

https://dspace.org/

Using machine learning to predict low academic performance at a Nigerian university

Ekubo, Ebiemi Allen; Esiefarienrhe, Bukohwo Michael

2022-12-23
LINK Centre, University of the Witwatersrand (Wits), Johannesburg
https://hdl.handle.net/10539/34923
Ekubo, E. A., & Esiefarienrhe, B. M. (2022). Using machine learning to predict low academic performance at a Nigerian university.The African Journal of Information and Communication (AJIC), 30, 1-33. https://doi.org/10.23962/ajic.i30.14839

Downloaded from a DSpace repository. https://dspace.org/

Using machine learning to predict low academic performance at a Nigerian university

Ebiemi Allen Ekubo
Lecturer, Department of Computer Science, Niger Delta University, Amassoma, Nigeria
https://orcid.org/0000-0001-9348-5630

Bukohwo Michael Esiefarienrhe
Associate Professor, Department of Computer Science and Information Systems, and Researcher, Material Science Innovation and Modelling (MaSIM), Faculty of Natural and Agricultural Sciences (FNAS), North-West University, Potchefstroom, South Africa
https://orcid.org/0000-0002-1631-9759

Abstract
This study evaluates the ability of various machine-learning techniques to predict low academic performance among Nigerian tertiary students. Using data collected from undergraduate student records at Niger Delta University in Bayelsa State, the research applies the cross-industry standard process for data mining (CRISP-DM) research methodology and the Waikato Environment for Knowledge Analysis (WEKA) tool for modelling. Five machine-learning classifier algorithms are tested—J48 decision tree, logistic regression (LR), multilayer perceptron (MLP), naïve Bayes (NB), and sequential minimal optimisation (SMO)—and it is found that MLP is the best classifier for the dataset. The study then develops a predictive software application, using PHP and Python, for implementation of the MLP model, and the software achieves 98% accuracy.

Keywords
machine learning, educational data mining, student academic performance, university, cross-industry standard process for data mining (CRISP-DM), Waikato Environment for Knowledge Analysis (WEKA), classifier algorithms, J48 decision tree, logistic regression (LR), multilayer perceptron (MLP), naïve Bayes (NB), sequential minimal optimisation (SMO), Nigeria, Niger Delta University

Acknowledgements
The authors thank the Material Science Simulation and Modelling (MaSIM) research focus area in the Faculty of Natural and Agricultural Sciences (FNAS), the School of Postgraduate Studies, and the Department of Computer Science and Information Systems, all at North-West University, for their assistance throughout this research. We also thank Niger Delta University for making its data available for research purposes. This paper draws on the contents of the first-listed author's PhD thesis at North-West University (Ekubo, 2020).

DOI: https://doi.org/10.23962/ajic.i30.14839

Recommended citation
Ekubo, E. A., & Esiefarienrhe, B. M. (2022). Using machine learning to predict low academic performance at a Nigerian university. *The African Journal of Information and Communication (AJIC), 30*, 1-33. https://doi.org/10.23962/ajic.i30.14839



This article is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence: https://creativecommons.org/licenses/by/4.0

1. Introduction

Low academic performance is a challenge faced by many higher education institutions worldwide (Romero et al., 2010). The challenge is more serious in developing countries, where the phenomena undermining academic success are more numerous and more pronounced than in developed-world settings (Al-Zoubi & Younes, 2015). The UN has identified the challenges present in developing countries that ultimately result in low academic performance as poverty, unstable electricity, lack of funding, diverse health challenges, environmental degradation, level of social development (low school enrolment rate), vulnerability in employment, maternal mortality, and quality of life of slum dwellers (UN DESA, 2013). With regard to low academic performance of undergraduate students, the research shows that the effects on students can be long-term, resulting in low self-esteem, unease, and fear of failure (Aryana, 2010; Nsiah, 2017; Nurmi et al., 2003). For tertiary institutions, low academic performance limits their growth in terms of, inter alia, execution of operations and rankings in global indices (Serdyukov, 2017). Thus, the challenge needs continuous monitoring and the creation of incentives to help boost students' progress.

The issue of low academic performance is prevalent in Nigeria, where many universities record a high number of low-performing undergraduate students (Oyebade & Dike, 2013). Educational research in Nigeria concerned with low academic performance has considered numerous factors, including those related to the student, the environment, the institution, and the relevant level(s) of government (Farooq et al., 2011).

The discipline of data mining seeks to discover useful patterns in large sets of data, so as to predict future outcomes (Gullo, 2015). In education, the use of data mining has led to improvements in predicting factors that will render certain students prone to struggling with their studies, to failure, and to dropping out (Hughes & Dobbins, 2015). Results from these predictions can assist stakeholders in addressing the factors identified as contributing to poor student performance. Therefore, it is necessary for tertiary institutions to make use of this useful prediction method.

Some research carried out in Nigeria using data-mining techniques with educational data has already looked at predicting students' performance using selected attributes (Adeyemo & Kuye, 2006; Oyerinde & Chia, 2017). These studies predicted students' performance using various attributes, but did not develop software based on their prediction to monitor and track the progress of students' performance. The absence of a dedicated software tool that enables Nigerian tertiary institutions to track their students' performance, despite the abundance of packages for data mining, justified this study's efforts to identify the highest-performing data-mining algorithm and to implement it in the form of customised software. This study used a case study of a single university.

Universities need to document the attributes of low performers, classify the low performers using data-mining techniques, and develop a model that can identify likely low performers. The model developed can serve as a foundation for the design and implementation of systems that can provide potential low performers with the assistance that they need to perform better. The goal of this study was to develop such a model, based on data from a single Nigerian university—Niger Delta University in Bayelsa State—and with potential applicability to multiple universities.

This research sought to: (1) identify factors that are causing the poor performance of undergraduates in Nigeria; (2) collect and represent these factors as features in machine-readable format for data mining; (3) identify the best set of features from the total features collected for predicting low academic performance; (4) identify which machine-learning technique could best classify low-performing students based on the selected features; and (5) develop a customised software system that uses the identified best machine-learning algorithm to identify low-performing students and automatically recommend various interventions.

Using data collected from the university's undergraduate student records and applying the cross-industry standard process for data mining (CRISP-DM) research methodology and the Waikato Environment for Knowledge Analysis (WEKA) tool for modelling, five machine-learning classifier algorithms were used to determine which one was the most accurate in predicting poor student performance.

Section 2 of this article reviews related literature; section 3 discusses the methodology followed; section 4 presents the testing of the five classifier algorithms; section 5 presents the feature selection process; section 6 presents the development and evaluation of the predictive software application; and section 7 provides conclusions and a recommendation.

2. Literature review

Educational data mining is the multidisciplinary research area that applies data-mining techniques to educational data (Romero & Ventura, 2013). The educational environment regularly produces large amounts of data (Romero & Ventura, 2013), and with the use of data-mining techniques, stakeholders can gain knowledge to help them understand learners and improve their learning process (Algarni, 2016). Using educational data-mining techniques can also benefit society in general, as every society thrives socially and economically when its education system performs optimally (Mitra, 2011).

Yağcı (2022) proposes a model that predicts the final exam grades of students using their midterm exam grades. Six algorithms—random forest (RF), nearest neighbour, support vector machines (SVMs), logistic regression (LR), naïve Bayes (NB), and k-nearest neighbours (K-NN)—are used, and their accuracy falls within the range of 70 to 75%. Thus, the Yağcı (2022) study contributes to identifying high-risk students who are likely to fail their exams. Our study was different from that that of Yağcı (2022), because we used the cumulative grade point average (CGPA), which is technically a student's average of all courses, instead of a single exam grade for a subject. Dhilipan et al. (2021) use four data-mining algorithms to predict the performance of students. The four algorithms used, and their levels of accuracy, are: LR (97.05%), decision tree (88.23%), entropy (91.19%), and K-NN (93.71%). While the Dhilipan et al. (2021) focuses on prediction of student performance in general, our study was focused more on identification of low-performing students. In our study, not all students' data was used—only the data related to low-performing students. Also, the algorithms we used differed from those used by Dhilipan et al. (2021).

Vergaray et al. (2022) present a stacking multi-classification model for mining academic performance of students, using five classification algorithms, with their output as a feeder to the stacking model. The five classification algorithms with their accuracies are: extra trees (ET) (57.41%), RF (61.96%), decision tree (91.44%), adaptive boosting (AdaBoost) (59.65%), and extreme gradient boosting (XGBoost) (83.3%). The proposed stacking model is then applied, combining the five algorithms, enabling accuracy of 92.86%.

