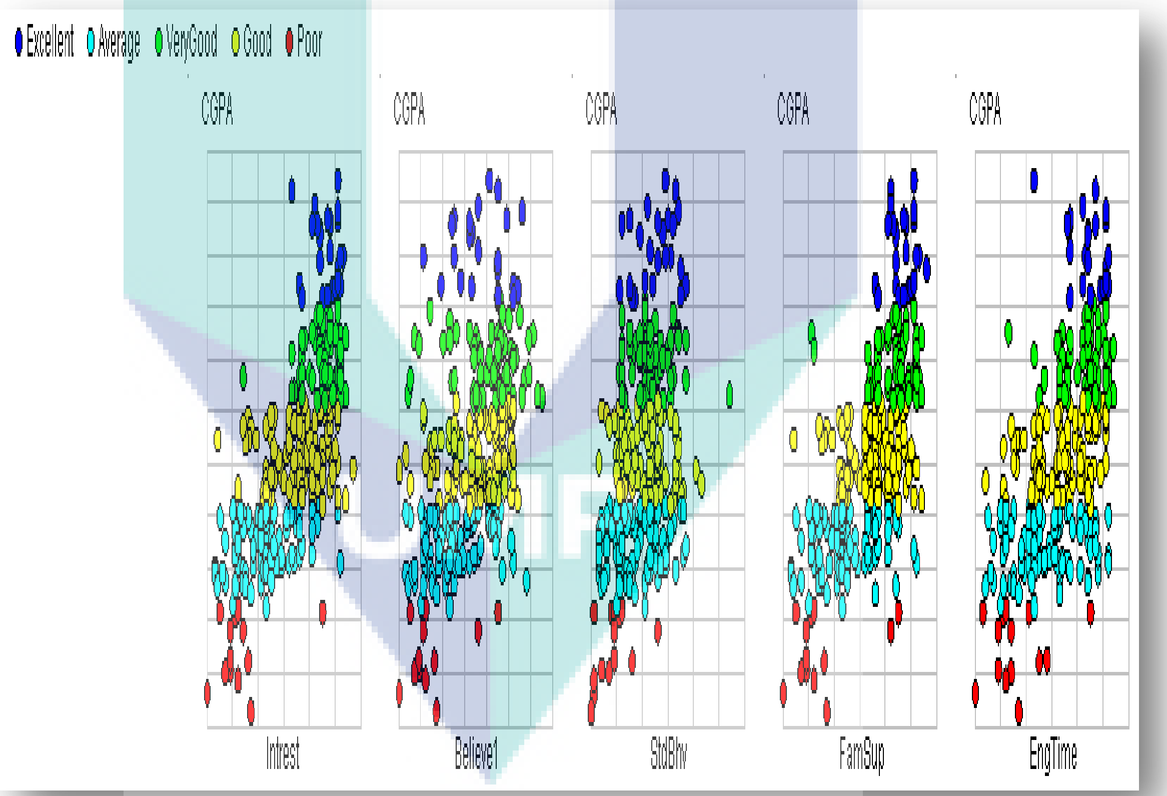


Figure 4.5: Student’s Cluster based on CGPA and performance predictors

The data is grouped based on their performance related to variables predictors. It can be seen that the performance predictors affected the performance of students. Figure 4.5 is shown to explain that the variables “Interest”, “Family Support, and “Engaged Time” have a strong correlation with the performance of the students, while the variables “Believe” and Study Behavior” have a weak correlation with students’ performance.

The clusters of students is presented in Figure 4.5, where every data point labeled by colors. The ‘blue’ data point presents the “excellent" performance in the group. The ‘red colors’ of data point presents the "poor" performance, while the ‘green’ and the ‘yellow colors’ present the "very good" and "good" performance. The students’ performance "average” is represented by the ‘light blue’ color. The cluster of students described the current condition of student performance. This information can be used by academic planners in the process decision-making.

4.3.2 Comparison Results of Kernel K-Means Vs K-Means

The other experiment of the data was evaluated by using K-Means clustering algorithm. The mixed data sets were tested but the algorithm cannot classify the mixture dataset. The cluster model by K-Means algorithm is presented in Figure 4.6.

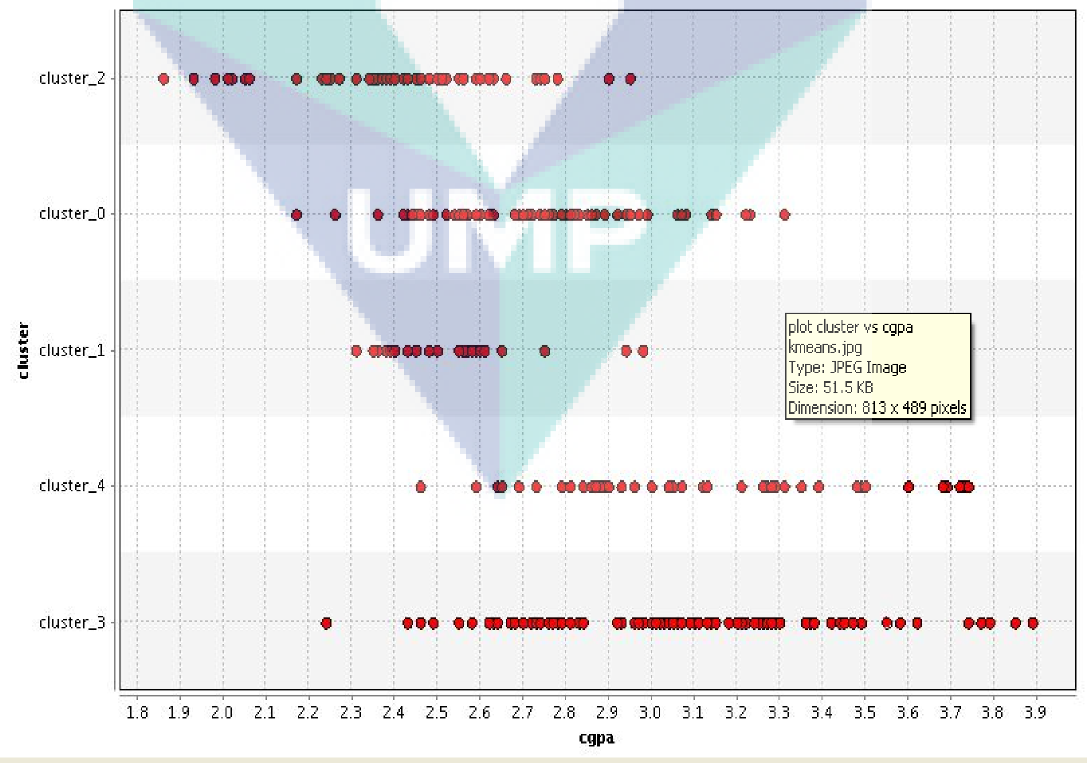


Figure 4.6: Cluster Model by K-Means Algorithm

The cluster size obtained of each cluster is too different and very difficult in interpretations the characteristics of memberships cluster (cluster analysis). All cluster size and performance index is generated by K-Means clustering and presented in Table 4.8.

Table 4.8 Cluster size and Performance Index

Cluster	Cluster size	Performance Index
Cluster-0	77	1.86-2.95
Cluster-1	71	2.19-3.31
Cluster-2	38	2.31-2.95
Cluster-3	33	2.25-3.89
Cluster-4	85	2.47-3.75

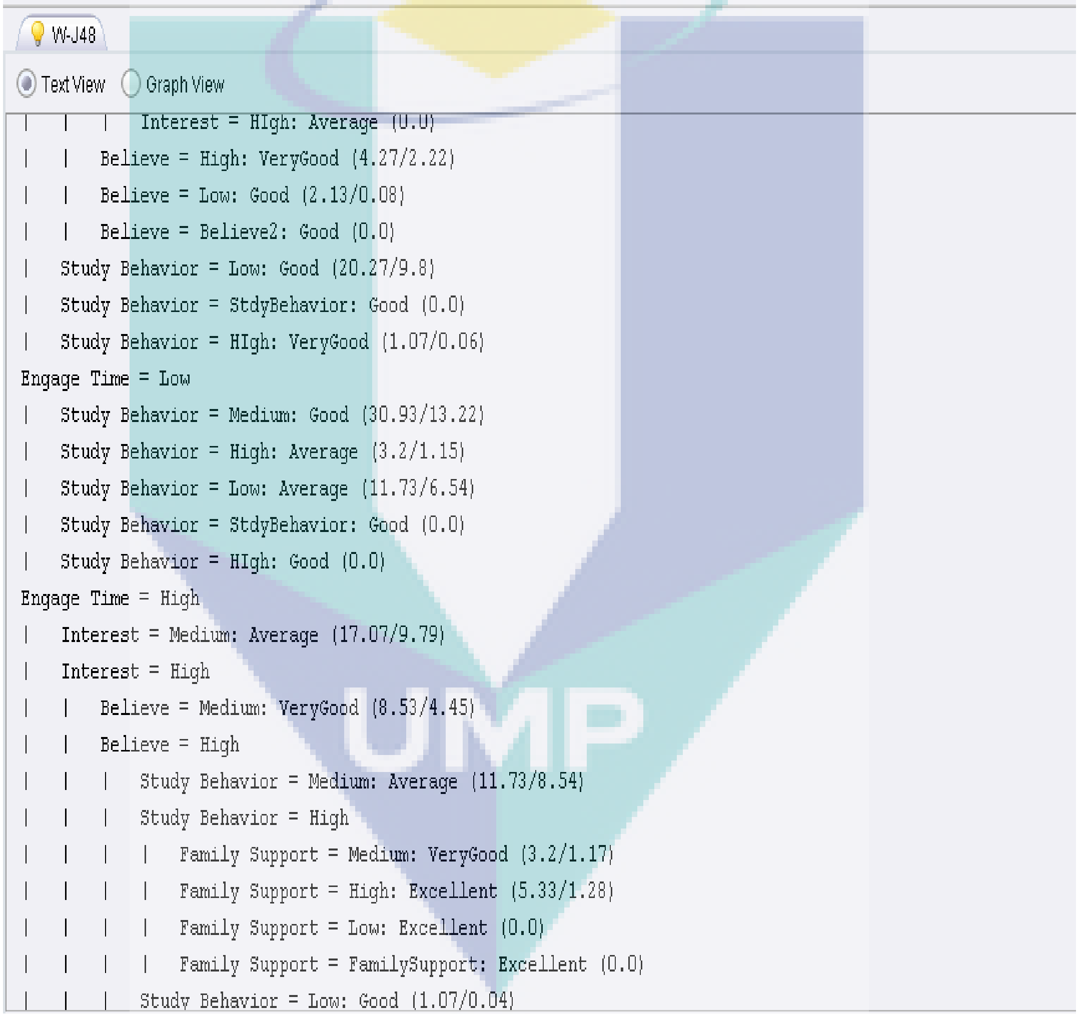
As presented, the cluster size of cluster model generated by the Kernel K-Means algorithm is almost the same in each cluster, while the cluster size of the cluster model generated by K-means algorithm is very different. It means that Kernel K-Means is better than K-Means in the data segmentation. Kernel K-Mean is a powerful application to classify high dimensional data and ability to identify nonlinear separable cluster in input space. As seen in Figure 4.4, the cluster model generated by Kernel K-Means is coupled with the label of student performance, which could not be done by K-Means clustering. K-means cannot process the mixture dataset (numerical & categorical) on this case. It could be a weakness of K-Means algorithm. The comparison of cluster size obtained between K-Means and Kernel K-Means clustering algorithms in grouping the dataset is as seen in Table 4.9.

