# A Comparative Study of Two Data Mining Classification Techniques in the Prediction of Student Performance in Post Unified Tertiary Matriculation Examination (PUTME)

**Chiemeke, S.C & Akinlade, O.A.**
Department of Computer Science
University of Benin
Benin City, Nigeria
E-mails: schiemeke@yahoo.com, akinladetoba@yahoo.com

**ABSTRACT**

Two data mining classification techniques; J48 Decision Tree and Artificial Neural Network Multilayer Perceptron, were used in this research to study trends of student performance in the post unified tertiary matriculation examination (PUTME). Student records for three years of over 100,000 instances were collected from the CRPU unit of the University of Benin, Benin City and were divided into training and test sets. The training set was uploaded to build models and the test to validate the performance of the models. The J48 decision tree had a slightly better performance accuracy of 73% on test data while the multilayer perceptron had a performance accuracy of 72.86%. The decision tree algorithm was a better choice as it also built models faster. Other performance evaluations were also put into consideration that showed both models to actually be good models for prediction.
From this study, it was observed that the decision tree can be a better choice algorithm over the neural network when it involves large amount of data sets for analysis. It builds models faster and presents results in tree constructions that can be easily interpreted and understood by humans.

**Keywords**: Education Data Mining (EDM), Classification, Decision Tree (DT), Multilayer Perceptron (MLP), Post Unified Tertiary Matriculation Examination (PUTME)

**CISDI Journal Reference Format**
Chiemeke, S.C. & Akinlade, O.A. (2017): A Comparative Study of Two Data Mining Classification Techniques in the Prediction of Student Performance in Post Unified Tertiary Matriculation Examination (PUTME). Computing, Information Systems, Development Informatics & Allied Research Journal. Vol 8 No 1. Pp 29-42. Available online at www.cisdijournal.net

## 1. INTRODUCTION

Huge amount of data reside in databases that has led to problems and difficulty in analyzing such huge amount for knowledge extraction and proper decision making, hence the need for a way out of the ordeal. Information technology has made significant improvement in the education system. Nowadays, the universities databases contain so much information about prospective candidates and students. Zhang et al, states that the information has kept increasing rapidly by the times, but there is no action taken to gain knowledge from it. To determine significant relationship among variables in data sets has become slow and subjective (Erdogan et al, 2005). A possible solution that has emerged is Data Mining (DM) which is an information extraction activity whose primary or major goal is to uncover hidden or unknown information not apparent in databases, with huge amount of isolated, structured or unstructured data sets, but potentially useful using automated methods to extract new or previously unknown information. Data mining is a field in computer science that has helped enormously to proffer solutions as it uncovers previously unknown patterns and trends from huge amount of data.

Data mining has then been applied in diverse fields, with the educational sector inclusive and is termed Educational Data Mining (EDM). Data mining classification techniques are used majorly in educational data mining to predict and improve the academic performance of students. From the foregoing, a problem that persists is access to knowledge for accurate prediction of student academic outcome based on trends and patterns of performance so that proactive intervention can be made. There is need to uncover various patterns and study the trend of performance and as well compare classifications algorithms and select the algorithm with best performance accuracy for prediction. The patterns and trends extracted will aid early identification of students who are likely to fail; this insight or knowledge will help put measures in place to improve such student's performance. Azwa et al, (2014) upholds that obtaining patterns in student academic performance will assist the students, lecturers, and university administrators to take appropriate actions to improve the performance of students in various courses or programme of study.

The specific objectives of this research are to:

- Implement classification data mining technique using the J48 Decision Tree (DT) and Artificial Neural Networks Multi Layer Perceptron (MLP) algorithms to analyze data sets.
- Predict the performance of students in the PUTME.
- Comparatively analyze the best classification algorithm on student data using the WEKA tool.

### 1.1 Data Mining Algorithms and Techniques

Data mining is an interdisciplinary subfield in computer science. It is a computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. (Fayyad et al, 1996)

Data mining is about processing data to uncover hidden facts contained in databases, set and transform the extracted information into structures and extract patterns for future knowledge driven decisions. Several core techniques used in data mining describe the type of mining and data recovery operation (Maninder, 2014). Bharati (2010) further states various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc are used for knowledge discovery from databases.

### 1.2 Educational Data Mining

Romero et al, (2008) defines Educational data mining is an emerging sector in the field of data mining, which caters for mining data that reside in databases from the education sector to extract trends and relationships in data sets basically to improve student academic performance. Since its conception, it has drawn the attraction of several researchers as to its relevance and the opportunity to provide answers to diverse unresolved issues regarding improving the academic performance of students and improved planning by the university administration.

Though manual data analysis has been around for some time now, it creates a bottleneck for large data analysis. Baker and Yacef (2009) explains educational data mining is geared towards advancing and improving the traditional educational process through data mining technology for better decision making in the educational system. Educational data mining can help to detect early the problem of low academic performance of students, such weak students can be identified and more measures can be put in place to assist them.

