International Journal of Multidisciplinary and Current Research *ISSN: 2321-3124*
Research Article *Available at: http://ijmcr.com*

A Multi-class Support Vector Machine Approach for Students Academic Performance Prediction

Mojisola G. Asogbon[1], Oluwarotimi W. Samuel[2,3], Mumini O. Omisore[2,3], and Bolanle A. Ojokoh[1]

[1]Department of Computer Science, Federal University of Technology Akure, Ondo State, P.M.B. 704, Nigeria
[2]Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, CAS, Shenzhen, Guangdong 518055, China
[3]University of Chinese Academy of Sciences, Beijing, 100049, China

Accepted 28 Feb 2016, Available online 03 March 2016, **Vol.4 (March/April 2016 issue)**

Abstract

To date, students' educational data is still one of the most importance resources in institutions of higher learning. One way to achieving qualitative education standard by institutions of higher learning is to properly evaluate and predict the performance of entrant students and suggest appropriate faculty programmes for them based on their educational data. Several attempts have been made to predict the performance of students and placement into appropriate faculty programmes prior to admission without much success. In an attempt to correctly predict students' performance in order to admit them into appropriate faculty programmes, a multi-class support vector machine (MSVM) predictor was built in this study. The performance of the MSVM predictor was examined using educational dataset of students from the University of Lagos, Nigeria. Findings from our experiment show that the MSVM with K-fold (K=7) cross validation adequately predicted the performances of students across all categories.

Keywords: *Educational data, Student performance, Predictive model, Support vector machine*

1. Introduction

Institutions of higher learning are rapidly becoming more competitive as a result of the rapid increase in their numbers globally. In other to remain relevant in the education business, many are focusing more on ways to improve the quality of education offered to their students. This is because qualitative education would most likely lead to excellent records and achievements on the part of students. Also, it will empower students with appropriate skills and knowledge thereby leading to socioeconomic advancement of the society. A critical step towards achieving qualitative education standard by institutions of higher learning is via proper predictive model which could be used to evaluate and predict the performance of entrant students. This model can accurately identify performing and non-performing students and further suggest means of improvement for non-performing students. Prediction of students' success is crucial because the primary objective of teaching and learning is to develop and produce self-sufficient students who can learn to measure up with academic expectations. However, effective prediction of students' performance remains a major challenge in educational systems, which may be due to the large volume of unorganized educational data that contains relevant information that cannot be easily extracted. In time past, educational data have been shown to exhibit hidden knowledge that can be used to predict the performance of students at different levels. Such hidden knowledge could be of great importance during admission of students to higher institutions and in their placement into appropriate faculty programmes. Such knowledge can also provide the management of academic institutes with information for identifying weak students and help improve their performance during the period of their studies [1]. Therefore, to enhance the quality of education through the evaluation of students' performance, institutions of higher learning need to extract substantial amount of hidden knowledge in their educational dataset [2]. This knowledge extraction task can be achieved through a knowledge discovery technique called data mining, as this will assist them in making quality decisions on their educational activities [3].

Prior to the emergence of data mining techniques, conventional methods for predicting students' performance in institutions of higher learning was based on subjective evaluations which are often carried out by the same person responsible for tutoring the students. These subjective gateways had a number of limitations among which is their inability to predict the performance of inferior students in the institutions, thereby leading to woeful results [4]. In addition, they generally do not provide adequate means of analysing and monitoring students' progress and performances [5]. Research efforts in the times past had proven data mining as a veritable technique that can be applied to study available data in educational field in order to extract hidden knowledge [6]. In other words, the application of data mining techniques to education system is concern with the development of methods that can help discover knowledge from a large volume of data and also uncover hidden information that are potentially useful in predicting student's academic performance [7].

In the last few decades, several data mining techniques have been proposed to mitigate the problems associated with conventional methods: Bayesian networks, decision trees, artificial neural networks, K-nearest neighbour, discriminant analysis, and Sequential pattern mining among others. With the application of these techniques, large amount of educational data have been traversed to discover useful knowledge for optimal decision making [8]. Multi-class support vector machine (SVM), a robust and accurate technique for pattern classification and knowledge mining has attracted little research attention in educational domain. This technique has been reported to have sound theoretical foundation and highly reliable. Unlike other algorithms, the application of multi-class support vector machine (SVM) technique to educational data is yet to be fully explored. Hence, in this study, a decision support system based on multi-class SVM technique was built to predict the performance of students in institution of higher learning.

