Received: 14 October 2022 | Revised: 15 December 2022 | Accepted: 29 December 2022 | Published online: 12 January 2023

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2023, Vol. 00(00) 1–10
DOI: 10.47852/bonviewJCCE3202470

BON VIEW PUBLISHING

A Machine Leaning Approach in Predicting Student's Academic Performance Using Artificial Neural Network

Rufai Aliyu Yauri[1], Hassan Umar Suru[2], James Afrifa[3,*] and Hannatu G. Moses[3]

[1]*Department of Information Technology, Kebbi State University of Science and Technology, Nigeria*

[2]*Department of Computer Science, Kebbi State University of Science and Technology, Nigeria*

[3]*Department of Computer Science, Adamawa State Polytechnic, Nigeria*

Abstract: The rate at which students succeed in their academic pursuits contributes significantly to the academic achievement of their educational institutions because it is used as a measure of the institution's performance. Many factors could be responsible for students' academic performance and student success. Quick understanding of weak students and providing solutions to improve their performance will significantly increase their academic success rate. Educational data mining using artificial neural network plays a crucial role in determining their likely performance and helps them to initiate measures that can reposition the students' performance in the future. This study developed a model that predicts students' failure and success rates with the aid of a machine learning algorithm. The study sampled 720 students from three selected tertiaries institutions in Adamawa State, Nigeria. Three hundred students were selected from Modibbo Adama University, Yola, 300 students were selected from Adamawa State University, Mubi, and 120 students were selected from Adamawa State Polytechnic, Yola. The research makes use of descriptive statistics to identify the variables that likely affect students' academic performance. The collected data were preprocessed, cleaned, and modeled using Jupyter Notebook, a Python Anaconda development platform for artificial neural to build the student's academic performance predictive model. The neural network is modeled with 12 input variables, two layers of hidden neurons, and one output layer. The dataset is trained using the backpropagation learning algorithm. The performance of the neural network is evaluated using k-fold cross-validation. The neural network model has achieved a good accuracy of 97.36%.

Keywords: academic performance, students, educational data mining, schools, prediction, artificial neural network

1. Introduction

The ability of a child to learn is a very important moment in the child's life. Kubiatko et al. (2018) pointed out that the student's ability to assimilate quickly has an important role to play in his academic success, and the student's academic success affects the general satisfaction of the student and society at large. This means that the student's level of assimilation is among the most required activities in the life of the student. However, numerous attributes affect students' assimilation process and academic achievement. The numerous attributes are otherwise referred to as assimilation factors. These attributes are connected with the student's physiological, psychological, parental, social situations, and environmental conditions. It is of important note to understand that these assimilation factors contribute to the academic success of the student in different ways (Uluğ, 2000). The level of a student's academic assimilation alongside the student's academic success has an impact to the fullest on the student; for this very important reason, educational institutions globally would understand it as a vital factor in students' achievement academically.

It is important to consider student's academic performance to be excellent in tertiary institutions of learning as a yardstick for the academic success of both the students and the institutions. It is of important fact to note that the socioeconomic development and industrial growth of any nation depend solidly on the quality of education, academic performance, and success of its citizens. High-quality education, excellent student performance, and making sure that there is professionalism in the education sector are among the key objectives in the education system of any country.

Student performance is an essential aspect in higher institutions of learning and hence determines the student's academic achievement and that of the institution (Amo & Santelices, 2017). Students' performance is mostly used as a criterion for high-quality

*Corresponding author: James Afrifa, Department of Computer Science, Adamawa State Polytechnic, Nigeria. Email: afrifajames@gmail.com

education because of their excellent record of academic achievements. It is important to note that majority of the tertiary institutions in Nigeria uses grade as the main measure to assess students' academic performance. Additionally, the structure of students' courses, marking of students' assignments, scores grades for final exams, and extracurricular activities will affect the student's academic performance. Comprehending students' academic success is very important to link with the understanding of the centric learning environment of the students. Early prediction of students' academic status in any tertiary institution of learning will go a long way in curtailing students' failure rate and dropout. This will also assist in improving the success rate in the institutions by making the students identify their areas of weakness and take necessary measures to improve their performance (Sandoval et al., 2019). Heredia et al. (2015) result findings indicated that the poor performance of the student is mostly affected by varieties of variables that the students encounter during the days of their study. Therefore, a clear identification of these variables responsible for students' poor performance and success needs to be investigated. This is necessary because the results of the investigation will indicate poor and weak-performing students in which appropriate measures will be taken to improve their academic performance to achieve the desired success in their academic pursuits (Di Caudo, 2015).

The quantity of student information gathered and kept in various colleges, universes, and polytechnic is large, and this information could be used to predict students' academic performance. The expert that is engaged in data science no longer uses the old method of computing data which makes use of paper and pen. Data experts use an automated method to analyze data. These huge amounts of data collected are kept unused. Educational data mining (EDM) happens to be an interesting field that is gaining more attention in data science, particularly when working on educational data. With EDM, educational institutions can make use of students to predict their academic achievement and also serve as a decision support system (DSS) for the institution. The final goal of EDM is to bring out good policies through data that will serve as a basis for making good decisions for quality education in any institution (Silva & Fonseca, 2015).

Jamala and Elizabeth (2022) considered EDM as an area of study that is used to improve education. EDM has contributed significantly to data mining and has numerous areas of application. EDM can be defined as the use of different techniques in the application of data collection applied in data analysis in the education sector, which are aimed at solving challenges faced by institutions of learning (Chakraborty et al., 2016). EDM application areas include the formulation of e-learning systems, clustering educational data as well as making student performance predictions among others (Eakasit, 2015). The focus of EDM is to explore educational data to gain insights into how individuals learn (Meghji et al., 2018).

