Vol 8. No. 1 Issue 2 – May, 2015
African Journal of Computing & ICT

www.ajocict.net

Predicting Postgraduate Performance Using Resample Preprocess Algorithm and Artificial Neural Network

M.D. Gbolagade
Department of Physical Sciences
Computer Science Program
Al-Hikmah University
Ilorin-Kwara State.

***M.A. Hambali & A.A. Akinyemi**
Computer Science Department
Federal University Wukari
Wukari-Taraba State, Nigeria
E-mail: hambali.moshood.a@ieee.org
Phone: +234-7065868393

ABSTRACT

Presently, there is an increasing interest in data mining and educational systems, making educational data mining as a new growing research community. From the literature, different analysis has been carried out on university data, which includes student's university entrance examination and Ordinary level results but the relationship between these entry results and students' final graduation grades has been in isolation. Therefore, in this paper, classification data mining (DM) technique Weka tools is employed on enrolment data of master students of a department in University of Ilorin, using multilayer perceptron (MLP), radial basis functions (RBF) and Sequential minimal optimization algorithm (SMO) to classify the data, effect of data preprocess algorithm is also revealed, Attribute selection was also used to evaluate the factors that are very important to predict possible student that will graduate with PhD grades. The developed system could be very useful in predicting masters postgraduate student result grade, even from the point of starting their masters program in the university. This will help management staff, academic planners to properly counsel students that wish to proceeds in their academics after their masters program to improve their overall performance.

Keyword: Artificial Neural Network, Education Data Mining, Prediction, Student's Performance

African Journal of Computing & ICT Reference Format:
M.D. Gbolagade, M.A. Hambali & A.A. Akinyemi (2015): Predicting Postgraduate Performance Using Resample Preprocess Algorithm and Artificial Neural Network. Afr J. of Comp & ICTs. Vol 8, No. 1, Issue 1. Pp 145-158.

1. INTRODUCTION

The focus is on the development of a domain specific language (DSL) for modeling oil and gas pipeline systems. Traditionally the aim is to describe a roadmap to developing a cohesive tool for the design of such artefacts through the use of disparate tools [7]. The approach to the development of the DSL is based upon Model Driven Engineering (MDE) technologies. The two main schools of MDE are Model Driven Architecture (MDA), and Domain-Specific Modelling (DSM). MDA language specifications restrict the user to diagram definition standards (e.g. UML), whereas DSM languages identify the problem and the goal to be reached during the process provided. A master degree has become an essential and basic skill for various jobs today, most especially academia. This has leads to the necessity of providing enhance master's program related curriculums for postgraduate students. Universities must design means of helping students to develop proficiency of good researches and tap the advantage of postgraduate program.

For example, students are required to learn how to conduct, report and present their research work (thesis) during postgraduate program; for them to be able publish some of their researches, in order to enhance their academic profession. Therefore, postgraduate students who are seeking to proceed on their academic career should be certified for some of the courses offered together with their theses. In Nigerian University, for a student to proceed further (after master degree) in his/her academic career, he must obtain 60% and above (PhD grade) before he can further their studies to PhD level. Therefore, governments and educational institutions in many countries have paid much attention to enhancing the master degree program for postgraduate students. Some universities have developed specific grade which a student can obtain before he can be considered for having PhD grade which is 4.0 out of 5.0 points in some institutions while some use 60% out of 100%.

These considerations motivate this research work, in order to develop an early warning mechanism for students, by analyzing the characteristics of passed students (with PhD grade) and failed (below PhD grade) students using data mining (DM) techniques. Educational institutions maintain data repositories collected over a period of time about students, courses offered and student performances etc. These data are increasing tremendously, provides a rich resource for analysing and gaining interesting knowledge that can be used for decision-making in order to improve the quality of the educational programme. Despite this fact, decision makers in educational system rarely exploit DM techniques to reveal interesting patterns from the vast amount of data in their data repositories [1]. There are increasing research interests in using DM in education. This new emerging field, called education data mining (EDM), concerns with developing methods that discover knowledge from data originating from educational environments [2],[3].

Knowledge can be extracted from historical and operational data that resides in the educational institution's database using DM techniques such as Decision Trees, Neural Networks, Naïve Bayes, K- Nearest neighbour, Support Vector Machine etc. Using these techniques, many kinds of knowledge can be discovered such as association rules, classifications and clustering. The discovered knowledge can be used for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about students' performance and soon.In general, DM techniques provide a lot of opportunities to mine huge amount of data to discover hidden patterns and relationships that help in decision making [4], [5], [1], [3].

Classification is one of the most frequently used DM technique, which employs a set of pre-classified examples to develop a model that can classify large record of data [6]. Classification is a method of finding a set of models (or functions) which define and differentiate data classes (or concepts), for the aim of using the model to predict the class of objects whose class label is not defined before (unknown). The derived model can be denoted in numerous forms, such as classification (IF-THEN) rules, decision trees, mathematical formulae, or neural networks. Classification is usually used for predicting the class label of data objects. However, in some situation, predicting missing or unavailable data values is what is required rather than class labels. In this case, the predicted values are usually numerical data, and is often referred to as prediction. The data classification process involves two stages: learning and classification. In learning stage, the training data are analyzed by classification algorithm; while in classification, test data are used to evaluate and determine accuracy of the classification rules. If the accuracy is satisfactory the rules generated can be applied to the new data tuples [7].

