2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS) 978-1-7281-3126-9/20/$31.00 ©2020 IEEE 10.1109/ICMCECS47690.2020.240888

Performance Evaluation Of Machine Learning Techniques For Prediction Of Graduating Students In Tertiary Institution

Ajinaja Micheal Olalekan
Department of Computer Science,
Federal Polytechnic Ile-Oluji,
Ondo State, Nigeria
ajinajalekan@gmail.com

Ojonukpe Sylvester Egwuche
Department of Computer Science,
Federal Polytechnic Ile-Oluji,
Ondo State, Nigeria
ojoegwuche@fedpolel.edu.ng

Sylvester Oluyemi Olatunji
Department of Computer Science,
Federal University of Technology,
Akure, Nigeria
solatunji@futa.edu.ng

***Abstract* -- Near accurate prediction of students' future performance based on their historical academic records is important for effective pedagogical interventions. It is imperative to provide an enhanced prediction system that can assist educational institutions to identify and monitor students at different threshold and to focus on improving students that their threshold is less than graduation at early stage. Studies on the prediction of graduating students using data mining techniques have been widely carried out in the existing literature. The paper applied Baye's theorem and Artificial Neural Networks (ANN) to build a predictive model for the likelihood of students' graduation in a tertiary institution. The prediction was performed on four variables – Unified Tertiary Matriculation Examination (UTME), Number of sittings for O'level (NOS), Grade Points of O'level (Grade) and Mode of Entry (PreND). The implementation was carried out in R-studio environment. The results showed that ANN had higher accuracy compared to Bayesian Classification. ANN performed better because of the learning rules it contains.**

Keywords: Model-Based Approach, predicting graduating students, Education data mining, Naive Bayes, Artificial Neural Networks.

I. INTRODUCTION

Every higher institution aims at providing quality education to its students. Evaluating and enhancing the educational institution requires adequate knowledge for assessing, planning and taking decisions especially during the admission process. In any tertiary institutions, students are regarded as the major assets. The task of evaluating and predicting students' ability to graduate is part of educational quality assurance [1-5].

In cross disciplinary perspective such as computer science, statistics and psychometrics, it has been demonstrated that data mining can function in business practices as a knowledge discovery tool for transforming historical and operational data into high level knowledge that supports decision making [10-11]. From an administrative perspective, the ability of data mining for prediction can enhance admission processing because of adequate understanding admission officers will derive from historical data.

Meaningful knowledge in educational institution of learning can help education administrators in achieving their stated objectives. Data mining technology has the potentials to bridge the knowledge gaps in higher educational institution. Therefore, the association, anomalies and unknown patterns that exist among the large data set can be used to improve the efficiency, effectiveness and the speed of the processes. The discovered unseen patterns can guide the educational institution to take informed decisions and can provide advance plan in guiding and counselling of students. In educational domain, each education related problem has specific objectives with unique characteristics and different approaches are required in solving the problem [12].

In existing literature, many analysis have been performed on tertiary institutions students' data, which include students' entrance examination and higher level results. However, there is a need to establish the relationship between these entry results and students' final graduation grades among education administrators. With data mining, hidden pattern in a large data set can be successfully extracted. It has been widely used in different areas including the educational institution to enhance learning processes [13].

[9] Presented predicting students' academic performance using their pre-university marks of first and second year courses in a four-year university program. In the study, decision tree was used to identify indicators to low students' performance.

Existing requirements that qualify prospective candidates for admission into higher educational study do not depict the true academic prospects of such candidates. Therefore, there is a need to discover objectively the relationship among students' entry data variables and the impact of these variables on students' graduation in their terminal year. The study aims at evaluating the performance of Baye's theorem and Neural Networks to predict students' graduation in a tertiary institution.

The organization of the paper is as follows. Section II reports a survey of related works in the published literature. Section III describes the research methodology. Section IV reports the results and the discussion. The conclusion is given in section V.