So as to predict student performance at a very early stage, Li et al. (2022) propose an end-to-end deep-learning model that can automatically (with no manual intervention) extract features from datasets to form a two-dimensional convolutional neural network (CNN) of behaviours reflective of the dataset. Their experiment is conducted using university students' data from Beijing, and the results show improvement over traditional data-mining algorithms. Ofori et al. (2020) present findings from a review of literature related to using machine-learning algorithms

in the prediction of students' academic performance and their learning outcomes. The reviews show various algorithms used in the prediction of students' academic performance and the highest levels of accuracy achieved.

Abu Zohair (2019) mines data on postgraduate students' performance using 50 students' datasets obtained from a university. The authors split their data into dataset1 and dataset2, and use five data-mining algorithms to train their datasets, with the following results in terms of accuracy: multilayer perceptron (MLP) (60.5%), NB (71.1%), SVMs (76.3%), K-NN (65.8%), and linear discriminant analysis (LDA) (71.1%). Flanagan et al. (2022) use the linear kernel SVMs model to predict the performance and engagement of students' behaviour using digital textbooks during an open book test. A good level of accuracy is achieved in identifying, at an early stage, students who are low performers.

Factors that influence low academic performance generally vary from society to society, and from individual to individual. In many developed countries, low academic performance often relates to personal issues such as a lack of inspiration to study or emotional trauma resulting from certain situations (Banerjee, 2016). However, in many developing countries, factors emanating from institutions and the government, in addition to individual factors, also contribute to poor academic performance. Thus, the causes of poor academic performance in developing countries can broadly be divided into individual, institutional, and governmental factors.

The individual factors relate to students' capacity to focus and concentrate on their academic work. Some individual factors that affect the performance of students in the Arab countries of Jordan and Oman, as outlined by Al-Zoubi and Younes (2015) and Alami (2016), are: poor motivation, lack of planning, low self-confidence, fretfulness about exams, poor examination practices, low opinion of the course, lack of interest in the course, laziness, and lack of future plans. Other individual factors identified as affecting low student performance are medical or mental problems, and the family's financial background (Al-Zoubi & Younes, 2015). Factors associated with low tertiary performance in Nigeria include students' uncertainty about their future because of high levels of unemployment; the large number of low-salary earners; lack of funding (e.g. bursaries); a lot of competition for scholarships; self-sponsoring students' distractions due to their engagement in temporary jobs; engaging in small-scale business activities such as buying and selling of goods and services, which then occupies most of their time, leaving little for studies; lack of planning; poor study habits; negative peer influence; family crises; lack of support from guardians or parents; students' view of a course as difficult; and the family's financial or psychological background (Oyebade & Dike, 2013).

Contributing factors associated with tertiary institutions, according to Frimpong et al. (2016), writing in the Ghanaian context, include poor conditions in the learning environment and educators' inadequate knowledge of educational and psychological matters.

Government factors include unstable electricity within and outside the school environment; insufficient security in learning environments; failure to adequately address high levels of poverty in the country; insufficient budgets for the education sector; and pay disputes between the academic staff of higher education institutions and government, which result in regular strike activities (Adeyemi & Adeyemi, 2014; Longe, 2017; Ugar, 2018).

3. Methodology

The process model followed in this study was the widely used CRISP-DM. It involves six steps (Wirth & Hipp, 2000):

- *Business understanding:* Scouting relevant stakeholders and assembling vital information to make sure the goals of the research are achievable.
- *Data understanding:* Data gathering and investigation to ensure that the data has quality and shows useful patterns.
- *Data preparation*: Cleaning and converting gathered data into a suitable format for mining and ensuring the dataset fits the selected modelling tool.
- *Modelling:* Use of the cleaned dataset and application of selected algorithms for knowledge mining.
- *Evaluation:* Examination of the output of the modelling step to ensure the knowledge gained is in line with the proposed goals of the project.
- *Deployment:* Display of the discovered knowledge by either integrating it into an existing system or developing a new system to assist stakeholders in benefiting from the knowledge gained (Wirth & Hipp, 2000).

In line with the CRISP-DM process model, the study consisted of the following steps: a survey of literature (as summarised above in section 2) on data-mining of educational data and on the causes of low performance in Nigerian universities; the gathering of student details from Niger Delta University and the storage of the details in Microsoft Excel; the testing of five machine-learning algorithms to determine the best algorithm for identifying low student performance based on collected data; using four feature selection techniques to choose the optimal features from the dataset; and designing and testing a predictive system, using the PHP programming language, for the identification of potential low performers.

The CRISP-DM methodology was used in this study because it provided all the phases necessary for our research endeavour, including a phase for data collection and pre-processing (necessary for us as our data was mostly in the form of physical

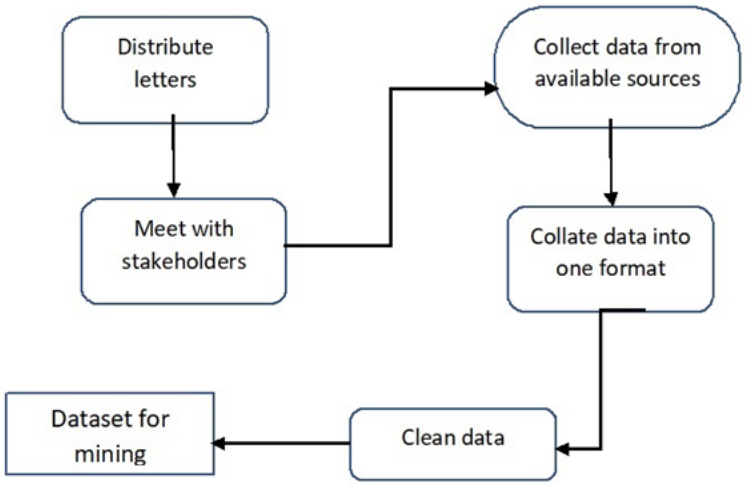
files obtained from the university administrator). The methodology also provided a phase for data mining and then construction of a solution for implementation. Using the CRISP-DM methodology saved us from having to combine numerous methodologies for the different parts of the research, particularly the software development (we relied on the CRISP-DM deployment phase to develop the software). The WEKA tool was used for modelling because of its simplicity and its ability to present results in a concise form.

The study used data stored in the Niger Delta University repository, and sampled 2,348 low-performing students' data, i.e., data on students with a CGPA of less than 3.00 (CGPA < 3.00).

Data collection

Figure 1 shows the complete data collection process followed in this study. The process involved letters being distributed to key stakeholders for permission to obtain data, which led to a meeting with stakeholders to agree on terms regarding the data-collection process and privacy issues. The next steps were the collection and collation of existing data from the available sources into one format, cleaning data to rid it of inaccurate or incomplete information, and preparation of the data for mining.

Figure 1: The data collection process



Source: Ekubo (2020, p. 36)

The lead researcher (the first-listed author, Ekubo) had a meeting with the Deputy Vice-Chancellor (Academic) of the Niger Delta University, who oversees the handling of all information related to students' academic results. The DVC assisted the researcher by sending out memos to all faculty deans, heads of department,

faculty officers, and faculty examination officers to assist the researcher in data-gathering from the university repositories. The faculty officers stored the students' details in spreadsheet files and gave the researcher the data on USB flash drives. The collection of students' results data involved meeting with examination officers within departments. Where the examination officers were unavailable, the heads of department provided the data, which was in either pdf or hard copy files. The researcher transferred the files on USB flash drives immediately to his laptop and stored all hard copy files in a file jacket.

Data preparation

The data collected was incomplete and inaccurate. The students' details collected from the university comprised 10,472 records. After manually inserting CGPAs from the pdfs and hard copies, the total number of students with assigned CGPAs comprised 5,631 records, of which students with CGPAs less than 3.00 comprised 3,481 records. This formed the entire population of low-performing undergraduate students used in this research.

Attribute selection

At the end of the data preparation process, one class attribute (CGPA, the core attribute used for the prediction) and 24 dependent attributes were selected for mining. Table 1 shows the class attribute and 24 dependent attributes, the attributes' variable codes, and the codes' corresponding values. In the table, "JAMB score" refers to the student's score on the exam provided by the Joint Admissions and Matriculation Board (JAMB)—the exam called the "UTME" (the Unified Tertiary Matriculation Examination). The "course from JAMB" attribute indicates whether the course the student was admitted to study was the actual course the student chose to study from JAMB. The "post-UTME score" is the score the student obtained on the university's internal entrance examination.

Also in Table 1, "SSCE" refers to the Senior Secondary School Certificate Examination. The student's "average SSCE score" was calculated as the average of the student's pre-university academic performance. Nigerian high school students write exams in a minimum of seven subjects and a maximum of nine subjects, and earn grades A, B, C, D, E, or F, based on their performance. For the purpose of the research, the values 6, 5, 4, 3, 2 or 1 replaced the respective grades, and the sum for each student divided by the number of subjects the student wrote gave the value of the average SSCE score.