Therefore in this paper, classification DM technique is employed on enrolment data of masters' students of a department in University of Ilorin, using multilayer perceptron (MLP), radial basis functions (RBF) and Sequential minimal optimization algorithm (SMO) to classify the data, effect of data preprocess algorithm is also revealed, attribute selection was also used to evaluate the factors that are very important to predict possible student that will graduate with PhD grades. This paper is organized as follows, Section 2 review of related works. Section 3 presents the research methodology, including the collected dataset, the procedure for developing different classifier models for comparisons, etc. Section 4 shows the experimental results and the conclusion is provided in Section 5.

2. RELATED WORK

In recent years, the problem of predicting students' performance and failure rate among the students is one of the prominent research area in EDM. Using data mining in higher education is a recent research field; a lot of work has been done recently on the application of machine learning to educational databases to solve educational problems that has to do with prediction of student's performance. Some researchers have used various DM techniques to help instructors and administrators to improve e-learning systems [8], [9], [10], [11], [12], [4], [13], [14], [5], [15].

Guruler et al. [9] employed data mining techniques to explore some factors that have impact on the success of university students. Similarly, Lee et al. [11] analysed some important factors which can influence the preferences of learners from diverse backgrounds. For web based systems, Hamdi [10] presented a method for extracting and inferring useful knowledge for student learning using web mining techniques.

Further, a hybrid DM technique is proposed by Shih et al. [5] to evaluate the significant characteristics of study strategy scales and their inter-relationships for freshmen students in a web-based self-assessment system. Romero et al. [14] proposed an architecture of recommender system that utilizes web usage mining to recommend the next pages to visit in an adaptive, web-based educational system in order to aid the instructor to carry out the web mining process. For teaching and learning content, Wang et al. [15] applied a decision tree model to identify the most adaptive learning sequences based on students' profiles for a particular teaching content. Guo and Zhang [8] provided a method for representing and extracting a vibrant learning process and learning patterns to support students' deep learning, effective tutoring and collaboration in a web-based learning environment.

Hijazi & Naqvi [16] presented a study on the student performance by randomly sample 300 students (225 males and 75 females) from a group of colleges affiliated to Punjab University of Pakistan. They try to identify which of the following factors: Students' attitude towards attendance in class, hours spent in study on daily basis after college, students family income, students mother's age and mother's education has significant effect related to student performance. By means of simple linear regression analysis, it was found that the factors like mother's education and student's family income were highly correlated with the student academic performance.

Al-Radaideh et al. [17]applied a decision tree algorithm to predict the final grade of students who studied the C++course in Yarmouk University, Jordan in the year 2005. Three different classification algorithms namely ID3, C4.5 and the Naïve Bayes were used. The outcome of their results shows that Decision Tree model outperform other prediction models. Bray [18]observed during his research on private tutoring and its implications that, the percentage of students receiving private tutoring in India was relatively high than that of Malaysia, Singapore, Japan, China and Sri Lanka. Oladipupo & Oyelade [19] analysed a data using undergraduate students' result in the department of Computer Science from a university in Nigeria.

The department offers two programmes; Computer Science and Management Information Science. A total number of 30 courses for 100 level and 200 level students were considered as a case study. The analysis revealed that there is more to students' failure than the students' ability. It also reveals some hidden patterns of students' failed courses which could serve as bedrock for academic planners in making academic decisions and an aid in the curriculum re-structuring and modification with a view to improving students' performance and reducing failure rate.

Kovacic [20] presented a case study on EDM to identify up to what extent the enrolment data can be used to predict students' success. The algorithms CHAID and CART were applied on student enrolment data of information system of students of Open Polytechnic of New Zealand into two class of successful and unsuccessful students. The accuracy obtained with CHAID and CART were 59.4 and 60.5 respectively. Yadav et al. [21]used the university students data such as attendance, class test, seminar and assignment marks from the students' database, to predict the performance at the end of the semester using three decision tree algorithms ID3, C4.5 and CART and showed that CART is the best algorithm for classification of data. Surjeet & Saurabh [22] used student's past academic performance to generate a model using ID3 decision tree algorithm for prediction of student's enrolment in MCA course. Shreenath & Madhu [23] build up a model for the placement database, for institutions to use it to discover some interesting patterns that could be analyzed to plan their future activities.

Abdulsalam et al [3] presented a report on three decision tree algorithms for predicting students' performance in a computer programming course taken in 200 level based on their ordinary level results in Mathematics and Physics and their 100 level results in Mathematics and Physics courses. One hundred and thirty one (131) students' records from computer science programme at Kwara State University (KWASU) Malete, Nigeria between 2009 and 2013 were used. The attributes used including students' ordinary level scores in Mathematics and Physics, 100 level results in Mathematics and Physics courses and the score in a 200 level computer programming course (CSC 203). J48, Classification and Regression Tree (CART), and Best-First Tree (BF Tree) decision tree algorithms were used in WEKA data mining software to generate three classification models employed in predicting students' performance in CSC 203.

Their results showed that J48 tree has the highest prediction accuracy of 70.37% and least execution time of 0.02 seconds while CART and BFTree has prediction accuracy of 60.44% and 60.30% respectively and both having execution time of 0.22 seconds based on the data set used in this study. This study also revealed that previous knowledge of Mathematics and Physics both at Ordinary level and 100 level are essential determinants of students' performance in a computer programming course.

3. METHODOLOGY

Data were collected from the University database using sampling method, 105 datasets of Master (M. Sc.) program were selected from a department in University of Ilorin, which include 2003/04, 2004/05, 2005/06, 2007/08, 2008/09, 2009/10, 2010/11 and 2011/12 sessions.Students' demographics,Undergraduate grade, state of origin, year of graduation (B. Sc./PGDS), age at year of admission, higher qualification, year of entry and year spent were collected for the data mining study.Table 1 below shows description of students' dataset.