2. Literature review

The application of data mining to educational data involves the extraction of meaningful knowledge from large repositories of data generated from learning activities of students [2]. This kind of application usually provide means of identifying students' preferences towards course choices, their selection of specialization, predicting students' level of understanding, and grades [9]. The following are some notable contributions in the field of educational data mining. A survey in [10] describes educational data mining as a leading way to determine academic performance in institutions of higher learning. This can be backed by an expert system developed in [11] to predict students' success in gaining admission into higher institution in Greece. In the study, prediction was made at three points with different variables considered at each point and the system demonstrates accuracy and sensitivity potencies of 75% and 88% respectively. In [12], the poor quality of graduates from Nigerian Universities was attributed to inadequacies of admission systems. The study employed artificial neural network to predict the performance of candidates considered for admission into University of Ibadan, Nigeria. The model operates with a promising accuracy rate of 74% using 10 variables from demographic and educational backgrounds information of prospective students with no psychometric factors were considered in the study. In [13], the promising behaviour of three layer neural network based on backpropagation algorithm was adopted to predict students' final grade. Experiment was conducted with 1,407 profiles of students at Waubonsee Community College, Chicago, USA. The proposed method achieved an average predictability rate of 68%. In [14], the academic performance of 600 students from different colleges of Awadh University in India was predicted by means of Bayes classification. The study considered variables from background qualification of students and language used in teaching them and it was concluded that performance of new intake students will likely be low. Furthermore, [15] applied kernel mining method to analyse the relationships between students' behaviour and their success. The method was developed using smooth support vector machine to cluster final year students at University Malaysia Pahang, and the obtained results shows a strong correlation between mental conditions and academic performance of students. And it was concluded that psychometric information can be taken as an important factor while predicting students' academic performance. [16] developed a data mining model based on Bayesian classification method to predict students division on the basis of previous year students' database. It was concluded from the study that academic performances of students do not always depend on their own effort alone but others factors have significant influence on students' academic performance. [17], employed simple linear regression analysis to predict the performance of students and experimental results shows a correlation between students' academic performance and factors such as mother's education and students' family income. However, from the above review, none has applied a multi-class support vector machine algorithm based on Gaussian Radial Basis Function (RBF) kernel to predict students' academic performance which is the focus of our study.

3. Materials and methods

3.1 Data collection

Psychometric and physiological data of students were elicited by means of a designed questionnaire and direct link with educational database of students in the department of Computer Science, University of Lagos, Nigeria. The dataset comprises of information of 300 student records from all levels in 2013/2014 academic session. The data sources provided information about students' demographics, current and previous academic standings, departmental structure, and family backgrounds in which 17 significant attributes of students were extracted for experimentation in the study. Table 1 presents the dataset attributes, description, and grading of attributes into various categories while Table 2 shows the description of students dataset outputs grading with their class categories.

Table 1 Attributes of student performance prediction

Group code	Group name	S/N	Attributes	Attributes grading (From Class 1- 6)					
				1	2	3	4	5	6
A	Demography Information	1	Gender	Female	Male				
		2	Age (years)	14-20	21-25	>26			
		3	Entry Type	DE	UTME				
		4	Campus Location	Off-Campus	Shared-Hostel	Own-Hostel			
B	Personal Information	1	Department Management	Poor	Average	Good	Best		
		2	Academic Advisor	Poor	Average	Good	Best		
		3	Curriculum	Poor	Average	Good	Best		
		4	Courses	Poor	Average	Good	Best		
C	Academics Information	1	Secondary School Grade						
		2	Entrance Score Average	<60	60-70	>70			
		3	Grade Point Average	0–1.49	1.50–2.39	2.40-4.49	4.5-5.0		
		1	Family Size	<=2	3-4	>=5			
D	Family Information	2	Annual Income						
		3	Fathers Qualification	N/A	Elementary	Secondary	Bachelor	Masters	PhD
		4	Mother's Qualification	N/A	Elementary	Secondary	Bachelor	Masters	PhD
		5	Father's Occupation	N/A	Personal	Private	Civil	Public	
		6	Mother's Occupation	N\A	Personal	Private	Civil	Public	

Table 2 Classification of student's output performance

S/N	GPA range	Class categories	Code
1	4.50 - 5.00	Distinction	C1
2	3.50 - 4.49	Good	C2
3	2.40 - 3.49	Average	C3
4	1.50 - 2.39	Weak	C4
5	0.00 -1.49	Hapless	C5

3.2 Multi-class support vector machine model for students' performance prediction

This study adopted a multi-class SVM algorithm based on Gaussian RBF kernel proposed by Crammer and Singer [18] to predict students' performance in higher institutions. The algorithm works by solving a single optimization problem through maximizing the margins between all designed classes simultaneously.

