2011 International Conference on Management and Artificial Intelligence
IPEDR vol.6 (2011) © (2011) IACSIT Press, Bali, Indonesia

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

[1]Sajadin Sembiring,
[1]Faculty of Computer System & Software Engineering,
Universiti Malaysia Pahang, Malaysia
[1] Teknik Informatika STT Harapan Medan, Indonesia
[1]sajadinbiring@gmail.com,

[2]M. Zarlis, [3]Dedy Hartama, [4]Ramliana S, [5]Elvi Wani
[2,3,4,5]Program Magister (S2) Teknik Informatika
USU, Medan, Indonesia
[2]m.zarlis@usu.ac.id,[[3]dedyhartama,[5]wani_elvi]
@yahoo.com

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 students'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 reported a 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 predicated 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 dataset 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,γ^i), stop if the gradient objective function

$$\min_{w,\gamma} \quad \frac{v}{2}\left\|\left(e-D(Aw-e\gamma)\right)_+\right\|_2^2+\frac{1}{2}(w'w+\gamma^2) \quad \text{is zero}$$

That is $\nabla\Phi_\alpha\left(w^i,\gamma^i\right)=0$. *Else compute* $\left(w^{i+1},\gamma^{i+1}\right)$ *as follows:*

*i). **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 variables :*

$$\nabla^2\Phi_\alpha\left(w^i,\gamma^i\right)d^i=-\nabla\Phi_\alpha\left(w^i,\gamma^i\right)'$$

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

$$\left(w^{i+1},\gamma^{i+1}\right)=\left(w^i,\gamma^i\right)+\lambda_i d^i$$

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

$$\Phi_\alpha\left(w^i,\gamma^i\right)-\Phi_\alpha\left(\left(w^i,\gamma^i\right)+\lambda_i d^i\right)\geq-\delta\lambda_i\nabla\Phi_\alpha\left(w^i,\gamma^i\right)d^i$$

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 $g(C_k)$.*

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

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

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

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

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, $m_k = \text{Arg min } D(\Phi(x_i), z_k)$.*

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 major in faculty of computer system and software engineering UMP. The sources of collected data were: personal records, academic record of students and course records.

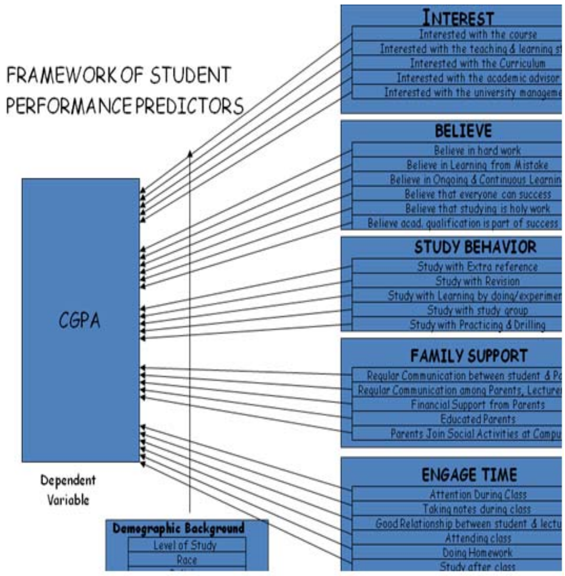
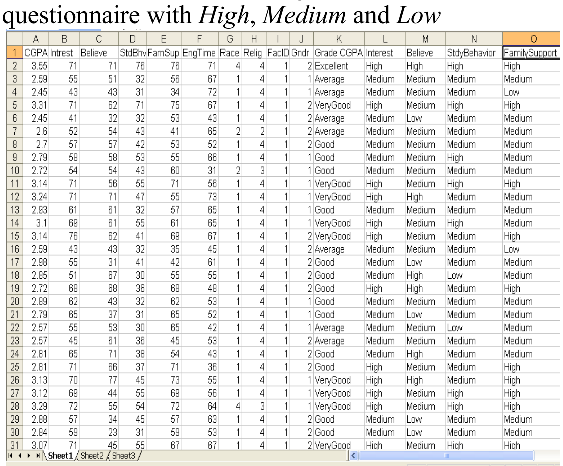


Figure 1. Framework of Student Performance Predictors

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 files, to increase interpretation and comprehensibility, we discretized the attributes to categorical ones. For examples, we grouped all grades into five groups' e*xcellent, 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*