3.1 Pre-processing

This stage involve dataset preparation before applying data mining techniques. At this stage, traditional pre-processing methods such as data cleaning, transformation of variables and data partitioning were applied. Also, other techniques of data pre-process from WEKA such resample were employed in order to reduce the size of datasets. Resample filter produces a random subsample of a dataset using sampling with replacement. In this work, random seed of 1 and sample size percentage of 100 were used.

African Journal of Computing & ICT

www.ajocict.net

Table 1: The Students' dataset description table

Variable Name	Description	Variable coding values
Sex	Students' gender	Male = M Female = F
Age	Students' age at the time of securing admission into master programme	Numeric range from 22 – 47 years
State of origin	State origin of individual students.	Any state in Nigeria
Higher Qualification	Last qualification obtained before M. Sc. program	First degree = B. Sc. Postgraduate Diploma = PGDS
Year of Grad	The year student obtained the last qualification	Date
Year of Entry	The year student gained admission for M. Sc. program	Date
B. Sc. Grad Grade	Graduate grade obtained in B. Sc. or HND program	B. Sc. grade First class = 1 Second class upper = 2 Second class lower = 3 Third class = 4 Pass = 5 HND Grade Distinction = 1 Upper credit = 2 Lower Credit = 3
Marital status	Marriage status of students	Married = M Single = S Divorce = D Married but Separate = MS
Year spent	Number of years spent by students before completed their M. Sc. program	Numeric
M. Sc. grade	The grade obtained by students at the end of M. Sc. program	Above 60% = Merit >= 50% but < 60% = Pass < 50% = Fail

Vol 8. No. 1 Issue 2 – May, 2015
African Journal of Computing & ICT

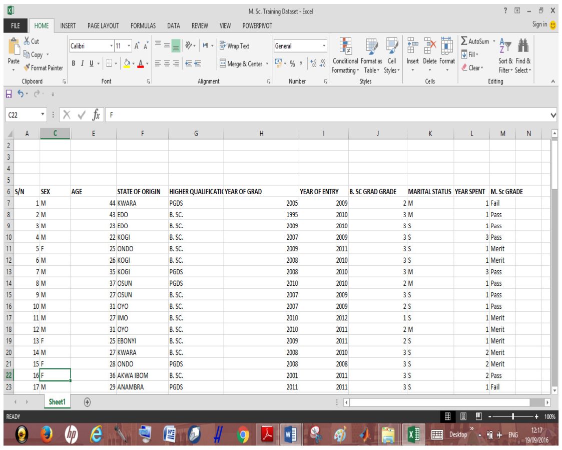


Fig. 1: Spreadsheet interface of the training data set

3.2 Data Mining

At this stage, data mining algorithms are applied in order to predict students' performance at the end of M. Sc. program. In doing this, classification algorithms such as MLP, SMO and RBF are employed and compared using WEKA data mining tool. Also, two test mode procedure were used, that is, percentage split (66% for training and 44% for testing) and 10 folds cross validation were used, and compared.

3.3 Artificial Neural Network

Artificial neural networks (ANNs) are mathematical forms of the human neural design, representing its "learning" and "generalization" capabilities [24]. As the human brain comprises millions of neurons that are connected together by synapses, ANNs are formed from several numbers of simulated neurons, linked to one another in a manner akin to brain neurons. Just like human brain, the neuron's strength interconnections may differ (or be changed by the training algorithm) in reaction to a presented stimulus or an obtained output, which enables the network to "learn". ANNs are widely used with different application areas in science, medicine, technology, finance and soon. ANNs have the incredible capability to infer meaning from complicated or imprecise data and also, can be employed to extract patterns and detect trends that are too complex to be detected by either humans or other computer algorithms.

3.4Multi-Layer Perceptron

A Multi-Layer Perceptron Feed-Forward Back Propagation Neural Network was one of the classifier employed in this work. ANN was preferred for this work out of numerous other algorithms due of its simplicity of use and capabilities for supervised learning. Multi-layer indicates that the network has three layers: input, hidden and output layers. The term, feed forward explains how the neural network processes the pattern and recalls patterns. When ANNs are described as "feed forward neural network", it means its neurons are connected forward only.

Backpropagation, or propagation of error, is a common way of training artificial neural networks on how to perform a given task. The back propagation algorithm is used to described feedforward ANNs mean that, the artificial neurons are organized in layers, send their signals "forward", and the errors are propagated backwards. The back propagation algorithm uses supervised learning, that is, the algorithm is provided with examples of the inputs and outputs of the data we want the network to compute; then the error (difference between actual and expected results) is estimated. The main aim of the back propagation algorithm is to minimize this error, until the ANN learns the training data.

Yashpal and Alok [25] summarized the ANN technique thus:

1. Present a training sample to the neural network.
2. Compare the network's output to the desired output from that sample. Compute the error in each output neuron.
3. For each neuron, compute the actual output and then a scaling factor, how much lower or higher the output must be adjusted to match the desired output. This is the local error.
4. Adjust the weights of each neuron to lower the local error.

3.5 RBF Neural Network

An RBF neural network has three layers as in other neural networks [24]. The first layer is an input layer; the second layer is a hidden layer that includes some radial basis functions, also known as hidden kernels; and the third layer is the output layer.