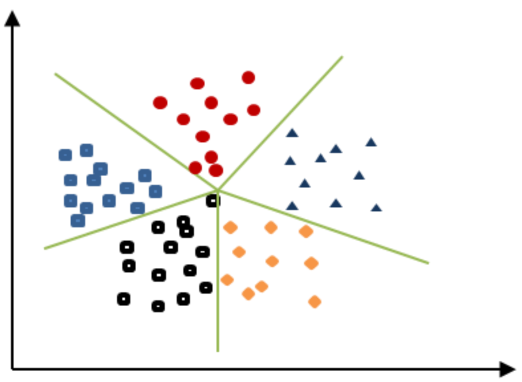

Fig.1 A five class problem based on CS classifier

CS in [19] proposed an approach that requires the solution of a single Quadratic Programming (QP) problem of size $(k-1)\ n$, which uses less slack variables in the constraints of the optimization problem. This approach considers all available training dataset at once and constructs k class categorization rules where k is the number of classes [20]. Given a set of n training dataset $X = \{(x_1, y_1), \dots\dots, (x_n, y_n)\}$ where $x_i \in R^d, i = 1, \dots, l$ are the input feature vectors, $y_i \in \{1, \dots, k\}$ is the class output associated with the training dataset x_i and d is the dimensionality of the input feature vectors. Solving this single optimization problem leads to the construction of k decision functions where the m^{th} decision surface $w_m^T \phi(x)$, determined by its normal vector $w_m \in R^d$, separates the training vectors of the m class from all the others, by minimizing the following primal problem expressed in equation (1):

$$\min_{\{w_m\},\{\xi_i\}} = \frac{1}{2}\sum_{m=1}^{k} w_m^T w_m + C\sum_{i=1}^{l}\xi_i \quad (1)$$
$$subject\ to \quad w_{y_i}^T\phi(x_i) - w_m^T\phi(x_i) \geq e_i^m - \xi_i \quad i = 1,2,\dots,l$$

where $\phi(.)$ denotes a function that maps the input feature vector x_i to an arbitrary-dimensional space $\mathcal{F}$,where the dataset are to be linearly separable, C denotes the parameter that penalizes the training errors, $\xi = [\xi_1, \dots, \xi_l]^T$ is the slack variable vector, w_m is the weight vector associated with class m, $e_i^m = 1 - \delta_{y_i}, m$ and

$$\delta_{y_i}, m \equiv \begin{cases} 1 & if\ y_i = m \\ 0 & if\ y_i \neq m \end{cases}$$

where δ_{y_i}, m denotes the Kronecker delta function.

In equation (1), the constraint $m = y_i$ corresponds to the non-negativity constraint, $\xi_i \geq 0$. Equation (2) represents the decision function of the primal optimization problem.

$$\arg \max_{m=1,..,k} (w_m^T \phi(x_i)) \tag{2}$$

The dual problem of equation (1) involves a vector α having dual variables $\alpha_i^m \forall_m, i$. the w_m get defined via α as shown in equation (3).

$$w_m(\alpha) = \sum_i \alpha_i^m x_i \ \forall_m \tag{3}$$

For:

$C_i^m \leq 0 \ \ if \ \ y_i \neq m$

$C_i^m \leq C \ \ if \ \ y_i = m$

The dual problem is expressed in equation (4)

$$\min_{\alpha} \quad f(\alpha) = \frac{1}{2}\sum_m \|w_m(\alpha)\|^2 + \sum_i \sum_m e_i^m \ \alpha_i^m$$

$$subject\ to \ \ (\alpha_i^m \leq C_i^m \ \forall_m, \ \sum_m \alpha_i^m = 0) \ \forall_i \tag{4}$$

The gradient of f is given in equation (5)

$$g_i^m = \frac{\partial f(\alpha)}{\partial \alpha_i^m} = w_m(\alpha)^T x_i + e_i^m \quad \forall i, m \tag{5}$$

Optimality of α can be checked using the quantity expressed in equation (6)

$$v_i = \max_m g_i^m - \min_{m:\alpha_i^m < C_i^m} g_i^m \ \forall i \tag{6}$$

where dual optimality holds when $v_i = 0 \ \forall i$

For a given i, the values of m that attain the maximum and minimum values in equation (6) are expressed in equation (7).

$$M_i = \arg \max_m g_i^m \ \ and \ \ m_i = \arg \max_{m:\alpha_i^m < C_i^m} g_i^m \ \forall i \tag{7}$$

The Gaussian BRF kernel was also employed alongside the CS algorithm in the training process to implicitly transform the input space (training dataset) into a linear separable feature space where linear classification are applicable, known as kernel trick. A kernel function K effectively computes the dot products in a higher-dimensional space represented by $\mathbb{R}^M$ while remaining in $\mathbb{R}^N$.