Table 4.9: Comparison of the cluster size

Clusters	Clustering Algorithms		Percentage	
	Kernel K-Means	K-Means	Kernel K-Means	K-Means
0	65	77	21.38%	25.33%
1	63	71	20.72%	23.35%
2	57	38	18.75%	12.50%
3	59	33	18.41%	10.86%
4	60	85	19.74%	27.96%

4.4 THE RULE MODEL IN PREDICTING STUDENT PERFORMANCE.

The data analyzed by using J48 decision tree algorithm to find the logical rules model in predicting students' performance. J48 decision tree used to represent the logical rules of students' performance prediction. The tree presented was larger and the variation of rule model was many. The researchers chose the best rule model presented. The results of rule learner generated by J48 decision tree can be seen in Figure 4.7.



```
|   |   |   Interest = HIgh: Average (0.0)
|   |   Believe = High: VeryGood (4.27/2.22)
|   |   Believe = Low: Good (2.13/0.08)
|   |   Believe = Believe2: Good (0.0)
|   Study Behavior = Low: Good (20.27/9.8)
|   Study Behavior = StdyBehavior: Good (0.0)
|   Study Behavior = HIgh: VeryGood (1.07/0.06)
Engage Time = Low
|   Study Behavior = Medium: Good (30.93/13.22)
|   Study Behavior = High: Average (3.2/1.15)
|   Study Behavior = Low: Average (11.73/6.54)
|   Study Behavior = StdyBehavior: Good (0.0)
|   Study Behavior = HIgh: Good (0.0)
Engage Time = High
|   Interest = Medium: Average (17.07/9.79)
|   Interest = High
|   |   Believe = Medium: VeryGood (8.53/4.45)
|   |   Believe = High
|   |   |   Study Behavior = Medium: Average (11.73/8.54)
|   |   |   Study Behavior = High
|   |   |   |   Family Support = Medium: VeryGood (3.2/1.17)
|   |   |   |   Family Support = High: Excellent (5.33/1.28)
|   |   |   |   Family Support = Low: Excellent (0.0)
|   |   |   |   Family Support = FamilySupport: Excellent (0.0)
|   |   |   Study Behavior = Low: Good (1.07/0.04)
```

Figure 4.7: Rule Model Generated by J48 Decision Tree

The result obtained in this experiment elicits the emergence of a strong rule. For example, "if Engaged Time = high, Interest= high, believe=high and Family support= high, then performance prediction is Excellent, but if "Family support" is Medium then the performance predictions will be “Very Good”. The entire rule generated will be used as the rule model for student performance prediction. The rule model of students’ performance prediction is presented in Table 4.10.

Table 4.10: The rule model of students’ performance prediction

	Interest	Study Behavior	Engage Time	Belief	Family Support	Performance Prediction
IF	H	H	H	H	H	Excellent
	H	M	M	H	H	Very Good
	M	M	M	M	M	Good
	L	M	M	L	M	Average
	L	L	L	L	L	Poor
H=High,		M=Medium,		L= Low.		

4.5 PERFORMANCE PREDICTION ACCURACY

The experiment in this step used the second datasets in txt format. In this dataset, there were 304 pieces of samples, and every sample was expressed by ten characteristic parameters. Five variables of performance predictors as proposed in this study and five variables demographic of students used in this experiment. Among 304 of samples data, there were 4 samples with incomplete parameters. Hence in this experiment, 300 samples were used for analysis.

Based on the logical rules generated by the decision tree, the SSVM algorithm on SSVM Tenfold Software was used to analyze the data, which runs on Linux operating system. The data was divided in two sets of data automatically by SSVM tenfold randomly; 90% of datasets, used for training and 10% of datasets used for testing data.

SVM is a binary classification. Therefore, in this experiment, the testing data had been done separately to each performance grade, which means that for every grade, the true value is represented by +1 and others represented by -1. The performance of SSVM depends on the combination of several parameters; capacity parameter ν, the kernel K and its corresponding parameters. In this study RBF kernel function used, because of its good general performance and a few number of parameters; (only two parameters: ν and γ) (Santi, and Embong,, 2008). The experiment used ν=2 and γ= 0.002.

To guarantee that the results presented are valid and can be generalized for making prediction regarding new data, the data sets were randomly partitioned into training and testing data sets via 10-fold CV, and was ran automatically in SSVM tenfold. The data set is divided into 10- subset, and the holdout method is repeated 10 times. Each time, one of the 10 subset is used as the test set and the other 9 subsets were put together to form a training set. The tenfold cross validation include training accuracy and testing accuracy. Training accuracy/testing accuracy are average of training accuracy/training accuracy in 10 trials. The result of experiment on the data sets is average test accuracy and average training accuracy for each performance index, as shown on Table 4.11, Table 4.12, Table 4.13, Table 4.14 and Table 4.15, as follows:

Table 4.11: Performance Accuracy of “Excellent” Prediction

Fold Number	Size of Training Set	Size of Testing Set	Training Accuracy (%)	Testing Accuracy (%)
Fold #01	270	30	99.63	100.00
Fold #02	270	30	99.63	99.33
Fold #03	270	30	99.63	93.33
Fold #04	270	30	99.63	86.67
Fold #05	270	30	100.00	90.00
Fold #06	270	30	99.63	93.33
Fold #07	270	30	99.63	86.67
Fold #08	270	30	99.63	96.67
Fold #09	270	30	99.63	86.67
Fold #10	270	30	99.63	93.33
Best Test Accuracy = 100% (Fold: #01) Best Training Accuracy = 100% (Fold: #05) Average Test Accuracy = 92% Average Training Accuracy = 99.67%				

Table 4.12: Performance Accuracy of "Very Good" Prediction

Fold Number	Size of Training Set	Size of Testing Set	Training Accuracy (%)	Testing Accuracy (%)
Fold #01	270	30	100.00	76.67
Fold #02	270	30	100.00	66.67
Fold #03	270	30	100.00	76.67
Fold #04	270	30	100.00	70.00
Fold #05	270	30	100.00	76.67
Fold #06	270	30	100.00	70.00
Fold #07	270	30	100.00	93.33
Fold #08	270	30	100.00	70.00
Fold #09	270	30	100.00	83.33
Fold #10	270	30	100.00	73.33
Best Test Accuracy = 93.33% Best Training Accuracy = 100% Average Test Accuracy = 75.67% Average Training Accuracy = 100%				

Table 4.13: Performance Accuracy of "Good" Prediction

Fold Number	Size of Training Set	Size of Testing Set	Training Accuracy (%)	Testing Accuracy (%)
Fold #01	270	30	100.00	56.67
Fold #02	270	30	100.00	53.33
Fold #03	270	30	100.00	60.00
Fold #04	270	30	100.00	56.67
Fold #05	270	30	100.00	46.67
Fold #06	270	30	100.00	63.33
Fold #07	270	30	100.00	60.00
Fold #08	270	30	100.00	66.67
Fold #09	270	30	100.00	73.33
Fold #10	270	30	100.00	73.33
Best Test Accuracy = 73.33% (Fold: #09. #010) Best Training Accuracy = 100% (Fold: All) Average Test Accuracy = 61% Average Training Accuracy = 100%				

Table 4.14: **Performance Accuracy of "Average" Prediction**

Fold Number	Size of Training Set	Size of Testing Set	Training Accuracy	Testing Accuracy
Fold #01	270	30	99.63	63.33
Fold #02	270	30	99.63	66.67
Fold #03	270	30	99.63	80.00
Fold #04	270	30	99.63	70.00
Fold #05	270	30	99.63	76.67
Fold #06	270	30	99.63	63.33
Fold #07	270	30	99.63	76.67
Fold #08	270	30	100.00	66.67
Fold #09	270	30	100.00	70.00
Fold #10	270	30	99.63	60.00
Best Test Accuracy = 80% (Fold: #03) Best Training Accuracy = 100% (Fold: #08.#09) Average Test Accuracy = 69.33% Average Training Accuracy = 99.70%				

Table 4.15: **Performance Accuracy of "Poor" Prediction**

Fold Number	Size of Training Set	Size of Testing Set	Training Accuracy (%)	Testing Accuracy (%)
Fold #01	270	30	99.63	90.00
Fold #02	270	30	99.63	96.67
Fold #03	270	30	99.63	90.00
Fold #04	270	30	100.00	96.67
Fold #05	270	30	99.63	93.33
Fold #06	270	30	100.00	93.33
Fold #07	270	30	99.63	93.33
Fold #08	270	30	99.63	96.67
Fold #09	270	30	99.63	93.33
Fold #10	270	30	99.63	93.33
Best Test Accuracy = 96.67% (Fold: #02,#04,#08) Best Training Accuracy = 100% (Fold: #04,#06) Average Test Accuracy = 93.67% Average Training Accuracy = 99.70%				

Table 4.11 to Table 4.15 shows the average testing accuracy where "Good" obtained the lowest, performance prediction that is 61% while "Poor" achieves the

highest percentage of 93.67% prediction performance. The results of the experiment can be used as basis to state that the rule prediction model of students' performance by using psychometric factors as predictors is acceptable and good enough to serve as predictors of students' performance. The Summary of experiment results as shown in the Table 4.16 below:

Table 4.16: **Overall Performance Prediction Accuracy**

Performance Prediction	Training		Testing	
	Best Accuracy (%)	Average Accuracy (%)	Best Accuracy (%)	Average Accuracy (%)
Excellent	100.00	99.67	100.00	92.00
Very Good	100.00	100.00	93.33	75.67
Good	100.00	100.00	73.33	61.00
Average	100.00	99.70	80.00	69.33
Poor	100.00	99.70	96.67	93.67

Based on the result obtained, the lowest prediction accuracy (61%) on "Good" performance prediction may be caused by a wide variation of the rule.. This case requires further testing to find the better rule for "Good" performance prediction. However, the overall results obtained are sufficient to prove that the predictor students' performance proposed is a prior predictor for predicting students' academic performance.

4.6 SUMMARY

The main objective of this study is to apply kernel methods in educational area and to create the model of student performance prediction. The results of this study proves that the Kernel K-Means clustering is a promising method in student segmentation based on their performance, with results showing the current condition of student performance. Information of student performance in the real time could be used to help academic planners in the process of decision making.

The finding of these studies revealed that Kernel K-Means clustering and Smooth Support Vector Machine classification were capable to explore the data that comes from educational area. This means that other kernel methods are very promising to be used as a methodology in Educational Data Mining.

The findings also displayed a trend suggesting that psychometric factors of student is good to serve as a predictor model in educational predictive data mining modeling. These findings revealed that the variable predictors proposed had a strong correlation (R^2= 0 .522) with students' performance. The performance predictors used in this study is a good variable used in application of student performance prediction.

This study identified the logical rules for prediction of students' performance generated by J48 decision tree and used SSVM to classify the students based on their performance index. Tenfold cross validation method was used to guarantee that the present results are valid and can be generalized for making predictions.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 INTRODUCTION

This chapter presents the conclusion of the study and the implication for future practice and research. This study utilizes data mining in the field of education area. The Kernel Methods; Kernel K-Means and Smooth Support Vector Machines were used as data mining techniques to extract the data. The steps of data mining process were carried out and explained in detail. The area of application was education, different from usual data mining studies. The use of the data mining technique in education may provide us with more varied and significant findings, and may improve the quality of decision-making in the higher education system.

5.2 CONCLUSIONS

The purpose of this study is to apply Kernel Methods; Kernel K-Means Clustering and Smooth Support Vector Machine classification to explore the data in educational area. Kernel K-Means clustering had been used in grouping the students based on their psychometric factors and performance.

This study has shown the significance of the relationship between the predictor variables with students' performance. The predictor variables used in this study are psychometric factors of student; *Interest*, *Study Behavior*, *Engage Time*, *Believe, and*

Family Support. These predictor variables contributes significantly in increasing or decreasing students' performance [52.2% (R^2=.522)].

The study also identified the cluster model of students based on their performance by using Kernel K-means Clustering techniques. Each member of the clusters was labeled with their performance index to describe the current condition of student performance.

This study generated the rule model in predicting student performance. The performance index of student was classified into five classes; *Excellent, Very Good, Good, Average, and Poor*. The rule model had been tested to classify the students based on their performance index. Smooth Support Vector Machine algorithm had been applied to predict the students' performance and tenfold Cross Validation (10-CV) method was used to guarantee the accuracy of the predicted results. The prediction accuracy of the predicting model has the lowest accuracy that is 61% on "Good" performance prediction index and the highest accuracy is 93.67% in "Poor" Performance prediction index. Furthermore, the prediction accuracy on "Excellent" performance index is 92%, and prediction accuracy is 75.67% for "Very Good", 69.33% for "Average" performance index.

Finally, this study is a case study in predictive performance modeling in educational data mining. Kernel Method: Kernel K-Means Clustering and Smooth Support Vector Machines Classification had been applied as data mining techniques in educational data mining area. Educational data mining still provides promising research directions combining more aspects from educational data. Particularly, the result of this study serves as a good benchmark in predicting models of students' performance in higher learning institution.