A machine learning (ML) algorithm is considered one of the most common aspects of EDM used in predicting students' academic performance. ML can be defined as the process that a machine or model is gained access to data and through it, the machine can learn on its own. Hussain et al. (2019) put it that ML is a computer programming that enhances perfect results with the help of example data. ML plays a significant role in EDM particularly when it comes to prediction. Artificial neural network (ANN) is a ML algorithm that is extensively in the field of data mining. ANN models are used in data analysis, and they can be classified as a semi-parametric method. This class of models can learn complex tasks, such as recognition, decision-making, or predictions, and can deal with nonlinear data.

Tertiaries institutions in Adamawa State have recorded a high poor student success rate and dropout due to poor student academic performance in recent years. This has become a serious challenge to students, parents, and school management. With the huge amount of student data collected from various tertiaries institutions in Adamawa State, this paper proposed the use of an ANN to predict students' academic performance in tertiary institutions in Adamawa State, Nigeria to curtail the problems of poor academic performance and student dropout.

2. Review of Related Literature

Many academic scholars have carried out different studies on students' academic performances that can be found and model that predicts students' academic performance has been conducted around the world so far. Student's academic performance plays an important role in student academic achievement and this has a positive impact on the student's success rate (Koç et al., 2018). Shahiri et al. (2015) study revealed that neural networks and decision tree (DT) among other algorithms are two commonly used classification techniques with prediction accuracy of 98% and 91%, respectively. The results also show that the support vector machine (SVM) and k-nearest neighbor (K-NN) have the same accuracy of 83%, while Naïve Bayes (NB) has lower prediction accuracy (76%). From the results of the study, it was concluded that the neural network's ability to learn and adjust is what makes it useful and powerful in EDM prediction.

A study conducted by Erdem et al. (2007), primarily, is to investigate the variables responsible for poor students' success rates putting into consideration the student's final grades of selected university population students in the area of their study. The study shows that a student's success rate is related to the final grade point of the student. However, the results revealed that students' performance is affected by other variables like sex of the students, previous students' academic history, students' accommodation type, students' parental financial stability and income, student's social surroundings, pre-tertiary school attended and location attended, student's intelligence quotient (IQ), students studying time duration, and students' university entrance examination grade. Hence, the study revealed that these factors contributed significantly to affect the cumulative grade point average (CGPA) of the student.

Kiu (2017) conducted a study to analyze selected variables and their impact on the success rate of the sampled population. These variables include the background from which the student came, the social activities of the student, and the academic success of the student. Their paper makes use of three variables to test for data exchange and data normalization in the proposed model. The study makes use of the Webca data mining tool on datasets for analysis. The study makes use of NB, multilayer perceptron (MLP), DT, and random forests (RF) as the data mining methods in the dataset. The results of the study revealed that the background from which the students came and the personality traits the students exhibit in society play a significant role in revealing the initial stages as well as identifying these students that are weak and are performing poorly in their studies.

Masood et al. (2017) undertook a study to assess the factors affecting students' academic performance and the negative effect on the students. To effectively achieve the objectives, the study makes use of experiments for data collection and modeling. The study used 11 different ML techniques to assess factors affecting students' performance in the predictive models. The results of the study revealed that DTs and RF have high accuracy of prediction and were able to identify significant variables responsible for students' academic success. Yakubu and Abubakar (2021) conducted a study making use of a linear regression model that will forecast the success of their institution. The results of the study explained how the predicted results could help students and teachers improve their education and institution. Altarki et al. (2021) conducted a study on predicting student performance using EDM for applying early interventions. The study sampled three hundred students from the Department of Computer Science and the Department of Information Science that is from a university in Saudi Arabia. The study considered the evaluation of six data mining techniques in forecasting students' success and came out with a result that shows which method is more appropriate. Ahmed and Elaraby (2014) in their study developed what is known as rules for classification in predicting the success rate of a sampled population. The study investigated previous students' information that was admitted in other programs interval between 2005 and 2010, and the results of the study predicted students' last grade to increase their success rate and decrease their rate of failure of students in the selected program. In addition, Romero and Ventura (2013) conducted research to investigate the relationship between student success and some variables of interest. The results of the study revealed that students learning techniques on campus, students reading styles, student's environment, and student's family relationships contribute significantly to the student's success in the study area. Mogadishu-Banik and Kumar (2019) carried out a study on variables that contribute to undergraduate students' achievement at Arba Minch University. The outcome of the study shows that there is statistical significance between sex variation, type of examination, and reading time. The results of the findings also revealed that these variables determine the student's CGPA. This means that these variables of interest contribute significantly to the academic success of the students. The study further shows statistical relationships among variables such as students' former academic background, studying hours, social life on campus, and social media attitude with the student's success rate. This means that these factors have a significant impact on determining students' academic performance. However, another study conducted by Vamshidharreddy et al. (2020) shows that there are four broad categories of variables that contribute significantly to students' academic achievement in the university. These four categories of variables identified by their study are: the category of program admitted into, the nature of the test administered to students, students' behavior, and the school environment. These variables are identified as variables that have an impact on the performance of the students.

Nebojša et al. (2021) carried out research making use of using ANN algorithm to forecast student's performance on a course. They make use of MLP with a backpropagation learning algorithm (BPP). The study also adopted the cross-validation methodology for the training and testing of the classifiers. The transformation process was performed on the test points of the students to get different categories related to performance. The results show that the forecasting algorithm attained an accuracy of 92.3%. For the study to implement the algorithm, the study designed a Web-based program to show an outcome.