Educational data mining uses many techniques such as Decision Tree (DT), Neural Networks (NN), Naïve Bayes (NB), k-Nearest neighbor, and many others data mining algorithms (Surjeet and Saurabh, 2012).

## 2. RELATED WORKS

Paulo and Alice (2008) carried out a study on secondary school student performance in Mathematics and Portuguese Language classes. Data was obtained from two public schools from the Alentejo region of Portugal during the 2005/2006 school year. The study predicted the student performance using past school grades, demographic, social and other school related data. In the study, the Decision Tree (DT), Random Forest (RF), Neural Network (NN), and Support Vector Machine (SVM) methods were tested on the data set. It was discovered in the study that though achievement is highly affected by previous performances, nevertheless, in some cases, there are other relevant features such as school related (e.g. number of absences, reason of school choice, extra educational school support), demographic (e.g. student age, parent job and education), and social (e.g. going out with friends, alcohol consumption) variables that affect student performance.

Brijesh and Saurabh (2011) conducted a study on student performance on the basis of previous database by selecting an initial sample size of 50 students obtained fromVBS Purvanchal University, Jaunpur (Uttar Pradesh) on the sampling method of Computer Applications Department of Course MCA (Master of Computer Applications) from session 2001 to 2010. The main objective of the study was to predict the performance of students at the end of the semester using the ID3 classification method. The research was able to help improve the division of the students and to identify students which needed special attention to reduce failure ratio and taking appropriate action for the next semester based on information like attendance, class test, seminar and assignment marks from the students' previous semester.

Ramesh et al, (2013) conducted a study to identify the factors influencing the performance of students in final examinations and to find out a suitable data mining algorithm to predict the grade of students so as to give a timely warning to students at risk. A total of 500 records were taken for the analysis. The main objective of the study was to identify highly influencing predictive variables on the academic performance of higher secondary schools. Another major aim was to find the best classification algorithm on student data, and lastly to predict the grade at higher secondary examination. The WEKA bench work was used and the data set was tested with five different classification algorithms namely; Naïve Bayes, Multi Layer Perceptron (MLP), SMO, J48 and REP Tree. It was found out that of all classification algorithms used; the Multi Layer Perceptron (MLP) had a relatively higher accuracy with 72.38% with the J48 Decision Tree following with a performance accuracy of 64.88%.

MLP was considered the most appropriate for predicting student performance because of its higher performance accuracy. The obtained results from hypothesis testing reveals that type of school does not influence student performance and on the other hand, parents occupation plays a major role in predicting grades.

Otobo et al, (2013) conducted a research on performance evaluation of students using data mining techniques and cluster checking. Data was obtained from the admissions office; Delta Pack University of Port Harcourt and result of students were collected from the Department of Computer Science, University of Port Harcourt. The main objective of the research was to examine the performance of students who gained admission into the university through the University Matriculation Examination (UME) and through Basic Studies programs, with the aim of finding out variations in their performance when they graduate from the university. The research was recommended for use in making informed decisions on the ration of admission between Basic and UME mode of intake.

Adeyemo and Kuyoro (2013) conducted a study, using data mining techniques to investigate the effect of socio-economic or family background on the performance of students. A total of 240 records of students were obtained from Babcock University, Nigeria from 2002 to 2009. The decision tree classification method using CART, REPTree, C4.5, Ripple Down Rule Learner (RIDOR), SeeRule, SeeTree, BoostSeeTree algorithms on the WEKA workbench was employed. CART was said to have the lowest accuracy of 0.40 while Boost Decision Tree had the best accuracy level. The Boosted Decision Tree was then used to design a framework for Student Academic Performance. The study identified parental conditions such as parent's education and occupation, marital status, parent's socio-economic background as influencing socio-economic factors affecting performance of students.

Osofisan et al, (2014) conducted a research on an empirical study of decision tree and artificial neural network algorithm for mining educational database. The data comprised of 411 records of M.Sc. students from the department of computer science, University of Ibadan. The distribution was divided into training and test set where 288 instances served as training set and 123 instances served as test set. The aim of the research was to use data mining techniques to study student performance to discover appropriate knowledge and extract useful patterns from existing stored data of students. The Multilayer Perceptron (MLP) and Decision Tree classification algorithms were used. It was observed from the study that MLP had a superior performance of 98.3% over decision tree 85.4%. In the research, the Multilayer Perceptron (MLP) gave the best classification results as well as prediction capability.

Raheela et al, (2015) conducted a study on predicting performance of students at the end of a university degree at an early stage of the degree program to help the university focus on bright students and assist students with low academic achievement. The data for the study was obtained from four academic batches of Computer Science and Information Technology (CSIT) department at NED University, Pakistan comprising altogether 347 undergraduate students. The RapidMiner tool was used and the decision tree, naïve bayes, and neural networks classification algorithm were used for mining of the data. It was observed from the study that it is possible to predict with reasonable accuracy the performance of students with no socio-economic or demographic features.