Table 1: The attributes used in the data mining

Attribute	Variable codes	Values
Class attribute: CGPA	HL LL	2.50 – 2.99 0.01 – 2.49
Sex	M F	Male Female
Age	B30 30A	Below 30 years 30 years and above
Marital status	S M	Single Married
Attended primary school	NO YES	No Yes
Secondary school type	PRI PUB	Private Public
Secondary school area	URB RUR	Urban Rural
Sponsor type	GUAD PAR SELF	Guardian Parents Self-sponsor
Sponsor qualification	DEG NODEG NOEDU	Educated with degree Educated without degree No formal education
Sponsor income (per month)	LOW MED HIGH	Below N50,000 N50,000 – N100,000 Above N100,000
Sponsor support	LOW MED HIGH	Little support Average support Great support
Family size	SMALL MED LAR	1 – 4 5 – 9 Above 9
Work and study	YES NO	Yes No
University accommodation	CMPS OFFCMPS	Campus Off-campus
Years before admission (gap between completing secondary school and commencing university)	NONE B5 5A	None Below 5 years 5 years and above
Course from JAMB	YES NO	Yes No
Course interest	LOW AVE HIGH	Little interest Average interest High interest

Attribute	Variable codes	Values
Weekly study time	LOW AVE HIGH	Less than 10hrs 10 – 20hrs Above 20hrs
Postgraduate degree	NO YES NS	No Yes Not sure
Own smart phone	YES NO	Yes No
Smart phone assistance	ASGMT STUDY NONE	For assignment For studying None
Sports activeness	LOW HIGH	A little active Very active
JAMB score	LOW AVE HIGH	Below 180 180 – 250 Above 250
Post-UTME score	LOW AVE HIGH	Below 180 180 – 250 Above 250
Average SSCE score	LOW AVE HIGH	Less than 4.00 4.00 – 4.99 5.00 and above

The dataset was split into two parts: one for training and the other for testing of the models to be built with the five machine-learning classifier algorithms chosen for testing:

- J48 decision tree;
- logistic regression (LR);
- multilayer perceptron (MLP);
- naïve Bayes (NB); and
- sequential minimal optimisation (SMO).

These algorithms were selected for testing based on their frequent use in existing similar studies (see literature reviewed in section 2) and their high rates of accuracy in such studies. The dataset also underwent a feature selection process, to retrieve the optimal features from the dataset for prediction of low academic performance, using the following four feature selection techniques:

- correlation;
- gain ratio;
- information gain; and
- ReliefF.

The next phase of the study was development of a predictive software application, using the best-performing machine-learning model (among the five tested) and the optimal features (as identified in the feature selection exercise). Finally, the study evaluated this predictive application with the test dataset to determine its effectiveness.

4. Testing of the five classifier models

In line with the objectives of this research, the study developed five machine-learning models. It started by splitting the entire dataset into two parts, using a 70:30 ratio. The first part, with 70% of the dataset, was for training, while the other 30% was for testing, i.e., testing the model built to ensure that the model performed well with unseen data. This study employed the resample filter available in WEKA for the splitting purpose. This filter ensured that the dataset split has no duplicates.

For the modelling process, the study used the WEKA tool to build models with the training dataset for the five selected machine-learning algorithms, namely J48 decision tree, LR, MLP, NB and SMO. Figures 2 to 6 below provide snapshots of the results from the modelling processes.

Figure 2: J48 decision tree model for training dataset

```
Time taken to build model: 0.07 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1566               95.3135 %
Incorrectly Classified Instances        77                4.6865 %
Kappa statistic                          0.8979
Mean absolute error                      0.0643
Root mean squared error                  0.2085
Relative absolute error                 13.8717 %
Root relative squared error             43.299  %
Total Number of Instances             1643

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.912    0.023    0.958      0.912    0.934      0.899    0.961     0.941     HL
               0.977    0.088    0.951      0.977    0.964      0.899    0.961     0.960     LL
Weighted Avg.  0.953    0.064    0.953      0.953    0.953      0.899    0.961     0.953

=== Confusion Matrix ===

   a    b   <-- classified as
 547   53 |    a = HL
  24 1019 |    b = LL
```

Figure 3: Logistic regression (LR) model for training dataset

```
Time taken to build model: 0.32 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1578               96.0438 %
Incorrectly Classified Instances        65                3.9562 %
Kappa statistic                          0.9143
Mean absolute error                      0.063
Root mean squared error                  0.1835
Relative absolute error                 13.5776 %
Root relative squared error             38.1076 %
Total Number of Instances             1643

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.935    0.025    0.956      0.935    0.945      0.914    0.982     0.979     HL
                 0.975    0.065    0.963      0.975    0.969      0.914    0.982     0.986     LL
Weighted Avg.    0.960    0.050    0.960      0.960    0.960      0.914    0.982     0.984

=== Confusion Matrix ===

    a    b   <-- classified as
  561   39 |    a = HL
   26 1017 |    b = LL
```

Figure 4: Multilayer perceptron (MLP) model for training dataset

```
Time taken to build model: 51.48 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1596               97.1394 %
Incorrectly Classified Instances        47                2.8606 %
Kappa statistic                          0.9381
Mean absolute error                      0.0302
Root mean squared error                  0.156
Relative absolute error                  6.5135 %
Root relative squared error             32.3991 %
Total Number of Instances             1643

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.952    0.017    0.969      0.952    0.960      0.938    0.992     0.991     HL
                 0.983    0.048    0.972      0.983    0.978      0.938    0.992     0.993     LL
Weighted Avg.    0.971    0.037    0.971      0.971    0.971      0.938    0.992     0.992

=== Confusion Matrix ===

    a    b   <-- classified as
  571   29 |    a = HL
   18 1025 |    b = LL
```

Figure 5: Naïve Bayes (NB) model for training dataset

```
Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1461               88.9227 %
Incorrectly Classified Instances       182               11.0773 %
Kappa statistic                          0.7601
Mean absolute error                      0.1164
Root mean squared error                  0.2901
Relative absolute error                 25.0931 %
Root relative squared error             60.2594 %
Total Number of Instances             1643

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.838     0.081     0.855       0.838   0.847      0.760    0.945     0.939     HL
               0.919     0.162     0.908       0.919   0.913      0.760    0.945     0.958     LL
Weighted Avg.  0.889     0.132     0.889       0.889   0.889      0.760    0.945     0.951

=== Confusion Matrix ===

   a   b   <-- classified as
 503  97 |   a = HL
  85 958 |   b = LL
```

Figure 6: Sequential minimal optimisation (SMO) model for training dataset

```
Time taken to build model: 1.19 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1580               96.1656 %
Incorrectly Classified Instances        63                3.8344 %
Kappa statistic                          0.9164
Mean absolute error                      0.0383
Root mean squared error                  0.1958
Relative absolute error                  8.2693 %
Root relative squared error             40.6697 %
Total Number of Instances             1643

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Class
               0.923     0.016     0.970       0.923   0.946      0.917    0.954     0.924     HL
               0.984     0.077     0.957       0.984   0.970      0.917    0.954     0.952     LL
Weighted Avg.  0.962     0.055     0.962       0.962   0.961      0.917    0.954     0.942

=== Confusion Matrix ===

    a    b   <-- classified as
  554   46 |    a = HL
   17 1026 |    b = LL
```

After successful modelling of the training dataset, the study used the testing dataset to test the models built to confirm how well each built model performed with the unseen dataset. Figures 7 to 11 are snapshots of the results from this process.

Figure 7: J48 decision tree model for testing dataset

```
=== Re-evaluation on test set ===

User supplied test set
Relation:     NewDataset2-weka.filters.unsupervised.instance.Resample-S1-Z70.0-no-replacement-V
Instances:     unknown (yet). Reading incrementally
Attributes:   25

=== Summary ===

Correctly Classified Instances         684               97.0213 %
Incorrectly Classified Instances        21                2.9787 %
Kappa statistic                          0.9349
Mean absolute error                      0.0443
Root mean squared error                  0.1685
Total Number of Instances              705

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.934    0.009    0.984      0.934    0.958      0.936    0.976     0.967     HL
                 0.991    0.066    0.963      0.991    0.977      0.936    0.976     0.977     LL
Weighted Avg.    0.970    0.046    0.971      0.970    0.970      0.936    0.976     0.973

=== Confusion Matrix ===

   a   b   <-- classified as
 239  17 |   a = HL
   4 445 |   b = LL
```