Dhilipan et al. (2021) carried out a study that used a DT, SVM, and NB algorithms to predict the performance of the students. The results of their study revealed that the NB algorithm happens to have the best result of 77.0% in terms of the accuracy of the prediction model, the DT was second with 71.0% accuracy, and then SVM with 38.0% accuracy. Vamshidharreddy et al. (2020) studied different ML algorithms like binomial logical regression, DT, entropy, and K-NN to predict the student's success rate. The results of the study revealed that binomial logical regression came out with an accuracy of 97.05%, entropy produced at 91.19%, K-NN has 93.71%, and the DT model has an accuracy of 88.23%.

Oferi et al. (2020) conducted a literature-based review. The study focused on the application of ML techniques to forecast students' success to come out with an improved success rate. The study compares different types of ML algorithms to test for prediction accuracy. The results of the study came out with different results based on the model outcome. The results of the study could not identify the best ML model suitable for prediction. The study concludes that various variables are responsible for students' academic success (Oferi et al., 2020).

The main objective of the study is to design a model that will predict students' performance using a different types of ML algorithms (Mehil et al., 2019). The final results of the study show that the graduate booting algorithm has the highest level of prediction with 93.8% accuracy.

Zacharis (2016) carried out a study to predict the success of the student in a particle course; he sampled students' information kept in a Moodle server to predict students' performance in a course using four different learning activities stored in the server. The study trained a MLP neural network which he used to predict student performance in a blended learning course environment. The results of the model predicted the performance of students with a correct classification rate of 98.3%. This shows that the MPPNN has a significant impact on predicting student performance.

3. Methodology

3.1. Study area

The study was conducted in some selected tertiary institutions in Adamawa State, namely Adamawa State University, Mubi, Adamawa Polytechnic, Yola, and Modibbo Adama University, Yola.

3.2. Population

The population of this study comprises students admitted to the Modibbo Adama University, Yola, Adamawa State University, Mubi, and Adamawa State Polytechnic, and data used for this study were collected from 2017 to 2020.

3.3. Data collection process and sampling

The goal of this paper is to predict students' academic performance; it means that the study will apply a classification model for the prediction relying on a dataset from some selected

educational institutions in Adamawa State. To obtain the necessary data, a questionnaire was organized and distributed to students in the selected educational institutions manually.

The study sampled 720 students from three selected tertiaries institutions in Adamawa State, Nigeria. Three hundred students were selected from Modibbo Adama University, Yola, 300 students were selected from Adamawa State University, Mubi, and 120 students were selected from Adamawa State Polytechnic, Yola. Random sampling was used as a method of sampling techniques across schools in the study. The study makes use of a multistage sampling procedure to select students for the study. In the first stage, the schools (faculties) in each institution were selected using probability equals to one sampling strategy. In the second stage, three departments in each school (faculty) were selected using a simple random sampling technique (balloting with replacement). In the third stage, 720 respondents comprised of students were administered questionnaires in the selected tertiary institutions in Adamawa State. Information was elicited from these students using a questionnaire that was administered to them.

3.4. Dataset description

The dataset requirement for this research is fulfilled through a questionnaire and student grades from the universities database. The data attributes include the student's sex, pre-tertiary school status, parent socioeconomic status, personality traits, reading style, hours of study per day, health, IQ, and CGPA. Table 1 shows the description of dataset variables.

4. Data Preparation

After collecting the data, the next step was the data preparation stage. The collected data that were made up of both string and numeric variables cannot be modeled by the ANN; therefore, the data were prepared to make it more amenable for the algorithm to make predictions. The data that have been collected through the questionnaire and from the student's academic records were exported into an Excel sheet. The exported dataset was coded into a MS Excel sheet and converted into a comma-separated values (csv) file. Then, the data were reviewed and modifications were made where necessary on the dataset.

5. Data Preprocessing

Preprocessing plays a very important role in data mining. Its purpose is to convert raw data into a suitable form that can be used by a ML algorithm. Data preparation tasks include data integration, data cleaning, and discretization.

Two hundred and seventy datasets that were collected from the three tertiary institutions had 43 columns. The datasets that were 43 columns initially were cleaned to 13 columns and were discretized. The collected data were preprocessed. The final data items used in predicting students' academic performance are shown below in Table 2. The independent variables for academic performance prediction were gathered from both students and exam officers of the various departments; these variables include the sex/gender of the students, pre-tertiary school status of the students, the parent's socioeconomic status, study style, the nature and quality of the learning environment, the number of hours out into study by students per day, personality traits, festivity, regularity to class, the health of the students, the student intelligent quotient, and the student CGPA. The dependent variable or the target class is the predicted academic performance of the students which are ranked in four levels; the four levels are poor, low high, and exceptional performance. The predicted variable has four classes in which the students are classified – exceptional, high, low, and poor. Since the objective is on predicting students at risk of delay in graduation or dropping out, students with a CGPA below 2.50 are classified as poor and will need strong intervention, students with a CGPA between 2.50 and 3.00 are classified as low, needing less support, and students with CGPA between 3.00 and 4.50 classified as high, they are fine but may need encouragement to be exceptional students while students with CGPA above 4.50 are classified as exceptional students, they are fine without any intervention.