Fadhilah et al, (2015) proposed a framework for predicting students' academic performance of first year Bachelor students in Computer Science course. The data was collected from an 8 year period intake from July 2006/2007 until July 2013/2014 that contains the students' demographics, previous academic records and family background information. The study conducted a comparative analysis of the Decision Tree (DT), Rule Based (RB), and Naïve Bayes (NB) classification techniques using the WEKA tool to discover the best technique to develop a predictive model for student academic performance. The research results showed that the Rule Based has the best classification accuracy of 71.3% compared to Naïve Bayes (67%) and Decision Tree (68.8%).

## 3. METHODOLOGY

In this study, the KDD process model which is an iterative process with the following steps namely, data collection, selection, preprocessing, transformation, data mining, interpretation and knowledge (Fayaad et al, 2006) was adopted and used appropriately as relating to data from the educational sector.

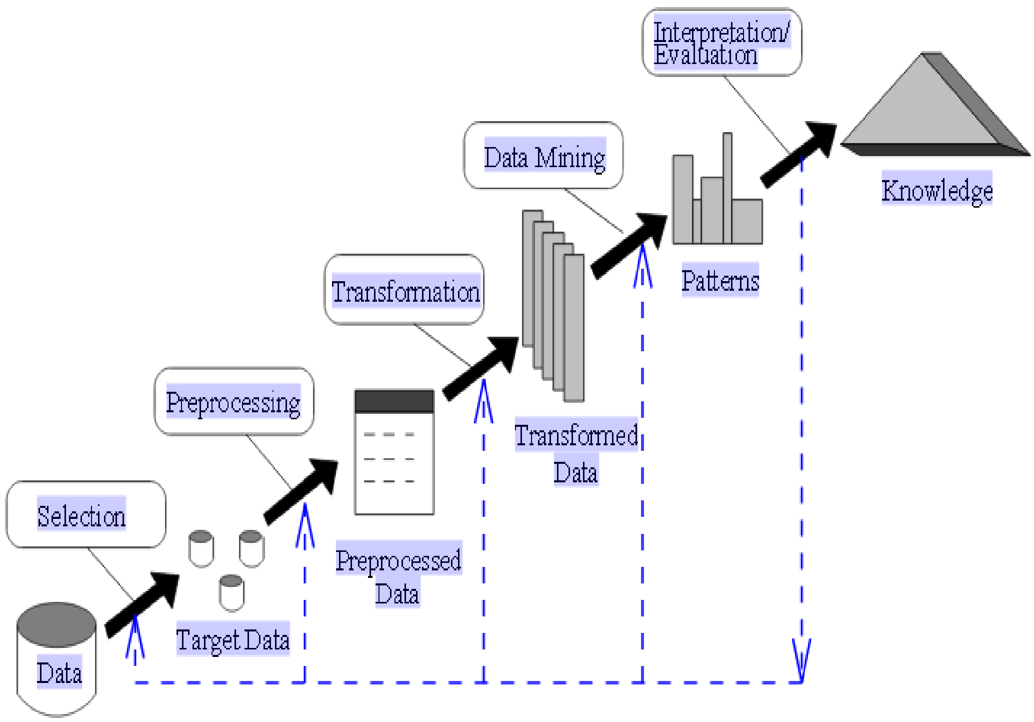


**Figure 1: A Typical Knowledge Discovery Process**

### 3.1 Data Collection

The PUTME student records of the University of Benin, Benin City; from the 2009/2010 academic year until 2014/2015 intakes were obtained for the study. The data were collected from the Central Record Processing Unit (CRPU) database of the University.

### *3.2 Data Selection*

The target data is created from the data set available and necessary attributes are selected for the study. This stage is crucial as the data mining algorithms learn and build model based on available data. Attributes obtained include jamb registration number, sex, state of origin, faculty, course of study, jamb score, and post utme score. It is important to remove or ignore input features with little effect on the output so as to keep the size of the model small, avoid over fitting and reduce computational cost. The data sets for this research was subjected to two attribute selection techniques to determine attributes that with most relevance in the prediction process. The two attribute selection techniques used are the GainRatio AttributeEval and the InfoGainAttributeEval. Both attribute selection techniques ranked the jamb score highest followed by the faculty code, state of origin and sex as the least relevant a shown in Figure 2.