Figure 8: Logistic regression (LR) model for testing dataset

```
=== Re-evaluation on test set ===

User supplied test set
Relation:     NewDataset2-weka.filters.unsupervised.instance.Resample-S1-Z70.0-no-replacement-V
Instances:     unknown (yet). Reading incrementally
Attributes:   25

=== Summary ===

Correctly Classified Instances         680               96.4539 %
Incorrectly Classified Instances        25                3.5461 %
Kappa statistic                          0.9227
Mean absolute error                      0.0608
Root mean squared error                  0.1724
Total Number of Instances              705

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.934    0.018    0.968      0.934    0.950      0.923    0.988     0.986     HL
                 0.982    0.066    0.963      0.982    0.972      0.923    0.988     0.992     LL
Weighted Avg.    0.965    0.049    0.965      0.965    0.964      0.923    0.988     0.990

=== Confusion Matrix ===

   a   b   <-- classified as
 239  17 |   a = HL
   8 441 |   b = LL
```

Figure 9: Multilayer perceptron (MLP) model for testing dataset

```
=== Re-evaluation on test set ===

User supplied test set
Relation:     NewDataset2-weka.filters.unsupervised.instance.Resample-S1-Z70.0-no-replacement-V
Instances:     unknown (yet). Reading incrementally
Attributes:   25

=== Summary ===

Correctly Classified Instances         693               98.2979 %
Incorrectly Classified Instances        12                1.7021 %
Kappa statistic                          0.9631
Mean absolute error                      0.0195
Root mean squared error                  0.1205
Total Number of Instances              705

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.969    0.009    0.984      0.969    0.976      0.963    0.998     0.997     HL
               0.991    0.031    0.982      0.991    0.987      0.963    0.998     0.999     LL
Weighted Avg.  0.983    0.023    0.983      0.983    0.983      0.963    0.998     0.998

=== Confusion Matrix ===

   a   b   <-- classified as
 248   8 |   a = HL
   4 445 |   b = LL
```

Figure 10: Naïve Bayes (NB) model for testing dataset

```
=== Re-evaluation on test set ===

User supplied test set
Relation:     NewDataset2-weka.filters.unsupervised.instance.Resample-S1-Z70.0-no-replacem
Instances:     unknown (yet). Reading incrementally
Attributes:   25

=== Summary ===

Correctly Classified Instances         629               89.2199 %
Incorrectly Classified Instances        76               10.7801 %
Kappa statistic                          0.7673
Mean absolute error                      0.1168
Root mean squared error                  0.2918
Total Number of Instances              705

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC A
               0.855    0.087    0.849      0.855    0.852      0.767    0.944     0.938
               0.913    0.145    0.917      0.913    0.915      0.767    0.944     0.956
Weighted Avg.  0.892    0.124    0.892      0.892    0.892      0.767    0.944     0.950

=== Confusion Matrix ===
```

Figure 11: Sequential minimal optimisation (SMO) model for testing dataset

```
=== Re-evaluation on test set ===

User supplied test set
Relation:     NewDataset2-weka.filters.unsupervised.instance.Resample-S1-Z70.0-no-replacement-V
Instances:     unknown (yet). Reading incrementally
Attributes:   25

=== Summary ===

Correctly Classified Instances         681               96.5957 %
Incorrectly Classified Instances        24                3.4043 %
Kappa statistic                          0.9256
Mean absolute error                      0.034
Root mean squared error                  0.1845
Total Number of Instances              705

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.930    0.013    0.975      0.930    0.952      0.926    0.958     0.932     HL
                 0.987    0.070    0.961      0.987    0.974      0.926    0.958     0.957     LL
Weighted Avg.    0.966    0.050    0.966      0.966    0.966      0.926    0.958     0.948

=== Confusion Matrix ===

   a   b   <-- classified as
 238  18 |   a = HL
   6 443 |   b = LL
```

Model performance

This sub-section presents the performance of the trained and tested models by exploring the number of correctly classified datasets and looking at performance in terms of six metrics.

Table 2: Comparison of models' performance on training dataset, based on correctly and incorrectly classified student data

Metric	J48	LR	MLP	NB	SMO
Correctly classified students	1566	1578	**1596**	1461	1580
Incorrectly classified students	77	65	**47**	182	63
Correctly classified HL students	547	561	**571**	503	554
Incorrectly classified HL students	53	39	**29**	97	46
Correctly classified LL students	1019	1017	1025	958	**1026**
Incorrectly classified LL students	24	26	18	85	**17**

Table 2 compares the performance of the classifier models based on correctly and incorrectly classified student data for the training dataset. The acronyms "LL" and "HL" in the table stand for "LowLow" and "HighLow", which were this study's two hyperparameters, i.e., parameters whose values controlled the learning process. Each machine-learning model had to deliver either an LL or HL classification for each student record. The focus in this study was on classifying low-performing students

into two groups—HL and LL—with the idea being that students in the LL group would be the ones to receive urgent assistance to improve their chances of performing well. As seen in the table, the MLP algorithm correctly classified the highest number of students (1,596) for the entire training dataset and misclassified the lowest number of students (47). SMO had the next best performance, with 1,580 records classified correctly and 63 misclassifications. LR followed SMO with a classification margin of two less than SMO: 1,578 records were correctly classified, and 65 records were incorrectly classified. J48 correctly classified 1,566 student records and misclassified 77 records. NB showed the poorest performance, with 1,461 correctly classified records and 182 misclassified records.

For the performance of the algorithms in correctly classifying students in the HL class, MLP outperformed the other algorithms, while for the performance of the algorithms in correctly classifying students in the LL class, SMO performed the best. However, the difference in performance between SMO and MLP in correctly classifying students in the LL class was just one record.

Table 3: Comparison of the models' performance on the training dataset, using six selected metrics

Model	recall	specificity	ROC curve	F-measure	kappa	RMSE
J48	0.977	0.912	0.961	0.964	0.8979	0.2085
LR	0.975	0.935	0.982	0.969	0.9143	0.1835
MLP	0.983	**0.952**	**0.992**	**0.978**	**0.9381**	**0.1560**
NB	0.919	0.838	0.945	0.913	0.7601	0.2901
SMO	**0.984**	0.923	0.954	0.970	0.9164	0.1958

Table 3 shows the five models' performance on the training dataset in terms of six selected metrics:

Recall is the proportions of positive instances that are correctly classified. It is often called sensitivity.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Specificity is the proportion of true negatives that are correctly predicted by the model.

$$\text{Specificity} = = \frac{TN}{TN + FP}$$

The **receiver operating characteristic (ROC) curve** gives the performance of a classification model at all classification thresholds.

True positive rate

False positive rate

True positive rate (TPR) is a synonym for recall and is therefore defined as follows:

$$\text{True positive rate (TPR)} = \frac{TP}{TP + FN}$$

False positive rate (FPR) is defined as follows:

$$\text{False positive rate (FPR)} = \frac{FP}{TF + TN}$$

Therefore, the ROC curve plots TPR vs. FPR at different classification thresholds.

F-measure is a measure that conveys the balance between recall and precision and can be said to be the harmonic mean of recall and precision.

$$\text{F-measure} = \frac{2 * Precision * Recall}{Precision + Recall}$$

The **kappa value** is a statistic used to control only those instances that may have been correctly classified by chance.

$$\text{Kappa can be calculated as: Kappa} = \frac{total\ accuracy - random\ accuracy}{1 - random\ accuracy}$$

The **root mean square error (RMSE)** is a standard way to measure the error of a model in predicting quantitative data. RMSE is calculated as:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

where,

$\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$ are predicted values

$y_1, y_2, \ldots, y_n$ are observed values

n is the number of observations

As seen in the results in Table 3, SMO had the best performance (98.4%) in terms of the recall metric, followed closely by MLP (98.3%). For all of the other five metrics—specificity, ROC curve, F-measure, kappa and RMSE—MLP performed best. For all six metrics, the weakest performer was NB.

After modelling the training dataset, the study compared the performance of the five algorithms on the testing dataset. Table 4 shows the results of the comparison of the performance based on correctly and incorrectly classified data, and Table 5 shows the performance in terms of the six metrics.

Table 4: Comparison of models' performance on test dataset, based on correctly and incorrectly classified student data

	J48	LR	MLP	NB	SMO
Right All	684	680	**693**	629	681
Wrong All	21	25	**12**	76	24
Right HL	239	239	**248**	219	238
Wrong HL	17	17	**8**	37	18
Right LL	**445**	441	**445**	410	443
Wrong LL	**4**	8	**4**	39	6

As seen in Table 4, the MLP algorithm performed best, correctly classifying the entire dataset, and achieved the best performance for classifying students in the HL and LL groups. Overall across the metrics, as seen in the findings presented up to this point, MLP was the best performer.