Table 1
Dataset description

S/N	Variable	Description of the variable with property
1	Sex	Male (1), female (2)
2	Pre-ter sch st	Urban, suburban, or rural (U,S,R)
3	Parent socio-econ st	1 (high), 2 (moderate), 3 (low)
4	Studying style	Group (0), single (1), both (2)
5	Qual of Edu Env	From 1 (very low) to 5 (very high)
6	Hrs study per day	1 (<3 hrs), 2 (3 to 6 hrs), 3 (7 to 9 hrs), 4 (>9 hrs)
7	Regularity to class	1 (high) 2 (moderate) 3 (low)
8	Festivity	From 1 (very low) to 5 (very high)
9	Personality traits	From 1 (very low) to 5 (very high)
10	Health	From 1 (very bad) to 5 (very good)
11	IQ	From 1 (very low) to 5 (very high)
12	CGP	From 1 (poor) to 4 (exceptional)

Table 2
Preprocessed students' records for academic performance prediction

```
In [1]: import pandas as pd
        dataframe = pd.read_csv("PerformancePredictData.csv")
        #dataframe[dataframe.Class=='Poor']
        dataframe
```

Out[1]:

	Sex	Pre-ter_Status	Parent_socio_econ	Studying Style	Qua.Edu. Evironment	HRs Study per Day	personlity traits	Festivity	Regularity to Class	Health	I.Q	CGPA	Class
0	0	0	3	2	3	2	2	4	2	4	4	2.36	Poor
1	1	0	2	2	3	3	2	5	3	5	5	1.71	Poor
2	0	0	3	2	3	1	2	2	1	5	1	1.73	Poor
3	0	0	1	2	5	1	1	3	2	5	5	2.80	Low
4	0	1	3	2	3	1	2	3	2	3	3	2.66	Low
5	1	0	3	2	3	3	1	5	1	5	4	2.19	Poor
6	0	0	1	0	1	2	1	1	2	5	3	2.55	Low
7	1	0	2	1	5	2	2	1	3	4	5	3.22	High
8	1	1	2	1	2	2	2	3	3	5	5	4.19	High
710	1	1	3	2	5	2	2	2	2	4	5	3.35	High
711	1	0	2	0	3	3	2	5	2	5	3	2.63	Low
712	1	0	3	2	4	3	2	1	2	5	3	2.60	Low
713	0	0	3	1	3	2	2	5	3	5	5	4.00	Exceptional
714	1	1	3	2	3	1	2	3	3	4	5	3.99	High
715	0	1	3	0	2	2	2	4	3	5	5	4.26	Exceptional
716	0	0	3	1	3	1	2	4	3	4	4	4.00	Exceptional
717	1	0	3	0	5	1	2	5	3	5	5	4.05	Exceptional
718	0	0	2	1	5	3	2	4	3	5	5	3.67	High
719	0	1	3	1	3	1	2	3	3	3	4	3.32	High

720 rows x 13 columns

6. Modeling and Experiment

Jupyter Notebook, a Python Anaconda powerful developmental wide-range classification and prediction platform for ANN, was used to build the model for predicting students' academic performance. The building of the model was done after the dataset collected was preprocessed, and missing data were removed. The neural network library known as Keras was used for the experiment. It is a sequential model of stacks layers inbuilt into the python programming language that supports the multi-layer network. After the datasets were partitioned and initial parameters set, the model was trained at 50 epochs in order to obtain the performance

Figure 1
The neural network training behavior of the datasets at optimum performance

Layer (type)	Output Shape	Param #
dense_346 (Dense)	(None, 55)	715
activation_1 (Activation)	(None, 55)	0
dropout_201 (Dropout)	(None, 55)	0
dense_347 (Dense)	(None, 40)	2240
dropout_202 (Dropout)	(None, 40)	0
dense_348 (Dense)	(None, 60)	2460
dropout_203 (Dropout)	(None, 60)	0
dense_349 (Dense)	(None, 4)	244
activation_2 (Activation)	(None, 4)	0

Total params: 5,659.0
Trainable params: 5,659.0
Non-trainable params: 0.0

of the model. The performance of the neural network reached optimum at a performance accuracy of 97.36%. As the model trains through the epochs, the ANN learns and becomes familiar with the patterns of the datasets. Consequently, the loss function value starts reducing while the accuracy starts increasing. This continues steadily to the end of the epoch. The model training values are consistent with the good learning behavior of the neural network.

Moreover, from the Keras sequential model, and based on the training and performance accuracy, the final model is as shown in Figure 1 below.

Generally, the ANN architecture is usually in layers and each layer is made up of neurons. The pictorial representation of the final model of the neural network is shown below. This also represents the neural network model for students' academic performance as shown in Figure 2 below.

Figure 2
The artificial neural network architecture for students' academic performance

