A Multidisciplinary Journal Publication of the Faculty of Science, Adeleke University, Ede, Nigeria
2015 Edition Vol 2.

Comparative Analysis of Decision Tree Algorithms for Predicting Undergraduate Students' Performance in Computer Programming

Abdulsalam Sulaiman O.[1], **Babatunde Akinbowale N.**[2]**& Babatunde Ronke S.**[4]
[1,2,4]Department of Computer Science
Kwara State University
Malete, Kwara State, Nigeria
abdulsalamny@gmail.com, drealak@gmail.com, ronkebabs711@gmail.com

Hambali Moshood A.[3]
[3]Department of Computer Science
Federal University, Wukar
Wukari, Taraba State, Nigeria
hamberlite@gmail.com

ABSTRACT
Educational data are increasing tremendously with little exploration by the education managers. Hidden knowledge can be discovered from the huge data sets available in educational databases through the use of data mining techniques. Educational data mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that reside in educational databases. One of the significant areas of the application of EDM is the development of student models for predicting students' performances in their educational institutions. The focus of this work is to identify the optimal 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) between 2009 and 2013 were used. The attributes used are 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). C4.5 (known as J48 in WEKA) Classification and Regression Tree (CART), and Best-First Tree (BF Tree) decision tree algorithms were used in Waikato Environment for Knowledge Analysis data mining software to generate three classification models employed in predicting students' performance in CSC 203. The results of these algorithms were compared using 10-fold cross validation method in terms of prediction accuracy and computational time. Our 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.

Keywords: Data Mining, Educational Data Mining, Decision Tree algorithm, Students' Performance

1. INTRODUCTION

Few decades ago, analysis of data for effective and efficient decision making in our organizations was through the use of Mathematical and Statistical methods where tools such as charts, regression method, etc. were used for decision making. It was however, observed that the estimated amount of information in the world doubles every twenty (20) months leaving the size and number of the databases increasing even faster (Ann, 1997) and because of this large data size, extraction of useful and meaningful information has proven so herculean and challenging as the primitive tools can no longer analyze these data sets. The large data made readily available in our information industry are not useful until being worked upon and converted into useful and meaningful information. The analysis performed on these data to make it uncover interesting, important and meaningful patterns thereby allowing its applicability in many areas of our lives is called Data Mining (DM).

DM refers to extracting knowledge from huge amounts of data. It is the process of discovering interesting knowledge from large amount of data stored in databases, data warehouses or other information repositories (Han and Kamber, 2006). DM is the non-trivial extraction of valid, implicit, potentially useful and ultimately understandable information in large databases with the help of modern computing devices (Fayyad, Piatesky, Smyth and Uthurusamy, 1996). It allows users to analyze data from various dimensions, categorize them and summarize the relationship identified. Data mining techniques are used to operate on huge amount of data to discover hidden patterns and relationships helpful for decision making (Abdulsalam, Adewole, Akintola and Hambali, 2014).

Knowledge discovery process can be described as an iterative process being made up of all the processes involved from the collection of raw data either from a centralized or distributed database to the point at which information or a new knowledge is derived. According to Osmar (1999), the processes involved in the knowledge discovery process are: Data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation. DM is the most important step in the knowledge discovery process. The pattern or knowledge mined can be applied in medicine, scientific discovery, market analysis, crime detection, customer retention, science exploration, production control, education etc.

Educational institutions maintain data repositories collected over time about students, courses offered and student performances etc. These data which are increasing tremendously provides a rich resource for analyzing and gaining interesting knowledge that can be used for decision-making in improving the quality of the educational programme. Despite this fact, decision makers in educational system rarely utilize data mining techniques to uncover interesting patterns from the vast amount of data in their data repositories (Abdulsalam *et al.*, 2014). There are increasing research interests in using DM in education. This new emerging field, called EDM, concerns with developing methods that discover knowledge from data originating from educational environments (Romero, Ventura and Garcia, 2007). Knowledge can be extracted from historical and operational data that resides in the educational institution's database using data mining 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 so on.

The focus of this work is to identify the optimal 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. Three decision tree algorithms (J48, CART and BF Tree) were applied on the data set and their results were compared using 10-fold cross validation in terms of classification accuracy and execution time.

The rest of this paper is organized as follows. Section 2 presents related work to this study. In section 3, we described the methodology used. Section 4 contains discussion of the results. Finally, conclusions and future work are presented in section 5.

2. RELATED WORKS

The problem of predicting students' performance and failure rate among the students is one of the famous applications of 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 challenges that has to do with prediction of student's performance. The following are some of the related work on prediction of students' performance.

Bhardwaj and Pal (2011) used data mining technique called classification to evaluate students' performance; they made use of decision tree method for their classification. They considered the student's previous semester marks, class test grade, seminar performance, assignment, general proficiency, attendance, lab work and end semester marks as their variables. The aim of their work was to extract knowledge that describes students' performance at end of each semester examination.

Cortez and Silva (2008) predicted the secondary student grades from two core subjects (Mathematics and Portuguese) using two schools in Portugal as a case study. 29 predictive variables were used, including past school grade, demographics, social and other school related data. The results were obtained using Decision trees, Random Forests, Neural Networks and Support Vector Machine. They achieved high level of predictive accuracy when the past grades were included. In some cases their models included also the school related features, demographics (such as student's age, parent's job and education) and social variables. Kovacic (2012) applied educational data mining technique to identify up to what extent the enrolment data can be used to predict student's success in higher institution. The CHAID and CART algorithms were applied on students' enrolment data of open polytechnics of New Zealand information system to get two decision tree classifying: successful and unsuccessful students. The accuracy obtained with CHAID and CART were 59.4% and 60.5% respectively.

Oyelade, Oladipupo and Obagbuwa (2010) predicted students' academic performance using k-means clustering algorithm and combined with the deterministic model based on the data set of a private institution in Nigeria. A total number of 79 students that offered nine courses each in a semester was sampled and a system for analyzing students' results based clustering analysis and uses of statistical algorithms to arrange their scores data according to the degree of their of performance was demonstrated. Ayinde, Adetunji, Odeniyi and Bello (2013) applied Naïve Bayes and Decision Stump algorithms in predicting the students' final year grade at Ladoke Akintola University of Technology, Ogbomoso. A total of 473 students comparing 10 attributes namely, gender, age, mode of admission, religion, pre-degree score, student matriculation number, state of origin, 200level CGPA, 500level GPA and students' final grade were used. The results of their study divulged that the Naïve Bayes with average precision of 77% performs better than decision stump algorithm with average precision of 57%.

