University of Ibadan
Journal of Science and Logics in ICT Research

Predicting Postgraduate Students' Performance using Decision Tree Algorithms

Adeyemo, A. B. **Awokoya, E. A.**
sesanadeyemo@gmail.com ayodeleawokoya@gmail.com

Department of Computer Science,
University of Ibadan, Ibadan, Nigeria

Abstract

In this study, a data mining model that predicts postgraduate students' performance using decision tree algorithms was developed. Postgraduate student data collected from the postgraduate school, University of Ibadan and a case study department were pre-processed adequately. Seven different feature selection techniques in Waikato Environment for Knowledge Analysis (WEKA) were used to determine the major attributes that contribute to the prediction of Postgraduate students' performance. The highest ranked attributes were used for the analysis using RandomTree, RepTree andJ48 decision tree algorithms in WEKA. The work was evaluated using the AUROC performance metrics for the major classes of interest. Results obtained gave insight into the optimal algorithm for the analysis and rules that could predict postgraduate students' performance were generated from the developed model.

Keywords: Educational data mining, Student performance prediction, Class imbalance problem

I. INTRODUCTION

Education is the best legacy one can have, is a regular saying. Nowadays in this part of the world, having only undergraduate degree is not sufficient to say one is adequately educated. Therefore the quest for a postgraduate degree has become paramount in the mind of both parents and their wards. To this end, different higher institutes of learning now have various courses they offer at the postgraduate level and the performance of students is a key role to determining the success and sustenance of the institutions. The quality of education is measured by the academic performance of students and the results produced [20]. Students, being the key assets of any institute of learning receive more attention in order to enhance the quality of the institute. Due to this fact, institutions of learning keep track of their students' academic performance which serves as evidence of their achievement over their years of existence. The ability to predict students' performance is very important in educational environments because it plays an important role in producing the best quality graduates and postgraduates who will become great leaders [15].

Academic institutions are increasingly required to monitor their performance and the performance of their students. This gives rise to need to collate, analyze and interpret data in order to have evidence to inform academic polices that are aimed at, for example, improving student retention rate, allocating teaching and support resources, or creating intervention strategies to mitigate factors that may affect student performance adversely [3]. There are a number of reasons for these: (1) Tertiary education should aim to maximize the potential of each student. Therefore, a careful examination of student outcomes against some benchmark or expected outcome may provide evidence as to whether student potential is being realized. Such insights may also help the institutions to prioritize scarce resources, to focus them on specific problem areas, (2) Institutions have an obligation to deliver value for money to the bodies that fund them, (3) Institutions are often judged by the quality of the awards they provide; for example the more honours level graduate a course provides, the better the course is perceived to be. This provides additional incentives for institutions to take proactive steps to support students [3]. According to [17] there is need for deep and enough knowledge for better assessment, evaluation, planning and decision making.

II. LITERATURE REVIEW

Educational Data Mining (EDM) is concerned with "developing, researching and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of

Adeyemo, A. B. and Awokoya, E. A. (2017). Predicting Postgraduate Students' Performance using Decision Tree Algorithms, *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR)*, Vol. 1, pp. 15 - 24

data within which they exist" [18]. Reference [6] identified four goals of EDM as:

1. Predicting Student's future learning behaviour: with the use of student modelling, this goal can be achieved by creating student models that incorporate the learners' characteristics, including detailed information such as their knowledge, behaviours and motivation to learn. The user experience of the learner and their overall satisfaction with learning are also measured.

2. Discovering or improving domain models through the various methods and application of EDM, discovery of new and improving to existing models is possible. Examples include illustrating the educational content to engage learners and determining optimal instructional sequence to support the students learning style.

3. Studying the effects of educational support data can be achieved through learning systems.

4. Advancing scientific knowledge about learning and learners by building and incorporating student models into the field of EDM research and the technology and software used.

The academic performance of students is a key factor in making institutions stand out and it also determines how students enroll into the school. In order to give quality education, which leads to good performance from students, the tutors need to know the academic ability of students they are dealing with.

Several research works have been carried out on students' academic performance which gives insight into the cause of differences among student's performance. Some of these are by [2], [16], [1], [11], [7], [5], [4], [13], [3], [15], [14], [12], [9], [20], and [10]. Many studies included a wide range of potential predictors, which include personality factors, intelligence and aptitude tests, academic achievements, previous college achievements, demographic data, socio-economic/family background and lots more. While some of these factors seemed to be stronger than others, there is no consistent agreement among the different studies. However, all studies show that academic success is dependent on many factors where grades and achievements, personality and expectation, as well as sociological background all play significant roles [3].

It has been observed from research that several data mining methods and statistical methods have been used by scholars and researchers to attempt looking into the social status, economical status and educational background of students in order to predict their academic performance and also understand the factors that influence their academic performance. These researches have been carried out majorly on primary, secondary and undergraduate students.

