PREDICTION OF STUDENT ACADEMIC PERFORMANCE
BY AN APPLICATION OF DATA MINING TECHNIQUES

Abstract—One of the significant facts in higher learning institution is the explosive growth educational data. These data are increasing rapidly without any benefit to the management. The main objective of any higher educational institution is to improve the quality of managerial decisions and to impart quality education. Good prediction of student’s success in higher learning institution is one way to reach the highest level of quality in higher education system. There are many prediction model available with difference approach in student performance was reported by researcher, but 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 aims of this study was to apply the kernel method as data mining techniques to analyze the relationships between student’s behavioral and their success and to develop the model of student performance predictors. This is done by using Smooth Support Vector Machine (SSVM) classification and kernel k-means clustering techniques. The results of this study support the model of student academic performance predictors by employing psychometric factors as variables predictors.

Keywords: DM, SSVM, kernel-k-means, Student Performance.

I. INTRODUCTION

The topic of explanation and prediction of academic performance is widely researched. The prediction of student success in tertiary institution is still the most topical debates in higher learning center. In the older studies, the model of Tinto [18] is the predominant theoretical framework for considering factors in academic success. The Tinto’s model considers the process of student attrition as a sociopsychological interplay between the characteristics of the student entering university and the experience at the institute. The use of data-mining techniques in this field is relatively new. There are many data mining techniques was used in this field, such as neural networks, decision tree, Bayesian network, k-means clustering, and so on[5], but poor application of kernel methods in data mining on this area.

There are increasing research interests in education field using data mining. Application of Data mining techniques concerns to develop the methods that discover knowledge from data and used to uncover hidden or unknown information that is not apparent, but potentially useful [1].

The discovered knowledge can be used to better understand students’ behavior, to assist instructors, to improve teaching, to evaluate and improve e-learning system, to improve student academic performance; to improve curriculums and many others benefits [5].

This study investigates the educational domain of data mining using a case study from data that come from student’s behavior. It showed what kind of data could be collected, how could we preprocess the data, how to apply kernel method in data mining on the data, and finally how can we benefited from the discovered knowledge.

In this study, university students were predicted his/her final grade by using SSVM classification and grouped the students according to their similar characteristics, forming clusters. The clustering process was carried out using kernel k-means algorithm technique.

The rest of the paper is organized as follows: Section 2 summaries related work in an application of data mining techniques in educational environments. Section 3 describes the data mining task on educational system and brief review the methods of SSVM classification and kernel k-means clustering. Section 4 gives a general description of data were used in our case study and describes the preprocessing step to analyze the data. Section 5 reports our experiments results. In section 6 we conclude this paper with a summary and describe an outlook for future work. Finally, we put the references on section 7.

II. RELATED WORK

Data mining in higher education is a new emerging field, called Educational Data Mining [4],[5]. There are many works in this area, because of its potentials to educational institutes. Romero.C and Ventura.S, have a survey on educational data mining between 1995 to 2005[5]. They conclude that educational Data mining is promising area of research and it has a specific requirements not presented in other domains. Merceron.A and Yacep.K, gave a case study that used data mining to identify behavior of failing students to warn students at risk before final exam [4]. Also, Data mining in educational area used by Beikzadeh,M and Delavari N, to identify and then enhance educational process in higher educational system., which can improve their decision making process [7]. Finally, Waiyamai, K. used data mining to assist in development of new curricula and to help engineering students to select an appropriate major [3].
III. DATA MINING TASK IN EDUCATIONAL SYSTEM

Today, data collecting and storing are no longer expensive and difficult task. As a result, datasets are growing explosively. To extract the knowledge and information from these massive datasets has attracted a great deal of scientific attention and has become an important research area [2],[6]. Data mining is a flourishing research field and has become a synonym for the process of extracting hidden and useful information from datasets,[10]

Data mining used advanced techniques to discover pattern from data. The data mining task are the kinds of patterns that can be mined. In this study we focused to use both Smooth Support Vector Machine (SSVM) classification and kernel k-means clustering algorithms. In the following sections describes the brief review and the results of applying data mining techniques to the data of our case study for each of algorithms.

1.1. Smooth Support Vector Machine (SSVM) for Classification

Classification is data mining task that predicts group memberships for data instances [9]. In educational area application of the classification method, given works of a student, one may predicate his/her final grade.

The SSVM is further development of Support Vector Machine (SVM) [9],[11],[17]. The SSVM generated and solve an unconstrained smooth reformulation of the SVM for pattern classification using completely arbitrary kernel [11]. SSVM is solved by a very fast Newton-Armijo algorithm and has been extended to non linear separation surfaces by using non linear kernel techniques. The numerical results show that SSVM is faster than other methods and has better generalization ability [9].

The fast Newton-Armijo Algorithm for SSVM [9],[11],[17] is described in the C/C++ style pseudo-code. Start with any \(( w_0,\gamma_0) \in R^{n+1}\).