Table 5: Comparison of the models: Performance on the test dataset, using the six selected metrics

Model	recall	specificity	ROC curve	F-measure	kappa	RMSE
J48	**0.991**	0.934	0.976	0.977	0.9349	0.1685
LR	0.982	0.934	0.988	0.972	0.9227	0.1724
MLP	**0.991**	**0.969**	**0.998**	**0.987**	**0.9631**	**0.1205**
NB	0.913	0.855	0.944	0.915	0.7673	0.2918
SMO	0.987	0.930	0.958	0.974	0.9256	0.1845

As seen in Table 5, for the recall metric, MLP and J48 both achieved the highest value (99.1%). For the other five metrics—specificity, ROC curve, F-measure, kappa and RMSE—MLP was the best performer. For all six metrics, NB scored worst.

5. Feature selection

Feature selection in data mining helps models by looking out for the most important features in the dataset, reducing complexities and increasing the accuracy of the model (Neumann et al., 2016). This study ranked the attributes in order of importance using four feature selection techniques available in WEKA, namely: correlation, gain ratio, information gain, and ReliefF. Figures 12 to 15 are snapshots of the results.

Figure 12: Results from correlation feature selection technique

```
=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

average merit      average rank  attribute
 0.51  +- 0.007      1   +- 0        8 SponQual
 0.426 +- 0.005      2.4 +- 0.49     5 SecType
 0.424 +- 0.009      2.6 +- 0.49    12 WorkStudy
 0.381 +- 0.007      4.4 +- 0.49    13 UniAcc
 0.382 +- 0.007      4.6 +- 0.49     6 SecArea
 0.321 +- 0.006      6   +- 0       21 SptAc
 0.293 +- 0.006      7.4 +- 0.49    14 BeAdmYrs
 0.293 +- 0.006      7.6 +- 0.49     7 SponType
 0.273 +- 0.004      9.6 +- 0.49    10 SponSup
 0.27  +- 0.008      9.8 +- 0.87    15 JambCou
 0.261 +- 0.009     10.8 +- 0.87     2 Age
 0.252 +- 0.004     11.8 +- 0.4     17 WkStud
 0.222 +- 0.004     13.4 +- 0.49    24 AveSc
 0.219 +- 0.007     13.7 +- 0.64     9 SponInc
 0.203 +- 0.005     15.3 +- 0.46    22 JambSc
 0.199 +- 0.008     15.7 +- 0.78    19 SmPhn
 0.188 +- 0.008     16.9 +- 0.3     18 PgDeg
 0.162 +- 0.004     18   +- 0       23 PumeSc
 0.151 +- 0.004     19   +- 0        1 Sex
 0.125 +- 0.002     20.1 +- 0.3     16 CouInt
 0.119 +- 0.004     20.9 +- 0.3     11 FamSize
 0.088 +- 0.004     22   +- 0       20 SmPhnAss
 0.045 +- 0.007     23   +- 0        4 AtPri
 0.01  +- 0.007     24   +- 0        3 MarStat
```

Figure 12 gives the ranking of the features (attributes) in order of contribution to the low performance of students, ranked from highest to lowest, as determined by the correlation feature selection technique. The four most-correlating attributes were: sponsor qualification, secondary school type, work and study, and university accommodation. The four least-correlating attributes were: family size, smart phone assistance, attended primary school, and marital status.

Figure 13: Results from gain ratio feature selection technique

```
=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

average merit      average rank  attribute
 0.194 +- 0.006      1   +- 0        8 SponQual
 0.142 +- 0.006      2.4 +- 0.49    12 WorkStudy
 0.136 +- 0.003      3.2 +- 0.98     5 SecType
 0.132 +- 0.002      3.5 +- 0.67    24 AveSc
 0.126 +- 0.003      5   +- 0.45    17 WkStud
 0.119 +- 0.005      6   +- 0.45    13 UniAcc
 0.109 +- 0.004      6.9 +- 0.3      6 SecArea
 0.097 +- 0.003      8.2 +- 0.4      7 SponType
 0.092 +- 0.003      8.9 +- 0.54    22 JambSc
 0.085 +- 0.003      9.9 +- 0.3     10 SponSup
 0.077 +- 0.003     11.2 +- 0.4     21 SptAc
 0.075 +- 0.003     11.9 +- 0.54     9 SponInc
 0.069 +- 0.004     12.9 +- 0.3     18 PgDeg
 0.061 +- 0.002     14.2 +- 0.4     14 BeAdmYrs
 0.054 +- 0.004     15.4 +- 0.66     2 Age
 0.053 +- 0.003     15.5 +- 0.81    15 JambCou
 0.047 +- 0.004     17.4 +- 0.8     19 SmPhn
 0.044 +- 0.002     18   +- 0.63    23 PumeSc
 0.042 +- 0.002     18.5 +- 0.81    16 CouInt
 0.019 +- 0.002     20.1 +- 0.3     20 SmPhnAss
 0.016 +- 0.001     20.9 +- 0.3      1 Sex
 0.012 +- 0.001     22.1 +- 0.3     11 FamSize
 0.009 +- 0.003     22.9 +- 0.3      4 AtPri
 0     +- 0.001     24   +- 0        3 MarStat
```

Figure 13 provides the feature rankings generated by the gain ratio technique. The strongest (most-correlating) four attributes were: sponsor qualification, work and study, secondary school type, and average SSCE score. The weakest (least-correlating) four attributes were: sex, family size, attended primary school, and marital status.

Figure 14: Results from information gain feature selection technique

```
=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

average merit      average rank  attribute
 0.223 +- 0.006      1   +- 0        8 SponQual
 0.194 +- 0.005      2   +- 0       17 WkStud
 0.18  +- 0.004      3   +- 0       24 AveSc
 0.145 +- 0.004      4.1 +- 0.3      7 SponType
 0.135 +- 0.005      5.8 +- 1.17    22 JambSc
 0.135 +- 0.005      5.8 +- 0.75    10 SponSup
 0.132 +- 0.003      6.8 +- 0.98     5 SecType
 0.129 +- 0.006      7.5 +- 0.67    12 WorkStudy
 0.113 +- 0.004      9.1 +- 0.3      9 SponInc
 0.106 +- 0.004     10.3 +- 0.46     6 SecArea
 0.103 +- 0.004     10.8 +- 0.87    13 UniAcc
 0.092 +- 0.006     11.8 +- 0.4     18 PgDeg
 0.078 +- 0.003     13.3 +- 0.46    14 BeAdmYrs
 0.077 +- 0.003     13.7 +- 0.46    21 SptAc
 0.067 +- 0.003     15.2 +- 0.4     23 PumeSc
 0.062 +- 0.003     15.8 +- 0.4     16 CouInt
 0.053 +- 0.003     17.2 +- 0.4     15 JambCou
 0.048 +- 0.003     17.8 +- 0.4      2 Age
 0.028 +- 0.002     19   +- 0       20 SmPhnAss
 0.027 +- 0.002     20   +- 0       19 SmPhn
 0.019 +- 0.001     21.1 +- 0.3     11 FamSize
 0.016 +- 0.001     21.9 +- 0.3      1 Sex
 0.002 +- 0         23   +- 0        4 AtPri
 0     +- 0         24   +- 0        3 MarStat
```

Figure 14 presents the features ranked by the information gain technique. The four strongest attributes were found to be: sponsor qualification, weekly study time, sverage SSCE score, and sponsor type. The weakest attributes were: family size, sex, attended primary school, and marital status.

Figure 15: Results from ReliefF feature selection technique

```
=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

average merit      average rank  attribute
 0.375 +- 0.009      1   +- 0       11 FamSize
 0.353 +- 0.008      2   +- 0        7 SponType
 0.309 +- 0.008      3.6 +- 0.66    17 WkStud
 0.3   +- 0.01       4.1 +- 1.04     8 SponQual
 0.3   +- 0.005      4.4 +- 0.66    10 SponSup
 0.283 +- 0.009      6.5 +- 0.92     2 Age
 0.278 +- 0.006      6.9 +- 0.7     14 BeAdmYrs
 0.274 +- 0.007      7.5 +- 0.67    22 JambSc
 0.26  +- 0.006      9.2 +- 0.6      9 SponInc
 0.251 +- 0.008     10.7 +- 1.1     24 AveSc
 0.246 +- 0.006     11.1 +- 0.94    23 PumeSc
 0.241 +- 0.008     11.8 +- 1.08    16 CouInt
 0.234 +- 0.007     12.9 +- 1.45    18 PgDeg
 0.228 +- 0.008     14.2 +- 1.54    21 SptAc
 0.219 +- 0.009     15.8 +- 1.33    20 SmPhnAss
 0.218 +- 0.01      15.8 +- 1.25     5 SecType
 0.219 +- 0.007     15.9 +- 1.3      6 SecArea
 0.199 +- 0.007     17.9 +- 0.94    13 UniAcc
 0.193 +- 0.006     18.8 +- 0.6      1 Sex
 0.181 +- 0.005     20.1 +- 0.54    12 WorkStudy
 0.168 +- 0.009     20.8 +- 0.4     15 JambCou
 0.075 +- 0.003     22   +- 0       19 SmPhn
 0.057 +- 0.006     23   +- 0        3 MarStat
 0.017 +- 0.003     24   +- 0        4 AtPri
```

As shown in Figure 15, use of the ReliefF technique generated these four strongest attributes: family size, sponsor type, seekly study time, and sponsor qualification. The weakest features were: course from JAMB, own smart phone, marital status, and attended primary school.