In this study, the prediction of postgraduate student's academic performance was based on their previous academic achievements and their personal details, this was influenced by the data available at the postgraduate school. This prediction can thus help tutors and decision makers in institutions make strategic decisions with regards to these students. This study first identified the set of attributes that contribute towards postgraduate student's academic performance and then developed a model using decision tree algorithm to predict the student's academic performance hence generating prediction rules from the developed model.

III MATERIALS AND METHODS

The dataset used for the study was obtained from the Information and Communication Technology unit of the Postgraduate School, University of Ibadan and the records office of Computer Science department, University of Ibadan. The Computer Science department has a very productive and successful postgraduate program which is over-subscribed with an annual intake of about a hundred students per session (from amongst about 500 applicants annually). The data collected initially for the study had about 2000 records from the academic M.Sc. Computer Science programme from the year 2009-2014. The attributes of the dataset were: Name of Students, Name of University Attended for Undergraduate, Course Studied during Undergraduate, Class of Degree from Undergraduate, Marital Status, Sex of Student, Date of Birth, Name of Sponsor, Session of Admission, Matriculation Number, O/L result in Mathematics, O/L result in English Language, O/L result in Physics, O/L result in Chemistry and Class of Degree after completing master's program.

After preprocessing the data to remove noisy data which will have no impact in the classification of attributes and generation of rules, such as "Name of Students", "Matriculation number" because names of student and matriculation number are unique and a student's name cannot influence his/her grade, and duplicated data records, the dataset was reduced to two hundred and fifty three (253).

In the process of developing the model it was discovered that the data set was highly skewed (imbalanced) and after the application of a novel hybrid of oversampling and undersampling method called SMOTE+ENN [8] which was used to resolve the class imbalance problem, the dataset was reduced to one hundred and twenty-eight (128). Some of the data collected cannot be used by the data mining software except they were transformed into formats that can be used by the data mining tool. For example, the attribute 'Name of University Attended for Undergraduate' was transformed into 'Type of School' while its original data which was made of different school names were categorized into Private, Federal and State schools. The field 'Date of Birth' was transformed into 'Age' and its initial data which was of the format day/month/year of birth were transformed into age in numbers. The 'Name of Sponsor' field which initially contained different sponsor's name were transformed to 'Sponsor' which described the different type of sponsor and the data contained in it were transformed into Parent/Guardian, Self, Fellowship/Scholarship. The Mathematics, English, Physics and Chemistry Ordinary Level (O/L) results were transformed and categorized into Distinction, Credit, and Pass with respect to the students' grade in the exam. The data file was then converted from a Comma Separated Value (CSV) file format to an Attribute Relation File Format (ARFF) which is used by WEKA data mining software. A summarized description of the data set is presented in Table 1.

The data selection phase of the research which involved an understanding of the available data and selecting the attributes which will produce the necessary data set needed to infer the knowledge been sought for was then carried out.

Attributes contributing more to the development of the 'best' model (in relation to the collected data) were derived using Correlation Feature Subset evaluator alongside BestFirst and GreedyStepwise search methods, CorrelationAttributeEvaluator (CO), GainRatioAttributeEvaluator (GR), InfoGainAttributeEvaluator (IG), SymmetricalUncertAttributeEvaluator (SU) and ReliefFAttributeEvaluator (RF).

Table 2 presents a summary of the attributes and how they were ranked by each of the evaluators and their respective search methods. The derived data subset was hence selected from the entire dataset and used for the data mining task.

Table 1: Summarized description of the data set

Variables	Description	Possible Values
Type of School	Nominal	{SU,FU,PU}
Age on Admission	Numeric	Between 20 and 60
Previous UG Grade	Nominal	{First Class, Second Class Upper, Second Class Lower}
Course Studied	Nominal	{Computer Science, Computer Engineering, Physics with Electronics, Electrical Electronic Engineering, Mathematics/ Computer Science, Agricultural Engineering, Computer Education, Petroleum Engineering, Computer Information System, Telecom Science, Computer Electronics}
Session Resumed	Nominal	{2008/09, 2009/10, 2010/11, 2011/12, 2012/13, 2013/14}
Sponsor	Nominal	{Parent, Self, Scholarship/Fellowship}
Sex	Nominal	{Male, Female}
Marital Status	Nominal	{Single, Married}
O/L Mathematics	Nominal	{Distinction, Credit, Pass}
O/L English	Nominal	{Distinction ,Credit, pass}
O/L Chemistry	Nominal	{Distinction ,Credit, pass}
O/L Physics	Nominal	{Distinction ,Credit, pass}
Grade	Nominal	{PHD, MPHIL/PHD, MPHIL, TERMINAL, Non Graduate (NG)}