Kuyoro, Goga, Awodele and Okolie (2013) discovered the optimal algorithm for appropriate predicting first-year tertiary students' academic performance based their family background factors and previous academic achievements. Their considered records of 1500 computer science programme students at Babcock University, Nigeria between 2001 and 2010. Ten classification algorithms (five decision tree algorithms -Random forest, Random tree, J48, Decision stump and REPTree, and five rule induction algorithms – JRip, OneR, ZeroR, PART, and Decision table) and a multilayer perception were used in building classification models. The generated models were compared using both 10-fold cross-validation and holdout methods in terms of accuracy level, confusion matrices and execution time. It was revealed that Random tree algorithm outperform other algorithms used in their study.

Akinola, Akinkunmi and Alo (2012) used artificial neural network to develop a data mining model for predicting performance of computer science students in computer programming course taken at 200 level using their pre-qualification ordinary level results and their results obtained in 100 level Mathematics and Physics courses. 200 records of students in Department of Computer Science, University of Ibadan were used for the study and it was discovered that previous background knowledge of mathematics and physics are very vital in order for a student excel in computer programming.

Al-Radaideh, Al-Shawakfa and Al-Najjar (2006) developed classification models using the decision tree algorithms namely, ID3 and C4.5 and the Naïve Bayes classifier, to predict the final grade of students who studied C++ course in Yamouk University, Jordan. Their results showed that decision tree model had better prediction accuracy than the other models.

Ayesha, Mustafa, Sattar and Khan (2010) applied k-means clustering algorithm to predict students' learning activities. The variables used include class quizzes, mid and final exam and assignments. The information generated after the implementation of data mining technique help instructor as well as students to reduce the failure rate by taking necessary steps before time of exam and improve the students' performance. El-Halees (2008) used educational data mining technique to analyze students' learning behavior. The objective of his study is to illustrate how useful data mining can be used in higher education to improve students' performance. He make uses of students' data available in the course database and collected all accessible data including personal records and academic records of students, course records and data generated from e-learning system. Subsequently, he applied data mining techniques such as association rules and classification rules using decision tree to discover many kinds of knowledge. Also he clustered the students into groups using EM clustering, and detected all outliers in the data using outlier analysis. Finally, he provided how we can benefit from the discovered knowledge to improve the students' performance.

Kotsiantis, Pierrakeas and Pintelas (2004) used key demographic variables and assignment marks in the supervised machine learning algorithms (such Decision trees, Artificial Neural Networks, Naïve Bayes Classifier, Instance-based learning, Logistic regression and Support Vector Machines) to predict performance of computer science students at the Hellenic Open University. When only the demographic variables were used the prediction accuracy varied from 58.84% (when using Neural Network) to 64.47% (when using Support Vector Machines). However, when other variables beside demographic were included, the naïve Bayes classifier was found to be the most accurate algorithm for predicting students' performance.

Vandamme, Meskens and Superby, (2007) used Decision trees, Neural Networks and Linear Discriminant Analysis for the early identification of three categories of students: low, medium and high-risk students. The background information (demographics and academic history) of the first-year students in Belgian French-speaking universities were used for the work. The outcome of the experiment revealed that demographics information such as previous education, number of hours spend on mathematics work, financial independence, and age were highly correlated to students' academic performance, while gender, parent's education and occupation, and marital status were not significantly related to the academic success.

Boero, Laureti and Naylor (2005) established that gender is one of the principal determinants of the probability of dropping out. In the Binomial probit model they employed, males have a higher probability of dropping out relative to the reference group of females. They also found that age has a significant positive effect. The variable was entered in a quadratic form to allow the effect of age to have diminishing effect on the dropout probability. With regard to pre-university educational qualifications, the type of school attended had a significant effect on the probability of dropping out. Herrera (2006) concluded that many variables vary in their success at predicting persistence, depending on the academic level. In other words variables that affect persistence at one academic level won't necessarily affect persistence at a different academic level. This means that different models which differentiate between dropout and persistent student should be constructed for each academic level. The same results could be expected at the course levels. That would mean that we would get different probabilities of leaving or staying on the course even for the same student depending upon the course. Yadav, Bhardwaj and Pal (2012) obtained the university students data like attendance, class test, seminar and assignment marks from the students' database, to predict the performance at the end of the semester using three algorithms ID3, C4.5 and CART and establish that CART is the best algorithm for classification of data.

3. METHODOLOGY

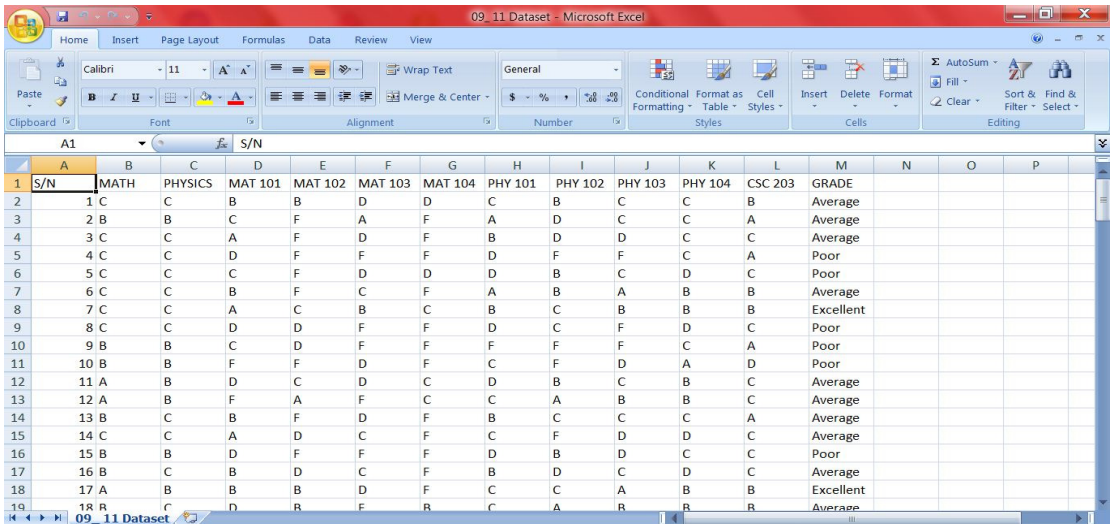
The proposed methodology used in this work for predicting students' performance in computer programming using decision tree algorithms belongs to the process of Knowledge Discovery and Data Mining. The stages in the process include the following:

- **Data gathering:** At this stage information were collected about the students. For doing this, some set of factors that can affect the students' performance must be identified and collected from the students' database in the university and finally integrate all the information into dataset. The data set used in this work was obtained from records of Computer Science students' database kept at the Computer Science Department, Kwara State University, Malete, Nigeria. Data sets obtained were 2009/2010, 2010/2011, 2011/2012, and 2012/2013 sessions. Students' demographics, their scores in Physics and Mathematics pre-university grade, first year Physics ({General Physics PHY 101}, {General physics II PHY 102}, {Experimental Physics I PHY 103}, {Experimental Physics II PHY 104}) and Mathematics ({Elementary Mathematics I MAT 101}, {Elementary Mathematics II MAT 102}, {Elementary Differential and Integral Calculus MAT 103}, {Elementary Algebra and Trigonometry MAT 104}) and as well as their scores in Principle of Programming I (CSC203) at their 200 levels were collected for the data mining study. Table 1 below shows description of students' dataset. For the purpose of this study, a total number of 131 students' data were collected for the Mining Process.
- **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 such as attributes selection and re-balancing of data were employed in order to solve the problems of high dimensionality and imbalanced data that may be present in the dataset.

Table 1: The Students' dataset description table

Variable Name	Description	Variable coding values
Math	Ordinary Level result obtained in Mathematics by the students	6 = A1 5 = B2 4 = B3 3 = C4 2 = C5 1 = C6
Physics	Ordinary Level result obtained in Physics by the students	6 = A1 5 = B2 4 = B3 3 = C4 2 = C5 1 = C6
Math 101 (Elementary Mathematics I)	Grade Point Result obtained in Math 101 at first semester (100 level) by the students	A = 70 – 100% B = 60 – 69% C = 50 – 59% D = 45 – 49% F = 0 – 44%
Math 102 (Elementary Mathematics II)	Grade Point Result obtained in Math 102 at second semester (100 level) by the students	A = 70 – 100% B = 60 – 69% C = 50 – 59% D = 45 – 49% F = 0 – 44%
Math 103 (Elementary Differential& Integral Calculus)	Grade Point Result obtained in Math 103 at first semester (100 level) by the students	A = 70 – 100% B = 60 – 69% C = 50 – 59% D = 45 – 49% F = 0 – 44%
Math 104 (Elementary Algebra & Trigonometry	Grade Point Result obtained in Math 104 at second semester (100 level) by the students	A = 70 – 100% B = 60 – 69% C = 50 – 59% D = 45 – 49% F = 0 – 44%
Physics 101 (General Chemistry)	Grade Point Result obtained in Physics 101 at first semester (100 level) by the students	A = 70 – 100% B = 60 – 69% C = 50 – 59% D = 45 – 49% F = 0 – 44%
Physics 102 (General Physics for Physics Science)	Grade Point Result obtained in Physics 102 at second semester (100 level) by the students	A = 70 – 100% B = 60 – 69% C = 50 – 59% D = 45 – 49% F = 0 – 44%
Physics 103 (Experimental Chemistry I)	Grade Point Result obtained in Physics 102 at first semester (100 level) by the students	A = 70 – 100% B = 60 – 69% C = 50 – 59% D = 45 – 49% F = 0 – 44%
Physics 104 (Experimental Physics)	Grade Point Result obtained in Physics 104 at second semester (100 level) by the students	A = 70 – 100% B = 60 – 69% C = 50 – 59% D = 45 – 49% F = 0 – 44%
Score	Score obtained in CSC 203 (Principle of Programming I) during first semester in their second year (200 level)	A = 70 – 100% B = 60 – 69% C = 50 – 59% D = 45 – 49% F = 0 – 44%

Figure 1 below shows spreadsheet interface of the data set used for training the decision tree algorithms.



S/N	MATH	PHYSICS	MAT 101	MAT 102	MAT 103	MAT 104	PHY 101	PHY 102	PHY 103	PHY 104	CSC 203	GRADE
1	C	C	B	B	D	D	C	B	C	C	B	Average
2	B	B	C	F	A	F	A	D	C	C	A	Average
3	C	C	A	F	D	F	B	D	D	C	C	Average
4	C	C	D	F	F	F	D	F	F	C	A	Poor
5	C	C	C	F	D	D	D	B	C	D	C	Poor
6	C	C	B	F	C	F	A	B	A	B	B	Average
7	C	C	A	C	B	C	B	C	B	B	B	Excellent
8	C	C	D	D	F	F	D	C	F	D	C	Poor
9	B	B	C	D	F	F	F	F	F	C	A	Poor
10	B	B	F	F	D	F	C	F	D	A	D	Poor
11	A	B	D	C	D	C	D	B	C	B	C	Average
12	A	B	F	A	F	C	C	A	B	B	C	Average
13	B	C	B	F	D	F	B	C	C	C	A	Average
14	C	C	A	D	C	F	C	F	D	D	C	Average
15	B	B	D	F	F	F	D	B	D	C	C	Poor
16	B	C	B	D	C	F	B	D	C	D	C	Average
17	A	B	B	B	D	F	C	C	A	B	B	Excellent

Figure 1: Spreadsheet interface of the training data set

- **Data Mining:** In this stage, data mining algorithms are applied in order to predict students' performance. In doing this, classification algorithms such as C4.5 (J48), CART and BF Tree are employed and compared.
- **Interpretation:** At this stage, the obtained models are analyzed to determine the students' performance. This section is done by observing the factors that appeared (in the rules and decision trees) and how they are related for consideration and interpretation.

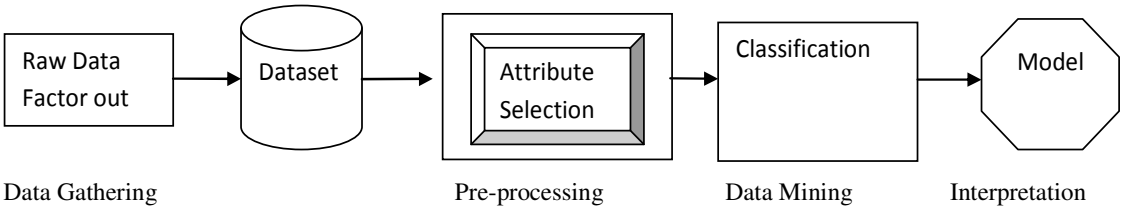


Figure 2: Proposed Methodology of Classification Model

3.1 Experimental Tool Used

The experimental tool used was WEKA. WEKA (Waikato Environment for Knowledge Analysis) is used for classifying data in this work. Weka is one of the popular suites of machine learning software developed at the University of Waikato. It is open source software available under the GNU General Public License. The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality.