5.3 RECOMMENDATION FOR FUTURE RESEARCH

Educational Data Mining is a research area that is still promising for researchers to discover or improve models of a domain's knowledge structure. Concern in predicting students' performance is an opportunity to use a combination of psychometric and soft skills variables as predictor variables. To obtain comparing results of prediction accuracy, the several methods of validation such as confusion matrix is reasonable.

5.4 FUTURE WORK

SSVM for multi-class classification is a very promising application in this case study to obtain a more accurate prediction results. In addition, the experiment could be done using more data mining techniques, such as Bayesian Networks, genetic algorithm, k-nearest neighbor and the others kernel methods. Soft Skills variables and psychometrics could be combined as predictors in predicting student performance.

REFERENCES

Abelson, R. 1968. *Theories of Cognitive Consistency: A Sourcebook*. Rand McNally,

Adriaans, P. and D. Zantige, 2000. *Data Mining.* Addison Wesley.

Alexander, P., and Judy, J. 1988. *The interaction of domain-specific and strategic knowledge in academic performance*. Review of Educational Research, (58)**:** 375-404.

Al-Radaideh, Q. A., Al-Shawakfa, E. M., and Al-Najjar, M. I. 2006. Mining student data using decision trees. *Proceedings of the 2006 International Arab Conference on Information Technology (ACIT'2006).*

Anil, K.J., Murty, M.N. and Flynn, P.J. 1999. Data Clustering: A Review. *ACM Computing Surveys*. Vol. 31 No.3.

Apte, C. and Weiss, S. 1997. Data Mining with Decision Trees and Decision Rules, *Future Generation Computer Systems*, pp.197-210

Armijo, L. 1966. Minimization of functions having Lipschitz-continuous first partial derivatives. *Pacific Journal of Mathematics*, (16): 1–3

Bach, F. R. and. Jordan, M. I. 2002. Kernel independent component analysis. *Journal of Machine Learning Research*, (3)**:**1–48.

Baker, R.S.J.D and Yacef, K. 2009. The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining,* (1)**:** 3-17.

Baker, R.S.J.D. 2008 Data Mining For Education. In *International Encyclopedia of Education (3rd edition)*, B. MCGAW, PETERSON, P., BAKER Ed. Elsevier, Oxford,UK.

Baker, R. 2007. Modelling and Understanding Student off-task behavior in intelligent tutoring system. *Proceeding Conference Humanity Factors Computer System*. San Jose, CA. pp. 1059-1068

Barker, K., Theodore, T and Teri, R.R. 2004. Learning From Student Data. *Proceedings of the Systems and Information Engineering Design Symposium*. Mathew H. Jones, Stephen D. Patek, and Barbara E. Towney eds.pp.79-86

Barnes, T. 2005. The q-matrix method: Mining student response data for knowledge.*Proceeding AAAI Workshop Educational Data Mining.*Pittsburgh, PA. pp. 1-8

Barnes, T. Desmarais, M., Romero, C., and Ventura, S. 2009. Presented at *the 2nd International Conference, Educational Data Mining*. Cordoba Spain.

Baylis, P.1999. *Better Health Care with Data Mining.* SPSS White Paper, UK.

Bean, J.P and Metzner, B.S. 1985. A Conceptual Model of Nontraditional Undergraduate Student Attrition. Review of Educational Research Winter **55** (4)**:** 485-540

Beck, J.E. and Mostow,J. 2008. How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, pp.353-362.

Bernhard, S and Smola, A,J. 2002, *Learning with Kernels — Support Vector Machines, Regularization, Optimization and Beyond*. The MIT Press Cambridge, Massachusetts London, England.

Bertsekas, D.P. 1999. *Nonlinear Programming*. Athena Scientific, Second Edition Belmont, MA.

Biehler, R.F and Snowman, J. 1996. *Psychology Applied To Teaching.* 8th Revised Edition Houghton Muffin. Boston

Bodea, C., Bodea, V., and Tudor, C.A. 2006. Data mining in higher education*, The 3rd International Workshop IE&SI*, Timisoara, pp.19-26

Boero, G., Laureti, T., and Naylor, R. 2005. An econometric analysis of student withdrawal and progression in post-reform Italian universities. Centro Ricerche Economiche Nord Sud - *CRENoS Working Paper 2005/04*.

Boyd, S.and Vandenberghe, L.2004. *Convex optimization*. Cambridge University Press, 2004.

Budoff, M. and Corman L. 1974. Demographic and psychometric factors related to improved performance on the Kohs learning-potential procedure. Am J Ment Defic.,**78** (5):578-85

Brachman, R.J. and Anand, T. 1996, *The process of knowledge discovery in databases: a human-centered approach*, *Advance in Knowledge Discovery and Data Mining.* AAAI/MIT Press, Menlo Park, CA/Cambridge, MA, pp. 33-58.

Bresfelean, VP, Bresfelean M, Ghi_oiu N.and Comes C-A. 2007. Data mining in continuing education. *INTED 2007, International Technology, Education and Development Conference,* Valencia, Spain.

Bresfelean, VP., Bresfelean, M., Ghi_oiu, N. and Comes, C-A. 2006. Continuing education in a future EU member, analysis and correlations using clustering techniques. *Proceedings of The 5th WSEAS International Conference EDU '06,* Tenerife, Spain, pp.195-200.

Brophy, J. 1983. Conceptualizing student motivation.*Journal of Educational Psychologist,* (18): 200-215.

Burges, C. J. C. 1998. A tutorial on support vector machines for pattern recognition.

Burlak G, Munoz J, Ochoa A, and Hernández JA. 2006. Detecting Cheats in Online Student Assessments Using Data Mining. *Proceedings of DMIN.* pp. 204-210

Bushara, I.C. and Rehana, M. 2010. Impact of Parental Support On the Academic Performance and Self Cencept of this student. *Journal of Research and Reflection in Education*. .**4** (1): 14-26

Byun, H. and Lee, S.W. 2003. A survey on Pattern Recognition Application of Support Vector Machines. *International Journal of Pattern Recognition and Artificial Intelligence*. **17 (**3): 459-486.

Chakrabarti, S., Cox., F. G, Han., Jiang., Kamber., Lighstones, Nadeau., Naepolitan., Pyle., Refaat., Schneider., Teorey, Witten, F. 2009. *Data Mining: Know it all*. Morgan Kaufman Publisher, Burlington, MA. 01803. ISBN 978-0-12-374629-0.

Chang, W.H.T and Lee Y. H, 2000 Telecommunications Data Mining for Target Marketing. *Journal of Computers,* **12** (4): 60-74.
Chapman & Hall/CRC.

Cherkassky, V.S. and Filip, M. 2007. *Learning From Data . Concepts, Theory, and Methods*.2nd Edition. John Wiley & Sons, Inc. New York.

Cherkassky, V.S. and Filip, M. 1998, *Learning From Data-Concepts, Theory and Methods*. Jhon Wiley & Son, Inc. New York.

Chonan, B.I. and Khan, R. M. 2010. Impact of Parental Support on the Academic Performance and Self Concept of the Student. *Journal of Research and Reflections in Education*. **4** (1):14-26.

Choquet, C., Luengo, V. and Yacef, K. 2005. Usage Analysis in Learning Systems. *Workshop Proceedings, held in conjunction with AIED 2005*, Amsterdam, The Netherlands, July 2005.

Christopher, B.1998. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery,* pp. 121–167.

Chunhui C and. Mangasarian. O. L. 1996. A class of smoothing functions for nonlinear and mixed complementarity problems. *Computational Optimization and Applications*, **5** (2): 97–138.

Chunhui, C and Mangasarian, O.L. 1995. Smoothing methods for convex inequalities and linear complimentarily problems, *Mathematical Programming*, **71** (1): 51-69.

Cocea, M. and Weibelzahl, S. 2007. Cross-System validation of engagement prediction from log files. *Proceeding of international conference. Technology Enhanced learn.* Crete Greece, pp. 14-25.

Corbett, A.T. 2001. Cognitive Computer Tutors: Solving the Two-Sigma Problem. *Proceedings of the International Conference on User Modeling*. pp. 137-147.

Cortez, P. and Silva, A. 2008. Using data mining to predict secondary school student performance. *Proceedings of 5th Annual Future Business Technology Conference*, Porto, Portugal, pp. 5-12.

Cristianini, N. and Taylor J.S. 2000. *An introduction to Support Vector Machines and other Kernel-Based Learning Methods*, Cambridge Press University.

Cunha, M.M., Putnik, G.D., Ávila, P. 2000. Towards Focused Markets of Resources for Agile /Virtual Enterprise Integration. *Advances in Networked Enterprises*: *Virtual Organisations, Balanced Automation, and Systems Integration*, Kluwer Academic. pp. 15-24.

David, H., Heikki, M and Padhraic, S. 2001. *Principles of Data Mining*, MIT Press Cambridge.

Dekker, G.W., Pechenizkiy, M., and Vleeshouwers, J. M. 2009. Predicting student drop out: A case study. *Proceedings of the 2nd International Conference on Educational Data Mining (EDM'09)*. Cordoba, Spain. pp. 41-50.

Delavari, N., Beikzadeh, M.R., and Amnuaisuk, S.P. 2005. Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System. *5th International Conference on Information Technology based Higher Education and Training: ITEHT '05,* July 7 – 9, 2005, Juan Dolio, Dominican Republic: F4B 1-6.

Dennis J. E. and. Schnabel R. B. 1983 *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, N.J.

Desmarais M.C. and PU, X. 2005. A Bayesian Student Model without Hidden Nodes and Its Comparison with Item Response Theory. *International Journal of Artificial Intelligence in Education* (15)**:** 291-323

Dhillon L.S, Y.Guan and B.Kullis, 2005. *A unified view of kernel k-means spectral clustering and graph partitioning*. Technical Report. Department of Computer Science. University of Texas Austin.

D'mello, S.K., Craig, S.D., Witherspoon, A.W., Mcdaniel, B.T. and Graesser, A.C.2008. Automatic Detection of Learner's Affect from Conversational Cues. *User Modeling and User-Adapted Interaction* (18): 45-80.

El-Halees, A. 2008. Mining Student Data To Analyzed Learning Behavior: A Case Study.(online): http://uqu.edu.sa/files2/tiny_mce/plugins/filemanager/files/30/papers/f158.pdf. (15 December 2009)

Fayyad, U. M, 1996, *Advances in Knowledge Discovery and Data Mining*. Camberidge, MA: The MIT Press.

Fer, S. 2004. Qualitative Evaluation of Emotional Intelligence In-Service Program for Secondary School Teachers. *Communications of the Qualitative Report* **9(***4)* December 2004.

Furqan, M., Embong, A., Suryanti, A., Santi W.P. and Sajadin, S. 2009. Smooth Support Vector Machine For Face Recognition Using Principal Component Analysis. *Proceeding 2nd International Conference on Green Technology and Engineering (ICGTE).*Faculty of Engineering Malahayati University, Bandar Lampung, Indonesia.

Gall, M., Borg, W and Gall, J. 1996, *Educational Research: An Introduction.* White Plains NY: Longman Publisher

Girolami, M. 2002. Mercer Kernel-Based Clustering in Feature Space. *IEEE. Transactions on Neural Networks.* **13** (3):780-784,

Glymour, C. 2001. *Mind's arrows: Bayes and Graphical causal models in psychology*. Cambridge, MA: MIT Press.

Golding, P and Donaldson. O. 2006, Predicting Academic Performance, *36th ASEE/IEEE Frontier In Education Conference*, TID-21.

Gong, Y., Rai, D., Beck, J and Heffernan, N. 2009. Does Self-Discipline Impact Students' Knowledge and Learning?. *Proceedings of the 2nd International Conference on Educational Data Mining*, pp. 61-70.

Guha. S, Rajeev, R and Kyuseok, S, 1998. ROCK : A robust Clustering Algorithm for Categorical Attributes. *Proceeding of 15th International Conference on Data Mining 1999.*

Guyon, and Ellisseeff,A. 2003. An introduction to variable and feature selection*, journal of Machine learning research,* pp.1157-1182.

Han J, and Kamber M, 2001. *Data Mining: Concepts and Techniques,* Simon Fraser University, Morgan Kaufmann publishers, ISBN 1-55860-489-8.

Han, J. and Kamber, M. 2006. *Data mining: Concepts and Techniques*, 2nd edition. The Morgan Kaufmann series in Data Management System, Jim Grey, series Editor.