Table 2: Summary of evaluator's ranking of each attribute

Attributes	SU	IG	RF	GR	CO	CFS+BF	CFS+GSW
Type of School	0.03632	0.075	0.06	0.053	0.147	30%	20%
Age on Admission	0.05539	0.059	0.008	0.089	0.176	30%	30%
Previous UG Grade	0.05879	0.105	0.065	0.086	0.154	100%	100%
Course Studied	0.06822	0.111	0.013	0.099	0.062	100%	100%
Session Resumed	0.10502	0.248	0.091	0.098	0.119	100%	100%
Sponsor	0.02441	0.042	0.022	0.043	0.076	0%	0%
Sex	0.00312	0.008	0.01	0.008	0.041	0%	0%
Marital Status	0.2267	0.035	0.014	0.047	0.144	0%	0%
O/L Mathematics	0.01392	0.025	0.027	0.024	0.053	0%	0%
O/L English	0.01873	0.032	0.008	0.039	0.07	0%	10%
O/L Physics	0.04105	0.068	0.037	0.064	0.092	90%	80%
O/L Chemistry	0.2063	0.039	0..031	0.035	0.055	0%	0%

The first six highest ranked attributes by most of the evaluators as best influencing the grade of postgraduate students are: Session Resumed, Previous UG Grade, Type of School, O/L Physics, Age on Admission and Course Studied. These best influencing attributes were selected from the loaded dataset and used by the WEKA data mining tool. The algorithms used for the study were: J48, RepTree and RandomTree algorithms.

IV RESULTS AND DISCUSSION

A. Descriptive Analysis of the Dataset

Figure 1 presents a chart which summarizes the entire dataset. From the chart it can be seen that for the class label (Grade) 47% of the students had PHD grade at the end of their program, 29% had MPHIL/PHD grade, 13% had MPHIL grade, 5% had Terminal and 5% did not graduate, that is, dropped out. Students from state university dominate the dataset with a percentage of 44% while the students from federal university end their program more with PHD grade than students from other type of institutions while most of the students from state university have MPHIL/PHD grade. Also, most of the students with TERMINAL grade are from state university while federal university students constitute the highest percentage of dropouts.

Students within the age range 23-32 are pre-dominant in the dataset. Students within the age range 20-23 constitute 10% of the entire dataset and they had either PHD grade or MPHIL/PHD at the end of their program. 51% of the students had second class upper division in their previous course of study at undergraduate level. Most of those with second class upper also turn out to have PHD grade at the end of their master program. 4.7% of the students had first class at the end of their undergraduate program and all the first class students also had PHD grades at the end of their master program. The previous course studied that dominated the dataset is Computer Science with 80% followed by Computer Engineering which is about 10% of the dataset.

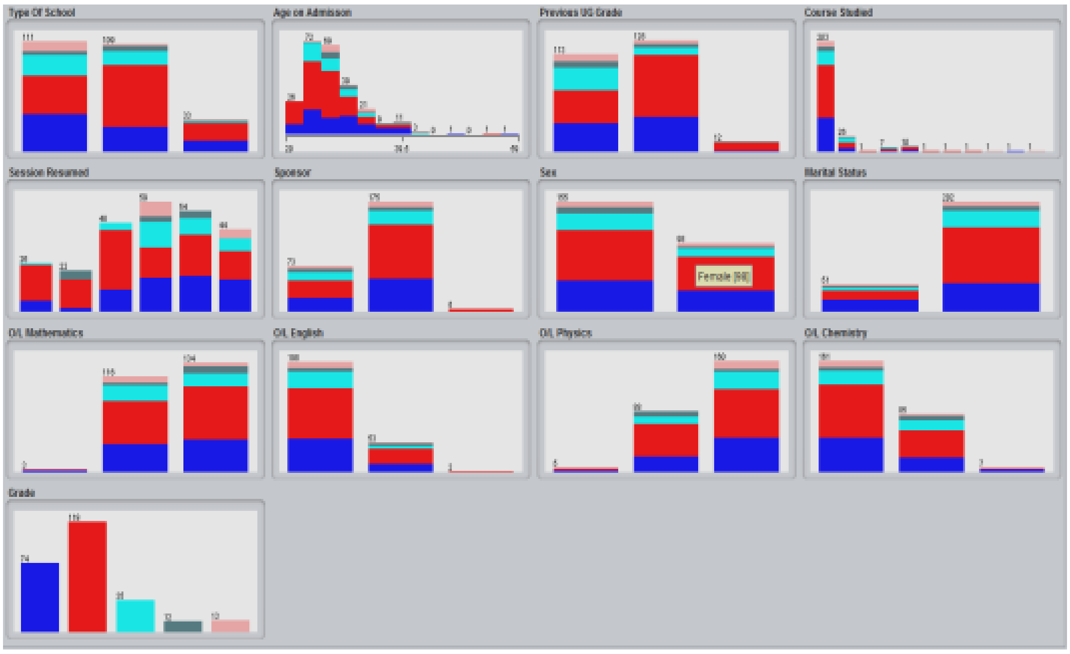

Figure 1: Summarized Chart of the dataset.