II. REVIEW OF RELATED LITERATURE

There are many attempts by researchers and administrators to mine students' data because of the need to utilize the volume of data that are available in education databases for enhanced educational service delivery.

[2][8] combined different and many variable selection indicators of undergraduate-level performance to formulate a model-based approach that predict graduate-level performance using linear regression. The study aims at showing the independence of the predictive power of undergraduate performance indicators and their aggregates. The work analysed 81 variables out of 171 records of students from a bachelor's and a master's program in Computer Science department. However, in the study, there was an unbiased estimate of the predictive value of undergraduate level indicators. The reported results show that the third-year grade point average stands as the explanatory variable that is most significant with higher influence than the grades earned in challenging first-year courses.

[3] Worked on data mining in higher education using university student dropout as case a study. In the paper, different data mining approaches was used to examine and predict cases of students' dropouts in university programs. The collected data captures the history of the students' study and the transcript of the courses taught within the first two years of computer science programme, student GPA and high school average with class label to indicate whether the student graduate or not. Decision Tree (DT) and Naive Bayes (NB) data mining techniques were used to formulate the model. The method was tested with 10-fold cross validation.

[1] used Adaptive Neuro-Fuzzy Inference system (ANFIS) for the prediction of student academic performance. It noted that ANFIS can serve as a support tool for decision making in higher learning that tends towards improving the academic success rates of students. The proposed work also provide guidelines that identify areas of enhancement in their educational processes.

[16][7] Carried out a comparative study of two data mining classification techniques in the prediction of student performance in post unified tertiary matriculation examination (PUTME). Two data mining classification techniques; J48 Decision Tree and Artificial Neural Network Multilayer Perceptron, were used in the research to study the trends of student performance based on the mode of entry of the students. The study reported that decision tree can be a better choice algorithm over the neural network when it involves large amount of data sets for analysis. It builds fast models and presents results in tree constructions for easy construction.

[17] Applied data mining approaches to examine the rate of students' dropouts in a university. The study also includes discovering hidden relationships between student dropout status and their mode of enrolment. Enrolment persistence by mining a frequent cases using FP-growth algorithm.

[9] Conducted a study on predicting student academic performance using data mining techniques. The study aimed at using data mining techniques for predicting the students' graduation performance in final year at university using pre-university marks and university early examination marks.

[4] Adopted data mining technique for the prediction of students' grade point averages during graduation. Selected data mining prediction techniques were implemented to assist educational institutions with likelihood of students' GPAs at graduation. As the system predicts chances of students graduation, students that have their GPA on the low curve would be encouraged early enough to invest more efforts towards their academic success. Neural networks and support vector machine algorithms were implemented on the data of computer education and instructional technology students to predict their possible GPAs at graduation [6].

[14] Presented a prediction model for students' academic performance and identified the courses that influence and affect students' low academic performance. [15] proposed a predictive model based on artificial neural networks to predict students' graduation time.

III. RESEARCH METHODOLOGY

A predictive data mining technique based on naïve bayes classifier was used in the prediction model. In the model, the likelihood of the students' graduation is forecast based on some variables with some level of transparency, noise outliers and good performance.

Probability based classifier is chosen because of its ability to estimate the probability of observing the positive class and to predict the observations that belong to the positive or negative class. In Figure 3.1, Layer 1 consist of Neural Networks and layer 2 consist of Bayes' Theorem. These two classifiers are selected because of their performances in many domains of application.