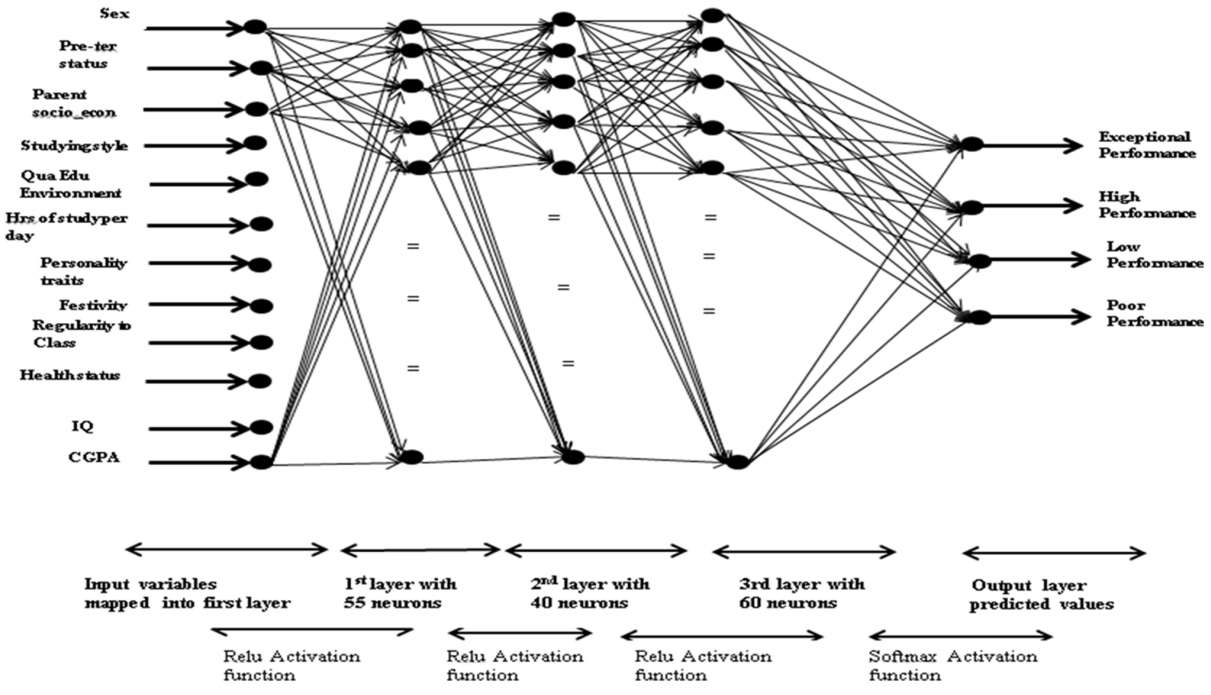
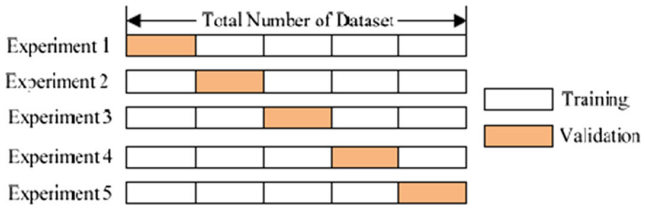


Figure 3
The cross-validation data partitions



K-fold cross-validation was used to evaluate the performance of the model in this study. Cross-validation methods are usually adopted as a performance measure of neural network models to assess the statistical relevance of the classifier.

$$\text{Accuracy} = \text{Number of predictions}/\text{Number of K-folds}$$

The procedure is to create a k-fold partition of the whole dataset, repeat K times to use K-1 folds for training and a left fold for validation, and finally average the error rates of K experiments. The $k = 5$ is statistically coherent in this study since the datasets have 720 data features or data items. The k-fold of 5 will give equal segmentation of the datasets before the final performance is obtained as shown in Figure 3 above.

$$\text{Accuracy} = \text{Number of predictions}/\text{Number of K-folds}$$

The number of predictions of the k-fold is 5; from the neural network setup, the performances of the algorithms in the neural network in the python code are: [0.97916667 0.97222222 0.97222222 0.97916667 0.96527778]

$$\text{Accuracy} = 0.97916667 + 0.97222222 + 0.97222222 + 0.97916667 + 0.96527778/5$$

$$\text{Accuracy} = 97.36\%$$

7. Model Prediction

The obvious reason for building a ML model is to make predictions; therefore, the last and important part of ANN model building is to predict unseen data called the test dataset. The academic performance prediction of students using the ANN model was a multi-class classification model that generates four outputs/labels based on the learning of the model through earlier training on input datasets. The four expected output samples from this study are

i. Exceptional performance
ii. High performance
iii. Low performance, and
iv. Poor performance

To make predictions, this work used the model designed to generate predictions on new data or the unseen dataset called the test dataset by calling the predict function such as predict () on the model.

There are four classes expected at the output layer of the ANN model. These are exceptional, high, low, and poor performance. These were the label in the datasets which were fed into the ANN; however, these strings in the output layer can be simulated and trained with the numerical data in the dataset as shown in Figure 4 below.