The session with the most instance in the dataset is the 2011/12 session making about 23% of the instances. About 69% of the students in the dataset were sponsored by their parents and less than 1% of the students in the dataset were sponsored through fellowship/scholarship.

All students sponsored through fellowship/scholarship had PHD grade at the end of their program. The dataset consists of more male than female with the male contributing to about 61% of the total instance of the dataset. Single students contribute to about 78% of the dataset. 53% of the students had distinction in their O/L mathematics examination. Less than 2% of the students had pass in their O/L mathematics and 66% of them had MPHIL/PHD at the end of their master program. About 78% of the students had credit in their O/L English language examination while less than 1% had Pass in the same subject. The entire students with pass had PHD grade at the end of their master program. About 63% of the students had credit in their O/L Physics examination while less than 2% of the entire students sampled had pass in the same subject. Those who had pass in their O/L physics had either MPHIL/PHD or PHD at the end of their master program. 64% of the students had credit in their O/L Chemistry examination while less than 3% of the students had pass in the same subject.

The dataset used in this study has class label of ratio 10:1 for PhD grade to Non-Graduating and almost of the same ratio for PhD grade to Terminal grade students. This type of dataset is called an imbalanced dataset [8]. For an imbalanced dataset, as the one used for this study, Accuracy, Precision rate and Recall rate of a model is not a good measurement because these measures are derived from TP, TN, FP, FN which can be highly influenced by the dominating class label in the dataset. AUC and AUPRC (Area Under Precision-Recall Curve) provide more robust and better performance estimates when comparing classifiers on imbalanced dataset. Therefore the classifiers used in this study were compared based on AUROC and AUPRC.

Figure 2 presents the Weighted Average of AUC and AUPRC based on the generated models:

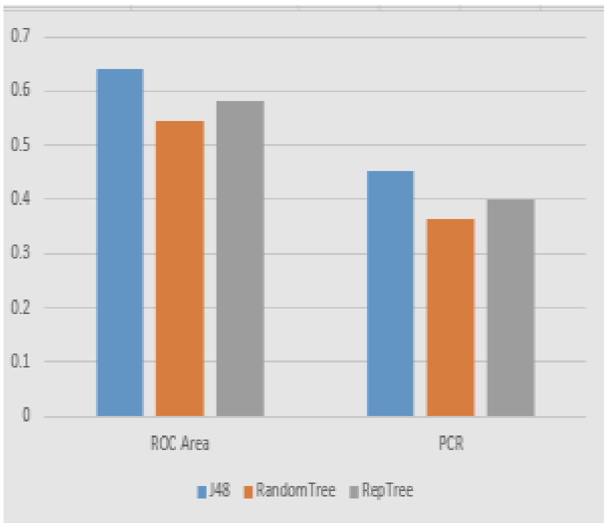

Figure 2: Chart visualizing Performance Measures in relation to the classifiers

It can be seen from figure 2 that the J48 classifier performed relatively better than other classifiers, yet it cannot concluded that the J48 model was a good predictor of postgraduate students' performance because its AUC was just a little above 0.5 which indicates that the model had only predicted a little above chance and its AUPRC is less than 0.5 which is not acceptable for a good model. Hence there was a need to further pre-process the data to reduce the class imbalance effect on the dataset. This was carried out using the Synthetic Minority Over-sampling Technique and Edited Neareast Neighbour (SMOTE+ENN) method which is a hybrid of SMOTE an Oversampling technique and ENN an Undersampling technique [8]. The SMOTE+ENN which had been implemented as a WEKA extension was applied on the imbalanced data. Figure 3 presents the summarized chart for the modified dataset.

The attributes that can best predict postgraduate students' performance were selected from the modified dataset using Correlation Feature Subset evaluator alongside BestFirst and GreedyStepwise search methods, CorrelationAttributeEvaluator (CO), GainRatioAttributeEvaluator (GR), InfoGainAttributeEvaluator (IG), SymmetricalUncertAttributeEvaluator (SU) and ReliefFAttributeEvaluator (RF). Table 3 presents the detailed description the ranking of the attributes in order of importance for predicting postgraduate students' performance.