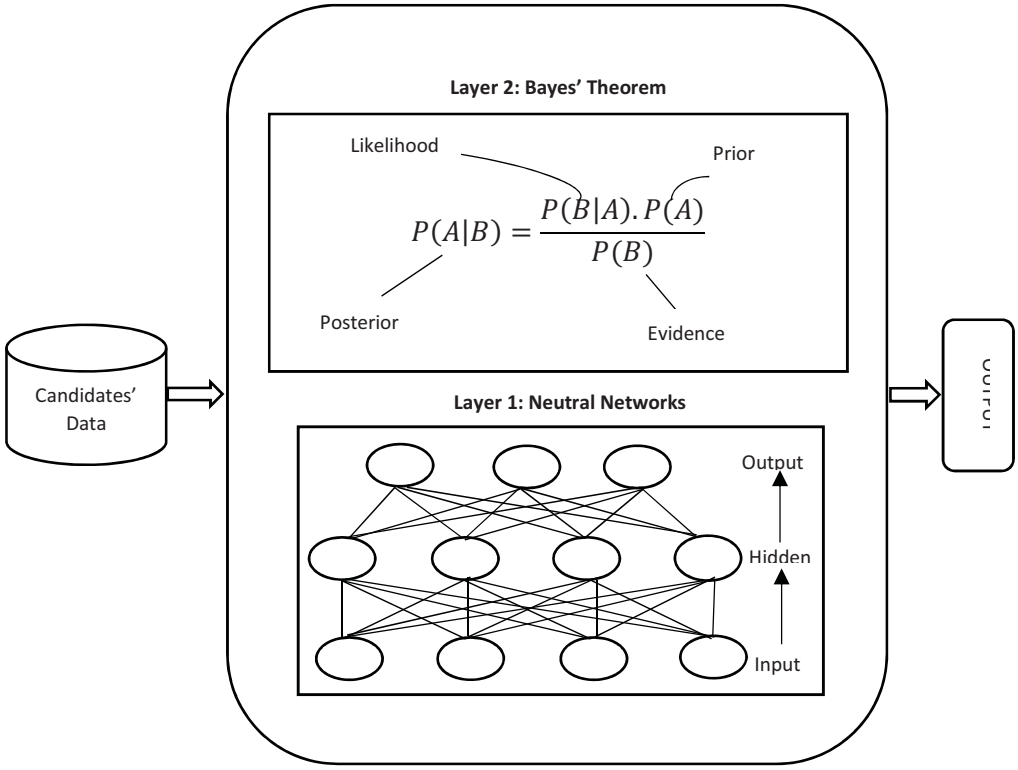

Figure 3.1 Two-layered classifier

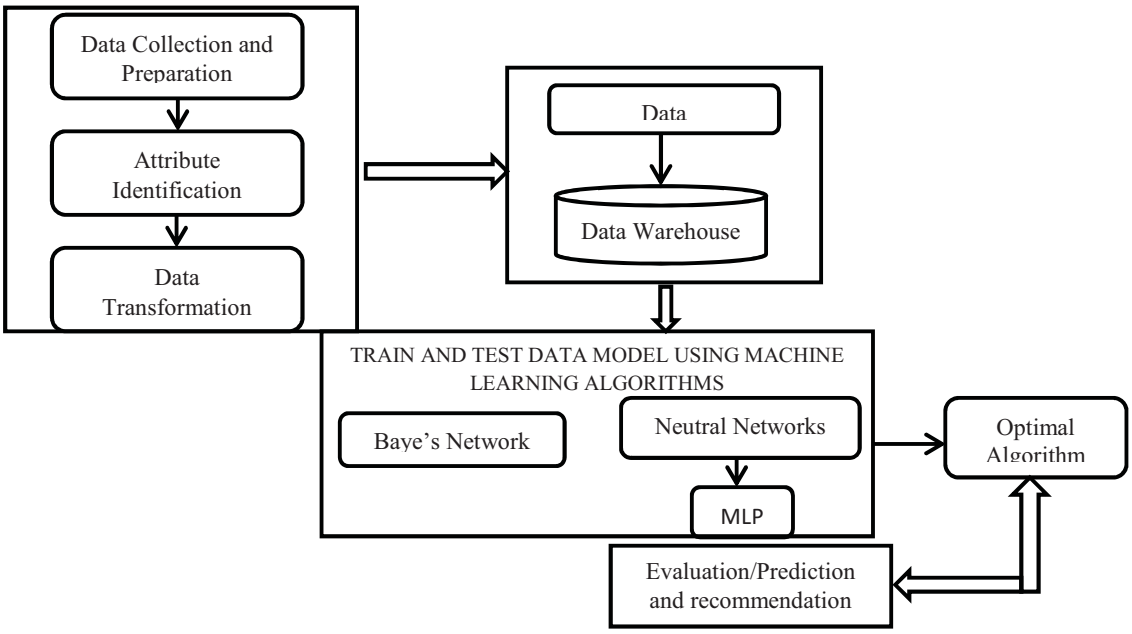

Figure 3.2 Architectural Framework of the Proposed System

The proposed system as shown in Figure 3.2 comprise five components. The first and second components are responsible for the acquisition and storage of the candidates' data. The model building is carried out in the third component which is responsible for obtaining knowledge about the candidates through selected classification model. The fourth component maps the pattern in the rules generated while the fifth component performs the evaluation/prediction task.

A. Artificial Neural Networks

Artificial Neural Networks are composed of many nodes connected to each other and function together by passing information. They consist of a number of layers called Multi-Layer Perceptron (MLP) and each layer performs a different function on the received data. The demand of ANN to solve complex problems is steadily on the increase with ease of adaptation as is being exposed to real time examples.

Neural network is chosen because of its ability to learn from examples and apply them when a similar event arises to work through real-time events, perform multiple tasks in parallel without affecting the system performance and its ability to detect fault and generate output with missing information. The learning rules it contains makes it easy for modification according to the input patterns of the connection.

B. Supervised Learning

Supervised learning in machine learning is a function that maps an input to an output on the principle of input-output pairs. Supervised learning technique is adopted to perform the required classification problems. Classification problems are based on a two-step process that includes model construction and model usage.

i. Model Construction describes a set of already specified classes of data. In the study, each tuple is assumed to belong to a pre-determined class label attribute.