The results as set out in Figures 11 to 14 indicated that, across all the techniques used, certain attributes were consistently identified as most relevant, meaning they were likely to contribute greatly to predicting a student's classification as HL or LL. At the same time, there were certain attributes that consistently emerged as the least relevant, suggesting that these features contribute least to predicting a student's classification.

Accordingly, we tested the performance of the five classifier algorithms through successive modelling of the features in each feature selection technique starting from the top four attributes. The ROC curve and RMSE values, which are two widely used metrics (Caruana & Niculescu-Mizil, 2004), were used to evaluate the performance of the models. Table 6 provides a summary of the results obtained.

Table 6: Performance summary of feature selection algorithms

Algorithm	ROC curve	RMSE	Best range
correlation	0.993	0.1410	15–21
gain ratio	0.994	0.1471	14–17
information gain	0.997	0.1382	15–16
ReliefF	0.995	0.1416	18–20

Table 6 shows that all algorithms achieved a higher level of performance on the reduced dataset than on the complete dataset. The information gain feature selection algorithm had the highest performance, with an ROC curve value of 99.7% and an RMSE value of 13.82%, and with the features identified as most relevant ranging between 15 and 16 of the features in Table 1.

Performance of MLP classifier using the 16 most relevant attributes

Since MLP had proven to be the best classifier, we built a new MLP model, now using only the 16 attributes determined to be most relevant by the information gain feature selection technique. The modelling process began by extracting the selected attributes in WEKA and then using the MLP classifier to build the model. Figure 16 shows the model built, and Table 7 presents the performance of the model using the 16 most relevant attributes identified.

Figure 16: MLP model built with 16 most relevant attributes, using the training dataset

```
Time taken to build model: 6.92 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances        1606               97.748  %
Incorrectly Classified Instances        37                2.252  %
Kappa statistic                          0.9513
Mean absolute error                      0.0269
Root mean squared error                  0.1382
Relative absolute error                  5.7956 %
Root relative squared error             28.7039 %
Total Number of Instances             1643

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.965    0.015    0.973      0.965    0.969      0.951    0.997     0.995     HL
                 0.985    0.035    0.980      0.985    0.982      0.951    0.997     0.998     LL
Weighted Avg.    0.977    0.028    0.977      0.977    0.977      0.951    0.997     0.997

=== Confusion Matrix ===

    a    b   <-- classified as
  579   21 |    a = HL
   16 1027 |    b = LL
```

Table 7: Performance of MLP using 16 most relevant attributes dataset

Item	Value	Metric	Value
Right All	1606	recall	0.985
Wrong All	37	specificity	0.965
Right HL	579	ROC curve	0.997
Wrong HL	21	F-Measure	0.982
Right LL	1027	kappa	0.9513
Wrong LL	16	RMSE	0.1382

In Table 7, all the metric values for the 16 most relevant attributes dataset show good performance. A comparison of these metrics values with those for the complete database of all attributes revealed the following: recall was 98.5% for the 16 most relevant attributes dataset, while it was 98.3% for the complete database; specificity was 96.5% for the most relevant attributes and 95.2% for the complete database; the ROC curve was 99.7% for the most relevant attributes and 99.2% for the complete database; the F-measure was 98.2% for the most relevant attributes and 97.8% for the complete database; the kappa value was 95.13% for the most relevant attributes and 93.81% for the complete database; and the RMSE was 13.82% for the most relevant attributes and 15.6% for the entire database.

The results showed improvement for every metric value using the 16 most relevant attributes when compared with using all 24 attributes. It was thus concluded that the MLP classifier, using the 16 most relevant features ranked with the information gain algorithm, was the best classifier for the dataset generated by this study. Accordingly, it was determined that it would be beneficial to design and implement the prediction application with the MLP classifier and the 16 identified attributes.

The 16 selected attributes

Table 8 shows the variable codes and values, and descriptions, of the 16 most relevant attributes identified from the modelling process. These were the features used for development of the predictive application.

Table 8: Selected attributes

Attribute	Variable codes	Values
Sponsor qualification	DEG NODEG NOEDU	Degree No degree No education
Weekly study time	LOW AVE HIGH	Below 10hrs 10 – 20hrs Above 20hrs
Average SSCE score	LOW AVE HIGH	Below 4.00 4.00 – 4.99 Above 5.00
Sponsor type	GUAD PAR SELF	Guardian Parents Self-sponsor
JAMB score	LOW AVE HIGH	Below 180 180 – 250 Above 250
Sponsor support	LOW MED HIGH	Little Average Great
Secondary school type	PRI PUB	Private Public
Work and study	YES NO	Yes No
Sponsor income (per month)	LOW MED HIGH	Below 50k 50k – 100k Above 100k
Secondary school area	URB RUR	Urban Rural
University accommodation	CMPS OFFCMPS	Campus Off-campus

Attribute	Variable codes	Values
Postgraduate degree	NO YES NS	No Yes Not sure
Years before admission (gap between completing secondary school and commencing university)	NONE B5 5A	None Below 5 5 above
Sports activeness	LOW HIGH	Little active Very active
Post-UTME score	LOW AVE HIGH	Below 180 180 – 250 Above 250
Course interest	LOW AVE HIGH	Little Average High

6. Development and evaluation of predictive software application

The predictive application was developed with the goal of creating a tool that could help Nigerian universities to identify potentially low-performing students—so as to be able to provide targeted intervention measures to address these students' needs.

Application development

The design process followed the following four steps: input students' features, which were the 16 most relevant attributes identified; use the best classifier algorithm identified, which was the MLP model; predict students' failure risk level as HL or LL; and output an intervention summary for students, with HL students requiring low intervention and LL students requiring high intervention. The main users of the software would be faculty officers in charge of gathering and storing students' information. Their role would be to use the application to obtain predictions and to forward results to the relevant authorities in the institution for the purpose of intervention in the form of one or more of: study load rescheduling, counselling, financial intervention in the form of bursary, and/or government assistance through various funding agencies for indigent students.

A 17th attribute was added (as seen below in Figure 17), called "matric number", to identify specific students whose information had been processed. This was the ID number of the student, unique to each student and eliminating the need to record student names in the application. The software interface was developed with the PHP programming language, while the part of the software that would handle the prediction used the Python programming language—incorporating the MLP model built in WEKA. Snapshots of the predictive application follow in Figures 17 to 19.

Figure 17 shows the welcome screen, which offers instructions to users on the relevant student details to gather, and informs users that only authorised persons can access the system.

Figure 17: Welcome screen

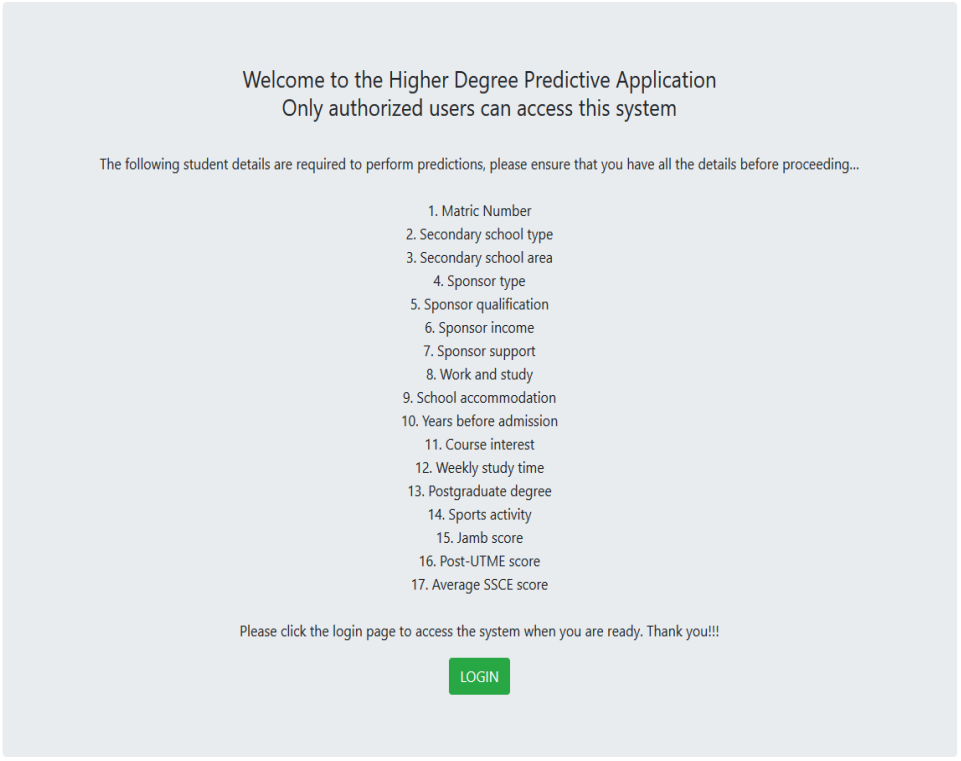


Figure 18 shows the user input page for the 16 most relevant attributes and the added "Matric number". Users select options from the dropdown menu for each feature. The page also has a button for viewing the predicted results, and another button for resetting the page.