For $\overrightarrow{x_i}, \overrightarrow{x_j} \in \mathbb{R}^N, K(\overrightarrow{x_i}, \overrightarrow{x_j}) = \langle \phi(\overrightarrow{x_i}), \phi(\overrightarrow{x_j}) \rangle M$,

where $\langle ., . \rangle M$ is an inner product of $\mathbb{R}^M, M > N$ and $\phi\ (\vec{x})$ transforms $\vec{x}$ to $\mathbb{R}^M (\phi: \mathbb{R}^N \rightarrow \mathbb{R}^M)$

Hence, equation (8) mathematically expressed the Gaussian RBF kernel

$$K(\overrightarrow{x_i}, \overrightarrow{x_j}) = exp^{-\frac{1}{2\alpha^2}\|\overrightarrow{x_i} - \overrightarrow{x_j}\|^2} \tag{8}$$

where $\alpha = \sqrt{5 * \boldsymbol{dim}(x_i)}$ and ***dim*** denotes the dimension of the training dataset.

4. Experimental study and evaluation

The objective of this study is to predict academic performance of students in institutions of higher learning to enhance appropriate steps towards improving quality of graduate students. This is established with an experimental study carried out as presented herein.

4.1 Experimental setup

The proposed model for the prediction of student performance in higher institutions was implemented using MSVMpack, a Multi-class support vector machine tool developed by Lauer and Guermeur [21], and MATLAB (Matric Laboratory). The acquired students' dataset from Computer Science Department, University of Lagos, Nigeria was analysed and pre-processed into the required format using SPSS statistical tool. The entire dataset has a dimension of 17x300 (representing the dataset of 300 students' records with each having 17 attributes) and it was partitioned into training and testing sets. The corresponding target output data has a dimension of 1x300, which was divided into 5 classes (class 1: Distinction, class 2: Good, class 3: Average, class 4: Weak and class 5: Hapless) as presented in Table (2). 83.3% (250 samples) of the dataset was used for training while the remaining 16.7% (50 samples) was used for testing.

The training process begins by selecting the training dataset using a working set selection strategy based on Karush-Kuhn-Tucker (KKT) conditions of optimality proposed in [22].The data points with maximal violation of the KKT conditions were selected and measured using equation (6), followed by the computation of the kernel sub matrix corresponding to the training dataset. The process waits for the CS M-SVM model to be available, before the model was trained, updated and then released to unblock waiting threads. As the training continues, kernel values were cached and the workload of kernel computations decreases. This procedure was repeated until training stopped and optimum accuracy with training error rate was computed. The trained model was tested by dividing the testing dataset into small sizes and the model outputs on each of them were computed. Also, the performance of the model was further evaluated using K-fold cross validation (K=7) evaluation metric.

4.2 Results and discussion

The performance of the multi-class SVM (MSVM) predictor in this study was observed by first training it via the use of a four processor(s) which was allocated 1MB of memory to cache 250 rows (100 %) of the kernel matrix. The training was achieved by arbitrarily selecting the first 250 dataset for training while the remaining 50 dataset was used to test the trained model as shown in Figure 2. From experimental results, an average test accuracy of 66% was achieved while an average training error rate of 9.2% was recorded after 7003 iterations in 21.73s.

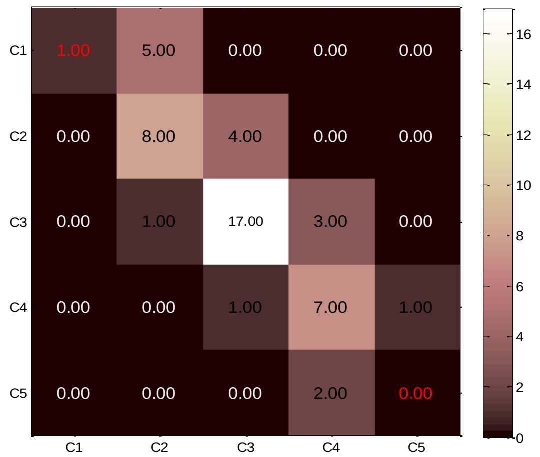

Fig. 2 Prediction result of the Non cross validation (CV) multi-class SVM predictor