ii. Model usage involves classifying future or unknown objects. Unknown attributes of the test sample is compared with the classified result from the model for accuracy check. If there is high accuracy level, then the model can be used for the classification of new data. The accuracy rate shows the percentage of the collected samples that are given correct classification.

Both Bayes' theorem and ANN employed in the study are supervised machine learning approaches. The ANN uses a 4-layered Multi-Level Perceptron algorithm.

C. Naïve Bayes' Classification model

Bayes' theorem popularly called Naive Bayesian classifier is based on the principle of independent assumptions between predictors.

The theorem provides a way of calculating the posterior probability, P(c|x), from P(c), P(x), and P(x|c). Naive Bayes classifier takes the position that the value effect of a predictor (x) on a given class (c) is independent of the values of other predictors within a given set of data [5].

Considering the provided training data, the following classifications were generated.

X = (PreND = NO, UTME = YES, NOS = YES, Grade = YES)

Will any student in the tuple X graduate?

Data X(x_1, x_2, x_3, x_4) = (PreND, UTME, NOS, Grade)

a. Determine the output classes

b. Determine the probability of each class

c. Determine the probability of each class given the label of the variable

d. Determine the probability of each class given a tuple, X

C. Data Collection and Selection

Final year students' data was collected from the department of Computer Science, Federal Polytechnic, Ile-Oluji in Ondo State, Nigeria. There were a total of 44 students who were expected to graduate in National Diploma II. Meaning that the dataset contains 44 instances with 5 attributes as shown in Table 3.1. The class distribution is coded as graduate and not graduated. There were four independent variables and one dependent variable. The mode of entry (PreND) can either be through PreND or direct entry, grade point (Grade) was calculated based on five compulsory subjects in their SSCE, number of seating states whether the student sat for SSCE more than once and UTME is the unified examination score of the students.