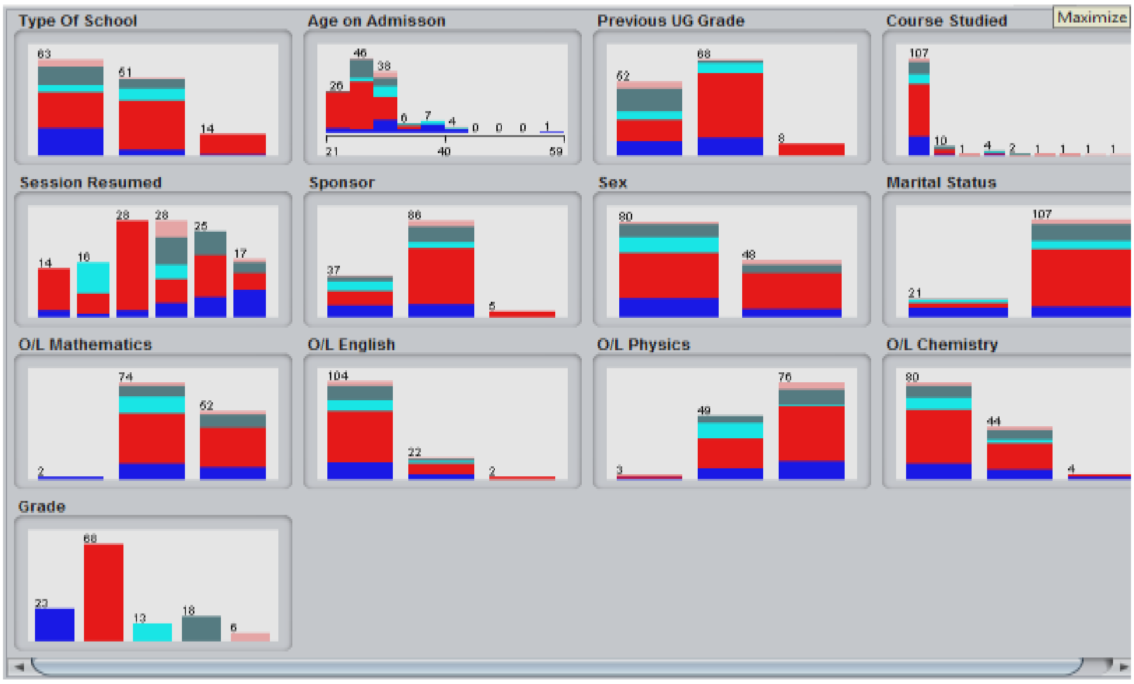

Figure 3: Summarized Chart for the modified dataset.

Table 3: Summary of evaluator's ranking of each attribute for the modified dataset

Attribute	SU	IG	RF	GR	CO	CFS+BF	CFS+GSW
Type of School	0.112	0.184	0.126	0.133	0.223	100%	80%
Age on Admission	0.172	0.239	0.045	0.3	0.352	100%	100%
Previous UG Grade	0.156	0.243	0.185	0.193	0.256	100%	100%
Course Studied	0.129	0.187	0.028	0.184	0.08	60%	40%
Session Resumed	0.267	0.586	0.233	0.233	0.203	100%	100%
Sponsor	0.081	0.120	0.073	0.111	0.171	0%	0%
Sex	0.072	0.102	0.087	0.106	0.139	80%	80%
Marital Status	0.128	0.161	0.079	0.249	0.268	70%	40%
O/L Mathematics	0.097	0.143	0.089	0.131	0.099	20%	10%
O/L English	0.018	0.024	0	0.032	0.039	0%	0%
O/L Physics	0.109	0.116	0.114	0.147	0.116	100%	90%
O/L Chemistry	0.031	0.046	0.031	0.041	0.069	0%	0%

From the Table 3 it can be observed that the first seven highest ranked attributes by most of the evaluators as best influencing the grade of postgraduate students are: Previous UG Grade, Session Resumed, Age on Admission, Type of School, O/L Physics, Marital Status and Sex. These attributes were then used in the model development phase using the modified dataset. The models were compared using the AUC value for each model under the class labels NG and TERMINAL alongside other performance measures. Figure 4 presents the comparison of AUC for each of the classifiers with respect to class labels TERMINAL and NG while figure 5 presents a chart showing Kappa Statistics, F-Measure, Precision and Recall for each classifier used on the modified dataset.

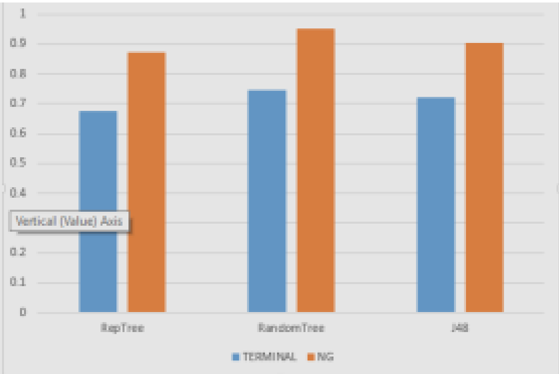

Figure 4: Chart showing comparison of AUC for the most important class labels

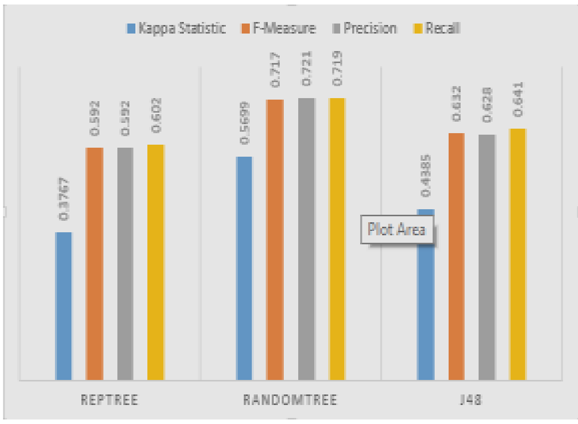

Figure 5: Chart showing Kappa Statistics, F-Measure, Precision and Recall for each classifier used on the modified dataset.