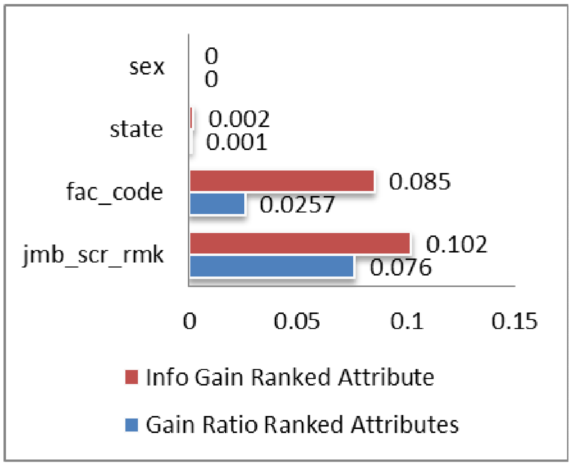


**Figure 2**: **A Chart Showing the Attribute Selection Evaluation Results**

### 3.3 Data Preprocessing

This is a stage in knowledge discovery that guarantees the reliability of the data to be used for analysis and the model built. This stage caters for eliminating imbalance and noise in data, inconsistency and data attributes of no interest. Major tasks in data preprocessing is data cleaning, data integration, data transformation, data discretization. Incomplete and inconsistent data records with instances such as applicants that failed to apply for the PUTME exam, blurred pictures, impersonation issues, absent from screening, wrong color background, no picture, wrong passport were all eliminated to cleanse the data and make the data set consistent. In all of the attributes selected and defined, the jamb registration number and course of study were screened out as they were of no interest because they presented personal information, as it is with the jamb registration number to ensure privacy, and had large variance of possible outcome.

Data discretization which describes a process of reducing possible values of attributes, especially for numeric data was performed on the PUTME and jamb scores. The jamb score was labeled by discretion; good (200-230), v. good (231-250) or excellent (251-400) based on the scores obtained by candidates. The target class to be predicted, the remarks which describes the PUTME score, was also by discretion labeled failure (0-54) and success (55-100). The student score range 0-54 describes students that failed and at risk of failing, and 55-100 to describe students with good standing. The state of origin also was labeled based on the six geopolitical zones in Nigeria to minimize the number of possible outcome. Table 1 shows a tabular representation of attributes, descriptions and possible values.

On completion of cleansing, the data sets were joined into a single worksheet and stored in the Microsoft Office Excel worksheet. The data set was then converted to a comma-delimited (.csv) data file and then finally converted to the Attribute Relation File Format (.arff) which is a format well understood by the toolkit chosen for mining.

### 3.4 Pattern Extraction

The two selected algorithms of choice for this study are the J48 Decision Tree (DT) and Artificial Neural Network Multi Layer Perceptron (MLP). In this stage, the data set distribution is divided into training and test sets. The training set is to build the model from the classification algorithms, and the test set to validate the performance and accuracy of the model. The models built by the algorithms which shows patterns and trends in data sets will be discussed and presented as knowledge. The two algorithms will be compared based on performance evaluation to find out the algorithm with the best accuracy on this data set domain. The WEKA toolkit is used to analyze the data.

**Table 1: Attribute Description and Possible Values**

| ATTRIBUTE | DESCRIPTION | DOMAIN |
|---|---|---|
| Sex | The sex of the candidates, either male or female | {M (Male), F(Female)} |
| State Of Origin | The state of origin provided by the candidates. The states are labeled based on the six geopolitical zones in Nigeria | {NC (north central), NE(north east), NW (north west), SE (south east), SS (south south), SW(south west)} |
| Faculty Code | The faculty under which the course applied for by the candidate is. | {Art, Psc, Lsc, Eng, Mgs, Agr, Bms, Den, Edu, Law, Med, Ssc, Pha} |
| Jmb Scr Rmk | The score obtained by the candidate in the examination | {Good (200-230), V. Good (231-250), Excellent (251-400)} |
| Remarks (Response Variable) | This describes the score obtained by the candidate in the post UTME. | {Failure (0-54), Success (55-100)} |

## 4. RESULTS AND DISCUSSION

The WEKA machine learning toolkit version 3.6.13 was used for analysis on student PUTME records. The confusion matrix is a table that presents the number of correctly and incorrectly classified instances of the actual and predicted class instances on the data set distribution. The Precision, TP rate, also referred to as Recall, are derived from the values given in the confusion matrix. The F-measure presents a proportion of the balance between the recall and precision of the algorithm for each class.

Tables 2, 3, and 4 present the output, confusion matrix and class performance of MultiLayer Perceptron algorithm on training and test sets respectively.