The following simplified assumptions were made in the study.

i. Absence of financial issues
ii. Zero accidental cases
iii. Absence of social vices
iv. No Post UTME

D. Data Pre-processing

In data pre-processing stage, the noises and inconsistencies that exist in the data were eliminated and data attributes that were of no interest were also removed. Table 3.1 shows a tabular representation of attributes, description and possible values. The data sets were combined into a single worksheets and saved as a CSV file in Microsoft Excel.

TABLE 3.1: THE TRAINING DATASET DESCRIPTION

S/N	Attributes	Values
1	Mode of Entry - **PreND**	1 (Yes, students was admitted through Pre-ND program), 0 (No, student was not admitted through Pre-ND program
2	JAMB score – **UTME**	≥120, <400
3	SSCE - **Grade**	≥1.0, <5.0
4	Number of sitting in SSCE - **NOS**	1 (Yes, students used more than one SSCE result), 0 (No, students used a single sitting SSCE result)
5	Grad	1 (Graduated), 0 (Not Graduated)

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The R statistical machine learning toolkit was used for the implementation. The confusion matrix was used to describe the performance of the classification model (or "classifier") on a known valued set of data. The table presents the number of correctly and incorrectly classified instances of the actual and predicted class instances.

Table 4.1, Table 4.2 and Table 4.9 show the output, confusion matrix and class performance measure of Bayesian classifier on training and test sets respectively.

TABLE 4.1: OUTPUT OF BAYESIAN ON TRAINING AND TEST DATA SET

PARAMETERS	OUTPUT	
	Training	Test
Time taken to build model	0.19 seconds	0.17 seconds
Correctly classified instances	27 (77.14%)	6 (66.67%)
Incorrectly classified instances	8 (22.86%)	3 (33.33%)
Total number of instances	35	9

TABLE 4.2: CONFUSION MATRIX FOR BAYESIAN ON TRAINING AND TEST SET

Training Set	**X**	**Y**	**<= Classified as**
	5	3	X = Not Graduated
	5	22	Y = Graduated
Test Set	**X**	**Y**	**<= Classified as**
	0	2	X = Not Graduated
	1	6	Y = Graduated

Table 4.3, Table 4.4 and Table 4.5 show the output, confusion matrix and class performance measures of ANN on training and test sets respectively.

TABLE 4.3: MISCLASSIFICATION OUTPUT OF ANN ON TRAINING AND TEST DATA SET USING 1, 2, 3 AND 4 HIDDEN LAYERS

	Hidden Layers			
	1	**2**	**3**	**4**
Training Set	20.69%	17.24%	13.79%	0.035%
Testing Set	26.67%	26.67%	0.4%	0.4%

TABLE 4.4: OUTPUT OF ANN ON TRAINING AND TEST DATA USING 1 HIDDEN LAYER

PARAMETERS	OUTPUT	
	Training	**Test**
Correctly classified instances	23 (79.31%)	6 (73.33%)
Incorrectly classified instances	11 (20.69%)	4 (26.67%)
Total number of instances	35	9

TABLE 4.5: OUTPUT OF ANN ON TRAINING AND TEST DATA USING 2 HIDDEN LAYER

PARAMETERS	OUTPUT	
	Training	**Test**
Correctly classified instances	23 (82.76%)	6 (73.33%)
Incorrectly classified instances	11 (17.24%)	4 (26.67%)
Total number of instances	35	9

TABLE 4.6: OUTPUT OF ANN ON TRAINING AND TEST DATA USING 3 HIDDEN LAYER

PARAMETERS	OUTPUT	
	Training	**Test**
Correctly classified instances	23 (86.21%)	6 (99.6%)
Incorrectly classified instances	11 (13.79%)	4 (0.4%)
Total number of instances	35	9

TABLE 4.7: OUTPUT OF ANN ON TRAINING AND TEST DATA USING 4 HIDDEN LAYER

PARAMETERS	OUTPUT	
	Training	**Test**
Correctly classified instances	23 (99.97%)	6 (99.6%)
Incorrectly classified instances	11 (0.035%)	4 (0.4%)
Total number of instances	35	9