Han, J., and Kamber M. 2003. *Data Mining: Concepts and Techniques* Morgan Kaufman Publishers, New Delhi.

Hardoon, D. R., Szedmak, S., and J. Shawe-Taylor.2003. Canonical correlation analysis: An overview with application to learning methods. Technical Report. CSD-TR-03-02.2003 Royal Holloway University of London.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. *The elements of statistical learning: Data mining, inference and prediction.* 2nd ed.. New York: Springer.

Haykin, S.1999. *Neural Networks: A Comprehensive Foundation, 2nd. Ed* Prentice Hall, Englewood Cliffs, NJ, USA.

Health, M.T. 2002. *Scientific Computing : An Introductory Survey*. McGraw-Hill,

Heathcote E and Dawson S. 2005. Data Mining for Evaluation, Benchmarking and Reflective Practice in a LMS. *In Proceedings E-Learn 2005* Vancouver, Canada.

Herrera, O. L. 2006. *Investigation of the role of pre- and post-admission variables in undergraduate institutional persistence, using a Markov student flow model*. PhD Dissertation, North Carolina State University,USA.

Hinnenburg and. Keim, D. A. 2003. A general approach to clustering in large databases with noise. *Knowledge and Information System (KAIS)*, **5**(4): pp. 387-415,

Hornik, K., Stinchcombe, M. and White. H. 1990.Universal approximation of an unknown mapping and its derivatives using multilayer feed forward networks. *Neural Networks*, **3 (**5): 551–560.

Jeong, H. and Biswas, G. 2008. Mining Student Behavior Models in Learning-by-

Jun, J. 2005. *Understanding dropout of adult learners in e-learning*. PhD Dissertation, The University of Georgia, USA.

Kaufman L, and. Rousseeuw, P.J. 1990. *Findings Groups in Data: An Introduction to Cluster*L.*Analysis*. New York: Jhon Wiley & Sons.

Kay, J., Maisonneuve, N., Yacef, K. and Reimann, P. 2006. The Big Five and Visualisations of Team Work Activity. *Intelligent Tutoring Systems*, M. IKEDA, K.D. ASHLEY and T.-W. CHAN Eds. Springer-Verlag, Taiwan, 197-206.

Klaus, R.M.. and Sebastian, M. 2001. An Introduction to Kernel-Based Learning Algorithms. *IEEE Trans. On Neural Networks*, **12** (2).

Kotsiantis S., Pierrakeas, C,. and Pintelas, P. 2004. Predicting Students' Performance in Distance Learning Using Machine Learning Techniques. *Applied Artificial Intelligence (AAI)*, **18**, (5): 411–426.

Larose. D.T. 2006, *Data Mining Methods and Models*, Jhon Wiley & Sons, Inc. Hoboken New Jersey.

Lee, Y.J. and. Mangasarian. O.L 2001. A Smooth Support Vector Machine, *Journal of Computational Optimization and Applications* (20)**:** 5-22.

Loing, B. 2005. ICT And Higher Education, *9th UNESCO/NGO Collective Consultation on Higher Education, Paris*.

Luan, J. 2001. Data Mining as Driven by. Knowledge Management in Higher Education. Persistence Clustering And Prediction. *SPSS Public Conference*. (online): http://nulan.mdp.edu.ar/381/1/00637.pdf (25 April 2010)

Luan, J. and Serban, A.M. 2002. *Data Mining and Its Application in Higher Education. Knowledge Management – Building a Competitive Advantage in Higher Education*. New directions for Institutional Research. Jossey-Bass, 2002. pp22-48

Luan, J. 2002. Data Mining and Knowledge Management in Higher Education:. Potential Applications. *Presentation at Air Toronto,* Canada.(online). http://eric.ed.gov/PDFS/ED474143.pdf. (25 January 2010).

Luan, J. 2004. *Data Mining Applications in Higher Education.* Executive Report @2004 SPPS Inc.

Lubinski, D., Webb, R. M., Morelock, M. J, and Benbow, C. P.2007.. Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology,* **86**, pp.718–729.

Ma, Y., Liu B., Wong C., Yu P. and Lee S. 2000. Targeting the right students using data mining. *Proceeding. of 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Boston, USA, pp. 457–464.

Madhyastha, T. and Tanimoto, S. 2009. Student Consistency and Implication for Feedback in Online Assessment Systems. *Proceedings of the 2nd International Conference on Educational Data Mining*, pp. 81-90.

Malhotra, N. K. 1999. *Marketing Research*, 3rd edition, Prentice Hall.

Mangasarian O.L, 1999. Arbitrary-norm separating plane. *Operations Research Letters*, (24)**:**15-23, ftp://ftp.cs.wisc.edu/math-prog/tech-reports/97-07r.ps

Mangasarian O.L, 2000, Generalized Support Vector Machines. In Smola, A., Bartlett, P., Scholkop, B., and Schurrmans, D. editors. *Advances in large Margin Classifiers,* Cambridge , pp. 135-146.

Mason, N.W.H., Mac.Gillivray, K., Steel, J. B. 2003. *An index of functional diversity.* J. Veg. Sci. (14): 571_/578

Mathews and Sicuranza, G.L. 2000. *Polynomial Signal Processing.* John Wiley & Sons, New York.

Mavrikis, M. 2008. Data-driven modeling of students' interactions in an ILE. *Proceedings of the 1st International Conference on Educational Data Mining*, pp. 87-96.

Mcquiggan, S., Mott, B. and Lester, J. 2008. Modeling Self-Efficacy in Intelligent Tutoring Systems: An Inductive Approach. *User Modeling and User-Adapted Interaction* (18): 81-123.

Merceron, A and Yacef, K. 2005. Educational Data mining: A case study. *Proceedings of the 12th International Conference on Artificial Intelligence in Education AIED 2005,* Amsterdam, The Netherlands, IOS Press.

Micchelli, C., Xu, Y, and Zhang, H. 2006. Universal Kernels. *Journal of Machine Learning Research*, (7**):** 2651–2667

Mika, S., Rätsch, G., Weston, J., Schölkopf, B. and Müller, K.R. 1999. Fisher discriminant analysis with kernels. *Neural Networks for Signal Processing IX. Proceedings of the 1999 IEEE Signal Processing Society Workshop*. Madison, WI, USA. pp. 41–48.

Minaei,B.B, Kashy D., Kortemeyer G., and Punch W. 2003. Predicting student performance: an application of data mining methods with an educational web-based system. *Proceeding. of IEEE Frontiers in Education*. Colorado, USA. pp. 13–18.

Mirkin, B. 2005. *Clustering For Data Mining: A Data Recovery Approach*.

Mitchell, M and Janina, J. 1999. *Research Design Explained*. New York: Holt, Rinehart and Winston.

Mitra, S., and Acharya, T. 2003. *Data Mining. Multimedia, Soft Computing, and Bioinformatics.* John Wiley & Sons, Inc., Hoboken, New Jersey;

Mohd. Noor, F and Mohd Afifi, A. M. 2005. A-Charting Students' Academic Performance. *Proceedings of the 2005 Regional Conference on Engineering Education.* Session 03-001 December 12-13, 2005, Johor, Malaysia.

Mohd. Noor, F., Jamaluddin, Mohd. Yatim, and Azmahani, A. A. 2004. Program Peningkatan Prestasi Akademik Pelajar di Fakulti Kejuruteraan Awam. *Conference on Engineering Education.* Kuala Lumpur.

Moore, A.W. 2006. Statistical Data Mining Tutorials. (online): http://www.autonlab.org/tutorials (20 January 2010).

Neil, M. R.S. 2007. *Beginning Linux Programming:, 4th edition*, Wiley Publishing inc.

Nelo, C and. Shawe, J.T. 2000. *Introduction to Support Vector Machines: And Other Kernel-Based Learning Methods*. Cambridge University Press, Cambridge, U.K., Ng, A. Y., Jordan, M. I and Weiss, Y. 2001. On spectral clustering: Analysis and an algorithm. *Proceedings of the 14th Annual Conference on Neural Information Processing Systems (NIPS 2001)*, pp. 849–856. Whistler, BC, Canada.

Nolan, J. 1998. A Prototype Application of Fuzzy Logic and Expert System in Education Assessment. *AAAI/IAAI Proceedings,* pp. 1134-1139.

Nugroho, A.S. 2008. *Pengantar Support Vector Machine*, Intitut Teknologi Telkom. (online): http://asnugroho.wordprress.com. (20 January 2010)

Nugroho, A.S., Witarto, A.B, and Handoko, D. 2003. Application of Support Vector Machine in Bioinformatics. *Proceeding of Indonesian Scientific Meeting in Central japan, Gifu-Japan.*

Oded, M and Lior, R. 2005. *Data mining and knowledge discovery handbook*, Springer.

Ogor. E N. 2007. Student Academic Performance Monitoring and Evaluation

Olson, D.L and Delen, D. 2008. *Advanced Data Mining Techniques*. Springer-Verlag Berlin Heidelberg.

Ones, D., Kuncel, N and Hezlett, S. 2004. Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All?" *Journal of Personality and Social Psychology.* American Psychological Association, Inc. 86,(1): 148–161

Oyelade, O.J, Oladipupo, O. and Obagbuwa, O.I.C. 2010. Application of K-means Clustering Algorithm for prediction of student Academic Performance. *International Journal of Computer Science and Information Security*.7 **(**1): 191-200.

Pardos, Z., Heffernan N,. Anderson B,. and Heffernan C. 2006. Using Fine-Grained Skill Models to Fit Student performance with Bayesian Networks. *Proceedings. of 8th Int. Conf. on Intelligent Tutoring Systems*. Taiwan.

Parr, R.O. 2001. *Data Mining Cookbook. Modeling Data for Marketing, Risk, and Customer Relationship Management*. John Wiley & Sons, Inc.

Pascarella, E. T., Duby, P. B., and Iverson, B. K.1983. A test and reconceptualization of a theoretical model of college withdrawal in a commuter institution setting. *Sociology of Education,* (56): 88-100.

Pavlik, P., Cen, H. and Koedinger, K. 2009. Learning factors transfer analysis using Learning curve analysis to automatically generate domain models. *Proceeding international Conference Educational Data Mining.* Cordoba Spain. pp. 121-130

Pavlik, P., Cen, H., WU. L. and Koedinger, K. 2008. Using Item-type Performance Covariance to Improve the Skill Model of an Existing Tutor. In *Proceedings of the 1st International Conference on Educational Data Mining*, pp. 77-86.

Perera, D., K.A.Y, Koprinska,,J.I., Yacef, K. and Zaiane, O. 2009. Clustering and sequential pattern mining to support team learning. *IEEE Transactions on Knowledge and Data Engineering* (21): 759-772

Pintrich, P.R and Groot, De., Elisabeth V. 1990. Motivational and Self-Regulated Learning Components of Classroom Academic Performance. Journal of Educational Psychology. 8 (1): 33-40

Pintrich, P. R. 1988. A process-oriented view of student motivation and cognition. *Improving teaching and learning through research. New directions for institutional research,* In J. S. Stark & L. Mets (Eds.), San Francisco: Jossey-Bass (57): pp. 55-70.

Pintrich, P. R. 1989. The dynamic interplay of student motivation and cognition in the college classroom. *Advances in motivation and achievement* In C. Ames & M. Maehr (Eds.). Vol. 6. Motivation enhancing environments. Greenwich, CT: JAI Press. pp. 117-160.

Pritchard M. and Wilson S., 2003. Using Emotional and Social Factors To Predict Student Success. *Journal of College Student Development*, 44 (1): 18–28.

Quinlan J.R. 1989, Unknown attribute value in induction. *Proceeding International Conference Machine Learning (ICML'89)*. NY. pp. 164-168.

Quinlan, J.R.1986. *Induction of decision trees*. Machine Learning, volume 1. Morgan Kaufmann.pp. 81-106.

Quinlann, J.R.. 1993. *C4.5 Programs for machine learning*. Morgan Kaufman.

Romero, C. and Ventura, S. 2007. Educational Data mining: A survey from 1995 to 2005, *Expert systems With Application.* (33): 135-146.

Romero, C., Ventura, S., Hervas, C.and Gonzales, P . 2008. Data mining algorithms to classify students. *Proceeding International Conference Educational Data Mining.* Montreal Canada. pp. 8-17.