**Table 2:** Output of MLP on training and test set

| PARAMETERS | OUTPUT | |
|---|---|---|
| | **Training** | **Test** |
| **Time taken to build model** | 797.18 seconds | 438.96 seconds |
| **Correctly Classified Instances** | 57103 (73.1452%) | 29309 (72.8753%) |
| **Incorrectly Classified Instances** | 20965 (26.8548%) | 10909 (27.1247%) |
| **Kappa Statistics** | 0.3994 | 0.3952 |
| **Mean Absolute Value** | 0.3634 | 0.3637 |
| **Root Mean Squared Error** | 0.4248 | 0.4253 |
| **Relative Absolute Error** | 76.1057% | 76.1358% |
| **Root Relative Squared Error** | 86.9463% | 86.997% |
| **Total Number of Instances** | 78068 | 40218 |

**Table 3: Confusion Matrix for MLP on training and test set**

| | | | |
|---|---|---|---|
| **Training Set** | A | B | ← Classified as |
| | 42408 | 4917 | A = FAIL |
| | 16048 | 14695 | B = SUCCESS |
| **Test Set** | A | B | ← Classified as |
| | 21697 | 2636 | A = FAIL |
| | 8273 | 7612 | B = SUCCESS |

**Table 4: Class Performance of MLP on training set and test set**

| | | **TP Rate** | **FP Rate** | **Precision** | **Recall** | **F-Measure** | **ROC Area** | **Class** |
|---|---|---|---|---|---|---|---|---|
| **Training Set** | | 0.896 | 0.522 | 0.725 | 0.896 | 0.802 | 0.786 | FAIL |
| | | 0.478 | 0.104 | 0.749 | 0.478 | 0.584 | 0.786 | SUCCESS |
| | Weighted Average | 0.731 | 0.357 | 0.735 | 0.731 | 0.716 | 0.786 | |
| **Test Set** | | **TP Rate** | **FP Rate** | **Precision** | **Recall** | **F-Measure** | **ROC Area** | **Class** |
| | | 0.892 | 0.521 | 0.724 | 0.892 | 0.799 | 0.786 | FAIL |
| | | 0.479 | 0.108 | 0.743 | 0.479 | 0.583 | 0.786 | SUCCESS |
| | Weighted Average | 0.729 | 0.358 | 0.731 | 0.729 | 0.714 | 0.786 | |

A graphical representation of the performance accuracy of both algorithms is shown in Figure 3. It is observed that with the large data sets for this study, both algorithms have good prediction accuracy but with the J48 decision Tree (DT) classification algorithm with a slightly, better performance accuracy (73%) over the neural network MultiLayer Perceptron classification algorithm performance accuracy (72.86%), and also builds prediction models at a faster speed. This performance accuracy of the J48 decision tree obtained shows that on large data sets, the decision tree had slightly better performance.

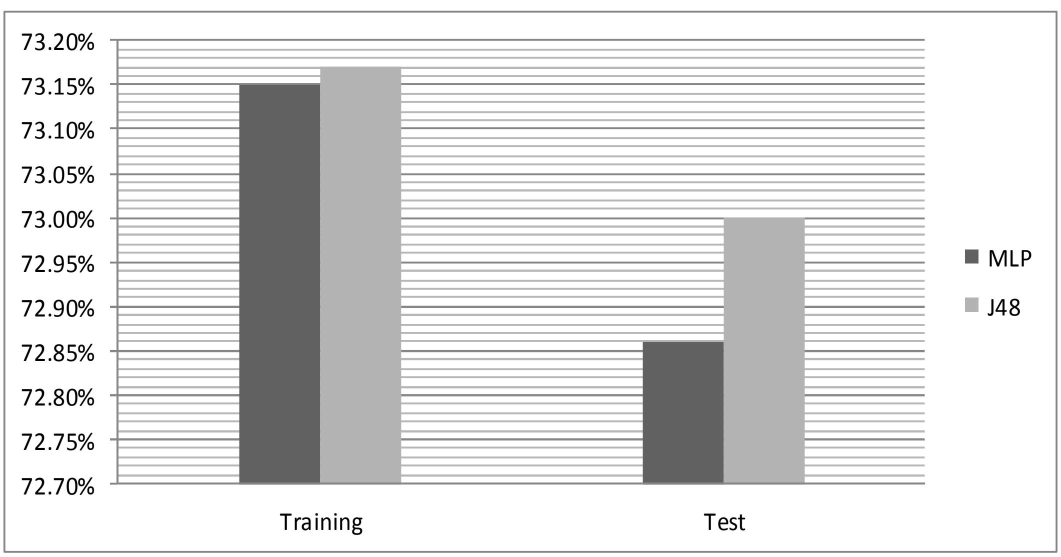


**Figure 3:** Graphical representation of Performance Accuracy for both algorithms

The output of the J48 decision tree algorithm can also be visualized as a tree construction as shown in Figure 4. This is a major advantage of the decision tree as it presents results such that it can be easily understood.