3.2 Decision Tree Algorithms

A decision tree is a flow-chart tree structure, where each internal node is denoted by rectangles, and leaf nodes are denoted by ovals. All internal nodes have two or more children node and the internal nodes contain splits, which test the value of an expression of the attributes. Arc from an internal node to its children are labeled with district outcomes of the test. Each leaf node has a class label associated with it.

The decision tree classifier has two phrases:

i. Growth phase or Build phase.
ii. Pruning phase.

The tree is built in the first phase by recursively splitting the training set based on local optimal criteria until all or most of the records belonging to each of the time partitions bearing the same class label. The tree may overfit the data (Han and Kamber, 2006).

The pruning phase handles the problem of overfitting the data in the decision tree. The prune phase generalizes the tree by removing the noise and outliers. The accuracy of the classification increases in the pruning phase. Pruning phase accesses only the fully grown tree. The growth phase requires multiple passes over the training data. The time needed for pruning the decision tree is very less compared to build the decision tree.
The decision tree algorithms used in this work are briefly described as follows:

- **Best-First Decision Trees (BF Tree)**

BF Tree is a decision tree algorithm in which we can evaluate boosting algorithms (Friedman et al., 2000), in order to expand nodes in best-first order instead of a fixed order. This method adds the "best" split node to the tree in each step. The "best" node is the node that maximally reduces impurity among all nodes available for splitting. Although this results in the same fully-grown tree as standard depth-first expansion, it enables us to investigate new tree pruning methods that use cross-validation to select the number of expansions. Both pre-pruning and post-pruning can be performed in this way, which enables a fair comparison between them.

- **C4.5**

C4.5 known as J48 in WEKA is a successor of ID3 developed by Quinlan (1992). It is also based on Hunt's algorithm. J48 handles both categorical and continuous attributes to build a decision tree. In order to handle continuous attributes, J48 splits the attribute values into two partitions based on the selected threshold such that all the values above the threshold as one child and the remaining as another child. It also handles missing attribute values. J48 uses Gain Ratio as an attribute selection measure to build a decision tree. It removes the biasness of information gain when there are many outcome values of an attribute.
At first, calculate the gain ratio of each attribute. The root node will be the attribute whose gain ratio is maximum. J48 uses pessimistic pruning to remove unnecessary branches in the decision tree to improve the accuracy of classification.

$$Gain\ (S,A) = Entropy\ (S) - \sum_{ve\ value(A)}^{a} \frac{|S_v|}{|S|}\ Entropy|S_v| \qquad (1)$$

Where *Values (A)* is the set of all possible values for attribute *A*, and *Sv* is the subset of *S* for which attribute *A* has value *v* (i.e., $S_v = \{s \in S \mid A(s) = v\}$).

The first term in the equation for *Gain* is just the entropy of the original collection *S* and the second term is the expected value of the entropy after S is partitioned using attribute *A*. The expected entropy described by this second term is simply the sum of the entropies of each subset, weighted by the fraction of examples
That belong to *Gain (S, A)* is therefore the expected reduction in entropy caused by knowing the value of attribute A.

$$\text{Gain Ratio(S, A)} = \frac{Gain\ (S,A)}{Split\ Information\ (S,A)} \qquad (2)$$

- **Classification and Regression Tree (CART)**

CART was introduced by Breiman et al. (1984). It is also based on Hunt's algorithm. CART handles both categorical and continuous attributes to build a decision tree. It handles missing values. CART uses Gini Index as an attribute selection measure to build a decision tree. Unlike ID3 and J48 (C4.5) algorithms, CART produces binary splits. Hence, it produces binary trees. Gini Index measure does not use probabilistic assumptions like ID3, C4.5. CART uses cost complexity pruning to remove the unreliable branches from the decision tree to improve the accuracy.

$$\text{Gini Index:} \quad 1 - \sum j\ p^2 j \qquad (3)$$

Gini index of a pure table which consist of single class is zero because the probability is 1 and 1-12 = 0. Similar to Entropy, Gini index also reaches maximum value when all classes in the table have equal probability

4. RESULTS AND DISCUSSION

In this study, three decision tree classification models were proposed for the purpose of predicting students' performance in a computer programming course and performance evaluation of three models were made using both 10-fold cross validation method based classification accuracy and execution time. Figures 3 and 5 depict the decision generated by J48 tree and CART respectively. Also, Figures 4, 6 and 7 shows the WEKA results interface of J48 tree, CART and BF tree respectively.