Figure 4
The encoded prediction code and corresponding meaning

```
========== Encoded Prediction Codes ======
[3 1 2 3 2 2 2 3 3 3 1 3 1 2 3 2 2 3 3 1 1 3 3 2 1 3 3 3 2 3 3 2 2 3 3 3 3
 3 3 3 2 1 2 2 2 3 1 3 3 3 2 3 3 2 3 2 1 3 2 2 3 1 3 3 3 3 1 3 1 2 3 1 1 2
 1 1 3 2 3 2 1 2 1 1 2 1 3 3 2 3 2 3 3 2 3 2 3 3 2 2 1 2 2 1 2 1 3 2 2 3 2
 3 1 1 1 1 2 2 2 2 3 2 3 2 2 3 1 3 2 2 1 1 3 3 2 2 2 2 3 2 2 3 1 2 2 3 1 2
 3 3 1 3 3 3 3 1 3 3 1 3 3 1 3 3 3 3 1 2 3 3 3 3 3 2 2 3 2 3 2 3 2 2 3 3 1
 2 2 1 1 3 3 2 2 1 3 1 1 3 3 3 3 2 1 1 2 1 2 2 2 1 1 2 2 2 2 1]
===== Corresponding Meaning of Codes======
['Poor' 'High' 'Low' 'Poor' 'Low' 'Low' 'Low' 'Poor' 'Poor' 'Poor' 'High'
 'Poor' 'High' 'Low' 'Poor' 'Low' 'Low' 'Poor' 'Poor' 'High' 'High' 'Poor'
 'Poor' 'Low' 'High' 'Poor' 'Poor' 'Poor' 'Low' 'Poor' 'Poor' 'Low' 'Low'
 'Poor' 'Poor' 'Poor' 'Poor' 'Poor' 'Poor' 'Poor' 'Low' 'High' 'Low' 'Low'
 'Low' 'Poor' 'High' 'Poor' 'Poor' 'Poor' 'Low' 'Poor' 'Poor' 'Low' 'Poor'
 'Low' 'High' 'Poor' 'Low' 'Low' 'Poor' 'High' 'Poor' 'Poor' 'Poor' 'Poor'
 'High' 'Poor' 'High' 'Low' 'Poor' 'High' 'High' 'Low' 'High' 'High' 'Poor'
 'Low' 'Poor' 'Low' 'High' 'Low' 'High' 'High' 'Low' 'High' 'Poor' 'Poor'
 'Low' 'Poor' 'Low' 'Poor' 'Poor' 'Low' 'Poor' 'Low' 'Poor' 'Poor' 'Low'
 'Low' 'High' 'Low' 'Low' 'High' 'Low' 'High' 'Poor' 'Low' 'Low' 'Poor'
 'Low' 'Poor' 'High' 'High' 'High' 'High' 'Low' 'Low' 'Low' 'Low' 'Poor'
 'Low' 'Poor' 'Low' 'Low' 'Poor' 'High' 'Poor' 'Low' 'Low' 'High' 'High'
 'Poor' 'Poor' 'Low' 'Low' 'Low' 'Low' 'Poor' 'Low' 'Low' 'Poor' 'High'
 'Low' 'Low' 'Poor' 'High' 'Low' 'Poor' 'Poor' 'High' 'Poor' 'Poor' 'Poor'
 'Poor' 'High' 'Poor' 'Poor' 'High' 'Poor' 'Poor' 'High' 'Poor' 'Poor'
 'Poor' 'Poor' 'High' 'Low' 'Poor' 'Poor' 'Poor' 'Poor' 'Poor' 'Low' 'Low'
 'Poor' 'Low' 'Poor' 'Low' 'Poor' 'Low' 'Low' 'Poor' 'Poor' 'High' 'Low'
 'Low' 'High' 'High' 'Poor' 'Poor' 'Low' 'Low' 'High' 'Poor' 'High' 'High'
 'Poor' 'Poor' 'Poor' 'Poor' 'Low' 'High' 'High' 'Low' 'High' 'Low' 'Low'
 'Low' 'High' 'High' 'Low' 'Low' 'Low' 'Low' 'High']
```

Figure 5
The summary of training and test datasets

```
-------- Summary of Datasets uded -----------
('Training dataset used out of 720,12:', (504, 12))
('Labels used corresponding to the trained sets', (504, 4))
('Test datasets used out of 720,12:', (216, 12))
('Labels used for testing', (216, 4))
```

From Figure 5 above, a total of 504 (70%) were used as training data, while 216 (30%) samples were used as test datasets.

The results in Table 3 show that variables such as number of hours studied per day, personality traits, festivity, regularity to class, health status, and IQ have a P value <0.05 which shows that statistically, those variables contribute significantly to the outcome of student's academic performance. On the other hand, variables such as sex or gender of the students, pre-tertiary status, parents' socioeconomic status, study style, and quality of educational environment have a correlation coefficient or the P value >0.05, this explains that those variables are statistically insignificant to the student's academic performance based on the dataset used in this ANN modeling, that is, level of significance (∞ >0.05).

8. Discussion

This study predicted students' academic performance with the use of ANN. The ANN model had four layers: the input layer, two middle layers, and the output layer. Twelve variables of the input layer were mapped into 55 neurons, the two middle layers had 40 and 60 neurons, respectively, and the output layer has 4 neurons that are representing the 4 classes of the expected outcome which are exceptional performance, high performance, low performance, and poor performance.

This academic performance prediction showed an optimal performance at three hidden layers of 55, 40, and 60 neurons, respectively. The prediction accuracy achieved optimum at 97.36% and a standard deviation of 0.52%. Two activation functions used were relu and softmax.

From the graphical representations above in Figure 6(a), the model started making correct predictions after epochs 10. The learning phase was smaller compared to the correct prediction phase. This implies that the model learns faster to be able to make correct predictions on the dataset. The two-colored graphs showed the performance of the dataset's partitions for training and testing datasets.

Figure 6(b) showed the model loss values in graphical form. The loss value of the model for the training datasets performs better and is consistent with the model learning behavior as expected for a good model. The testing dataset slants downward from the upper left at a value of about 1.2 model loss value to the downward right at almost zero loss value showing the reduction of incorrect predictions as the model encounters more data. However, the model loss for the test dataset started off from 0.4 and plateaued at about 0.8, denoting that the test dataset despite