TABLE 4.8: CONFUSION MATRIX FOR ANN ON TRAINING AND TEST SET

Training Set	**X**	**Y**	**<= Classified as**
	5	3	X = Not Graduated
	3	18	Y = Graduated
Test Set	**X**	**Y**	**<= Classified as**
	1	2	X = Not Graduated
	2	10	Y = Graduated

TABLE 4.9: CLASS PERFORMANCE MEASURE OF BAYESIAN ON TRAINING SET AND TEST SET

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**Class**
Training Set	0.625	0.185	0.500	0.625	0.556	NOT GRADUATED
	0.815	0.375	0.880	0.815	0.846	GRADUATED
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**Class**

Test Set	0.000	0.143	0.000	0.000	0.000	NOT GRADUATED
	0.857	1	0.750	0.857	0.800	GRADUATED

TABLE 4.10: CLASS PERFORMANCE MEASURE OF ANN ON TRAINING SET AND TEST SET

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**Class**
Training Set	0.625	0.143	0.625	0.625	0.625	NOT GRADUATED
	0.857	0.375	0.857	0.857	0.857	GRADUATED
	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**Class**
Test Set	0.333	0.25	0.333	0.333	0.333	NOT GRADUATED
	0.833	0.667	0.833	0.833	0.625	GRADUATED

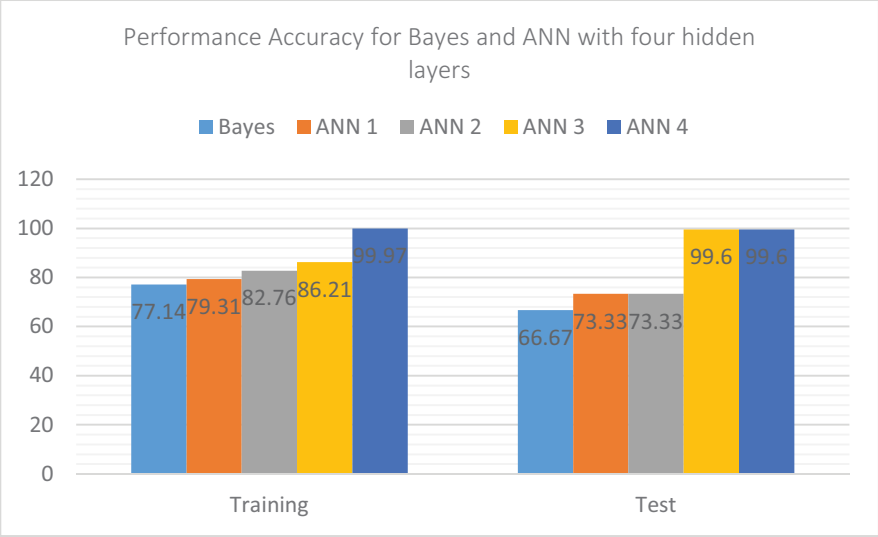

Figure 4.1 Performance analysis of Bayes and ANN

The graphical representation of the performance accuracy of the two algorithms (Bayesian and ANN) is shown in Figure 4.1. From the observation of the data, both algorithms have impressive prediction accuracy. Artificial Neural Network (ANN) algorithm with one hidden layer has a better accuracy of 79.31% than Bayesian Classification algorithm of 77.14%.

V. CONCLUSION

Educational data-mining is an efficient analytical tools for educational activities decisions making. It can be inferred from the study that both Bayes and Artificial Neural Network (ANN) have good accuracy on the data sets. However, the ANN shows a higher performance accuracy than Bayes theorem. Accuracy increased in ANN as the hidden layers increased. The results of the study can help education administrators to identify students with high risk of not graduating early for correctional steps. The study carried out a comparative analysis of two algorithm (Bayes and Artificial Neural Network) on students' data for predicting graduating students in tertiary institutions. The ANN and Bayes Classification are both suitable algorithms for mining the data set but with higher performance in ANN.