From figures 4 and 5, it can be observed that the model with the best AUC values and whose precision value was higher than its recall value and whose kappa statistics value and F-measure produced the highest value was the RandomTree classifier. Therefore, the RandomTree algorithm produced the best model for the study.

The decision tree constructed by RandomTree classifier is presented in figure 6 and some of the rules obtained are presented.

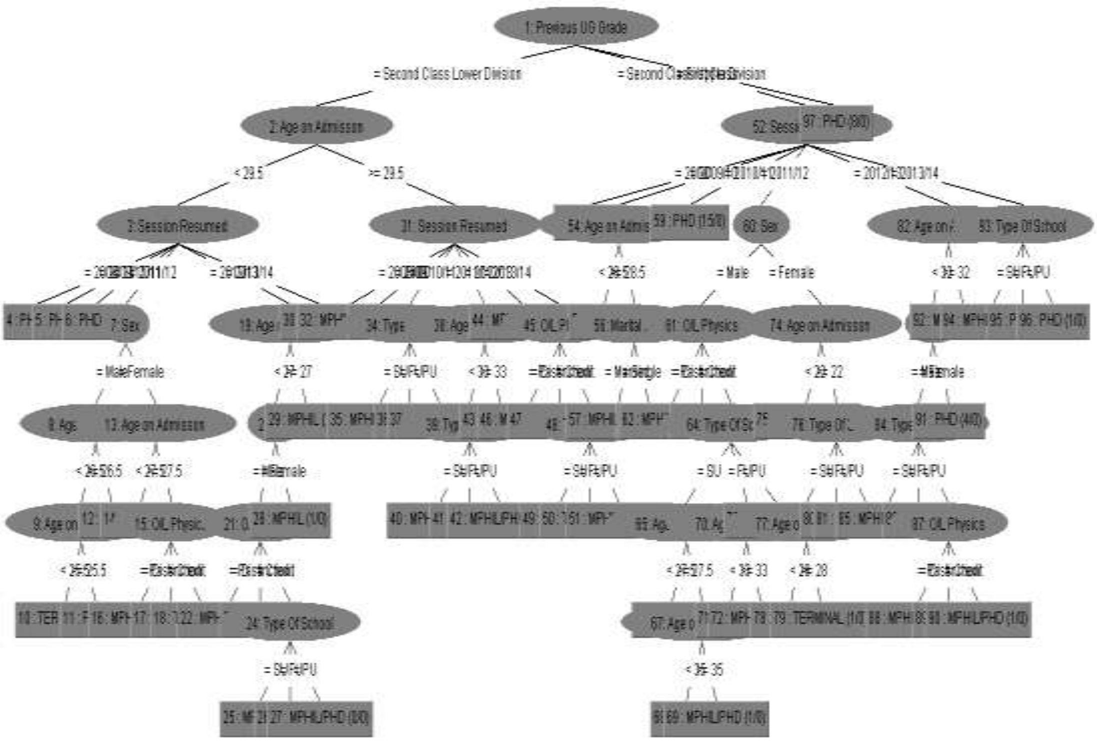

Figure 6: RandomTree classifier decision tree

B Rules from the RandomTree decision tree model

If PreviousUGGrade= First Class then Grade= PHD
If PreviousUGGrade= SecondClasssUpper and SessionResumed=2013/14 and TypeofSchool=FU then Grade=PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2013/14 and TypeofSchool=PU then Grade=PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2013/14 and TypeofSchool=SU then Grade=MPHIL/PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2012/13 and AgeonAdmission<32 and Sex=Male and TypeofSchool=PU and O/LPhysics=Distinction then Grade=PHD else if O/LPhysics=Credit then Grade=MHPIL/PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2012/13 and AgeonAdmission<32 and Sex=Male and TypeofSchool=FU then Grade=PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2012/13 and AgeonAdmission<32 and Sex=Male and TypeofSchool=SU then Grade=MHPIL/PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2012/13 and AgeonAdmission<32 and Sex=Female then Grade=PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2012/13 and AgeonAdmission>=32 then Grade=MHPIL/PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2011/12 and AgeonAdmission<28 and Sex=Female and TypeofSchool=SU then Grade=MHPIL
If PreviousUGGrade=SecondClassUpper and SessionResumed=2011/12 and AgeonAdmission<22 and Sex=Female then Grade=MHPIL/PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2011/12 and AgeonAdmission<33 and Sex=Male and TypeofSchool=FU and O/LPhysics=Credit then Grade=PHD else if AgeonAdmission>=34 then Grade=MPHIL/PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2011/12 and AgeonAdmission<35 and Sex=Male and TypeofSchool=SU and O/LPhysics=Credit then Grade=PHD else if AgeonAdmission>=35 then Grade=MHPIL/PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2009/10 and AgeonAdmission>28.5 and MaritalStatus= Married then Grade=MHPIL/PHD else if MaritalStatus= Single then Grade= NG
If PreviousUGGrade=SecondClassUpper and SessionResumed=2009/10 and AgeonAdmission<28.5 then Grade=PHD
If PreviousUGGrade=SecondClassUpper and SessionResumed=2008/09 then Grade=PHD
If PreviousUGGrade=SecondClassLower and SessionResumed=2013/14 and AgeonAdmission>29.5 and Sex=Male and TypeofSchool=SU and O/LPhysics=Credit then Grade=MPHIL else if TypeofSchool=FU then Grade=Terminal
If PreviousUGGrade=SecondClassLower and SessionResumed=2013/14 and AgeonAdmission>=29.5 and O/LPhysics=Distinction then Grade=MHPIL/PHD
If PreviousUGGrade=SecondClassLower and SessionResumed=2012/13 and AgeonAdmission>=29.5 then Grade=MHPIL/PHD
If PreviousUGGrade=SecondClasslower and SessionResumed=2011/12 and AgeonAdmission<33 TypeofSchool=FU then Grade=MHPIL
If PreviousUGGrade=SecondClassLower and SessionResumed=2011/12 and AgeonAdmission<33 and TypeofSchool=SU then Grade=MHPIL/PHD
If PreviousUGGrade=SecondClassLower and SessionResumed=2013/14 and AgeonAdmission<29.5 then Grade=MHPIL
If PreviousUGGrade=SecondClassLower and SessionResumed=2012/13 and AgeonAdmission<27 and Sex=Female then Grade=MPHIL