Figure 6
(a) and (b): The graph for model accuracy and model loss

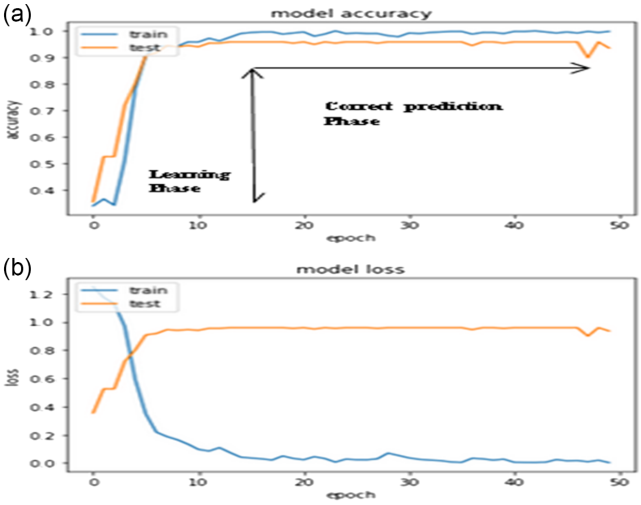


Table 3
Significant variables for student's academic performance

Variable	Description	Correlation coefficient	% of coefficient correlation
Sex	Gender (male or female)	0.058	2.45
Pre-ter status	The nature of pre-tertiary school attended (private or public)	0.032	1.35
Parental Socio.econ	Parent socio-economic background	0.000	00
Studying Style	Group or personal study style	0.060	2.54
Qua.Edu. Environment	The nature of the school environment and learning resources	0.003	0.13
HRs Study per Day	Hours of study by students	0.436**	18.42
Personality traits	Introvert or extrovert personality	0.111**	4.69
Festivity	Love for festivity activities and seasons	0.231**	9.76
Regularity to Class	Attending classes regularly	0.466**	19.69
Health status	Health status (whether the student's health condition is good or bad	0.099**	4.18
IQ	Intelligent quotient (the measure mental age of a student against the chronological age)	0.871**	36.79
CGPA	Cumulative grade point average	Outcome and basis for the predicted class	

having a prediction accuracy of 97% did not learn as faster as the training dataset. The result of the training and model loss agrees with the BPP which was used to train the datasets by default in the ANN.

The results obtained from this study revealed that developing ML models to make predictions is not achieved instantaneously. The process involves following laid down rules and conducting lots of experiments before a candidate model that optimizes a given problem is identified and selected.

Apart from the prediction accuracy of the ANN model, another important aspect of this study is to unravel the variables that contribute significantly to the overall prediction of the model. The study revealed that the variables that had a major significant impact on the overall predictive model are the student's IQ. The IQ had 36.79% of the total impact on student performance. The IQ variable stood out and has high significance in predicting the academic performance of the students. The next significant variable after IQ is the level of student attendance to classes coined regularity to classes, it accounts for about 19.69% of the total predictor. This means that class attendance contributes significantly to students' academic performance; hence, students need to attend class regularly. Other variables are the hour of study per day by the student which was 18.42% of the independent variables. Other variables that have a relatively significant high impact on a student's academic success are festivity, personality traits, and health status of the student, each having 9.76%, 4.69%, and 4.18%, respectively. This signifies that a festive inclined student may also have an outstanding personality trait, although that seems to inhibit performance of the student but it also accounts for a student's ability to collaborate with other students or the learning environment. These traits could positively impact students' academic performance. Lastly, the health status of a student can also be a major predictor of a student's academic success.

On the other hand, other variables like sex or gender of the students, pre-tertiary institution status, parental socioeconomic status, and quality of educational environment did not show any statistical relevance in the prediction of student academic performance and therefore could not have had a significant impact on the ANN model decision.

The application of this model will go a long way in curtailing the challenges faced by tertiaries institutions in Adamawa State, Nigeria. The ANN predictive model has identified significant factors that affect students' academic performance; hence, this will help the school management in taking decisions on intervening for the weak and poor-performing students. The application of this model will improve educational institutions' success rate and cut down the failure rate. The application of this predictive model will also improve the learning mechanism.

9. Conclusion

Accurate student academic performance prediction model plays an important role in any educational institution. The ability to handle data quality issues in a student's academic performance model has remained a major challenge. This paper presented a student's academic performance predictive model based on a supervised learning technique using an ANN. The performance of the predictive model is assessed on a dataset provided for modeling. The result shows that the model attained accuracy at 97%. The study also indicated that students' IQ and regularity in class are the two variables that contributed significantly to predicting students' academic performance in the model. This means that in real-life scenarios these significant variables contribute to students' success. Students should be encouraged in developing good IQ, attend class regularly, develop a steady reading habit, and have good personality traits. This will contribute significantly to the student's performance as revealed by the predictive model in this study.

In the future, the proposed model will be tested on a large dataset with more numbers of attributes.

10. Recommendations

The study revealed that there are factors that are responsible for the high rate of poor students' academic performance in educational institutions in Adamawa State. Hence, the study recommended that the identification of such factors should aid educational institutions in improving the learning mechanism of the students. Tertiaries institutions in Adamawa State consider using EDM in predicting students' academic achievement because of its benefits in uncovering the students' learning attributes that matter in academic success. The study also recommends that universities and institutions should make appropriate use of data collected from students at the point of entry to develop a model to serve as a DSS to predict students' performance.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

References

Ahmed, A. B. E. D., & Elaraby, I. S. (2014). Data mining: A prediction for student's performance using classification method. *World Journal of Computer Application and Technology*, *2*(2), 43–47. https://doi.org/10.13189/wjcat.2014.020203

Alturki, S., & Alturki, N. (2021). Using educational data mining to predict students' academic performance for applying early interventions. *Journal of Information Technology Education: Innovations in Practice*, *20*, 121–137. https://doi.org/10.28945/4835

Amo, C. & Santelices, M. V. (2017). Trayectorias universitarias: más que persistencia o deserción. University of Panama.

Banik, P., & Kumar, B. (2019). Impact of information literacy skill on students' academic performance in Bangladesh. *International Journal of European Studies*, *3*(1), 27–33. https://doi.org/10.11648/j.ijes.20190301.15

Chakraborty, B., Chakma, K., & Mukherjee, A. (2016). A density-based clustering algorithm and experiments on student datasets with noises using Rough set theory. In *2016 IEEE International Conference on Engineering and Technology*, 431–436.