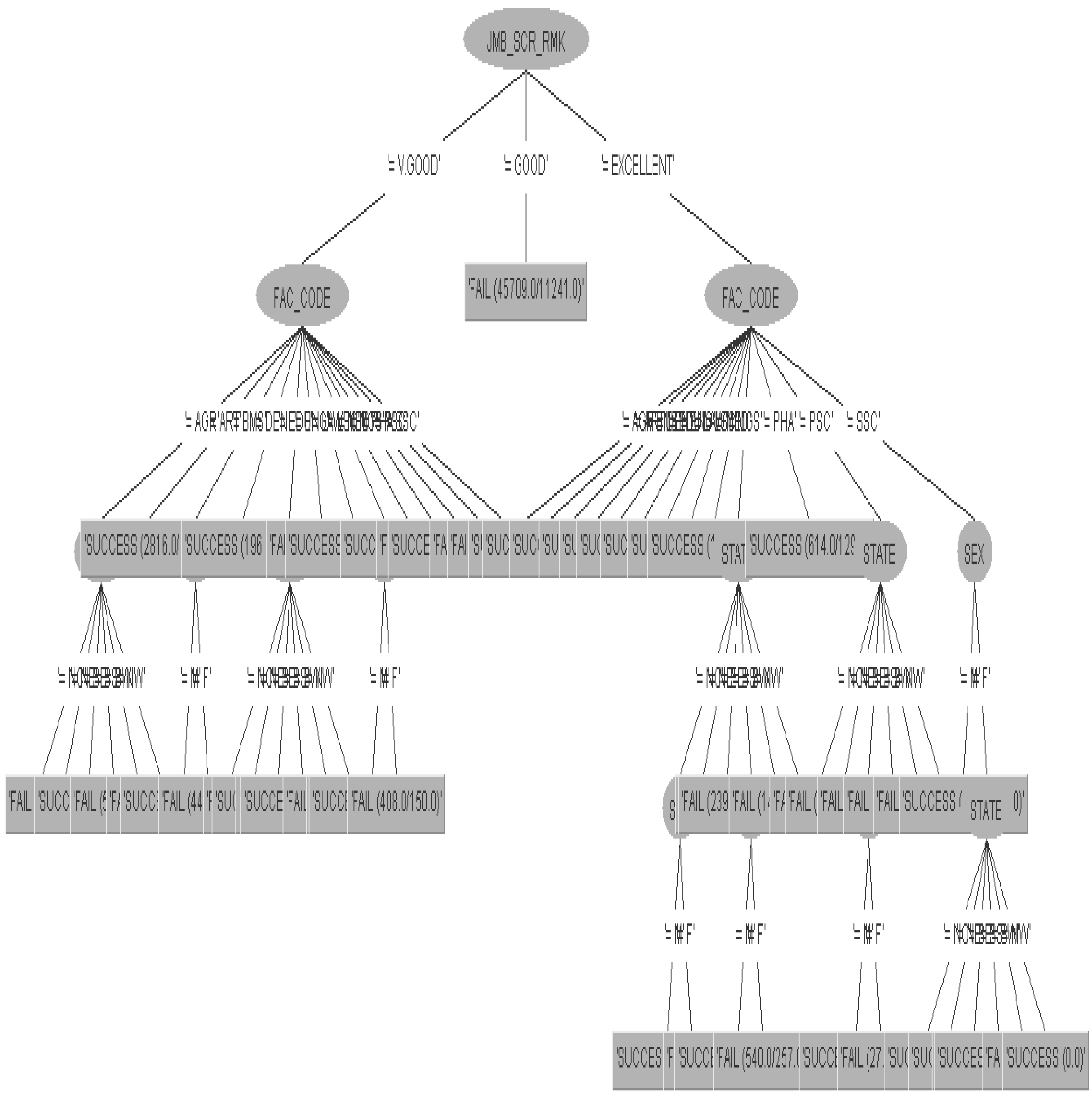


**Figure 4: Output Of The J48 Decision Tree Algorithm Presented As A Tree Construction.**

Tables 5, 6, and 7 present the output, confusion matrix and class performance on J48 Decision Tree algorithm on training and test data set respectively.

**Table 5: Output Of J48 Decision Tree On Training And Test Set**

| PARAMETERS | OUTPUT | |
|---|---|---|
| | **Training** | **Test** |
| **Time taken to build model** | 0.71 seconds | 0.68 seconds |
| **Correctly Classified Instances** | 57124 (73.1721%) | 29359 (72.9997%) |
| **Incorrectly Classified Instances** | 20944 (26.8279%) | 10859 (27.00037%) |
| **Kappa Statistics** | 0.4012 | 0.399 |
| **Mean Absolute Value** | 0.3815 | 0.3816 |
| **Root Mean Squared Error** | 0.4367 | 0.4371 |
| **Relative Absolute Error** | 79.8984% | 79.8836% |
| **Root Relative Squared Error** | 89.386% | 89.416% |
| **Total Number of Instances** | 78068 | 40218 |

**Table 6:** Confusion Matrix of J48 decision tree on training and test set

| | | | |
|---|---|---|---|
| **Training Set** | A | B | ← Classified as |
| | 42255 | 5070 | A = FAIL |
| | 15874 | 14869 | B = SUCCESS |
| **Test Set** | A | B | ← Classified as |
| | 21646 | 2687 | A = FAIL |
| | 8172 | 7713 | B = SUCCESS |