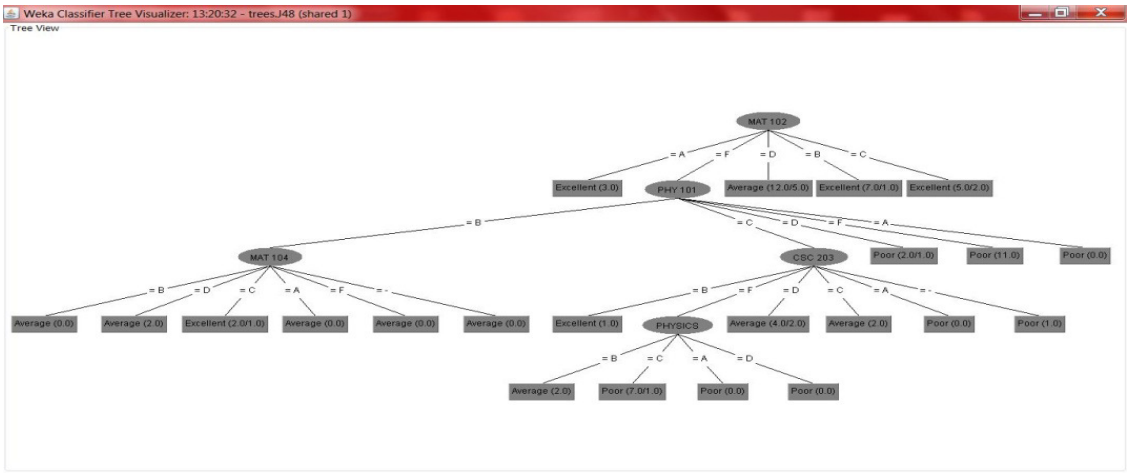


Figure 3: J48 Tree visualization

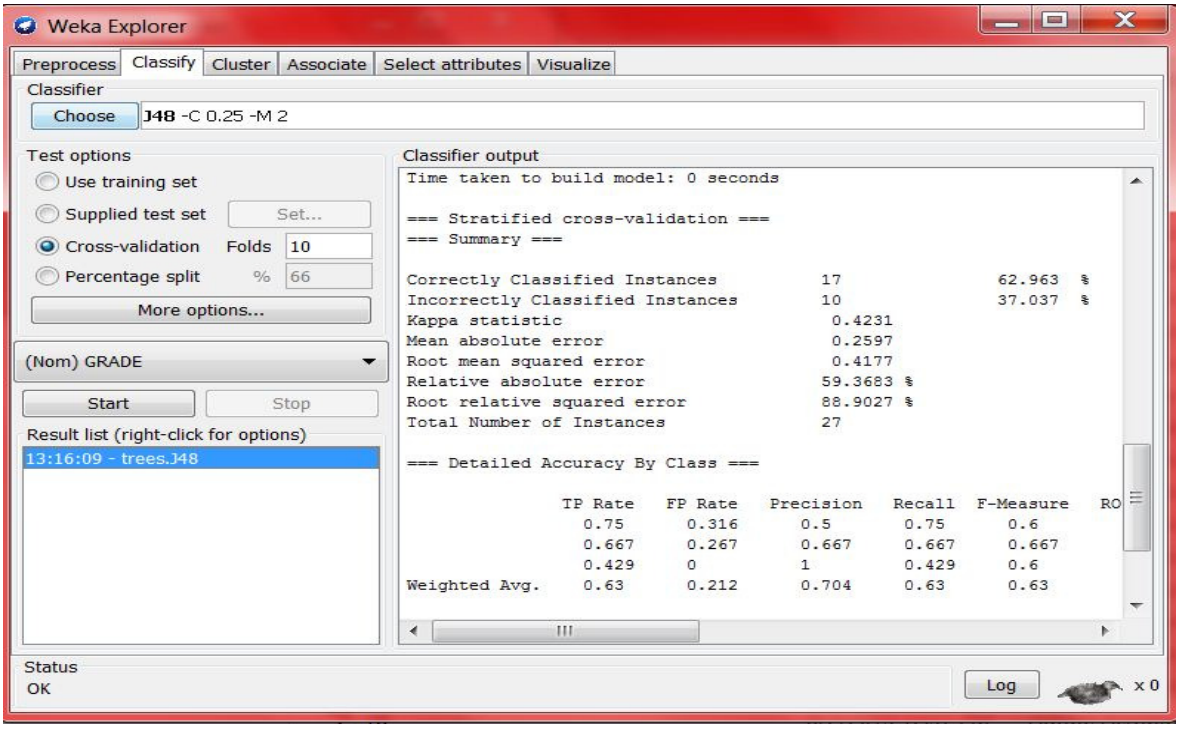


Figure 4: J48 Classifier Result

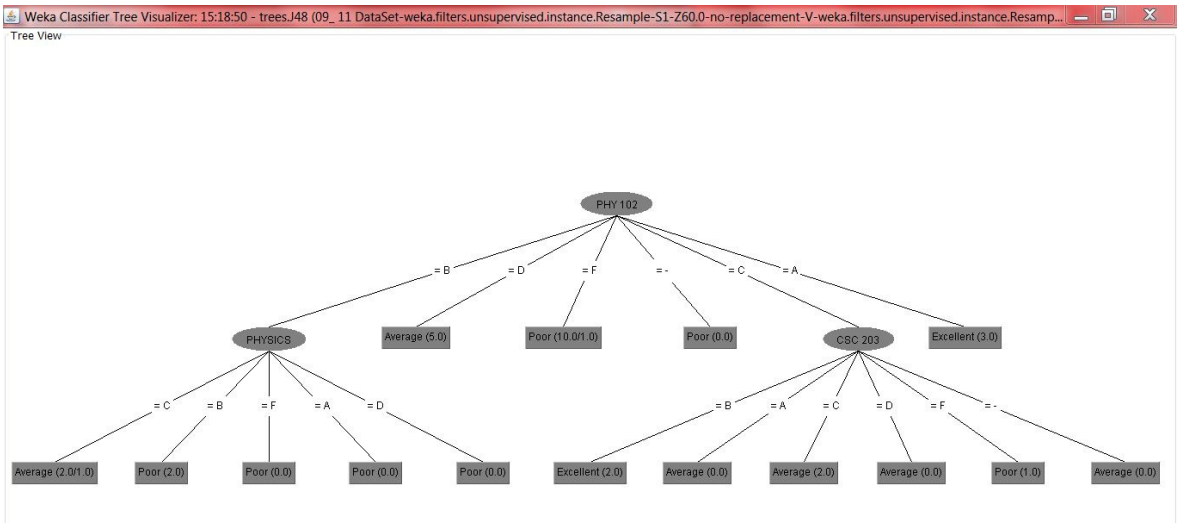


Figure 5: CART Visualization

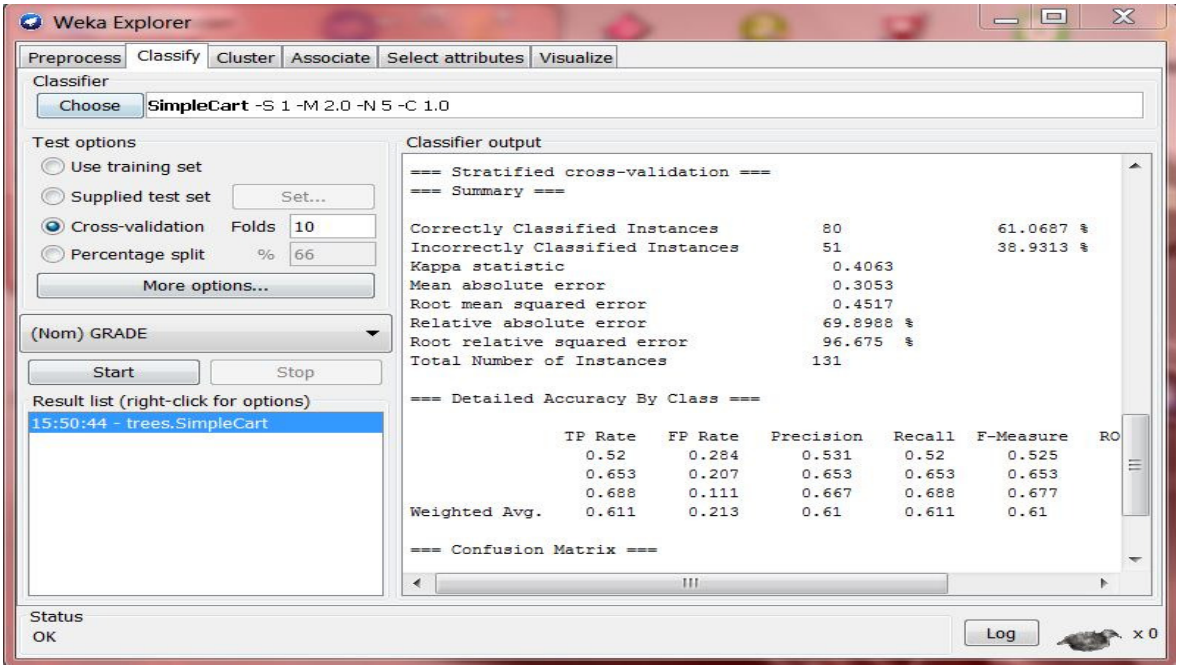


Figure 6: CART Classifier Result

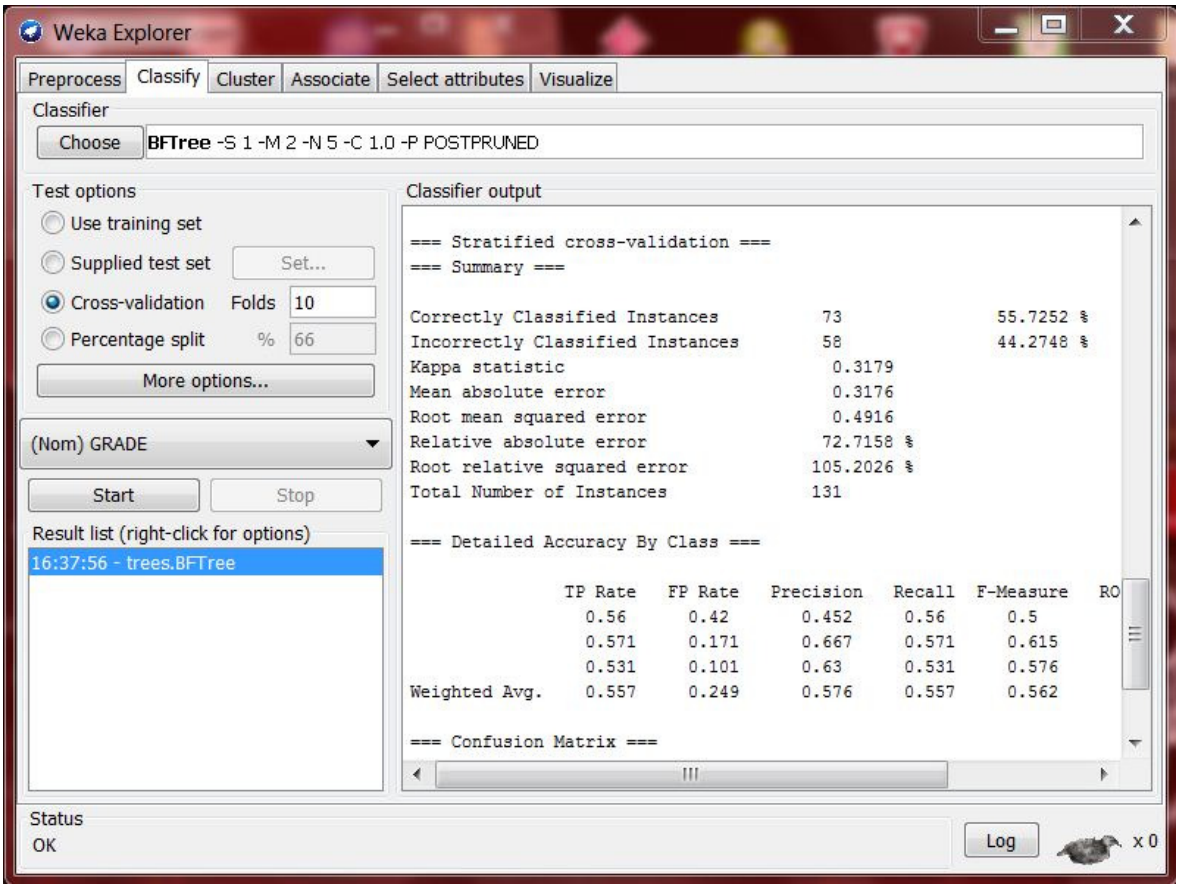


Figure 7: BF Tree Classifier Result

4.1 Experimental Results Obtained

The table 2 shows the accuracy of the J48 tree, CART and BF Tree classifiers for the three generated classification models using 10 – folds cross validation.

Table 2: Accuracy of the Classifier

Algorithm	Correctly Classified Instances	Incorrectly Classified Instances
J48	70.3704%	29.63%
CART	60.4421%	39.56%
BF Tree	60.30534%	39.70%

As shown in table 2, J48 tree has the highest percentage of classification accuracy with 70.37% correctly classified instances followed by CART with 60.44% and BF tree with the least classification accuracy of 60.31%. Hence, it can be deduced that J48 tree is the optimal classifier among the three decision tree algorithms based on the data set used in this study in terms of classification accuracy.

Table 3: Execution Time of the Classifiers

Algorithm	Execution Time (Sec.)
J48	0.02
CART	0.22
BF Tree	0.22

In addition, table 3 shows the execution time (in seconds) used by each classifier when building its model for the training data. J48 tree used the shortest time (0.02 seconds) for classification while CART and BF Tree used the same time (0.22 seconds) for their classification. Therefore, J48 tree also has the least execution time among the three algorithms under investigation.

Taking into consideration the classification accuracy and execution time of the three models, it is found that J48 tree takes very short computational time and outperforms all other two classifiers employed in this study. Figure 8 is the bar chart showing comparison of the three classifiers in terms of classification accuracy.