Rousseeuw, P.J. 1987. A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational Applied Math,* (20): 53–65.

Rusu, L and Bresfelean VP. 2006. Management prototype for universities. *Annuals of the Tiberiu Popoviciu Seminar, International Workshop in Collaborative Systems,* Mediamira Publisher, Cluj Napoca, Romania; pp. 287-295

Sahay, A and Karun, M. 2010. Assisting Higher Education in Assessing, Predicting, and Managing Issues Related to Student Success: A Web-based Software using Data Mining and Quality Function Deployment. *Academic and Business Research Institute Conference 2010, Las Vegas.*

Santi, W.P. and.Embong, A. 2008. Smooth Support Vector Machine for Breast Cancer Classification, *IMT-GT Conference on Mathematics, Statistics and Applications(ICMSA)*.

Schölkopf. B. and Smola, A. J. 2002. *Learning with Kernels*. The MIT Press, Cambridge, MA, USA.

Schölkopf, B., Smola, A and. Müller, K.-R. 1998. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10 **(**5): 1299–1319.

Shawe, T J. and Cristianini, N. 2004. *Kernel Methods for Pattern Analysis*. Cambridge University Press.

Shyamala. K., and. Rajagopalan, S.P. 2006. Data Mining for a Better Higer Educational System. *Information Technology Journal* 5 (3): 560-564.

Siraj, F., and Abdoulha, M. A. 2009. Uncovering hidden information within university's student enrolment data using data mining. *MASAUM Journal of Computing,* 1 (2): 337-342.

Soumen, C. 2009, *Data Mining Know It All*. Morgan Kaufmann Publishers.

Steinbach, M., Karypis, G. and Vivin, K. 1999. *A Comparison of Document Clustering.*

Strayhorn, T. L. 2009. An examination of the impact of first-year seminars on correlates of college student retention. *Journal of the First-Year Experience & Students in Transition,* 21 (1): 9-27.

Superby, J.F., Vandamme, J.P. and Meskens, N. 2006. Determination of factors influencing the achievement of the first-year university students using data mining methods. In *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems (ITS 2006)*, 37-44.

Tanimoto, S.L. 2007. Improving the Prospects for Educational Data Mining. *Proceedings of the Complete On-Line Proceedings of the Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling (UM 2007*)pp, 106- 110.

Tella, A. 2007. The Impact of Motivation on Student Achievement And Learning Outcomes in Mathematics among secondary Schools Student in Nigeria. *Eurosia Journal of Mathematics, Science and Technology Education.* **3**(2): 149-156

Tella, A. and Tella, A. 2003. Parental involvement, Home background, and School Environment as Determinant of Academic Achievement of Secondary School Students in Osun State, Nigeria. *African Journal of Cross-Cultural psychology and Sport Facilitation*, **5** (2): 42-48.

Tharp, J. 1998. Predicting persistence of urban commuter campus students utilizing student background characteristics from enrollment data. *Community College Journal of Research and Practice,* (22): 279-294.

Thompson, B. 1995. Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement,* 55 (4): 525-534.

Tsuda, K. 2000. Overview of Support Vector Machine, *Journal of IEICE.* 83 (6): 460-466.

Two Crows Corporation. 1999. *Introduction to Data Mining and Knowledge Discovery"*, TwoCrows Corporation, Third Edition, U.S.A.
Using Data Mining Technique. *4th Congress of Electronics, Robotics and Automotive Mechanics*. IEEE Computer Society,

Vandamme, J.P., Meskens, N. and Superby. J.F. 2007. Predicting academic performance by data mining methods. *Journal of Education Economics,* 15 (4): 405-419.

Vapnik, V.N.1999. *The Nature of Statistical Learning Theory*, 2nd edition, Springer-Verlag, New York Berlin Heilderberg.

Vapnik,V.N. 1995. *The Nature of Statistical Learning Theory*. Springer- Verlag New York, Inc., New York. USA.

Waiyamai, K. 2003. *Improving Quality Graduate Student by Data Mining*. Departement of Computer engineering. Faculty of Engineering. Kasetsart University, Bangkok Thailand.

Wang, M.C., Haertel, G.D., and Walberg, H. J. 1997. Toward a knowledge base for school learning. Review of Educational Research, (63): 249–294.

Witten, I.H. and Frank E. 2005. *Data Mining Practical Machine Learning Tools and Techniques*. Second Edition. The Morgan Kaufman Series in Data Management System.

Witten, I.H. and Frank, E. 1999. *Data mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Fransisco, CA.

Woodman, R. 2001. *Investigation of factors that influence student retention and success rate on Open University courses in the East Anglia region*. M.Sc. Dissertation, Sheffield Hallam University, UK.

Wu, G. 1999. On The shape of the probability Weighting Function. Journal of Cognitive Psychology. (38): 129-166.

Xiao, G and Yin, J. 2005. A New Clustering Algorithm Based On KNN and DENCLUE. *Proceedings of Fourth International Conference on Machine Learning and Cybernetics, Guangzhou*, Yu. C.H., Digangi, S., Jannasch, P. A.K.and Kaprolet, C.2008. Profiling students who take online course using Data Mining Methods. *Online Journal Distance Learning Administ.* 11 (2): 1-14

Zhang, R and Rudnicky, A I. 2006. *A large Scale Clustering Scheme for kernel-K-Means.* School of Computer Science, Carnegie Mellon University 5000 Forbes Avenue, Pittsburgh, PA 15213, USA.

Zlatko, J.K. 2009. Predicting Student Success by mining enrolment data. Research in Higher Education Journal (online): http://www.aabri.com/manuscripts/11939.pdf. (20 January 2010)

Zlatko, J.K. 2010. Early Prediction of Student Success : Mining Students Enrolment Data. *Proceedings of Informing Science & IT Education Conference (InSITE) 2010.*

APPENDIX A
RESEACRH QUESTIONNAIRE

Dear,
Participant

I am a Master candidate in Computer Systems & Software Engineering Faculty at University Malaysia Pahang. I am writing to request your participation in my thesis research, investigating the mental condition of students (Psychometric factors) as independent variables that influence student's academic performance at higher learning institution. To date, virtually no research study exists about mental conditions of student as performance predictors of academic performance in higher learning institution. Thus, your participation will help resolve a significant gap in the literatures.

You may take about 30 minute to fill all items. If you have completed the questionnaire, you can bring it into your faculty office, and I will collect it soon. Please be informed that all information given by you will be strictly confidential and will be use only for this research. I will make sure that your answers cannot be linked to you personally, when I send my results to publication or journal. There are no risks to you or to your privacy if you decide to join my study by filling out this survey. For each of the following statements, there is no right or wrong answer; there is no good answer or bad answer. The important thing to do is in what extent the statements or questions almost or nearest describe your condition right now.

Thank you for your participation

Sincerely

Master Candidate

Section A: Demografik Responden

Instruksi : silahkan beri tanda (X) untuk pada kotak jawaban

JENIS KELAMIN Pria □ Wanita □

USIA 18 □ 19 □ 20 □ 21 □
22 □ 23 □ > 24 □

SUKU Melayu □ India □ China □

AGAMA Muslim □ kristen □ Katholik □
Hindu □ Buddha □

FAKULTI Technology Menegment □ Mecahnical □ Chemical □
Manufactering □ Civil Engineering □
FSKKP □ □

SEMESTER 1 □ 2 □ 3 □ 4 □ 5 □ 6 □

CGPA □

Intruksi Umum :

Pilihlah jawapan daripada pernyataan di bawah ini yang mendekati atau paling mengambarkan keadaan dan diri anda yang sesungguhnya. Isi pilihan jawapan dengan 102ember tanda (x) dan semak lagi jika terdapat pernyataan yang belum terjawab.

Arahan Spesifik

Untuk skala Bahagian B hingga I Sila menjawab setiap soalan mengikut empat kategori di bawah ini:

Menurut anda, seberapa seringkah anda mengalami kejadian atau situasi seperti tertulis daripada soalan di bawah ini ?

① = Sangat tidak setuju 2 = Tidak setuju 3 = Setuju

④= Sangat setuju

Section B : Interest

		Frequently
1	Saya suka dengan mata kuliah yang saya turun	① ② ③ ④
2	Saya sangat bersemangat mengikuti kuliah di kelas	① ② ③ ④
3	Saya rajin mengikuti kuliah di kelas	① ② ③ ④
4	Saya memanglah tertarik dengan mata kuliah yang saya turun	① ② ③ ④
5	Kaedah pensyarah dalam memberikan kuliah menarik sangat	① ② ③ ④
6	Saya merasa senang dengan cara mengajar pensyarah di kelas	① ② ③ ④
7	Saya puas dengan kaedah belajar saya	① ② ③ ④
8	Saya sangat memerlukan untuk menguasai kaedah mempelajari mata kuliah yang saya turun	① ② ③ ④
9	Kurikulum pelajaran yang saya turun sangat saya sukai	① ② ③ ④
10	Saya merasa sesuai dengan kurikulum pelajaran di universiti saya	① ② ③ ④
11	Saya senang dengan kurikulum yang diajarkan di kelas	① ② ③ ④
12	Penasihat akademik saya sangat membantu saya selama di universiti	① ② ③ ④
13	Jika saya ada problem saya akan mendiskusikannya dengan penasihat akademik saya	① ② ③ ④
14	Saya suka berbual dengan penasihat akademik saya tentang problem yang saya alami	① ② ③ ④
15	Saya mengambil banyak manfaat dari penasihat akademik saya	① ② ③ ④

16	Pihak pengurus universiti tempat saya belajar banyak membantu proses kuliah saya	① ② ③ ④
17	Pihak pengurus universiti tempat saya belajar memberikan pelayanan yang berkualiti kepada pelajarnya	① ② ③ ④
18	Saya sangat selesa dengan perkhidmatan universiti	① ② ③ ④
19	Saya merasa dihargai oleh universiti dimana saya belajar dengan memberikan perkhidmatan terbaiknya	① ② ③ ④

Section C : Believe

		Frequently
1	Untuk berjaya saya harus bekerja dengan cekap	① ② ③ ④
2	Saya percaya keberjayaan merupakan hasil dari kerja keras	① ② ③ ④
3	Saya telah belajar dengan tekun hingga berjaya	① ② ③ ④
4	Saya gagal karena kurang belajar dengan sungguh-sungguh	① ② ③ ④
5	Kesalahan dalam belajar adalah bermanfaat	① ② ③ ④
6	Saya belajar dari kesalahan untuk Berjaya dalam pelajaran saya	① ② ③ ④
7	Melalui kesalahan dalam belajar maka saya boleh berjaya	① ② ③ ④
8	Kesalahan dalam belajar membuat saya semakin menguasai pelajaran	① ② ③ ④
9	Untuk berjaya saya mesti belajar sepanjang hayat	① ② ③ ④
10	Dengan belajar secara berterusan maka saya boleh Berjaya dalam pelajaran	① ② ③ ④
11	Untuk berjaya dalam peringkat universiti, maka saya mesti belajar dengan tekun	① ② ③ ④
12	Kejayaan saya dalam pelajaran merupakan hasil dari belajar secara berterusan	① ② ③ ④
13	Saya percaya bahawa setiap orang boleh berjaya	① ② ③ ④
14	Jika orang lain boleh berjaya maka saya yakin bahawa saya pun boleh berjaya	① ② ③ ④
15	Tidak ada hal yang membuat setiap orang tidak boleh berjaya asalkan ianya belajar dengan tekun	① ② ③ ④

16	Saya tidak percaya bahawa hanya orang-orang tertentu yang boleh berjaya	① ② ③ ④
17	Saya percaya bahawa belajar itu adalah ibadah kepada Tuhan	① ② ③ ④
18	Saya percaya bahawa belajar itu adalah perbuatan baik yang akan memberikan pahala	① ② ③ ④
19	Saya percaya bahawa belajar itu adalah perbuatan yang dialu-alukan dalam ajaran agama	① ② ③ ④
20	Belajar merupakan perbuatan yang suci	① ② ③ ④
21	Kejayaan dalam pelajaran akademik merupakan prasyarat untuk sukses di masa hadapan	① ② ③ ④
22	Saya yakin bahawa markah pelajaran studi saya akan menentukan kejayaan kerjaya saya di masa hadapan	① ② ③ ④
23	Saya mesti mendapatkan markah A untuk setiap mata pelajaran agar boleh menggalakkan kejayaan saya di masa hadapan	① ② ③ ④
24	Kejayaan kerjaya saya di masa hadapan salah satunya ditentukan oleh markah pelajaran studi saya	① ② ③ ④