An RBF neural network can be considered as a mapping of input domain *X* onto the output domain *Y*.

$$y_m(\ddot{x}) = \sum_{i=1}^{n} w_{mi}\, G(\| \ddot{x} - \ddot{t}_i \|) + b_m; \quad \ldots(1)$$

i = 1, 2, 3, …, N;
m = 1, 2, 3, …, M.

Here $\|\cdot\|$ stands for the Euclidean norm. M is the number of outputs. N is the number of hidden kernels. $y_m(\ddot{x})$ is output m corresponding to the input $\ddot{x}$. $\ddot{t}_i$ is the center of kernel i. w_{mi} is the weight between kernel i and output m. b_m is the bias on output m. $G(\| \ddot{x} - \ddot{t}_i \|)$ is the kernel function. The most commonly used kernel function for RBF neural networks is Gaussian kernel function as in equation 2:

$$G(\| \ddot{x} - \ddot{t}_i \|) = \exp\left(-\frac{\|\ddot{x} - \ddot{t}_i\|^2}{2\sigma_i^2}\right) \quad (2)$$

where σ_i is the radius of the kernel i.

www.ajocict.net

The main steps to construct an RBF neural network include [24]:

a) Determine the positions of all the kernels $\ddot{t}_i$,
b) ii. Determine the radius of each kernel, and
c) iii. Calculate the weights between the kernels and the output nodes.

3.6 Sequential Minimal Optimization

Sequential Minimal Optimization (SMO) is a simple algorithm that can solve the quadratic programming (QP) problem of support vector machine (SVM) very fast without any additional matrix storage and without employing numerical QP optimization steps at all. To ensure convergence, SMO beak down overall QP problem into QP sub-problems, using Osuna's theorem [26].

Compare SMO to the previous methods used to train SVM, it chooses to solve the smallest possible optimization problem at every step. For the typical SVM QP problem, the smallest possible optimization problem involves solving two Lagrange multipliers, because the Lagrange multipliers must satisfy a linear equality constraint. At every step, SMO combine two Lagrange multipliers together for optimization, evaluates the optimal values for these multipliers, and updates the SVM to reflect the new optimal values. Conversely, the full SMO algorithm contains many optimizations designed that can help in speeding up the process on large datasets and ensure that the algorithm converges even in degenerated situations.

The advantage of SMO lies in the fact that, solving two Lagrange multipliers can be done analytically and requires no extra matrix storage at all. Thus, it avoids numerical QP optimization entirely and also, very large SVM training problems can fit into an ordinary personal computer or workstation memory. Because no matrix algorithms are required in SMO, it is less prone to numerical precision problems.

4. RESULTS AND DISCUSSION

In this study, the results of proposed hybrid of Resample and ANN to predict the results of M. Sc. students were presented in this section. These results were compared with the standard neural network algorithms.

4.1 Experiment Results

The experiment was performed using Weka Explorer. Firstly, the dataset was uploaded into explorer, RESAMPLE preprocessing algorithm was used to reduce the size of datasets and second step involves the classification algorithms using MLP, RBF or SMO. We verified our experiment using a random average of 10-fold method (10-Fold cross validation) and percentage split test mode.

4.2 Evaluation Metrics

In selecting the appropriate algorithms and parameters that best model the predicted M. Sc. results variable were briefly discuss; the following performance metrics were used:

A classifier is evaluated by a confusion matrix as illustrated in table 2. The columns indicate the predicted class and the rows show the actual class. In the confusion matrix, True Negative (TN) is the number of negative samples correctly classified, False Positive (FP) is the number of negative samples incorrectly classified as positive, False Negative (FN) is the number of positive samples incorrectly classified as negative and True Positive (TP) is the number of positive samples correctly classified.

Table 2: Confusion Matrix

	Predicted Negative	**Predicted Positive**
Actual Negative	TN	FP
Actual Positive	FN	TP

Overall accuracy is defined in equation 3

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+TN+FN)} \qquad (3)$$

From the confusion matrix in table 2, the expressions for FP rate, Recall and Precision are derived and are presented in equations 4, 5 and 6.

$$\text{FP Rate} = \frac{FP}{(TN+FP)} \qquad (4)$$

$$\text{TP Rate} = \text{Recall} = \frac{TP}{(TP+FN)} \qquad (5)$$

$$\text{Precision} = \frac{TP}{(TP+FP)} \qquad (6)$$

Time: This is referred to as the time required to complete training or modeling of a dataset. It is represented in seconds

Mean Absolute Error (MAE): Mean absolute error is the average of the difference between predicted and the actual value in all test cases; it is the average prediction error.

Root Mean Squared Error (RMSE): Mean-squared error is one of the famous method for measures of success for numeric prediction. This value is computed by taking the average of the squared differences between each computed value and its corresponding correct value. The mean-squared error is simply the square root of the mean squared-error. The mean-squared error gives the error value the same dimensionality as the actual and predicted values. From the table 3, three Algorithms used were MLP, SMO and RBF with Percentage Split test mode.

For the prediction of postgraduate performance, when considering execution time, SMO algorithm with Resample pre-process has least time complexity of 0.58 Sec to build the classification model, while RBF algorithm without pre-process has least time of 1.02 Sec. RBF algorithm outperformed other algorithms when dataset is not pre-process while SMO performed very well with data pre-process with Resample algorithm. In general, correctly classified accuracy of MLP and SMO improved by over 50% and 20% respectively, while classified accuracy of RBF was reduced by over 20%. This indicate that Resample preprocess improve the performance of both MLP and SMO, but it reduce the performance of RBF from 64.71% to 50%. Figure 2 and 3 showing comparison of the three classifiers in terms of classification accuracy and other performance metrics using Percentage split.