If PreviousUGGrade=SecondClassLower and SessionResumed=2011/12 and AgeonAdmission<27 and Sex=Female then Grade=MPHIL
If PreviousUGGrade=SecondClassLower and SessionResumed=2012/13 and AgeonAdmission>=27 then Grade=MPHIL
If PreviousUGGrade=SecondClassLower and SessionResumed=2011/12 and AgeonAdmission>27.5 and Sex=Female and O/LPhysics=Distiction then Grade=MPHIL
If PreviousUGGrade=SecondClassLower and SessionResumed=2011/12 and AgeonAdmission>=26.5 and Sex=Male then Grade=NG
If PreviousUGGrade=SecondClassLower and SessionResumed=2009/10 and AgeonAdmission>=29.5 then Grade=NG
If PreviousUGGrade=SecondClassLower and SessionResumed=2008/09 or 2009/10 or 2010/11 then Grade= PHD else if AgeonAdmission>=29.5 and SessionResumed=2008/09 then Grade=MPHIL/PHD

C Discussion of Results

The results of the study show that it is not always possible to generate a good predictive model from an imbalance dataset. Attributes that can effectively predict any scenario are more dependent on the available dataset than on the methods used. From the rules generated from this study some interesting trends and observations were made. It was observed that grade obtained by students at the end of their undergraduate studies is a major determinant to what grade such student will have by the end of his postgraduate program. Also, each session is unique in itself and it is a major determinant of students' performance. It was observed that the academic performance of students with second class upper and lower divisions is more influenced by the session in which they were admitted in for the postgraduate program. It was further observed that grades in Ordinary Level (O/L) Physics has little or no impact in the prediction of postgraduate performance; though it is a major determinant in predicting undergraduate performance[4]. It was observed that in past years students who were less than thirty years of age had more PHD grades regardless of their sex and marital status, however in recent years especially from the 2013/14 session, age was not a key factor to students' performance at postgraduate studies but the type of school a student finished the undergraduate studies from. Also students that came from federal universities and whose ages were below thirty-two years of age and graduated with a second class upper division from their undergraduate study were likely to have PHD grade at the end of their program while their counterparts from state universities were likely to have a MPHIL/PHD grade at the end of their program. The research also indicates that any student who graduates with first class honors during the undergraduate program (irrespective of the institution) will likely finish with a PHD grade at the end of their postgraduate program in Computer Science.

V CONCLUSION

In this study the prediction of postgraduate students' academic performance using decision tree algorithm was carried out. The pre-processing of the data set used a novel hybrid data sampling technique developed in an earlier research called SMOTE+ENN for correcting the imbalanced data set [8]. The RandomTree classifier gave the best result which was able to model the data and generate rules that could be used to predict postgraduate students' performance. When working with an imbalanced dataset, all performance measures that are derived from True Positive, False Positive, True Negative and False negative may not be a good measure for the created model because most times they favour the class with the majority dataset. In these situations it is best to use performance measures based on AUC and AUPRC.