Section D : Study Behavior

		Frequently
1	Saya sering membaca buku-buku tambahan untuk memperkaya pemahaman saya akan pelajaran yang saya turun	① ② ③ ④
2	Saya tidak bergantung pada catatan selama di kelas untuk memahami pelajaran yang saya turun	① ② ③ ④
3	Saya sering meminjam buku-buku pelajaran tambahan di perpustakaan	① ② ③ ④
4	Saya mempelajari pelajaran yang saya turun melalui semakan yang diberikan pensyarah	① ② ③ ④
5	Saya sering mempelajari semakan yang diberikan oleh pensyarah	① ② ③ ④
6	Saya lebih memahami pelajaran jika mempelajari semakan	① ② ③ ④
7	Saya sering melakukan poercobaan untuk lebih memahami pelajaran yang saya turun	① ② ③ ④
8	Dengan melakukan latihan soal, saya lebih memahami pelajaran yang saya turun	① ② ③ ④
9	Saya sering mempraktikkan apa yang dipelajari di kelas secara mandiri	① ② ③ ④

10	Saya memiliki kumpulan belajar	① ② ③ ④
11	Saya sering mengikuti diskusi dalam kumpulan belajar di universiti saya	① ② ③ ④
12	Saya terbabit secara aktif dalam kumpulan belajar	① ② ③ ④
13	Saya sering berlatih soalan pelajaran secara berterusan	① ② ③ ④
14	Saya sering mencoba secara berulang latihan soalan yang terdapat di buku pelajaran	① ② ③ ④
15	Saya suka mengerjakan latihan soalan untuk menguasai pelajaran yang saya turun	① ② ③ ④

Section E : Family Support

		Frequently
1	Saya merasa selesa di dalam keluarga saya	① ② ③ ④
2	Saya sering berbual dengan orang tua saya tentang pelbagai topik dan hal	① ② ③ ④
3	Komunikasi antara saya dan orang tua cukup baik dan terbuka	① ② ③ ④
4	Orang tua saya mahu mengerti keperluan saya	① ② ③ ④
5	Setahu saya, orang tua saya tidak pernah berkomunikasi dengan pihak fakulti	① ② ③ ④
6	Orang tua saya jarang berkomunikasi dengan penasihat akademik saya	① ② ③ ④
7	Orang tua saya tidak punya masa untuk berdiskusi dengan pihak fakulti	① ② ③ ④
8	Orang tua saya tidak pernah berkomunikasi dengan pensyarah di fakulti saya	① ② ③ ④
9	Semua keperluan studi saya dipenuhi oleh orang tua	① ② ③ ④
10	Saya tidak pernah kekurangan dalam hal kewangan selama saya mengikuti studi di universiti setakat ini	① ② ③ ④
11	Orang tua saya secara rutin memberi saya keperluan kewangan selama studi	① ② ③ ④
12	Orang tua saya sangat memahami pentingnya studi di universiti	① ② ③ ④
13	Orang tua saya berhasil memperoleh peringkat sarjana atau pascasiswazah	① ② ③ ④
14	Orang tua saya sangat menggalakkan saya untuk berjaya di universiti	① ② ③ ④

Section F : Engage Time

		Frequently
1	Selama di kelas saya sentiasa memperhatikan arahan dari pensyarah	① ② ③ ④
2	Saya selalu berkonsentrasi mendengarkan semakan pengajar di kelas	① ② ③ ④
3	Saya mengambil perhatian penuh untuk memahami arahan pengajar di kelas	① ② ③ ④
4	Saya sentiasa membuat notis selama di kelas	① ② ③ ④
5	Saya sering mempersiapkan tulisan ringkasan selepas pelajaran di kelas	① ② ③ ④
6	Saya suka membuat ringkasan pelajaran untuk dipelajari di rumah/hostel	① ② ③ ④
7	Hubungan komunikasi saya dengan pengajar di kelas cukup baik.	① ② ③ ④
8	Saya sering berdiskusi dengan pengajar jika saya belum paham sesuatu hal	① ② ③ ④
9	Pengajar sentiasa memberikan kesempatan saya untuk mengajukan soalan selama di kelas	① ② ③ ④
10	Saya selalu menghadiri jadwal pelajaran di kelas	① ② ③ ④
11	Saya malas untuk menghadiri kelas	① ② ③ ④
12	Kalau tidak karena sakit atau ada hal, maka saya sentiasa hadir di kelas	① ② ③ ④
13	Saya sentiasa menyelesaikan homework yang diberikan pengajar di kelas	① ② ③ ④
14	Saya sekali waktu pernah tidak menyelesaikan homework	① ② ③ ④
15	Saya selalu menyelesaikan tugas pelajaran yang di berikan pensyarah	① ② ③ ④
16	Saya selalu belajar setelah pelajaran berakhir di hostel	① ② ③ ④
17	Di malam hari saya sering belajar	① ② ③ ④
18	Saya sering membaca lagi pelajaran tadi pagi ketika malamnya di hostel	① ② ③ ④

APPENDIX B

Instrument's Validity & Reliability

Reliability

Scale: INTEREST SCALE

Case Processing Summary

		N	%
Cases	Valid	100	100.0
	Excluded(a)	0	.0
	Total	100	100.0

a Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.926	19

Item Statistics

	Mean	Std. Deviation	N
VAR00001	1.6600	.78135	100
VAR00002	1.6500	.75712	100
VAR00003	1.7200	.71181	100
VAR00004	1.5700	.71428	100
VAR00005	1.5700	.72829	100
VAR00006	1.4600	.65782	100
VAR00007	1.4800	.65874	100
VAR00008	1.7100	.79512	100
VAR00009	1.6100	.69479	100
VAR00010	1.5900	.68306	100
VAR00011	1.7400	.67600	100
VAR00012	1.7200	.68283	100
VAR00013	1.5600	.74291	100
VAR00014	1.5200	.65874	100
VAR00015	1.6100	.82749	100
VAR00016	1.6600	.72780	100
VAR00017	1.6500	.75712	100
VAR00018	1.6800	.73691	100
VAR00019	2.1800	.82118	100

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
VAR00001	29.6800	74.038	.544	.923
VAR00002	29.6900	74.176	.553	.923
VAR00003	29.6200	73.794	.627	.921
VAR00004	29.7700	73.088	.685	.920
VAR00005	29.7700	74.785	.528	.923
VAR00006	29.8800	73.925	.672	.920
VAR00007	29.8600	74.404	.627	.921
VAR00008	29.6300	72.276	.670	.920
VAR00009	29.7300	74.017	.624	.921
VAR00010	29.7500	73.220	.708	.920
VAR00011	29.6000	75.374	.523	.923
VAR00012	29.6200	76.177	.446	.925
VAR00013	29.7800	72.416	.712	.919
VAR00014	29.8200	74.048	.660	.921
VAR00015	29.7300	71.997	.661	.920
VAR00016	29.6800	72.624	.710	.919
VAR00017	29.6900	73.691	.592	.922
VAR00018	29.6600	73.358	.639	.921
VAR00019	29.1600	76.075	.364	.928

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
31.3400	81.964	9.05340	19

Reliability
Scale: BELIEVE SCALE

Case Processing Summary

		N	%
Cases	Valid	100	100.0
	Excluded(a)	0	.0
	Total	100	100.0

a Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.772	13

Item Statistics

	Mean	Std. Deviation	N
VAR00002	2.0400	.60168	100
VAR00003	2.2900	.72884	100
VAR00004	2.9800	.63532	100
VAR00006	2.1200	.67090	100
VAR00007	2.4300	.67052	100
VAR00008	2.1500	.71598	100
VAR00010	2.5600	.60836	100
VAR00011	2.5800	.78083	100
VAR00012	2.6000	.66667	100
VAR00013	2.3800	.70754	100
VAR00014	2.8100	.64659	100
VAR00015	2.0400	.72363	100
VAR00016	2.3000	.73168	100

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
VAR00002	29.2400	18.770	.403	.757
VAR00003	28.9900	17.545	.517	.744
VAR00004	28.3000	18.556	.416	.755
VAR00006	29.1600	19.085	.290	.767
VAR00007	28.8500	18.634	.371	.759
VAR00008	29.1300	18.155	.421	.754
VAR00010	28.7200	18.547	.442	.753
VAR00011	28.7000	18.596	.301	.768
VAR00012	28.6800	17.876	.517	.745
VAR00013	28.9000	18.919	.295	.767
VAR00014	28.4700	19.181	.289	.767
VAR00015	29.2400	17.901	.459	.750
VAR00016	28.9800	17.919	.449	.751

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
31.2800	21.234	4.60803	13

Reliability

Scale: STUDY BEHAVIOR SCALE

Case Processing Summary

		N	%
Cases	Valid	100	100.0
	Excluded(a)	0	.0
	Total	100	100.0

a Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.936	15

Item Statistics

	Mean	Std. Deviation	N
VAR00001	1.7600	.79290	100
VAR00002	1.8100	.78746	100
VAR00003	1.9100	.76667	100
VAR00004	1.7500	.84537	100
VAR00005	1.7200	.79239	100
VAR00006	1.6400	.75905	100
VAR00007	1.6800	.77694	100
VAR00008	1.9300	.90179	100
VAR00009	1.8100	.82505	100
VAR00010	1.8200	.84543	100
VAR00011	1.8200	.68726	100
VAR00012	1.8200	.74373	100
VAR00013	1.6800	.81501	100
VAR00014	1.7300	.82701	100
VAR00015	1.7700	.91954	100

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
VAR00001	24.8900	69.432	.565	.935
VAR00002	24.8400	69.004	.604	.934
VAR00003	24.7400	67.790	.724	.931
VAR00004	24.9000	66.980	.710	.931
VAR00005	24.9300	69.076	.594	.934
VAR00006	25.0100	67.889	.724	.931
VAR00007	24.9700	67.949	.700	.931
VAR00008	24.7200	66.143	.720	.931
VAR00009	24.8400	66.742	.749	.930
VAR00010	24.8300	66.203	.771	.929
VAR00011	24.8300	71.011	.521	.936
VAR00012	24.8300	70.930	.482	.937
VAR00013	24.9700	66.575	.773	.929
VAR00014	24.9200	66.317	.781	.929
VAR00015	24.8800	65.541	.748	.930

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
26.6500	77.523	8.80470	15

Reliability

Scale: FAMILY SUPPORT SCALE

Case Processing Summary

		N	%
Cases	Valid	100	100.0
	Excluded(a)	0	.0
	Total	100	100.0

a Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.916	15

Item Statistics

	Mean	Std. Deviation	N
VAR00001	1.6400	.77225	100
VAR00002	1.6600	.75505	100
VAR00003	1.7300	.70861	100
VAR00004	1.5700	.71428	100
VAR00005	1.5700	.72829	100
VAR00006	1.4700	.65836	100
VAR00007	1.4900	.65897	100
VAR00008	1.7100	.79512	100
VAR00009	1.6200	.69311	100
VAR00010	1.6000	.68165	100
VAR00011	1.7400	.67600	100
VAR00012	1.7300	.67950	100
VAR00013	1.5500	.74366	100
VAR00014	1.5100	.65897	100
VAR00015	1.6100	.82749	100

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
VAR00001	22.5600	46.815	.568	.912
VAR00002	22.5400	47.200	.544	.913
VAR00003	22.4700	46.938	.615	.911
VAR00004	22.6300	46.155	.695	.908
VAR00005	22.6300	47.488	.538	.913
VAR00006	22.7300	46.926	.671	.909
VAR00007	22.7100	47.440	.610	.911
VAR00008	22.4900	45.404	.689	.908
VAR00009	22.5800	46.630	.666	.909
VAR00010	22.6000	46.424	.702	.908
VAR00011	22.4600	48.029	.526	.914
VAR00012	22.4700	48.898	.427	.917
VAR00013	22.6500	45.442	.740	.906
VAR00014	22.6900	46.984	.664	.909
VAR00015	22.5900	45.355	.662	.909