Dhilipan, J., Vijayalakshmi, N., Suriya, S., & Christopher, A. (2021). Prediction of Students Performance using Machine learning. In *IOP Conference Series: Materials Science and Engineering*, 1055(1), 012122.

Di Caudo, M. (2015). Política de cuotas en Ecuador: me gané una beca para estudiar en la Universidad. *Ponto-e-Vírgula: Revista de Ciências Sociais*, *1*, 196–218. https://doi.org/10.47212/tendencias_vii_2019_14

Eakasit, P. (2015). *An introduction to data mining techniques*. China: Asia Publisher.

Erdem, C., Senturk, I., & Arslan, C. K. (2007). Factors affecting grade point average of university students. *The Empirical Economics Letters*, *6*(5), 360–368.

Hussain, S., Muhsin, Z. F., Salal, Y. K., Theodorou, P., Kurtoğlu, F., & Hazarika, G. C. (2019). Prediction model on student performance

based on internal assessment using deep learning. *International Journal of Emerging Technologies in Learning*, *14*(8), 4–22. https://doi.org/10.3991/ijet.v14i08.10001

Heredia, D., Amaya, Y., & Barrientos, E. (2015). Student dropout predictive model using data mining techniques. *IEEE Latin America Transactions*, *13*(9), 3127–3134. https://doi.org/10.1109/tla.2015.7350068

Jamala J. M., & Elizabeth, S. I. (2022). Role of educational data mining in student learning processes with sentiment analysis: A survey. *International Journal of Knowledge and Systems Science*, *11*(4), 31–44. https://doi.org/10.4018/IJKSS.2020100103

Kiu, C. C. (2017). Data mining analysis student's academic performance through exploration of student's background and social activities. In *2018 Fourth International Conference on Advances in Computing, Communication & Automation*, 1–5.

Koç, Y., Terzioğlu, E. A., & Kayalar, F. (2018). Examination of individual achievement motivation and general self-efficacy of candidates entering into special talent exam in physical education and sports sciences. *Journal of Sports and Performance Researches*, *9*(2), 64–73. https://doi.org/10.17155/omuspd.356632

Kubiatko, M., Hsieh, M. Y., Ersozlu, Z. N., & Usak, M. (2018). The motivation toward learning among Czech high school students and influence of selected variables on motivation. *Revista de Cercetare si Interventie Sociala*, *60*, 79–93.

Masood, M. F., Khan, A., Hussain, F., Shaukat, A., & Zeb, B. (2017). Towards the selection of best machine learning model for student performance. In *2019 6th International Conference on Soft Computing & Machine Intelligence*, 12–17.

Meghji, A. F., Mahoto, N. A., Unar, M. A. & Shaikh, M. A. (2018). Analysis of student performance using EDM methods. In *2018 5th International Multi–Topic ICT Conference.*

Mehil B. S., Maheeka K., & Yogesh G. (2019). Student performance assessment and prediction using machine learning. *In 2019 4th International Conference on Information Systems and Computer Networks*, 386–390.

Ofori, F., Maina, E. & Gitonga, R. (2020). Using machine learning algorithms to predict students' performance and improve learning outcome: A literature-based review. *Journal of Information and Technology*, *4*(1), 33–55.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12–27. https://doi.org/10.1002/widm.1075

Sandoval, I., Sánchez, T., Naranjo, D., & Jiménez, A. (2019). Proposal of a mathematics pilot program for engineering students from vulnerable groups of Escuela politécnica Nacional. In *Proceedings of the LACCEI International Multi-Conference for Engineering, Education and Technology*, *1*, 387. https://doi.org/10.18687/LACCEI2019.1.1.387

Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, *72*, 414–422. https://doi.org/10.1016/j.procs.2015.12.157

Shah, M. B., Kaistha, M., & Gupta, Y. (2019). Student performance assessment and prediction system using machine learning. In *2019 4th International Conference on Information Systems and Computer Networks*, 386–390.

Silva, C., & Fonseca, J. (2017). Educational data mining: A literature review. Europe and MENA Cooperation Advances in Information and Communication Technologies, 87–94.

Stanković, N. L., Blagojević, M. D., Papić, M. Ž., & Karuović, D. (2021). Artificial neural network model for prediction of students' success in learning programming. *Journal of Scientific & Industrial Research*, *80*(3), 249–254.

Uluğ, F. (2000). *Success in school*. Remzi Bookstore.

Vamshidharreddy, V. S., Saketh, A. S., & Gnanajeyaraman, R. (2020). Student's academic performance prediction using machine learning approach. *International Journal of Advanced Science and Technology*, *29*, 6731–6737.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, *140*(4), 1174–1204. https://doi.org/10.1037/a0036620

Yakubu, M. N., & Abubakar, A. M. (2022). Applying machine learning approach to predict students'performance in higher educational institutions. Kybernetes, *51*(2), 916–934. https://doi.org/10.1108/K-12-2020-0865

Zacharis, N. Z. (2016). Predicting student academic performance in blended learning using artificial neural networks. *International Journal of Artificial Intelligence and Applications*, *7*(5), 17–29. https://doi.org/10.5121/IJAIA.2016.7502

How to Cite: Yauri, R. A., Suru, H. U., Afrifa, J., & Moses, H. G. (2023). A Machine Leaning Approach in Predicting Student's Academic Performance Using Artificial Neural Network. *Journal of Computational and Cognitive Engineering* https://doi.org/10.47852/bonviewJCCE3202470