ACKNOWLEDGMENT

The authors acknowledge the authority of Computer Science department, Federal Polytechnic, Ile-Oluji for providing the data of their students during the implementation stage of the study.

REFERENCES

[1] A. Altaher and O. Barukab (2017). Prediction of Student's Academic Performance Based on Adaptive Neuro-Fuzzy Inference, *International Journal of Computer Science and Network Security*. 17(1): 165-168

[2] J. Zimmermann et al (2015). A Model-Based Approach to Predicting Graduate-Level Performance Using Indicators of Undergraduate-Level Performance. *Journal of Educational Data Mining*. 7(3): 151-170

[3] S. A. Ghadeer and M. Alaa (2015). Data Mining In Higher Education: University Student Dropout Case Study. *(IJDKP) International Journal of Data Mining & Knowledge Management Process*. 5(1): 15-26

[4] Ahmet *et al*. (2014). Early Prediction of Students' Grade Point Averages at Graduation: A Data Mining Approach. *Eurasian Journal of Educational Research* 54: 207-226

[5] P. A. Idowu, O. Agbelusi and T. A. Aladekomo (2016). The Prediction of Paediatric HIV/AIDS Patient Survival: A Data Mining Approach. *Asian Journal of Computer and Information System.* 4(3): 88-90

[6] J. Prathipa, C. Sindhu, J. Sanyam and A. Khetan (2018). Predicting Students' Performance Using Bayesian Classification Algorithm. *International Journal of Pure and Applied Mathematics*. 118(22): 127-132

[7] P. A. Idowu, K. O. Williams, A. B. Jeremiah and I. O. Adeniran (2015). Breast Cancer risk Prediction Using Data Mining Classification Techniques. *(TNC) Transaction on Networks and Communications*. 3(2): 5-8.

[8] A.O. Osofisan, O. O Adeyemo and S. T. Oluwasusi (2014). Empirical Study of decision tree and Artificial Neural Network Algorithm for mining educational database. *(IEEE) African Journal of Computing and ICT*. 7(2): 187-196.

[9] R. Asif, S. Hina and S. I. Haque (2017). Predicting Student Academic Performance using Data Mining Methods. *International Journal of Computer Science and Network Security*. 17(5): 187-191.

[10] E. Garcia, C. Romero, S. Ventura, and C. de Castro (2011). A collaborative educational association rule mining tool. *Internet and Higher Education*, 14(2): 77-88.

[11] A. Raheela, H. Saman and H. Saba (2017). Predicting Student Academic Performance using Data Mining Methods. *(IJCSNS) International Journal of Computer Science and Network Security*. 17(5): 187-191

[12] G. Suganthi and M. V. Ashok (2017). Predicting Employability of Students Using Data Mining Approach. *International Journal of Information Research and Review*. 4(2): 3798-3801.

[13] K. Surjeet, and P. Saurabh (2012). Data Mining Application in Enrolment Management: A Case Study. *(IJCA) International Journal of Computer Application.* 41(5): 1-6.

[14] A. Yasmeen, A. Wejdan, A. Isra, and A. Muna. (2016). Predicting critical courses affecting students performance: A case study. *Procedia Computer Science,* 82:65–71,.

[15] J. S. Bassi, E. G. Dada, A. A. Hamidu and M. D. Elijah (2019). Students Graduation on Time Prediction Model Using Artificial Neural Network. *IOSR Journal of Computer Engineering (IOSR-JCE)*. Volume 21, Issue 3, PP 28-35.

[16] S. C. Chiemeke and O. A. Akinlade (2017). A comparative study of two data mining classification techniques in the prediction of students performance n post unified tertiary matriculation examination.

[17] *Computing, Information Systems, Development Informatics and Allied Journal*. PP 29 – 42.

[18] S. A. Ghadeer and El-Halees (2015). Data mining in higher education: *university student dropout case study.* Vol. 5. No 1.