Having \(( w_i^r,\gamma_i)\), stop if the gradient objective function

\[
\min_{w,\gamma} \frac{1}{2}||e-D(Aw-e\gamma)||^2 + \frac{1}{2}(w^r+w+\gamma)^2
\]

That is \( \nabla \Phi_a(w^r,\gamma^r) = 0 \). Else compute \(( w_i^{r+1},\gamma_i^{r+1})\) as follows:

i). Newton Direction : Determine direction \( d_i \in R^{n+1} \) by setting equal to zero the linearization of \( \nabla \Phi_a(w^r,\gamma^r) \) around \((w_i^r,\gamma_i^r)\) which gives \( n+1 \) linear equations in \( n+1 \) variables:

\[
\nabla^2 \Phi_a(w^r,\gamma^r)d_i = -\nabla \Phi_a(w^r,\gamma^r)
\]

ii). Armijo Stepsize : Choose a step size \( \lambda_i \in R \) such that:

\[
(w_i^{r+1},\gamma_i^{r+1}) = (w_i^r,\gamma_i^r) + \lambda_i d_i
\]

Where \( \lambda_i = \max\left\{ 1, \frac{1}{2}, \frac{1}{4}, \ldots \right\} \) such that:

\[
\Phi_a(w^r,\gamma^r) - \Phi_a\left((w_i^r,\gamma_i^r) + \lambda_i d_i\right) \geq -\delta \lambda_i \nabla \Phi_a(w^r,\gamma^r)d_i
\]

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

Basically, the theory of SVM classification for two class or binary classification [16].

1.2. An Effective Kernel K-Means for Clustering

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different the objects in another group [8]. In educational area, clustering will be used to grouping students according to their behavior and performance. In this study we used Kernel K-means algorithm to cluster the given data.

A drawback to original K-means is that it cannot separate cluster that are non-linearly separable input space. Kernel K-Means is one approaches have emerged for tackling such a problem. Kernel K-means, where, before clustering, points mapped to a higher dimensional feature space using a non-linear function, and then Kernel K-means partitions the points by linear separator in new space[13][14].

Kernel K-means has been extended to efficient and effective large scale clustering [8], 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 corpora. The new clustering scheme is a large scale clustering for Kernel K-means algorithm [8].

1. Assign \( \delta(x_i, C_k) \) \((1 \leq i \leq N, 1 \leq k \leq k) \) with initial value, forming \( K \) initial cluster \( C_1, C_2, \ldots, C_K \).
2. For each cluster \( C_k \), compute \( |C_k| \) and \( g(C_k) \).

\[
|C_k| = \sum_{i=1}^{N} \delta(x_k, C_k)
\]

\[
g(C_k) = \sum_{i=1}^{N} \sum_{j=1}^{N} \delta(x_k, C_k) \delta(x_i, C_k) H(x_i, x_j)
\]

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

\[
f(x_i, C_k) = \frac{2}{|C_k|} \sum_{j=1}^{N} \delta(x_k, C_k) H(x_i, x_j)
\]

\[
\delta(x_k, C_k) = \begin{cases} 
1, & f(x_i, C_k) + g(C_k) < f(x_i, C_k) + g(C_k) \\
0, & \text{otherwise}
\end{cases}
\]

4. Repeat step 2 and 3 until converge.
5. For each cluster \( C_k \), select the sample that is closest to the centre as the representative of \( C_k \);

\[
m_k = \arg \min D(\Phi(x_i), x_k)
\]

IV. DATA COLLECTION AND PREPARATION

In our case study we collected the student data from database management system course held at the Universiti Malaysia Pahang (UMP) in third semester of 2007/2008 and we used questionnaire to collect the real data that describing the relationships between behavioral of students (psychometric factors) and their final academic performance. The variable was used in questionnaire are Interest, Study...
Behavior, Engage Time, Believe, and Family Support. The number of students was 1000 with three different majors in faculty of computer system and software engineering UMP. The sources of collected data were: personal records, academic record of students and course records.

To get better input data for data mining technique, we did some preprocessing for the data collected. The data was maintained in different tables was joined in a single table. After we integrated the data into one file, to increase interpretation and comprehensibility, we discretized the attributes to categorical ones. Examples, we grouped all grades into five groups: excellent, very good, good, average, and poor. In this step the fields used in the study were determined and transformed if necessary. By using normal distribution method, we categorized the value of each item in questionnaire with High, Medium, and Low.

V. EXPERIMENT

The significances of correlation variables predictors were tested using multi variant analysis method and we have four of five variables proposed had significant correlation they are: Interest, Study Behavior, Engage Time and Family Support. The fourth variables gave 52.6% contribution in prediction of student academic performance.