Figure 18: User input page

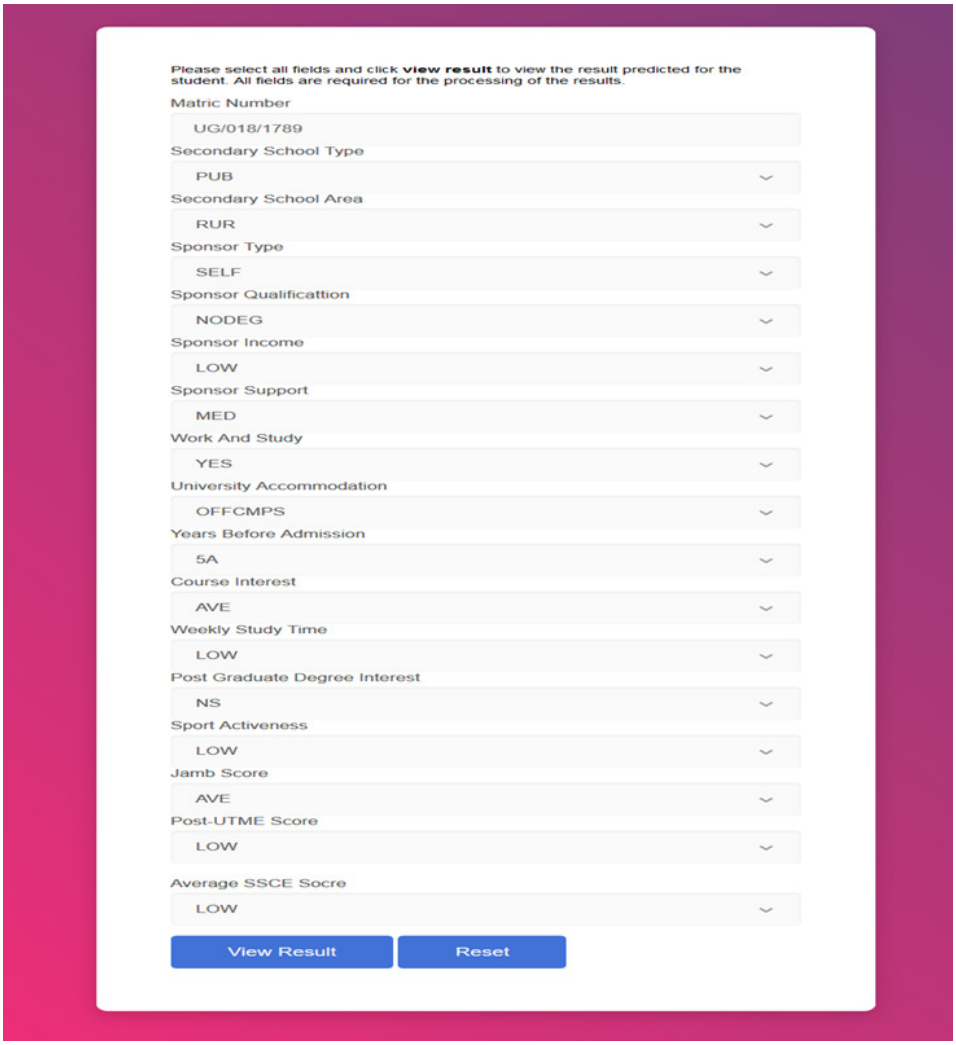
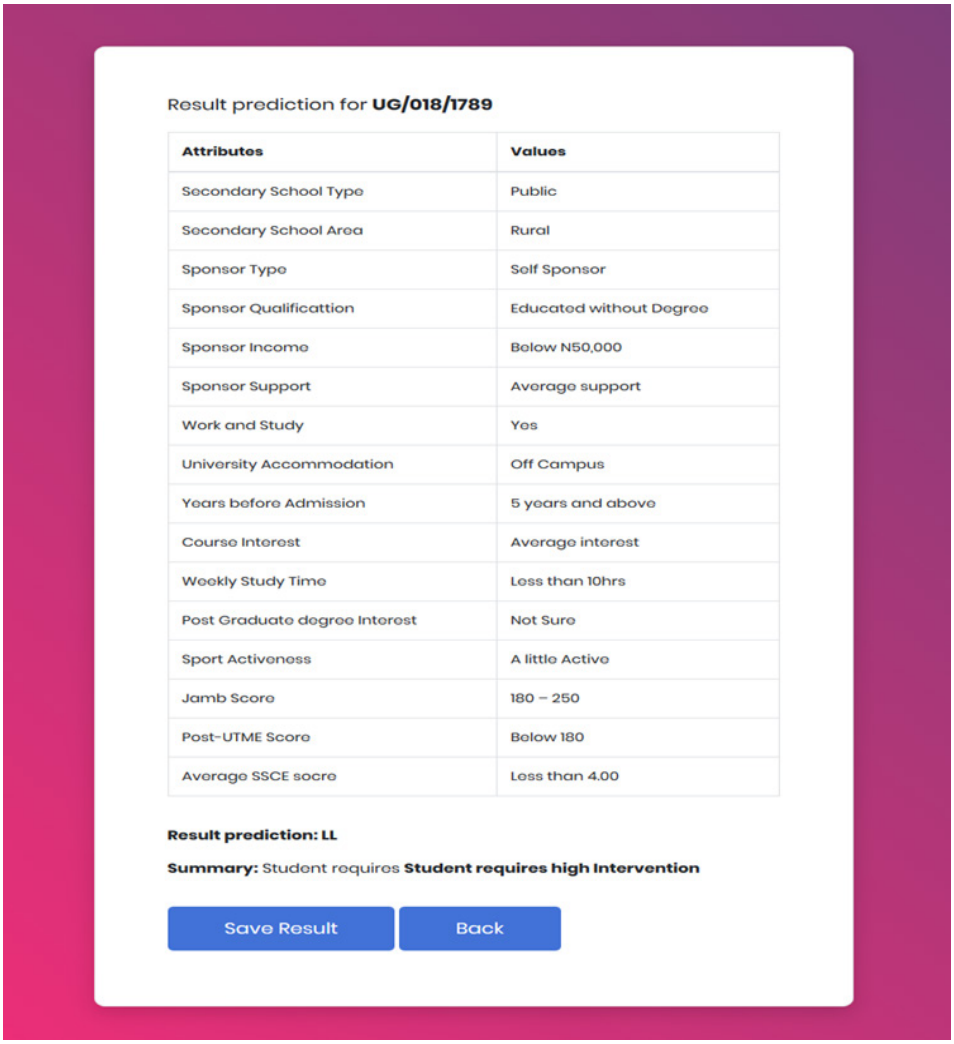


Figure 19 displays a "Result prediction" for a student, indicating (in this example case) an LL classification and specifying the need for "high intervention".

Figure 19: Prediction results page



Application evaluation

The software was evaluated in two phases. The first phase evaluated the software based on its requirements, and the second phase evaluated the software based on its performance. The first phase ensured that the application met the following requirements:

- Simple and interactive interface: Easy navigation between pages and straightforward instructions for users

- Secured system: Login interface to ensure that only authorised persons can access the software
- Avoidance of prediction error: Notifications to users of empty fields to ensure the application has received values for all 17 features before allowing the generation of a prediction
- Rapid results: Provision of results with a single click
- Provision of risk level: Offering a result prediction value for each student, and a risk intervention summary based on predicted value

The second phase of the evaluation involved testing the software using the 30% testing dataset, which contained 705 records. The process of evaluation began by entering the details of each record into the predictive application to obtain a prediction. Figure 20 shows a cross-section of the stored results in Excel, including "Actual" and "Predicted" values in terms of the "LowLow" and "HighLow" classifications.