**Table 7: Class Performance Of J48 Decision Tree On Training Set And Test Set**

| | | TP Rate | FP Rate | Precision | Recall | F-Measure | ROC Area | Class |
|---|---|---|---|---|---|---|---|---|
| **Training Set** | | 0.893 | 0.516 | 0.727 | 0.893 | 0.801 | 0.724 | FAIL |
| | | 0.484 | 0.107 | 0.746 | 0.484 | 0.587 | 0.724 | SUCCESS |
| | Weighted Average | 0.732 | 0.355 | 0.734 | 0.732 | 0.717 | 0.724 | |
| **Test Set** | | **TP Rate** | **FP Rate** | **Precision** | **Recall** | **F-Measure** | **ROC Area** | **Class** |
| | | 0.89 | 0.514 | 0.726 | 0.89 | 0.799 | 0.724 | FAIL |
| | | 0.486 | 0.11 | 0.742 | 0.486 | 0.587 | 0.724 | SUCCESS |
| | Weighted Average | 0.73 | 0.355 | 0.732 | 0.73 | 0.715 | 0.724 | |

IF-THEN rules are used to extract predictions for better understanding. Some rules extracted are shown in Figure 5.

1. IF Jmb_Scr_Rmk = Good THEN Remarks = Fail
2. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Art THEN Remarks = Success
3. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Eng THEN Remarks = Fail
4. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Med THEN Remarks = Success
5. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Mgs THEN Remarks = Fail
6. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Pha THEN Remarks = Success
7. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Psc THEN Remarks = Fail
8. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Ssc THEN Remarks = Fail
9. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Edu and State= nc THEN Remarks = Fail
10. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Edu and State= nw THEN Remarks = Success
11. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Agr and State= se THEN Remarks = Success
12. IF Jmb_Scr_Rmk = V. Good and Fac_Code = Agr and State= ss THEN Remarks = Fail
13. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Agr THEN Remarks = Success
14. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Bms THEN Remarks = Success
15. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Den THEN Remarks = Success
16. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Edu THEN Remarks = Success
17. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Law THEN Remarks = Success
18. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Eng THEN Remarks = Success
19. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Lsc THEN Remarks = Success
20. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Med THEN Remarks = Success
21. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Pha THEN Remarks = Success
22. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Mgs and State= se THEN Remarks = Fail
23. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Mgs and State= sw THEN Remarks = Fail
24. IF Jmb_Scr_Rmk = Excellent and Fac_Code = Mgs and State= nw THEN Remarks = Fail

.

.**Figure 5:** IF-THEN rule extracts from the tree construction by the J48 decision tree algorithm.

Table 8 presents a comparison on the output of the J48 and MLP algorithms. The J48 decision tree built prediction models faster at 0.68 seconds and had a performance accuracy of 73% while the neural network MultiLayer Perceptron built prediction models at 438.96 seconds, which is slower than the decision tree, and obtained a performance accuracy of 72.86%.

**Table 8: Comparative Analysis Of J48 And MLP Algorithms Based On Output**

| | **Training Set** | | **Test Set** | |
|---|---|---|---|---|
| **Parameters** | **MLP** | **J48** | **MLP** | **J48** |
| Time taken to build the model | 797.18 seconds | 0.71 seconds | 438.96 seconds | 0.68 seconds |
| Correctly Classified Instances | 57103 (73.15%) | 57124 (73.17%) | 29309 (72.86%) | 29359 (73%) |
| Incorrectly Classified Instances | 20965 (26.85%) | 20944 (26.83%) | 10909 (27.12%) | 10859 (27%) |
| Kappa Statistics | 0.3994 | 0.4012 | 0.3952 | 0.3990 |
| Mean Absolute Error | 0.3634 | 0.3815 | 0.3637 | 0.3816 |
| Root Mean Squared Error | 0.4248 | 0.4367 | 0.4253 | 0.4371 |
| Relative Absolute Error | 76.1057% | 79.8984% | 76.1358% | 79.8836% |
| Root Relative Squared Error | 86.9463% | 89.386% | 86.997% | 89.416% |
| Total Number of Instances | 78068 | 40218 | 78068 | 40218 |

The slight difference in the Mean Absolute Error and Relative Mean Squared Error values of each algorithm on the training and test sets as presented in Table 8 show that both models are good for prediction.

The recall and precision values obtained by both algorithms as presented in Table 9 show that prediction is successful for the fail and success class categories of the remarks attribute which represents the predicted performance of applicants in the post UTME.