Scale Statistics

Mean	Variance	Std. Deviation	N of Items
24.2000	53.414	7.30850	15

Reliability

Scale: ENGAGE TIME SCALE

Case Processing Summary

		N	%
Cases	Valid	100	100.0
	Excluded(a)	0	.0
	Total	100	100.0

a Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.803	13

Item Statistics

	Mean	Std. Deviation	N
VAR00002	2.1500	.57516	100
VAR00003	2.3600	.64385	100
VAR00004	2.0900	.58767	100
VAR00005	2.3500	.75712	100
VAR00006	2.4600	.59323	100
VAR00008	2.5100	.64346	100
VAR00009	2.4000	.68165	100
VAR00010	2.8300	.66750	100
VAR00011	2.0600	.70811	100
VAR00012	2.2200	.67540	100
VAR00013	2.3900	.76403	100
VAR00014	2.6300	.69129	100
VAR00015	2.3200	.80252	100

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
VAR00002	28.6200	21.268	.304	.800
VAR00003	28.4100	19.174	.642	.773
VAR00004	28.6800	20.280	.488	.787
VAR00005	28.4200	20.610	.295	.803
VAR00006	28.3100	21.125	.318	.799
VAR00008	28.2600	19.467	.586	.778
VAR00009	28.3700	20.316	.395	.794
VAR00010	27.9400	20.380	.395	.793
VAR00011	28.7100	19.743	.471	.787
VAR00012	28.5500	20.008	.454	.788
VAR00013	28.3800	19.511	.461	.788
VAR00014	28.1400	19.617	.509	.784
VAR00015	28.4500	19.806	.386	.796