The confusion matrix in Figure 2 shows the test results obtained when the trained MSVM model was used to predict the performance of students. The diagonal of the matrix shows that of the 50 students whose data were used for testing, 1 had distinction, 8 had good grade, 17 had average grade, 7 had weak grade, which means 33 out 50 students were correctly predicted. However, the model could not predict the academic performance of students that are hapless (C5) and as a result; there was no prediction result for students with hapless grade. This flaw could be attributed to the fact that there was some biasness (the first 250 dataset was used for training while the last 50 was used for testing) in the data partitioning method thereby leading to overfitting and unreliable model. Afterwards, the correlation coefficient between the observed and predicted classifications was measured using Matthew's Coefficient Correlation (MCC). MCC is a metric for measuring the quality of a machine learning classifier introduced by biochemist Brian W. Matthews [23]. MCC returns a value between −1 and +1, where a coefficient of +1 represents a perfect prediction, 0 no better than random prediction and −1 indicates no correlation between the prediction and observation. The MCC for the performance prediction obtained from the above confusion matrix is shown in Table 3.

Table 3: MSVM classification of student's output performance

SN	Class Type	MCC Values
1	C1	0.39
2	C2	0.48
3	C3	0.63
4	C4	0.59
5	C5	-0.03

From the MCC values presented in Table 3, it could also be observed that the multi-class SVM model was able to predict the performances of students in the first four classes correctly (C1, C2, C3, and C4). However, the model could not predict the performance of students in the fifth class (C5) correctly. In an attempt to improve the reliability of the model in terms of correctly predicting the performances of students in all the classes, a seven fold cross validation method was used with Gaussian RBF kernel and soft margin hyper-parameter (C = 10). Table 4 and Table 5 present the seven fold cross validation results and the overall MCC values across the seven folds respectively.

Table 4: Overall MSVM cross validation results

Number of fold	Training error	Testing accuracy rate	Processing time
1	11.16 %	71.43 %	20.59 sec
2	9.77 %	74.29 %	21.17 sec
3	8.37%	65.71%	28.03 sec
4	6.98 %	68.57%	22.73 sec
5	8.37 %	71.43%	21.43 sec
6	7.44%	74.29%	22.99 sec
7	9.05 %	80.00%	19.99 sec

Table 5: MSVM cross validation MCC classification results across all folds

SN	Class Type	MCC values
1	C1	0.64
2	C2	0.60
3	C3	0.66
4	C4	0.61
5	C5	0.50

From table 4, the average training error and accuracy rate was computed using equations (9) and (10) while the overall cross validation error estimate was computed using equation (11).

$$veager\ Training\ Error = \frac{\sum_{i=1}^{n} TE}{N} \quad (9)$$

$$Averager\ Training\ Error = \frac{61.14}{7} = 8.73\%$$

$$Averager\ Testing\ Accuracy\ Rate = \frac{\sum_{i=1}^{n} TAR}{N} \quad (10)$$

$$Averager\ Testing\ Accurracy\ Rate = \frac{505.72}{7} = 72.25\%$$

$$Cross\ Validation\ Error\ Estimate = \frac{sum\ (labels \cong Y)}{N} \quad (11)$$

$$Cross\ Validation\ Error\ Estimate\ = 27.75\%$$

Where TE denotes the training error, TAR is the testing accuracy rate, labels denote the vector of predicted outputs (concatenation of the labels predicted in each data subset without information about the subset during training) and N is the number of folds.

From the computed values, it can be seen that the cross validation MSVM performed better with a testing accuracy of 72.25% when compared to the non-cross validation MSVM whose average test accuracy is 66%. In addition, the cross validation model achieved a better performance by been able to predict the performance of students in all the 5 classes compared to the non-cross validation model which could not predict for all classes.

Conclusion

Students' education data, a veritable resource in institutions of higher learning from which useful knowledge could be extracted for proper managerial decisions, is constantly on the increase. The hidden knowledge in such a huge dataset could be used to predict the performance of students prior to admission in order to place them into the appropriate faculty programmes. An efficient method for predicting the performance of students based on their educational data in order to decide appropriate faculty programmes for them is still a challenge in institutions of higher learning. Hence, this study proposed the use of a Multi class support vector machine to predict the performance of students in order to place them on various faculty programmes. The performance of the multi-class support vector machine was further enhanced by using a 7-fold cross validation technique. Lastly, obtained results suggest that our proposed approach can help predict the performance of students and provide institution's managements with information that could be used to place students into various appropriate faculty programmes.