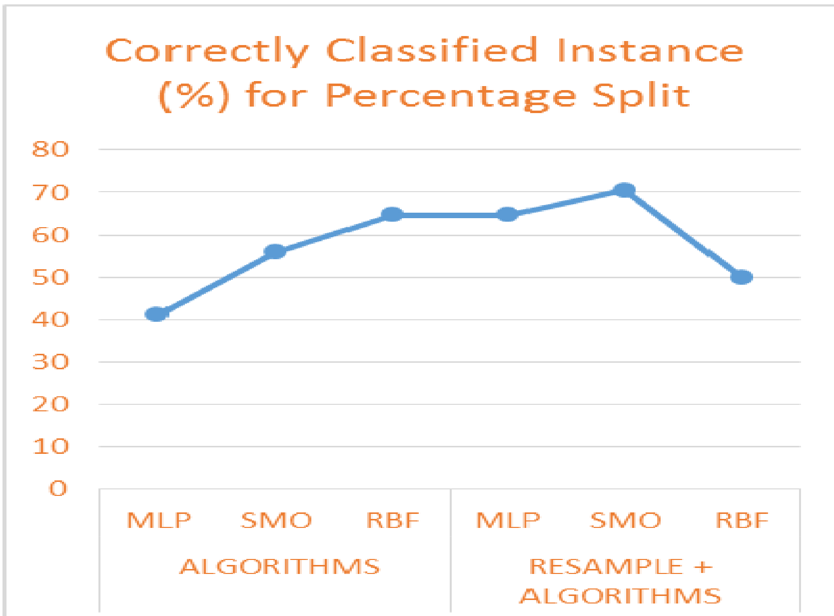


Fig. 2: Correctly Classified Accuracy for Percentage Split Test Mode

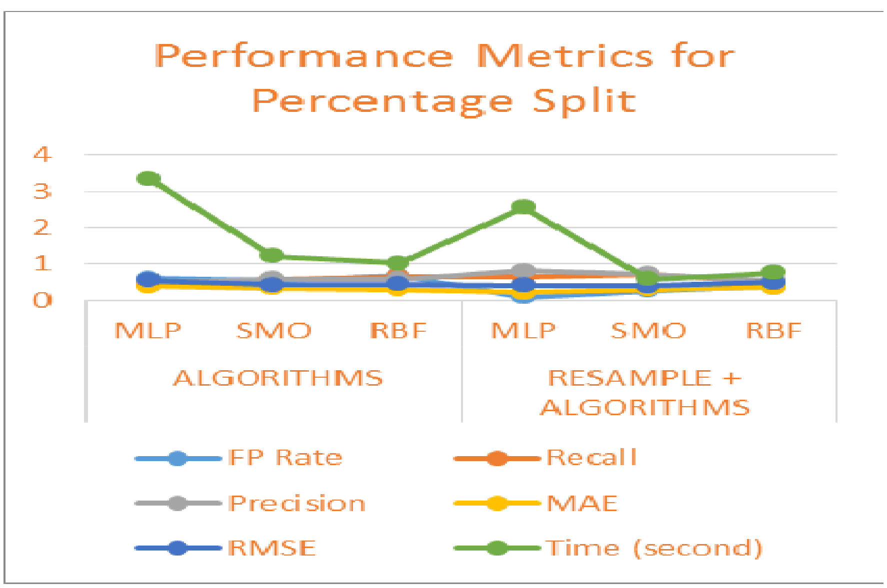


Fig. 3: Performance Metrics for Percentage Split Test Mode

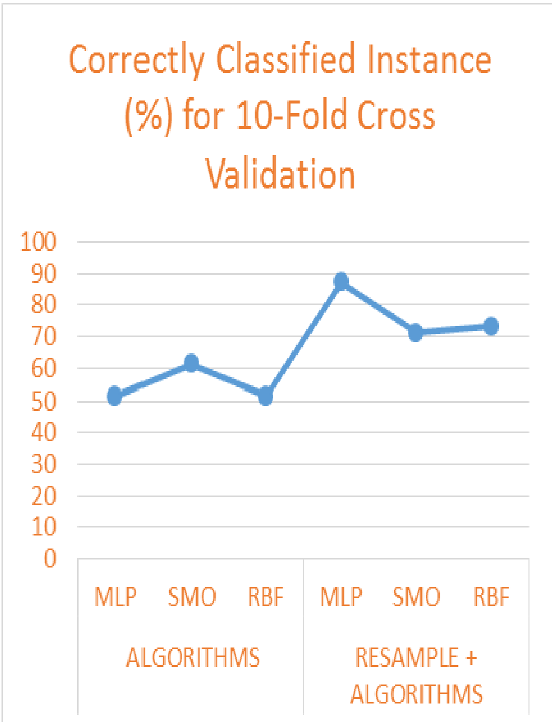
Table 3: Results of Experiment using Percentage Split Test mode

Performance Metrics	ALGORITHMS			RESAMPLE + ALGORITHMS		
	MLP	SMO	RBF	MLP	SMO	RBF
Correctly Classified Instance (%)	41.1765	55.8824	**64.7059**	64.7059	**70.5882**	50
Incorrectly Classified Instance (%)	58.8235	44.1176	35.2941	35.2941	29.4118	50
FP Rate	0.591	0.553	0.678	0.088	0.271	0.381
Recall	0.412	0.559	0.647	0.647	0.706	0.5
Precision	0.518	0.585	0.584	0.813	0.721	0.543
MAE	0.3798	0.3333	**0.297**	**0.2071**	0.2941	0.3552
RMSE	0.5495	0.4303	0.4335	0.3995	0.3792	0.4961
Time (second)	3.36	1.22	**1.02**	2.56	**0.58**	0.75