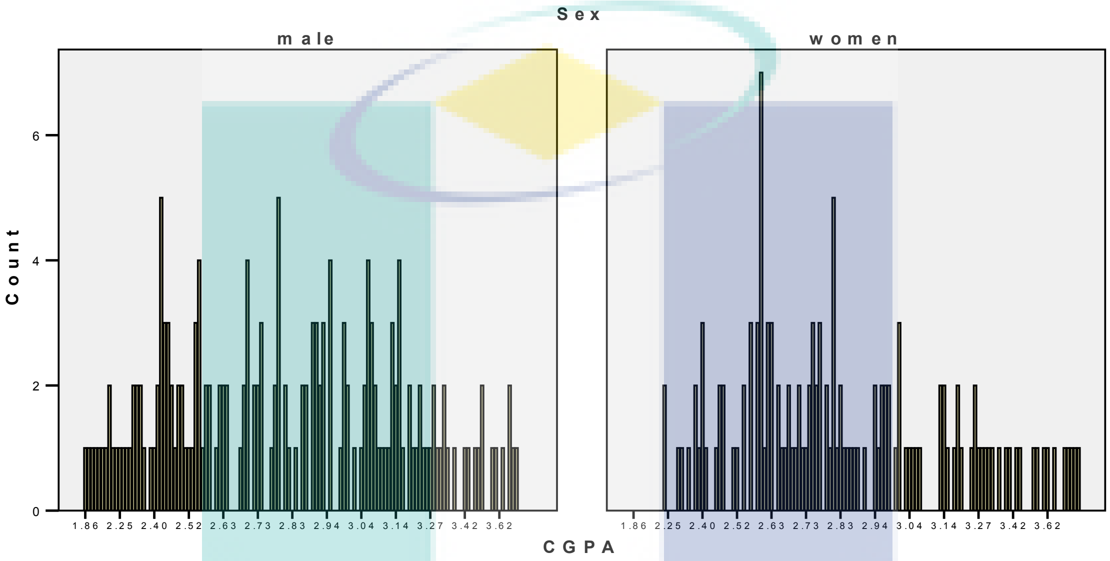
Scale Statistics

Mean	Variance	Std. Deviation	N of Items
30.7700	23.209	4.81759	13

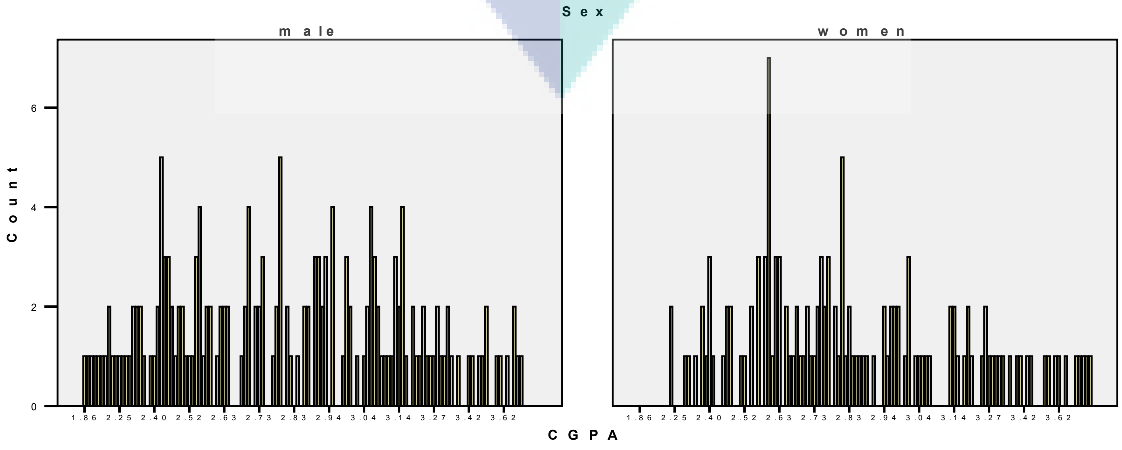
APPENDIX C

DATA DISTRIBUTION GRAPH

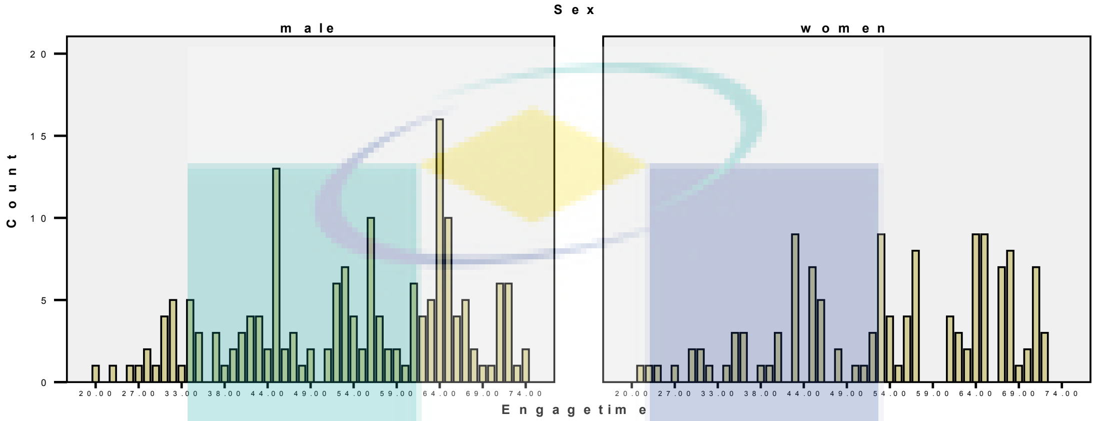
INTEREST ON GENDER



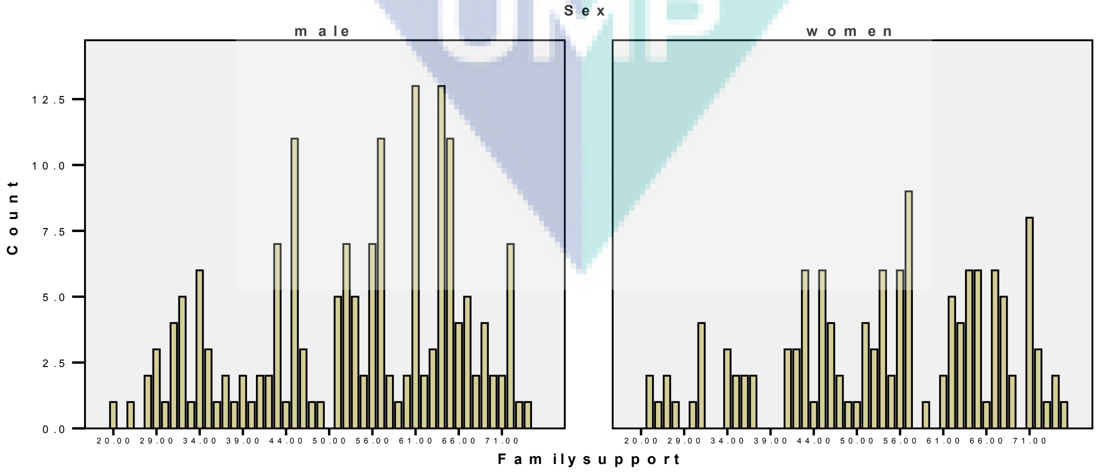
CGPA ON GENDER



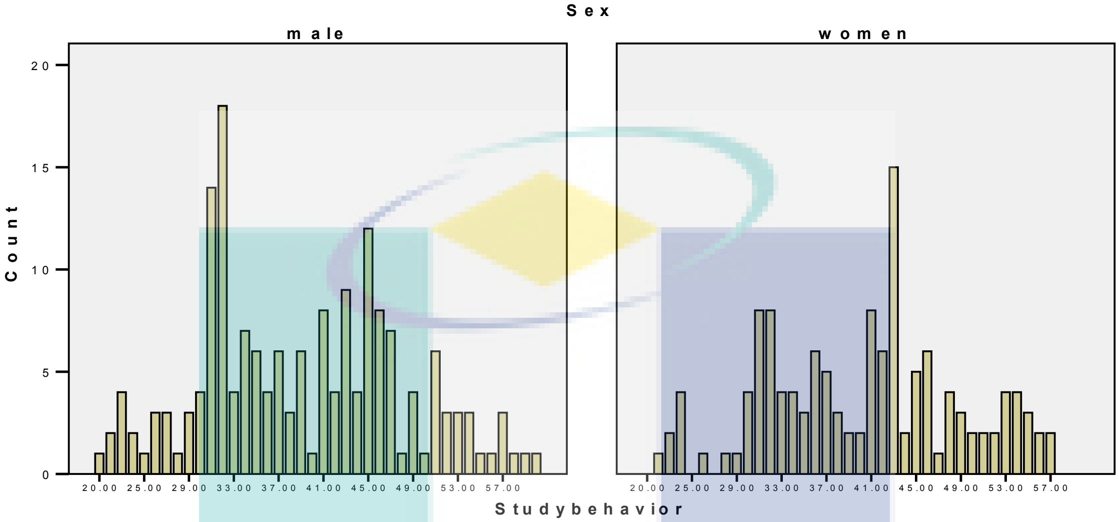
ENGAGE TIME ON GENDER



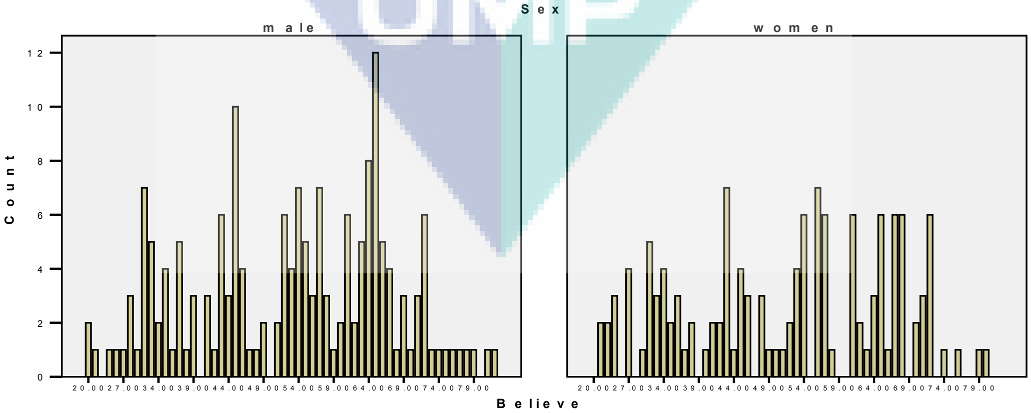
FAMILY SUPPORT ON GENDER

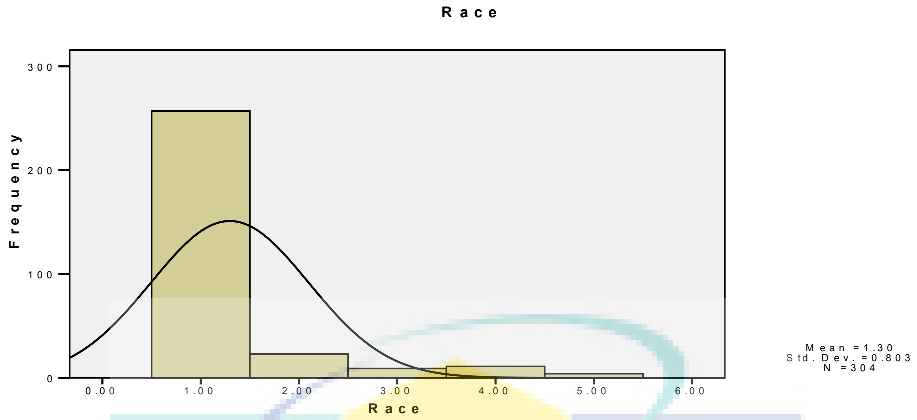


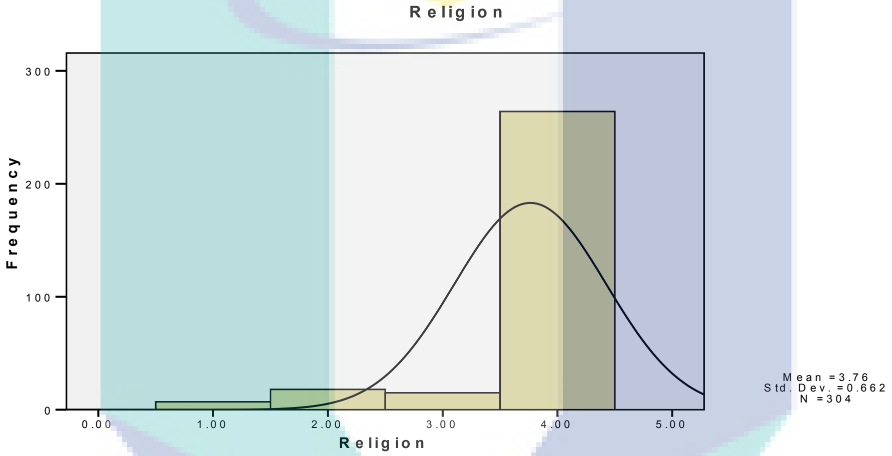
STUDY BEHAVIOUR ON GENDER

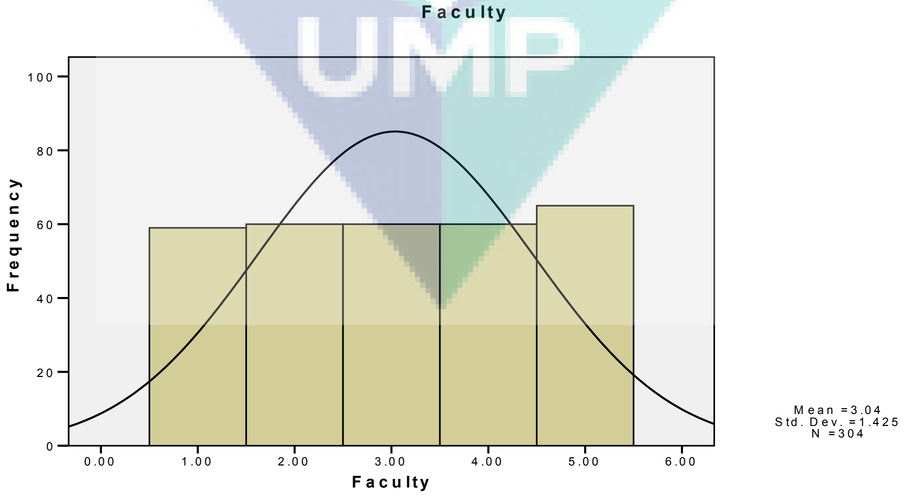


BELIEVE ON GENDER

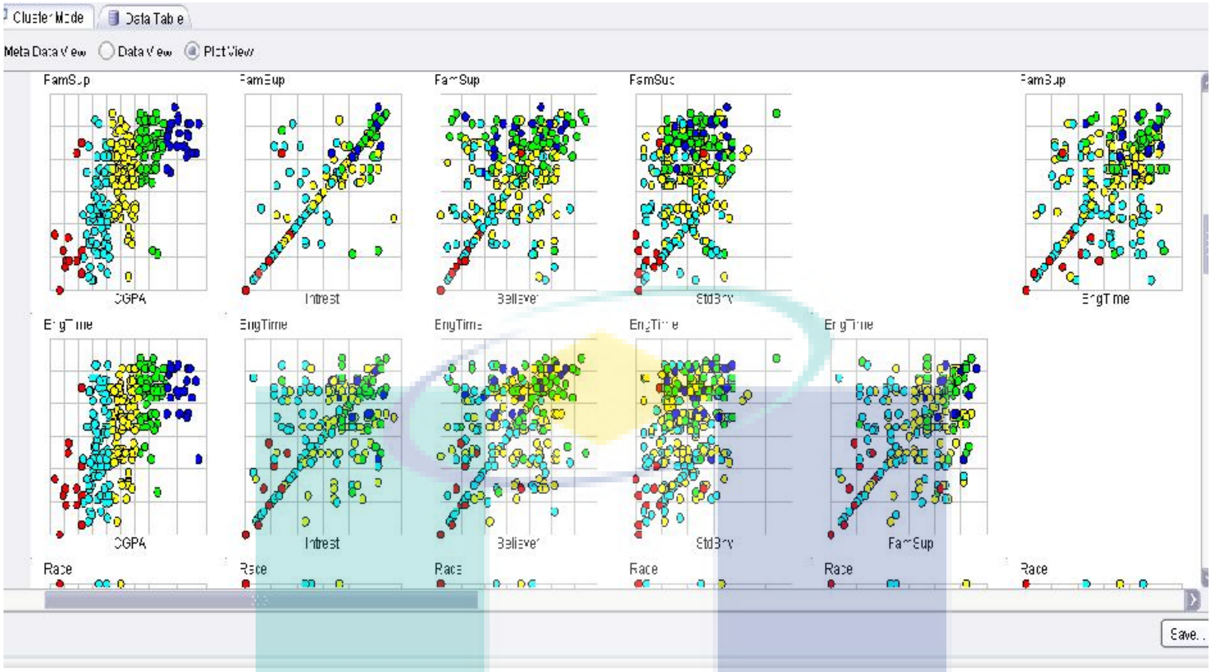



Race
Frequency
300
200
100
0
0.00
1.00
2.00
3.00
4.00
5.00
6.00
Race
Mean =1.30
Std. Dev. =0.803
N =304

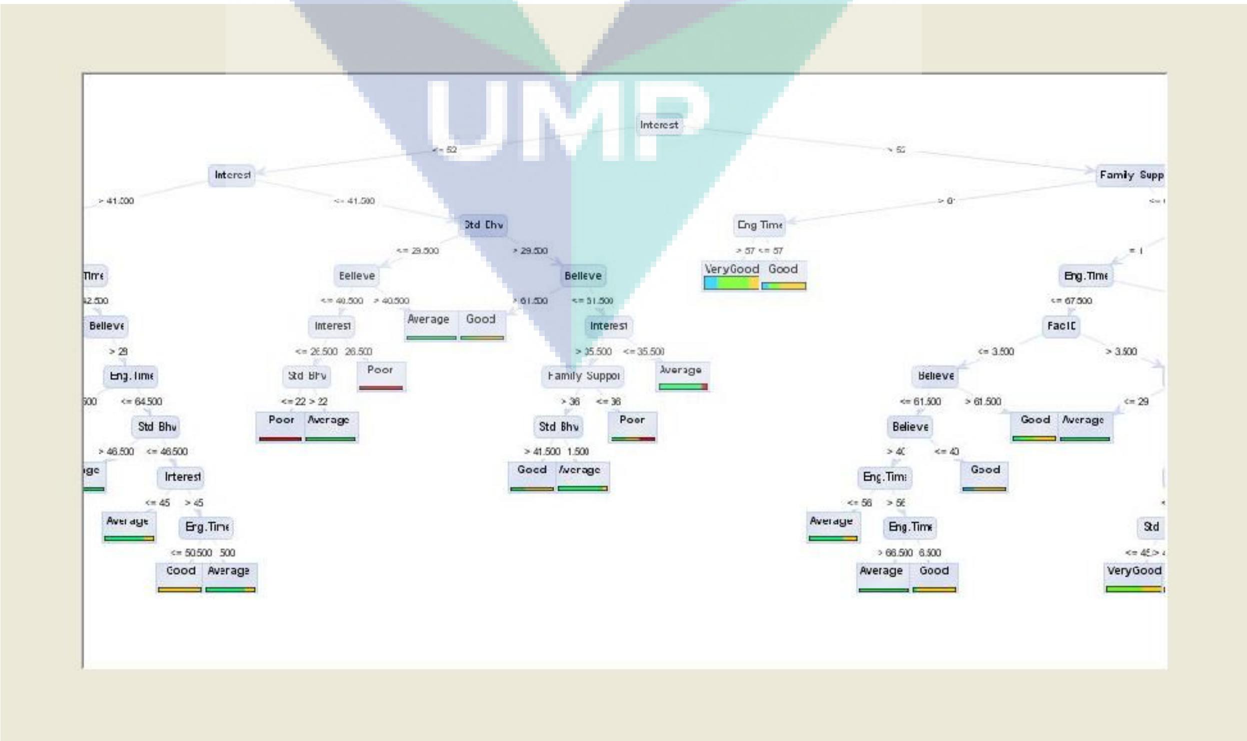

Religion
Frequency
300
200
100
0
0.00
1.00
2.00
3.00
4.00
5.00
Religion
Mean =3.76
Std. Dev. =0.662
N =304


Faculty
Frequency
100
80
60
40
20
0
0.00
1.00
2.00
3.00
4.00
5.00
6.00
Faculty
Mean =3.04
Std. Dev. =1.425
N =304

SCETTER MATRIX PLOT



Decision Tree



APPENDIX D
LIST OF PUBLICATIONS

1. Sajadin Sembiring, A.Embong, Mohd. Azwan M, 2008. *Academic Performance Monitoring System Using Data Mining Techniques.* Poster Publication in Research Plan Category on Postgraduate Research Poster Competition, Universiti Malaysia Pahang.

2. Sajadin Sembiring, Abdullah Embong, Mohd. Azwan Mohammad, Muhammad Furqan, "*Improving Student Academic Performance by An Application of Data Mining Techniques*", Proceeding The 5th IMT-GT International Conference on Mathematics, Statistics, and Their Application (ICMSA 2009), ISBN 978-602-95343-0-6, page 390-394.

3. Muhammad Furqan, Abdullah Embong, Suryanti Awang, S.W. Purnami, Sajadin Sembiring. "*Smooth Support vector Machine for Face Recognition using Principal Component Analysis*", Proceeding International Conference on Green Technology and Engineering (ICGTE 2009) vol.2, ISSN 1978-5933, page 293-298.

4. Muhammad Furqan, Abdullah Embong, Suryanti Awang, S.W. Purnami, Sajadin Sembiring. "*Face Recognition using Smooth Support Vector Machine Based on Eigenfaces*", Proceeding The 5th IMT-GT International Conference on Mathematics, Statistics, and Their Application (ICMSA 2009), ISBN 978-602-95343-0-6, page 708-714.

5. Sajadin Sembiring, M.Zarlis, Dedy H, Ramliana S, Elvi Wani " *Predicting Student Academic Performance by An Application of Data mining Techniques*" 2011 International Conference on Management and Artificial Intelligence IPEDR vol.6 (2011) © (2011) IACSIT Press, Bali, Indonesia

APPENDIX E
GRANT AND AWARD

1. 2008. Grant Research Skim (GRS) 070162, Research Management Center, Universiti Malaysia Pahang. (Main Researcher)

2. 2008. Research Grant Project Vot. RDU 07/03/62, 2008, Prof. Dr. Abdullah Embong, "*Applying Data Mining Technique to Classify Breast Cancer by Smooth Support Vector Machine.* Universiti Malaysia Pahang. (Research Assistant)

3. 2009. Second Prize Award in Postgraduate Research Poster Competition, Universiti Malaysia Pahang.