References

[1] Bienkowski M., Feng M., & Means B. (2012), Enhancing teaching and learning through educational data mining and learning analytics: An Issue Brief, Technical Report, Office of Educational Technology, United States.

[2] Jai Ruby & David K. (2014) .Predicting The Performance of Students in Higher Education Using Data Mining Classification Algorithms - A Case Study. International Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol. 2, Issue XI, ISSN No. 2321-9653.

[3] Althaf H. B., Ramesh S. K., Kumar, Y.R., Govardhan A. & Mohd. Z. A. (2012). Predicting Student Academic Performance Using Temporal Association Mining. International Journal of Information Science and Education, Vol. 2, No. 1, pp. 21-41.

[4] Alexander, H. (1996), Physiotherapy Student Clinical Education: The Influence of Subjective Judgments on Observational Assessment, Assessment and Evaluation in Higher Education, Vol. 21, No. 4, pp 357-366.

[5] Amirah M. S., Wahidah H. & Nur'aini A. R. (2015), A Review on Predicting Student's Performance using Data Mining Techniques. Procedia Computer Science, Elsevier, Vol. 72, pp 414 – 422.

[6] Kumar A. & Vijayalakshmi M. (2011). Implication of Classification Techniques in Predicting Student's Recitals. International Journal of Data Mining & Knowledge Management Process, Vol 1, No. 5, pp 41-51.

[7] Kash B., Theodore T. & Teri R. R. (2004). 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.. pp 79-86.

[8] Samuel O., Omisore M., & Atajeromavwo E. (2014). Real Time Fuzzy Based System for Human Resource Performance Appraisal. Journal of International Measurement Confederation, Vol. 55, pp 452 - 461.

[9] Mohd M. A. (2013). Role of data mining in education sector. International Journal of Computer Science and Mobile Computing, Vol. 2, Issue. 4, pp 374-383

[10] Romero C. & Ventura S. (2007). Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, Vol. 33, pp 135-146.

[11] Hatzilygeroudis I., Karatrantou A., & Pierrakeas C. (2004), PASS: An Expert System with Certainty Factors for Predicting Student Success in Negoita et al. (Eds.): KES 2004, LNAI 3213, pp 292-298.

[12] Oladokun V., Adebanjo A., & Charles-Owaba O. (2008). Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an

[13] Stamos T. & Andreas V. (2008). Artificial Neural Network for Predicting Student Graduation Outcomes. Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA.

[14] Pandey U. & Pal S. (2011). Data Mining: A Prediction of Performer or Underperformer using Classification. International Journal of Computer Science and Information Technology, Vol. 2. No. 2, pp 686-690.

[15] Sajadin S., Zarlis M., Dedy H., Ramliana S., & Elvi W. (2011). Prediction of Student Academic Performance by an Application of Data Mining Techniques. International Conference on Management and Artificial Intelligence, IACSIT Press, Bali, Indonesia, Vol. 6, pp 110-114.

[16] Shaeela A., Tasleem M., Ahsan R. S., Inayat K. M. (2010). Data mining model for higher education system. European Journal of Scientific Research, Vol.43, No.1, pp.24-29.

[17] Baradwaj B. K. & Pal S. (2011). Mining Educational Data to Analyze Students' Performance. International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 2, No. 6, pp. 63-69.

[18] Crammer K. & Singer Y. (2000). On the learnability and design of output codes for multiclass problems. In COLT.

[19] Crammer K. & Singer Y. (2002). On the learnability and design of output codes for multiclass problems. Machine Learning, Vol. 47(2–3), pp 201–233.

[20] Symeon N., Nikos N. & Ioannis P. (2010). Incremental Training of Multiclass Support Vector Machines. 2010 International Conference on Pattern Recognition, pp 4267-4270

[21] Lauer F. & Guermeur Y. (2011). MSVMpack: a Multi-Class Support Vector Machine Package. Journal of Machine Learning Research, vol.12, pp 2269-2272.

[22] Crammer K. & Singer Y. (2001). On the algorithmic implementation of multiclass kernel-based vector machines. Journal of Machine Learning Research, Vol. 2, pp265-292.

[23] Matthews B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA) - Protein Structure, Vol. 405, No. 2, pp 442–451.