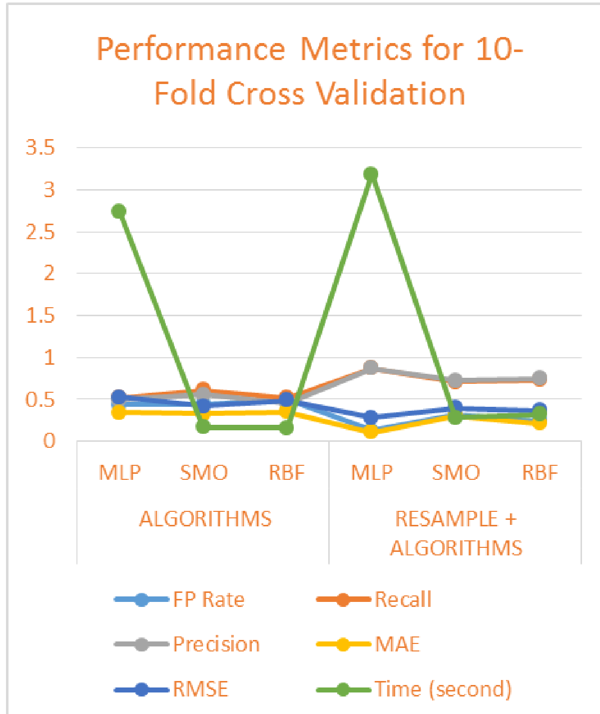
Table 4: Results of Experiment using 10-Fold Cross Validation Test mode

Performance Metrics	ALGORITHMS			RESAMPLE + ALGORITHMS		
	MLP	SMO	RBF	MLP	SMO	RBF
Correctly Classified Instance (%)	51.4851	**61.3861**	51.4851	**87.1287**	71.2871	73.2673
Incorrectly Classified Instance (%)	48.5149	38.6139	48.5149	12.8713	28.7129	26.7327
FP Rate	0.434	0.441	0.495	0.128	0.312	0.244
Recall	0.515	0.614	0.515	0.871	0.713	0.733
Precision	0.498	0.558	0.461	0.871	0.724	0.749
MAE	0.3465	**0.3256**	0.3538	**0.1087**	0.3036	0.2146
RMSE	0.5298	0.4213	0.4855	0.278	0.3934	0.3685
Time (second)	2.74	0.17	**0.16**	3.19	**0.28**	0.32

In table 4, 10-fold Cross Validation mode was employed for the three Algorithms used (MLP, SMO and RBF).When considering execution time, SMO algorithm with Resample pre-process has least time complexity of 0.28 Sec to build the model, while RBF algorithm without pre-process has least time of 0.16 Sec. SMO algorithm outperformed other algorithms when dataset is not pre-process while MLP performed very well with data pre-process with Resample algorithm. In general, correctly classified accuracy of all the three algorithms were improved when apply Resample pre-process before classification algorithms. Figure 4a and 4b showing comparison of the three classifiers in terms of classification accuracy and other performance metrics using 10-fold Cross Validation.



(a)



(b)

Fig. 4a: Correctly Classified Accuracy for 10-Fold Cross Validation Test Mode and Fig 4b: Performance Metrics for 10-Fold Cross Validation Test Mode

However, when compare the two test modes, performance of the three algorithms are excellent in 10-fold Cross Validation even without Resample pre-processing algorithm. Therefore, 10-fold Cross validation is best for students dataset used in this experiment for the prediction of Postgraduate performance.