This work has shown that postgraduate students' performance can be predicted even at the point of admission. The resulting rules generated are worthy of educational verification and testing. To improve the accuracy and predictive ability of model, the postgraduate school should collect more information from their prospective students which should include students' sociological background data, personality and their expectation alongside their previous academic performance so that further researches to be conducted in this area will put this information into consideration. Studies on the factors that could make sessions unique could also be looked into.

The scope of the dataset should be broadened, that is more instances of postgraduate students should be incorporated and other data mining methods such as Naïve Bayes, Neural Network and Support Vector Machine can also be used and their results compared for better predictive models.

This study can be used for strategic decision making by education managers when admitting students for postgraduate programmes. The extracted information that predicts student performance can be stored as intelligence knowledge for decision making to improve the quality of education in higher institutions.

REFERENCES

[1] Abubakar, R.B and Oguguo, O.D (2014). Age and Gender as Predictors of Academic Achievement of College Mathematics and Science Students, Proceedings of the 2011 International Conference on Teaching, Learning and Change (c) International Association for Teaching and Learning (IATEL).

[2] Adeyemo, A.B and Kuye, G (2006). Mining Students' Academic Performance Using Decision Tree Algorithms, Journal of Information Technology Impact Vol. 6, No 3, pp 161-170.

[3] 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), pp 61-78.

[4] Akinola, O.S, Akinkunmi, B.O and Alo, T.S (2012). A Data Mining Model for Predicting Computer Programming Proficiency of Computer Science Undergraduate Students, African Journal of Computing & ICT Proficiency Vol. 5, No. 1.

[5] Baradwaj, B.K and Saurabh, P (2011). Mining Educational Data to Analyze Students' Performance, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 2, No. 6.

[6] Baker R.S., Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions, Journal of Educational Data Mining (JEDM), 1 (1): 2017

[7] César A. Del Río and Julio A. Pineda Insuasti (2017). Predicting academic performance in traditional environments at higher-education institutions using data mining: A review, retrieved on 19/02/2018.

[8] Folorunso S.O and Adeyemo A.B (2013). Alleviating Classification Problem of Imbalanced Dataset, African Journal of Computing and ICT, Vol. 6, No. 1, pp.137-144

[9] Khushbu, N.S, Misika, R. P, Neha, V. T, Priyanka, N. G, Richa, H. S, Adhvaryu, N. (2015). Study of Data Mining in Higher Education-A Review, International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 6 (1), pp. 455-458.

[10] Kolo, K. D, Adepoju, S.A, Alhassan, J.K (2015). A Decision Tree Approach for Predicting Student's Academic Performance, International Journal

Education and Management Engineering, Vol. 5,pp. 12-19

[11] Kumar, V, and Chadha, A. (2011). An Empirical Study of the Applications of Data Mining Techniques in Higher Education, International Journal of Advanced Computer Science and Applications (IJACSA), Vol. 2, No.3

[12] Mashael A. Al-Barrak and Muna Al-Razgan (2016). Predicting Students Final GPA Using Decision Trees: A Case Study. International Journal of Information and Education Technology, Vol. 6, No. 7, July 2016

[13] Naik, N and Purohit, S. (2012). Prediction of Final Result and Placement of Students using Classification Algorithm, International Journal of Computer Applications (0975 – 8887) Vol. 56, No.12.

[14] Nnamani, C. N, Dikko, H. G and Kinta, L.M (2014). Impact of Students' Financial Strength on their Academic Performance: Kaduna Polytechnic Experience, International Multidisciplinary Journal, Ethiopia Vol. 8 (1), No. 32, pp. 83-98.

[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 & ICT, Vol 7. No. 2.

[16] Ramaswami, M. and Bhaskaran, R. (2010). A CHAID Based Performance Prediction Model in Educational Data Mining. IJCSIInternational Journal of Computer Science Issues,Vol.7, Issue 1,No.1, January 2010. ISSN (Online): 1694-0784, ISSN (Print):1694-0814

[17] Rashan, K,H. and Anushka Peiris, A. (2011). Data Mining Applications in the Education Sector, MSIT, Carnegie Mellon University, retrieved on 28/01/2011.

[18] Romero, C, Ventura, S and García, E. (2007). Data Mining in Course Management Systems: Moodle case study and tutorial, Elsevier Science.

[19] Ruby, J and David, K. (2014). A study model on the impact of various indicators in the performance of students in higher education, International Journal of Research in Engineering and Technology (IJRET), Vol. 3, Issue 5, pp.750-755.

[20] Ruby, J. and David, K. (2015). Analysis of Influencing Factors in Predicting Students Performance Using MLP - A Comparative Study, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 2.