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	CGPA	Intrest	Believe	StdBhv	FamSup	EngTime	Race	Relig	FacID	Gndr	Grade CGPA	Interest	Believe	StdyBehavior	FamilySupport
2	3.55	71	71	76	76	71	4	4	1	2	Excellent	High	High	High	High
3	2.59	55	51	32	56	67	1	4	1	1	Average	Medium	Medium	Medium	Medium
4	2.45	43	43	31	34	72	1	4	1	1	Average	Medium	Medium	Medium	Low
5	3.31	71	62	71	75	67	1	4	1	2	VeryGood	High	Medium	High	High
6	2.45	41	32	32	53	43	1	4	1	2	Average	Medium	Low	Medium	Medium
7	2.6	52	54	43	41	65	2	2	1	2	Average	Medium	Medium	Medium	Medium
8	2.7	57	57	42	53	52	1	4	1	2	Good	Medium	Medium	Medium	Medium
9	2.79	58	58	53	55	66	1	4	1	1	Good	Medium	Medium	High	Medium
10	2.72	54	54	43	60	31	2	3	1	1	Good	Medium	Medium	Medium	Medium
11	3.14	71	56	55	71	56	1	4	1	1	VeryGood	High	Medium	High	High
12	3.24	71	71	47	55	73	1	4	1	1	VeryGood	High	High	Medium	Medium
13	2.93	61	61	32	57	65	1	4	1	1	Good	Medium	Medium	Medium	Medium
14	3.1	69	61	55	61	65	1	4	1	1	VeryGood	High	Medium	High	Medium
15	3.14	76	62	41	69	67	1	4	1	2	VeryGood	High	Medium	Medium	High
16	2.59	43	43	32	35	45	1	4	1	2	Average	Medium	Medium	Medium	Low
17	2.98	55	31	41	42	61	1	4	1	2	Good	Medium	Low	Medium	Medium
18	2.85	51	67	30	55	55	1	4	1	2	Good	Medium	High	Low	Medium
19	2.72	68	68	36	68	48	1	4	1	2	Good	High	High	Medium	High
20	2.89	62	43	32	62	53	1	4	1	1	Good	Medium	Medium	Medium	Medium
21	2.79	65	37	31	65	52	1	4	1	1	Good	Medium	Low	Medium	Medium
22	2.57	55	53	30	65	42	1	4	1	1	Average	Medium	Medium	Low	Medium
23	2.57	45	61	36	45	53	1	4	1	2	Average	Medium	Medium	Medium	Medium
24	2.81	65	71	38	54	43	1	4	1	2	Good	Medium	High	Medium	Medium
25	2.81	71	66	37	71	36	1	4	1	2	Good	High	Medium	Medium	High
26	3.13	70	77	45	73	55	1	4	1	1	VeryGood	High	High	Medium	High
27	3.12	69	44	55	69	56	1	4	1	1	VeryGood	High	Medium	High	High
28	3.29	72	55	54	72	64	4	3	1	1	VeryGood	High	Medium	High	High
29	2.88	57	34	45	57	63	1	4	1	2	Good	Medium	Low	Medium	Medium
30	2.84	59	23	31	59	53	1	4	1	2	Good	Medium	Low	Medium	Medium
31	3.07	71	45	55	67	67	1	4	1	2	VeryGood	High	Medium	High	High

Sheet1 / Sheet2 / Sheet3

Figure 2. Data Preparation

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.

TABLE I. THE CORRELATION OF FOUR VARIABLES 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	
1	.694(a)	.482	.480	.27830	.482	280.553	1	302	.000	
2	.714(b)	.510	.506	.27113	.028	17.198	1	301	.000	
3	.720(c)	.519	.514	.26906	.009	5.637	1	300	.018	
4	.725(d)	.526	.519	.26751	.007	4.487	1	299	.035	1.259

a Predictors: (Constant), Interest
b Predictors: (Constant), Interest, Study behavior
c Predictors: (Constant), Interest, Study behavior, Engage time
d Predictors: (Constant), Interest, Study behavior, Engage time, Family support
e Dependent Variable: CGPA

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.