African Journal of Computing & ICT

www.ajocict.net

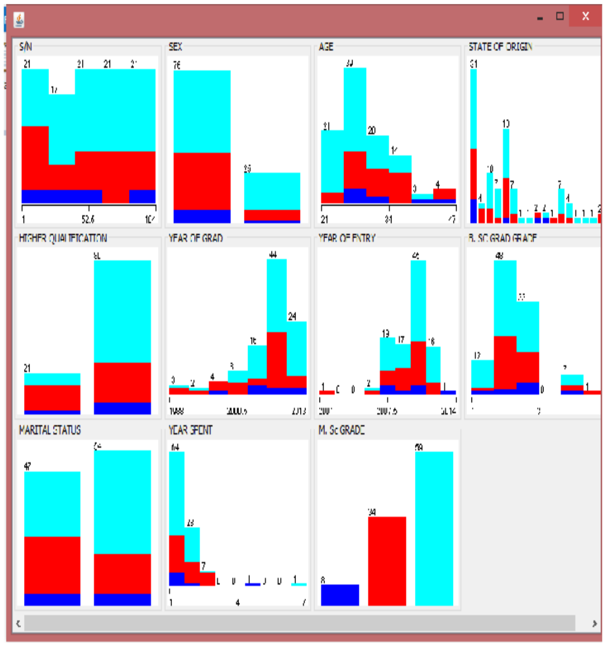


Fig. 6: Attributes visualizationBefore Resample Preprocessing

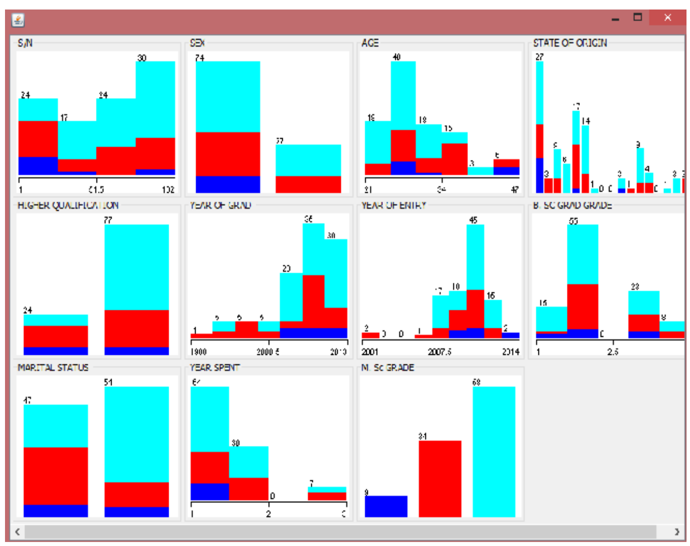


Fig. 7: Attributes visualization After Resample Pre-processing

Information Gain and Gain ratio are attribute selection measure which is used to rank attributes in a data set. The attribute with the highest ranked value indicates highly correlated to M. Sc. students' academic performance. Table 5 and 6 show the summary of Information Gain and Gain Ratio Attribute Evaluator respectively, and graphical representation in Figure 8 and 9.

Vol 8. No. 1 Issue 2 – May, 2015
African Journal of Computing & ICT

www.ajocict.net

Table 5: Attributes Ranking Using Information Gain

S/N	Attributes	Rank	Information Gain
1	STATE OF ORIGIN	4	0.3461
2	MARITAL STATUS	9	0.1009
3	HIGHER QUALIFICATION	5	0.0746
4	SEX	2	0.0438
5	YEAR OF GRAD	6	0
6	YEAR OF ENTRY	7	0
7	AGE	3	0
8	B.SC GRAD GRADE	8	0
9	YEAR SPENT	10	0

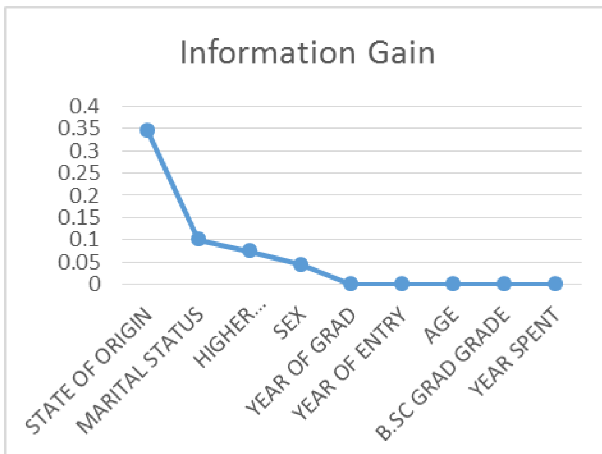


Fig 8: Information Gain of the Attributes Evaluator

Table 5: Attributes Ranking Using Information Gain Ratio

S/N	Attributes	Rank	Gain Ratio
1	STATE OF ORIGIN	4	0.1086
2	MARITAL STATUS	9	0.1012
3	HIGHER QUALIFICATION	5	0.0943
4	SEX	2	0.0523
5	YEAR OF GRAD	6	0
6	YEAR OF ENTRY	7	0
7	AGE	3	0
8	B.SC GRAD GRADE	8	0
9	YEAR SPENT	10	0

African Journal of Computing & ICT

www.ajocict.net

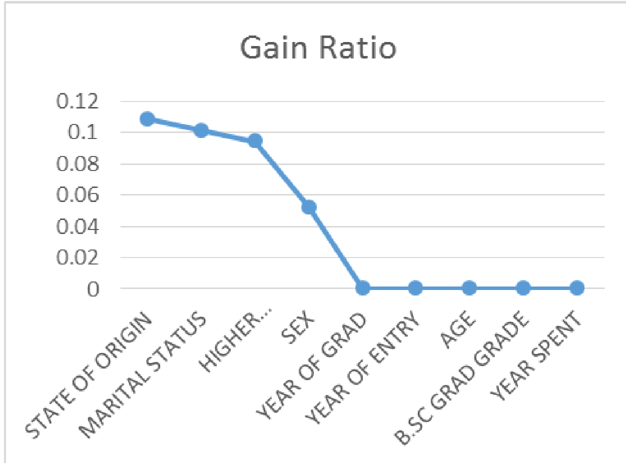


Fig. 9: Gain Ratio of the Attributes Evaluator

The attribute ranking with respect to the class label using Information gain and gain ratio criteria shows that Performance of M. Sc. students are determine by State in which student come from, marital status, higher qualification and sex of the student.

5 CONCLUSION

This study employed the use of three classification algorithms which are MLP, SMO and RBF in predicting the performance of M. Sc. Students in other to determine if they will have Ph.D. Proceeds Grade result or not. The study reveals that Performance of M. Sc. students are determined by State in which student comes from, marital status, higher qualification and sex of the student.

According to [27] Wong, Bodnovich, & Selvi (1997), about 95% of business applications reported in the literature used MLP type of neural model. In other words, they are appropriate for any functional mapping problem where we are interested in knowing how a number of input variables affect the output variable(s).Since most prediction and classification tasks can be treated as function mapping problems, the MLP networks are very motivating to data mining.

Result from this study will help the students to get fully prepared for their master's program especially if they are willing to proceeds in academics. This study will also help staffs to identify those students that will likely not meet up with the PhD grade.