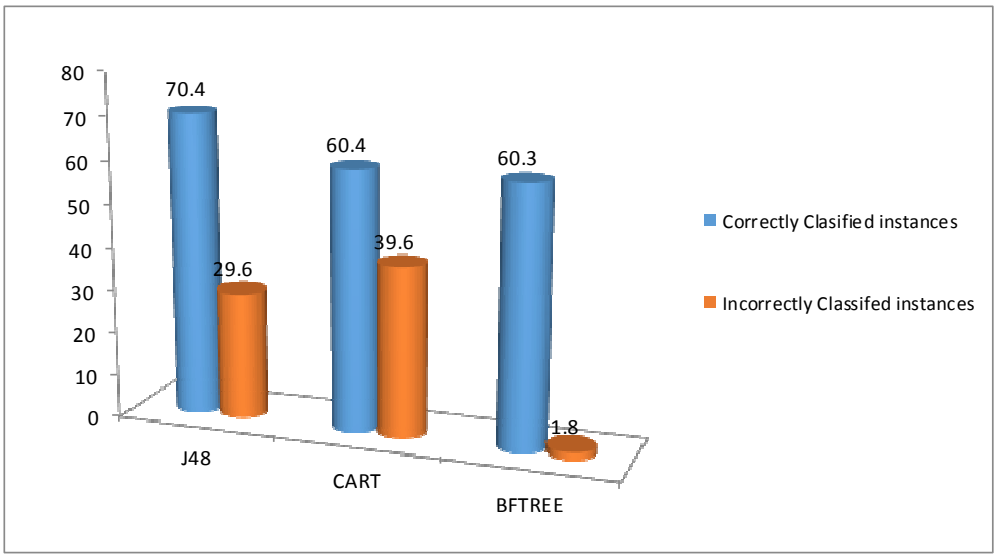


Figure 8: Classification accuracy on 10- folds Cross- validation method

Table 4 below shows the detailed accuracy of each of the classifier in terms of their True positive rate, False Positive rate, Precision, Recall, F- measure and Re-operating characteristic (ROC) area. The results in table 4 further established that J48 tree outperform the two other decision tree algorithms employed in this study. Additional remarks are used below the table in order to make this concept more comprehensive.

Table 4: Detailed Accuracy of Classifiers

Classifier	TP Rate	FP Rate	Precision	Recall	F- measure	ROC Area
J48	0.6	0.395	0.484	0.6	0.536	0.616
CART	0.42	0.272	0.488	0.42	0.452	0.574
BFTREE	0.36	0.284	0.439	0.36	0.396	0.48

TP Rate- True Positive Rate (Proportion of cases correctly classified to the actual class)
FP Rate- False Positive Rate (Proportion of cases belonging to another class misclassified as other class)
Precision- (Positive predictive value determined by (TP/[TP+FP])
Recall- Same as TP Rate
ROC- Receiver Operating Characteristics (graphical display of TPR vs FPR)

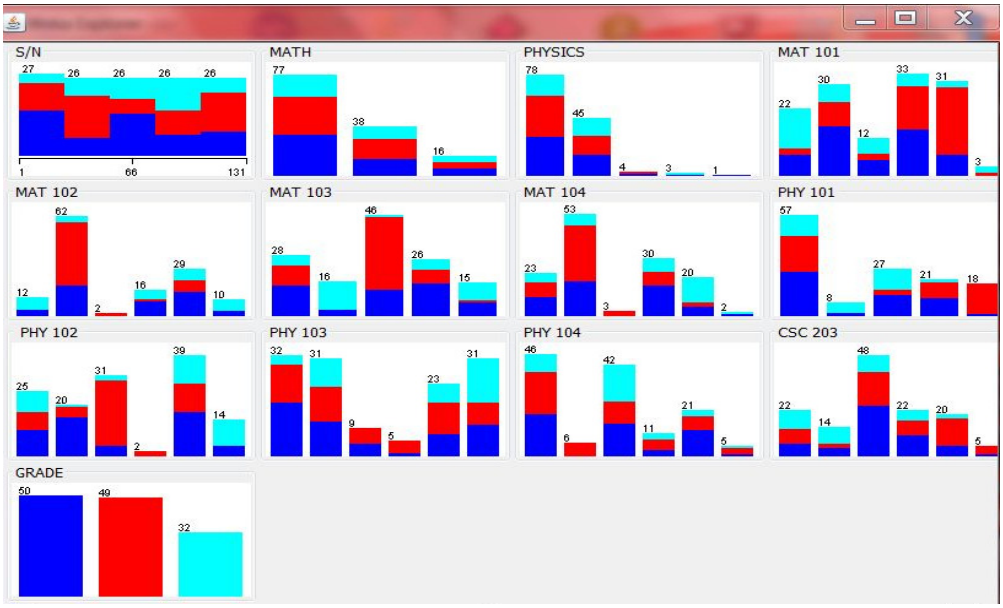


Figure 9: Attributes visualization

Table 5: Attributes ranking using gain ratio

S/N	ATTRIBUTES	RANK	GAIN RATIO
1	MAT 103	6	0.17918
2	MAT 102	5	0.17821
3	PHY 101	8	0.14488
4	PHY 102	9	0.1393
5	MAT 101	4	0.10259
6	PHY 103	10	0.08325
7	MAT 104	7	0.08177
8	PHYSICS	3	0.04105
9	PHY 104	11	0.03549
10	MATH	2	0.00403

Gain ratio is an attribute selection measure which is used to rank attributes in a data set. The attribute with the highest ranked value will be chosen as the root node which forms the decision tree. From table 5, MAT 103 has the highest gain ratio value (0.17918) compared to other attribute, and was selected automatically to be the root node. The rank value is just the position of the attribute based on how the data was arranged.

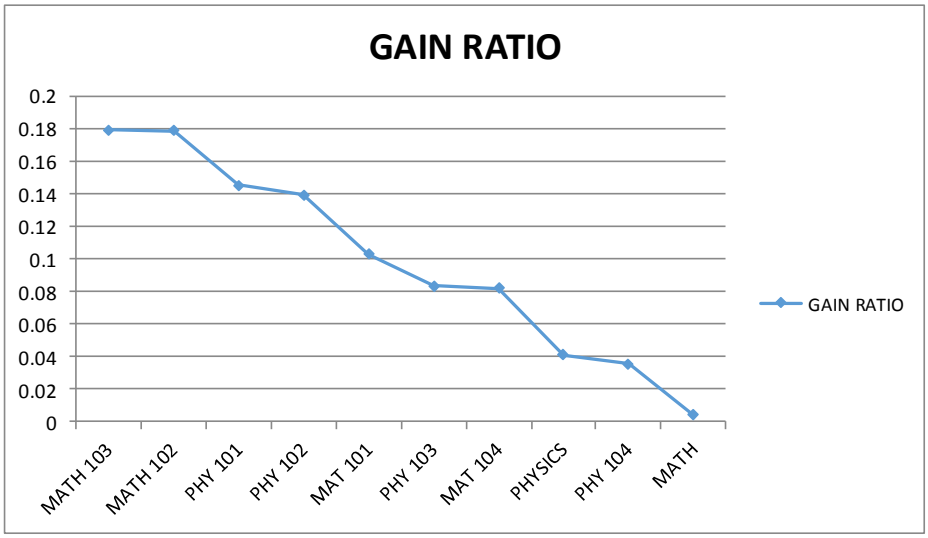


Figure 10: Gain ratio of the attributes

The attribute ranking with respect to the class label using gain ratio criteria shows that students with very good background knowledge in MAT 103, MAT 102, PHY 101, PHY 102 and MAT 101 will perform very well in computer programming courses. These courses are calculation-intensive and they require very sharp brains that can think very fast. Computer programming is one of the courses that involve developing resourceful algorithms and the ability to transform these algorithms to efficient working software. Sometime, these algorithms are mathematically based. The results of this work validate this fact.