**Table 9:** Comparative analysis of J48 and MLP algorithms based on recall and precision

| **Classifiers** | **MLP** | | | **J48** | | |
|---|---|---|---|---|---|---|
| | Success | Failure | Weighted Average | Success | Failure | Weighted Average |
| Recall | 0.478 | 0.896 | 0.731 | 0.484 | 0.893 | 0.732 |
| Precision | 0.749 | 0.725 | 0.735 | 0.746 | 0.727 | 0.734 |

## 5. FINDINGS

It was observed from this study that the decision tree and neural network have good performance accuracy on large data sets but the decision tree had slightly better performance accuracy as data sets were quite large. The decision tree as well built the prediction models faster and presented output in a pattern that can be easily interpreted and understood by humans.
The results of this study will help to identify students that are susceptible or at high risk of failing so that appropriate measures can be taken early enough to improve performance.

## 6. CONCLUSION

This research was able to study trends in student performance in PUTME, and further build models that can be used for prediction of student performance in the examination. By inference from this research study, the decision tree and artificial neural network are both suitable algorithms for mining huge amount of data set but with a preference for decision tree classification algorithm.

## 7. FUTURE WORKS

Educational Data Mining (EDM) is an emerging field, as such, more research work can be done on other forms of data set obtained from the education sector and other classification algorithms can be employed for mining and prediction.

**REFERENCES**

1. Zhang Y., Oussena S., Clark T., and Kim H. "Use Data Mining to Improve Student Retention in Higher Education – A Case Study" ICEIS - 12th International Conference on Enterprise Information Systems 2010, (2010).http://dx.doi.org/10.5220/0002894101900197
2. Erdogan Z. and Timor M. (2005) "A Data Mining Application in a Student Database" Journal of Aeronautics and Space Technologies, Vol. 2 No.2 (53-57)
3. Azwa A. A., Nor H. I., and Fadhilah A. (2014) "First Semester Computer Science Students' Academic Performances Analysis by Using Data Mining Classification Algorithms" Proceeding of the International Conference on Artificial Intelligence and Computer Science (AICS 2014), INDONESIA (e-ISBN978-967-11768-8-7).
4. Fayyad U., Piatetsky-Shapiro G., and Padhraic S. (1996) "From data mining to knowledge discovery in databases", American Association for Artificial Intelligence/The MIT press
5. Maninder S. (2014) "A Review on Data Mining Algorithms" International Journal of Computer Science and Information Technology Research, ISSN 2348-1196, Vol.2, Issue 2, pp: (8-14).
6. Bharati M. R. (2010) "Data Mining Techniques and Applications" Indian Journal of Computer Science and Engineering, Vol.1 No.4 pg 301-305 ISSN: 0976-5166.
7. Romero C., Ventura S. and Garcia E. (2008) "Data mining in course management systems: Moodle case study and tutorial" Computers and Education Vol.51, No.1,pp. 368-384.
8. Baker R. and Yacef K. (2009) "The State of Educational Data Mining: A review and future visions" Journal of Educational data mining.
9. Surjeet K. Y. and Saurabh P. (2012) "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification" World of Computer Science and Information Technology Journal (WCSIT) ISSN: 2221-0741, Vol.2, No 2, 51-56
10. Paulo C. and Alice S. (2008) "Using Data Mining to Predict Secondary School Student Performance", In EUROSIS, A. Brito and J. Teixeira (Eds.), pp. 5-12
11. Brijesh K. B. and Saurabh P. (2011) "Mining Educational Data to Analyze Students' Performance" International Journal of Advanced Computer Science and Applications, Vol.2, No.6
12. Ramesh V., Parkavi P., and Ramar K. (2013) "Predicting Student Performance: A Statistical and Data Mining Approach" International Journal of Computer Applications (0975-8887), vol.63, No. 8
13. Otobo F. N., Baah B., and Taylor O. E. (2013) "Evaluation of Student Performance Using Data Mining over a Given Data Space" International Journal of Recent Technology and Engineering (IJRTE), ISSN: 22-77-3878, vol.2, issue 4.
14. Adeyemo A. B. and Kuyoro S.O (2013) "Investigating the Effect of Students Socio-Economic/Family Background on Students' Academic Performance in Tertiary Institutions Using Decision Tree Algorithm" Journal of Life and Physical Sciences actaSATECH 4(2): 61-78.
15. Osofisan A.O, Adeyemo O.O, and Oluwasusi S.T. (2014) "Empirical study of Decision Tree and Artificial Neural Network Algorithm for Mining Educational Database" African Journal of Computing and ICT Vol.7 no.2 ISSN 2006-1781
16. Raheela A., Agathe M., and Mahmood K. P. (2015) "Predicting Student Academic Performance at Degree Level: A Case Study". I.J. Intelligent Systems and Applications, 49-61. Modern Education and Computer Science Press
17. Fadhilah A., Nur H. I., and Azwa A. A. (2015) "The Prediction of Students' Academic Performance Using Classification Data Mining Techniques" Applied Mathematics Sciences, Vol.9, No. 129, 6415-6426