![Figure 1. Framework of Student Performance Predictors](image)

![Figure 2. Data Preparation](image)

![Figure 3. Picture of student’s clusters](image)

### Table I. The Correlation of Four Variables Predictors

<table>
<thead>
<tr>
<th>Mode</th>
<th>R</th>
<th>R² Square</th>
<th>df</th>
<th>Sig F Change</th>
<th>Change Stat</th>
<th>F Change</th>
<th>Δ²</th>
<th>Sig F Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8460</td>
<td>0.7155</td>
<td>1</td>
<td>0.0001</td>
<td>152.66</td>
<td>1</td>
<td>0.0001</td>
<td>152.66</td>
</tr>
<tr>
<td>2</td>
<td>0.7140</td>
<td>0.51</td>
<td>1</td>
<td>0.0001</td>
<td>15</td>
<td>1</td>
<td>0.0001</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>0.7200</td>
<td>0.51</td>
<td>1</td>
<td>0.0001</td>
<td>10</td>
<td>1</td>
<td>0.0001</td>
<td>10</td>
</tr>
</tbody>
</table>

### A. Clustering

After the data preparation, the data selection and transformation process was performed. The prepared data was then put through the data mining process. The Kernel K-Means algorithm was used in this step. The number of clusters was determined as an external parameter. Different cluster numbers were tried, and successful partitioning was achieved with 5 clusters. In our case study, the cluster graph as a picture of students group according to their performance on figure 3 gives. For graphs, the Rapidminer software was used. The graphs are given in figure 4 is deviation plot five clusters of students. Using these results we can divide students into five groups and guide them according to their behavior.

### B. Classification

In our case study, by the training effort outside, we used J48 decision tree to represent logical rules of student final grade.
The represented tree is large; we generated some of the strong rule in tree as we can see on table 2:

Table II. Prediction Rule Model

<table>
<thead>
<tr>
<th>Interest</th>
<th>Study Behavior</th>
<th>Engage Time</th>
<th>Believe</th>
<th>Family Support</th>
<th>Performance Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>Excellent</td>
</tr>
<tr>
<td>H</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>H</td>
<td>Very Good</td>
</tr>
<tr>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>Good</td>
</tr>
<tr>
<td>L</td>
<td>M</td>
<td>L</td>
<td>M</td>
<td>L</td>
<td>Average</td>
</tr>
<tr>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>L</td>
<td>Poor</td>
</tr>
</tbody>
</table>

IF High, M = Medium, L = Low

In this research, the data source used was taken from numerical data set that was transformed into txt format. In this data sets, there are 300 students of samples and every samples is expressed by ten characteristics parameters. We used five performance predictors that proposed in this study and five characteristics demographic data of student. So the remained 300 samples data were used for analysis. Based on the logical prediction rule model as show on Table 2, we implemented the algorithm SSVM Classification’s software which run on Linux operating system. Experiment was conducted on two data sets; it’s randomly partitioned into training and testing data sets. The data sets for training was 90% of all data sets and 10% of all data sets used for testing.

Basically, SVM is binary classification, so that in this experiment we done the testing to each grade sparetely, which means for every grade on a label 1 as the true value and -1 for others. The performance of SSVM depend on the combination of several parameters. They are capacity parameters \( \nu \), the Kernel \( K \) and its corresponding parameters. In this study we used RBF kernel function, because of its good performance and a few numbers of parameters (only two parameters \( \nu \) and \( \gamma \)). The experiment was used \( \nu = 2 \) and \( \gamma = 0.002 \).

To guarantee that the present results are valid and can be generalized for making prediction regarding new data, the data sets randomly partitioned into training and testing datasets via 10-fold CV. In ten fold CV, the data sets 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 others 9 subsets put together to form a training sets. Training accuracy or testing accuracy are average or training or testing accuracy in 10 trials. The summary of experiment result we can see on table 3.

Table III. Summary of Experiment Result

<table>
<thead>
<tr>
<th>Performance Prediction</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Best Accuracy (%)</td>
<td>Average Accuracy (%)</td>
</tr>
<tr>
<td>Excellent</td>
<td>100.00</td>
<td>99.67</td>
</tr>
<tr>
<td>Very Good</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Good</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Average</td>
<td>100.00</td>
<td>99.70</td>
</tr>
<tr>
<td>Poor</td>
<td>99.70</td>
<td>96.67</td>
</tr>
</tbody>
</table>

As we saw in the Table3, that the average testing accuracy for the lowest 61% for prediction “good” performance and the highest 93.7% for the prediction “poor” performance.

Based on the results obtained they are sufficient to prove that the rule model of prediction student performance by using predictor’s of student performance proposed acceptable and good enough to serve as predictor of student performance.

VI. CONCLUSIONS AND FUTURE WORK

The result of this study indicates that Data Mining Techniques (DMT) capabilities provided effective improving tools for student performance. It showed how useful data mining can be in higher education in particularly to predict the final performance of student. We collected the data from student by using questionnaire to find the relationships between behavioral (psychometric factors) of student and their academic performance. We applied data mining techniques to discover knowledge. Particularly, we obtained the prediction rule model using decision tree. We implemented the rules into SSVM algorithm to predicate the students’ final grade. Also we clustered the student into group using kernel k-means clustering. This study expressed the strong correlation between mental condition of student and their final academic performance.

For future work, application of data mining techniques in educational field can be used to develop performance monitoring and evaluation tools system.

DMT has a potential in performance monitoring of High school and other levels education offering historical perspectives of students’ performances. The results may both complement and supplement tertiary education performance monitoring and assessment implementations.
REFERENCES