5. CONCLUSION

This study employed the use of three decision tree algorithms which are C4.5 (J48 in WEKA), CART and BF Tree in predicting the performance of students in computer programming course based on their ordinary level results in Mathematics and Physics and their 100 level results in Mathematics and Physics courses in order to discover the optimal classification model. It is reveal that J48 tree performs is better than that of the two other decision tree algorithms used in this study. The outcome of this work agrees with Akinola, *et al.* 2012, whose results revealed that prior good background knowledge in mathematics and physics is very vital for students' proficiency in computer programming. However, unlike Akinola, *et al.* 2012 that used artificial neural network which produces black box model, this work employed decision tree which produces white box model that can provide an explanation for the classification result and also be used directly for decision making due to their simplicity and comprehensibility. In this way, a non-expert user of DM such as a teacher can directly use the result obtained by these algorithms to identify students with deficiency in mathematics and physics courses and to make appropriate decisions on how to assist them and thwart their possible failure. Future work will involve using a larger educational data set, incorporating more than one computer programming course as the target class and application of other classification algorithms such as support vector machine, k-nearest neighbor and naïve bayesian classifier for the purpose of performance evaluation.

REFERENCES

1. Abdulsalam S.O., Adewole, K. S., Akintola, A. G. and Hambali, M. A. (2014). Data Mining in Market Basket Transaction: An Association Rule Mining Approach. International Journal of Applied Information Systems (IJAIS), Foundation of Computer Science FCS, New York, USA, 7(10), pp.15 – 20.
2. Akinola O.S., Akinkumni B.O., and Alo, T.S. (2012). A Data Mining Model for Predicting Computer Proficiency of Computer Science Undergraduate Students. African Journal of Computing and ICT, 5(1), pp. 43-52.
3. Ann, C. (1997). Data Mining Staking a Claim on your Privacy, Information and Privacy Commissioner, Ontario, Retrieved 21st June, 2014 from www.ipe.on.ca.
4. Ayesha, S., Mustafa, T., Sattar, A. and Khan, I. (2010). Data Mining Model for Higher Education System, European Journal of Scientific Research, 43(1), pp. 24-29.

5. Ayinde, A.Q, Adetunji A.B., Odeniyi O.A. and Bello M. (2013). Performance Evaluation of Naïve Bayes and Decision Stumps algorithms in mining students' educational data. International Journal of Computer Science Issues, 10(4), pp. 147-151.
6. AI-Radaideh, Q. A., AI-Shawakfa, E. W. and AI-Najjar, M. I. (2006). Mining Student Data Using Decision Trees. International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan.
7. Bhardwaj, B. K. and Pal, (2011). Mining Educational Data to Analyze Students' Performance. International Journal of Advanced Computer Science and Application, 2(6), 63 – 69.
8. Boero, G., Laureti, T., and Naylor, R. (2005). An Econometric Analysis of Student Withdrawal and Progression in Post-Reform Italian Universities. Centro Ricerche Economiche Nord Sud - *CRENoS Working Paper 2005/04*.
9. Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). Classification and Regression Trees. Wadsworth International Group, Belmont, CA.
10. Brijesh Kumar Baradwaj and Saurabh Pal (2011). Mining Educational Data to Analyze Students' Performance, (IJACSA) International Journal of Advanced Computer Science and Applications, 2(6), pp. 63 – 69.
11. Cortez P. and Silva A. (2008). Using Data Mining To Predict Secondary Student Performance. In EUROSIS, A. Brito and J. Teixeira (Eds.), pp.5 -12.
12. El-Halees, A. (2008). Mining Students Data to Analyze Learning Behavior: A Case Study, The 2008 International Arab Conference of Information Technology (ACIT2008) – Conference Proceedings, University of Sfax, Tunisia.
13. Fayyad, M., Piatesky-Shapiro, G., Smyth, P. &Uthurusamy (1996). Advances in knowledge Discovery and data Mining, *AAAI/MIT press.*
14. Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases; AI Magazine, 17; pp.37-54.
15. Friedman J. Hasties T. and Tibshirani R. (2000). Additive Logistic Regression: A Statistical View of Boosting, The Annals of Statistics, 38 (2).
16. Han, J. and Kamber, M. (2006). Data Mining: Concepts and Techniques, Second Edition, Morgan Kaufmann Publishers.
17. Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference and Prediction (2nd ed.), New York: Springer.
18. Herrera, O. L. (2006). Investigation of the Role of Pre and Post-Admission Variables in Undergraduate Institutional Persistence, Using a Markov Student Flow Model. PhD Dissertation, North Carolina State University, USA.
19. Kovacic, Z.J. (2012). Predicting Student Success by Mining Enrolment Data. Research in Higher Education Journal; 15, pp.1-20.
20. Kotsiantis, S., Pierrakeas, C., and Pintelas, P. (2004). Predicting Students' Performance in Distance Learning Using Machine Learning Techniques. Applied Artificial Intelligence, 18, pp.411-426.
21. Kuyoro S.O., Nicolae, G., Oludele, A. and Samuel, O. (2013). Optimal Algorithm for Predicting Students' Academic Performance, International Journal of Computers and Technology, 4(1), pp. 63-75.
22. Oyelade, O.J., Oladipupo, O.O., and Obagbuwa, I.C. (2010). Application of K-Means Clustering Algorithm for Prediction of Students' Academic Performance, International Journal of Computer Science and Information Security, 7(1), pp. 292-295.
23. Osmar, R.Z. (1999). Introduction to Data Mining, CMPUT 690 Principles of Knowledge Discovery in Databases.
24. Quinlan, J.R. (1992). C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, Inc.
25. Quinlan, J.R. (1986). Introduction of decision tree. Journal of Machine learning, pp. 81-106.
26. Romero, C., Ventura, S., and García, E. (2007): Data Mining in Course Management Systems: Moodle Case Study and Tutorial.Computers and Education; 51; 368-384.
27. Vandamme, J., Meskens, N. and Superby J. (2007). Predicting Academic Performance by Data Mining Methods. Education Economics, 15(4), 405-419.
28. Weka – Data Mining Machine Learning Software. Available at: http://www.cs.waikato.ac.nz/ml/weka/, (Accessed 20 April 2014).
29. Yadav, S.K., Bhardwaj B.K., and Pal, S. (2011). Data Mining Applications: A Comparative Study for Predicting Student's Performance, International Journal of Innovative Technology and Creative Engineering (IJITCE), 1(12), pp. 13-19.