Reference

[1] S. O. Abdulsalam, K.S. Adewole, A.G. Akintola and M.A. Hambali, "Data Mining in Market Basket Transaction: An Association Rule Mining Approach", *International Journal of Applied Information Systems (IJAIS), 7(10): 15-20, 2014*

[2] C. Romero, S. Ventura, and E. García, "Data Mining in Course Management Systems: Moodle Case Study and Tutorial", Computers and Education; 51; 368-384, 2007.

[3] S. O. Abdulsalam, A. N. Babatunde, M. A. Hambali and R. S. Babatunde, "Comparative Analysis of Decision Tree Algorithms for Predicting Undergraduate Students' Performance in Computer Programming", Journal of Advances in Scientific Research & Its Application (JASRA), 2, Pg. 79 – 92, 2015.

[4] C. Romero, and S. Ventura, "Educational data mining: A Survey from 1995 to 2005", Expert Systems with Applications (33), Pg. 135-146, 2007.

[5] C.-C Shih, D.-A Chiang, S.-W Lai, and Y.-W Hu, "Applying Hybrid Data Mining Techniques to Web-Based Self-Assessment System of Study and Learning Strategies Inventory. Expert Systems with Applications, 36 (3), 2009.

[6] S. Samrat, and K. Vikesh, Classification of Student's Data Using Data Mining Techniques for Training & Placement Department in Technical Education. International Journal of Computer Science and Network (IJCSN), 1 (4), 2012.

[7] M. Dash and H. Liu, "Feature Selection for Classification," An International Journal of Intelligent Data Analysis, 1 (3), Pg. 131-156, 1997.

[8] Q. Guo, and M. Zhang, (2009). "Implement Web Learning Environment Based on Data Mining", Knowledge-Based Systems, 22(6), Pg. 439-442, 2009.

[9] H. Guruler, A. Istanbullu, and M. Karahasan, "A New Student Performance Analysing System Using Knowledge Discovery in Higher Educational Databases", Computers & Education, 55(1), Pg. 247-254, 2010.

[10] M. S. Hamdi, "MASACAD: A Multi-Agent Approach to Information Customization for the Purpose of Academic Advising of Students", Applied Soft Computing, 7(3), Pg. 746-771, 2007.

[11] M. W. Lee, S. Y. Chen, K. Chrysostomou, and X. Liu, "Mining Students' Behavior in Web Based Learning Programs. Expert Systems with Applications, 36(2), 3459-3464, 2009.

[12] C. Romero, and S. Ventura, "Data Mining in e-Learning", Southampton. UK: WIT Press, 2006.

[13] C. Romero, S. Ventura, E. Garcia, "Data Mining in Course Management Systems: Moodle Case Study and Tutorial", Computers & Education, 51 (1), Pg. 368-384, 2008.

[14] C. Romero, S. Ventura, A. Zafra, and P. de Bra, "Applying Web Usage Mining for Personalizing Hyperlinks in Web-Based Adaptive Educational Systems. Computers & Education, 53(3), 828-840, 2009.

[15] Y.-H Wang, M.-H Tseng, and H.-C Liao, "Data Mining for Adaptive Learning Sequence in English Language Instruction. Expert Systems with Applications, 36 (4), 7681-7686, 2009.

[16] S. Hijazi, and R. Naqvi, "Factors Affecting Students' Performance: A Case of Private Colleges. Bangladesh e-Journal of Sociology, 3 (1), 2006.

[17] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar Mining Student Data Using Decision Trees. International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.

[18] M. Bray, "The Shadow Education System: Private Tutoring and Its Implications for Planners. (2nd ed.), UNESCO, PARIS, France, 2007.

[19] O. O. Oladipupo, and O. J. Oyelade, "Knowledge Discovery from Students' Result Repository: Association Rule Mining Approach", International Journal of Computer Science and Security, 4(2), Pg. 199-207, 2010.

[20] Z. Kovacic, "Early Prediction Of Student Success: Mining Student Enrollment Data", Proceedings of Informing Science & IT Education Conference, 2010.

[21] S. Yadav, B. Bharadwaj, and S. Pal, "Data Mining Applications: A Comparative Study for Predicting Students' Performance", International Journal of Innovative Technology and Creative Engineering (IJITCE), Vol. 1, No. 12, Pg. 13-19, 2011.

[22] K. Surjeet, and P, Saurabh, "Data Mining Application in Enrolment Management: A Case Study", International Journal of Computer Application (IJCA), 41 (5), Pg. 1-6, 2012.
[23] A. Shreenath, and N. Madhu, "Discovery of Students' Academic Patterns Using Data Mining Techniques. International Journal on Computer Science and Engineering (IJCSE), 4 (6), 2012.
[24] M. A. Hambali, and M. D. Gbolagade, "Ovarian Cancer Classification Using Hybrid Synthetic Minority Over-Sampling Technique and Neural Network. Journal of Advances in Computer Research (JACR), 7 (4), Pg. 109 – 124, 2016.
[25] S. Yashpal and S. C. Alok, "Neural Networks in Data Mining" Journal of Theoretical & Applied Information Technology, Pg 37-42, 2009.
[26] J. Platt, "Fast Training of Support Vector Machines Using Sequential Minimal Optimization, in Advances in Kernel Methods – Support Vector Learning", B. Scholkopf, C. Burges, A. Smola, eds., MIT Press, 1998.
[27] B. K. Wong, T. A. Bodnovich, and Y. Selvi, "Neural Network Application in Business: A Review and Analysis of the Literature (1988 - 1995)", Decision Support Systems, 19, Pg. 301 – 320, 1997.