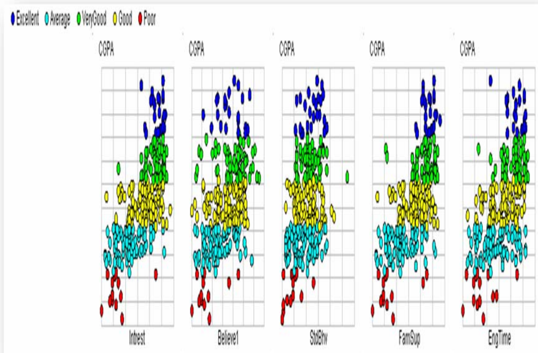


Figure 3. Picture of student's clusters

B. *Classification*

In our case study, by the training effort outside, we used J48 decision tree to represent logical rules of student final grade.

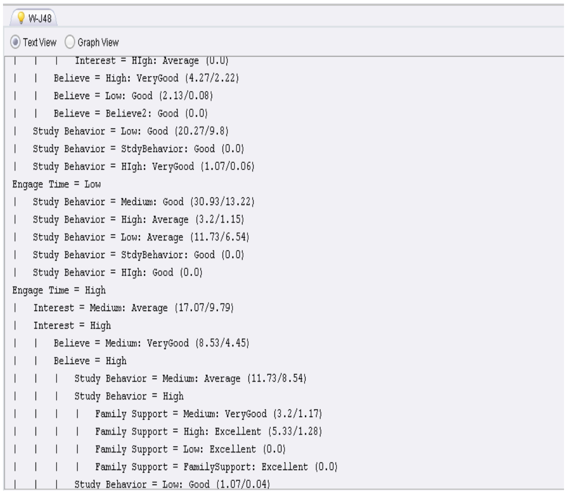


Figure 5. Rule model produced by J48 Decision Tree

The represented tree is large; we generated some of the strong rule in tree as we see on table 2 :

TABLE II. PREDICTION RULE MODEL

	Interest	Study Behavior	Engage Time	Believe	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.	

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 v, 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 v and γ). The experiment was used $v = 2$ 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 acuracy in 10 trials. The summary of experiment result we can see on table 3.

TABLE III. SUMMARY OF EXPERIMENT RESULT

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

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

[1] Kash Barker, Theodore Trafalis, and Teri Reed Rhoads "Learning From Student Data". Proceedings of the 2004 Systems and Information Engineering Design Symposium. Mathew H. Jones, Stephen D. Patek, and Barbara E. Towney eds. 2004. pp79-86

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

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

[4] Merceron, A and Ycef, K.,"Educational Data mining: A case study". In proceedings of the 12th International Conference on Artificial Intelligence in Education AIED 2005, Amsterdam, The Netherlands, IOS Press.2005

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

[6] Luan, J. "Chapter 2: Data Mining and Its Application in Higher Education. Knowledge Management – Building a Competitive Advantage in Higher Education." Serban, A. & Luan, J. (eds.) Jossey-Bass. 2002.

[7] Naeimeh Delavari and Mohammad Reza Beikzadeh and Somnuk Phon-Amnuaisuk, "Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System". ITHET 6th Annual International Conference. Juan Dolio, Dominican Republic. July 7 – 9, 2005,pp F4B-

[8] Rong Zhang and Alexander I. Rudnicky," A large Scale Clustering Scheme for kernel-K-Means"School of Computer Science, Carnegie Mellon University 5000 Forbes Avenue, Pittsburgh, PA 15213, USA.2006

[9] Y.J. Lee. And O.L Mangasarian, " A Smooth Support Vector Machine for classification",Journal of Computational Optimization and Applications.20, 2001, pp.5-22

[10] Ogor Emmanuel. N, " Student Academic Performance: Monitoring and Evaluation Using Data Mining Techniques". Fourth Congress of Electronics, Robotics and Automotive Mechanics. 2007. I EEE Computer Society.

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

[12] Christoper Burges. "A Tutorial on support vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 2(2), 1998

[13] Mark Girolami." Mercer Kernel Based Clustering in Feature Space" I EEE Trans. On Newral Networks.

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

[15] Cristianini N, Taylor,J.S., " An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods" Camberidge Press University, 2000.

[16] Nugroho,A.S, Witarto, A.B, Handoko,D., "Application of Support Vector Machine in Bioinformatics", Proceeding of Indonesian Scientific Meeting in Central japan, Dec,20, 2003, Gifu-Japan.

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

[18] V. Tinto, "Limits of theory and practice in student attrition," Journal of Higher Education no. 53, pp. 687-700, 1982.