Figure 20: Cross-section of results from testing dataset

MatNum	SecType	SecArea	SponType	SponQual	SponInc	Spon3_up	WorkStud	UniAcc	BeAdmYrs	CourIrt	WkStud	PgDeg	SptAc	JambSc	PumeSc	AveSc	Actual	Predicted	LowxLow	HighLow
UG/010/0301	P_b	Rur	Self	NoDeg	Med	Med	Yes	OffCmps	5A	Ave	Ave	Nis	Low	Low	Ave	Ave	LowLow	LL	A	
UG/010/0302	P_b	Rur	Self	NoDeg	Med	Low	Yes	OffCmps	5A	Ave	Ave	Nis	Low	Low	Low	Low	LowLow	LL	A	
UG/010/0303	P_b	Rur	Par	NoDeg	Med	High	No	OffCmps	B5	Ave	Ave	Nis	Low	Ave	High	Ave	LowLow	LL	A	
UG/010/0304	Pri	Urb	G_ad	Deg	High	High	No	OffCmps	B5	High	Ave	Yes	Low	High	Ave	High	HighLow	HL		A
UG/010/0305	Pri	Rur	Self	NoDeg	Med	Low	Yes	OffCmps	5A	Ave	Low	Nis	Low	Low	Ave	Ave	LowLow	LL	A	
UG/010/0306	Pri	Rur	G_ad	Deg	High	Med	Yes	Cmps	5A	Ave	High	Nis	High	Low	Low	Low	HighLow	HL		A
UG/010/0307	Pri	Rur	Self	NoDeg	Med	Low	Yes	OffCmps	B5	Ave	Low	Nis	High	Low	Ave	Low	LowLow	HL	D	
UG/010/0308	P_b	Rur	Self	NoDeg	Med	Low	Yes	Cmps	5A	Ave	Ave	Nis	High	Ave	Ave	Ave	LowLow	LL	A	
UG/010/0309	P_b	Rur	Self	NoDeg	Low	Low	Yes	OffCmps	5A	Ave	Low	Nis	Low	Low	Low	Low	LowLow	LL	A	
UG/010/0310	P_b	Urb	G_ad	Deg	High	High	No	OffCmps	B5	Ave	Ave	Nis	High	Ave	High	High	HighLow	HL		A
UG/010/0311	P_b	Urb	G_ad	Deg	High	High	No	OffCmps	B5	Ave	Ave	Nis	High	Ave	High	High	HighLow	HL		A
UG/010/0312	Pri	Rur	Self	NoDeg	Med	Med	Yes	OffCmps	5A	Low	Ave	Nis	Low	High	Ave	Ave	LowLow	LL	A	
UG/010/0313	P_b	Urb	G_ad	Deg	High	High	No	Cmps	B5	High	High	Yes	High	Ave	High	High	HighLow	HL		A
UG/010/0314	Pri	Urb	G_ad	Deg	High	High	No	Cmps	B5	High	High	Yes	High	Ave	High	High	HighLow	HL		A
UG/010/0315	Pri	Rur	G_ad	Deg	High	Med	Yes	Cmps	5A	Ave	High	Nis	High	Low	Low	Ave	HighLow	HL		A
UG/010/0316	Pri	Urb	G_ad	Deg	High	High	No	OffCmps	B5	High	Low	Yes	Low	Ave	Low	Ave	HighLow	HL		A
UG/010/0317	P_b	Rur	Self	NoDeg	Med	Med	Yes	OffCmps	5A	Ave	Ave	Nis	Low	High	Ave	Ave	LowLow	LL	A	
UG/010/0318	P_b	Rur	Self	NoDeg	Med	Med	Yes	Cmps	5A	Ave	High	Nis	High	Low	Low	Ave	LowLow	LL	A	
UG/010/0319	P_b	Rur	G_ad	NoDeg	Low	Low	No	OffCmps	B5	Ave	Low	Nis	Low	Low	Low	Low	LowLow	LL	A	
UG/010/0320	Pri	Urb	G_ad	Deg	High	Med	No	Cmps	B5	High	High	Yes	High	Ave	High	Ave	HighLow	HL		A
UG/010/0321	P_b	Urb	Self	NoDeg	Med	Low	Yes	OffCmps	B5	High	High	Nis	Low	Ave	Ave	Ave	LowLow	LL	A	
UG/010/0322	P_b	Rur	Par	NoDeg	Med	High	No	OffCmps	B5	Ave	Low	Nis	Low	Ave	Low	Ave	LowLow	LL	A	
UG/010/0323	Pri	Urb	G_ad	Deg	High	High	No	Cmps	None	High	High	Yes	High	High	Ave	Ave	HighLow	HL		A
UG/010/0324	P_b	Urb	G_ad	Deg	High	High	No	OffCmps	B5	High	Ave	Yes	Low	High	Ave	Ave	HighLow	HL		A
UG/010/0325	P_b	Rur	G_ad	Deg	Med	Med	Yes	OffCmps	B5	Ave	High	Nis	Low	Ave	Low	Ave	HighLow	HL		A
UG/010/0326	P_b	Rur	Par	NoEdu	Med	Med	No	Cmps	5A	Low	Low	Nis	High	Ave	Low	Low	LowLow	LL	A	
UG/010/0327	Pri	Urb	G_ad	Deg	Med	High	No	Cmps	B5	High	High	Nis	High	High	High	Ave	HighLow	HL		A
UG/010/0328	P_b	Rur	Par	NoDeg	Med	High	Yes	OffCmps	B5	Ave	Ave	Nis	Low	Ave	High	Ave	LowLow	LL	A	
UG/010/0329	P_b	Rur	Self	NoDeg	Low	Low	Yes	OffCmps	5A	Ave	Ave	Nis	Low	Low	Low	Ave	LowLow	LL	A	

From the data in the spreadsheet shown in Figure 20, a confusion matrix was constructed, as represented in Table 9, to determine the accuracy of the predictive application when engaging with the testing dataset.

Table 9: Confusion matrix

	Actual LL	Actual HL
Predicted LL	446	11
Predicted HL	3	245

The confusion matrix in Table 9 shows that 446 out of 449 records were correctly predicted for the LL group, and 245 out of 256 records were correctly predicted for the HL group. Furthermore, using the values obtained from the confusion matrix, the values obtained for recall and sensitivity were both 99.3%, and the value obtained for specificity was 95.7%. For prevalence, the software obtained 63.7% for LL students and 36.3% for HL students, which meant that high-risk students were found to be twice the number of low-risk students in the institution. The software achieved accuracy of 98%, which demonstrated a high level of dependability. For precision, the software achieved 99.3%, and its F-measure was 98.5%. Thus, the performance obtained from an evaluation of the software showed that the application could accomplish its objectives.

7. Conclusion and recommendation

This study was grounded in an understanding that low student performance is a global challenge that affects every tertiary academic institution in the world—and that the challenge is particularly prominent in developing countries, including Nigeria. In looking for ways to monitor and improve the performance of students, the study explored machine learning as a technique to assist Nigerian universities in identifying the potential for low academic performance among undergraduate students. With a dataset collected from Niger Delta University, which comprised 24 attributes, this study identified the 16 features (student attributes) that give optimal prediction results when subjected to machine learning. The study also identified the MLP classifier algorithm as the best model for generating such predictions. Furthermore, the study developed a predictive software application for potential deployment by Nigerian universities, using the 16 attributes and MLP model identified.

It is recommended that future research should study the performance of the MLP model, and of predictive applications, using datasets from other Nigerian higher institutions—so as to work towards enabling a robust and unified system for all institutions in the country.

References

Abu Zohair, L. M. (2019). Prediction of student's performance by modelling small dataset size. *Journal of Educational Technology in Higher Education*, *16*, 27. https://doi.org/10.1186/s41239-019-0160-3

Adeyemi, A.M., & Adeyemi, S.B. (2014). Institutional factors as predictors of students' academic achievement in colleges of education in South Western Nigeria. *International Journal of Educational Administration and Policy Studies*, *6*(8), 141–153. https://doi.org/10.5897/IJEAPS2014.0342

Adeyemo, A. B., & Kuye, G. (2006). Mining students' academic performance using decision tree algorithms. *Journal of Information Technology Impact*, *6*(3), 161–170.

Alami, M. (2016). Causes of poor academic performance among Omani students. *International Journal of Social Science Research*, *4*(1), 126–136. https://doi.org/10.5296/ijssr.v4i1.8948

Algarni, A. (2016). Data mining in education. *International Journal of Advanced Computer Science and Applications, 7*(6), 456–461. https://doi.org/10.14569/IJACSA.2016.070659

Al-Zoubi, S. M., & Younes, M. A. B. (2015). Low academic achievement: Causes and results. *Theory and Practice in Language Studies*, *5*(11), 2262–2268. https://doi.org/10.17507/tpls.0511.09

Aryana, M. (2010). Relationship between self-esteem and academic achievement amongst pre-university students. *Journal of Applied Sciences*, *10*(20), 2474–2477. https://doi.org/10.3923/jas.2010.2474.2477

Banerjee, P. A. (2016). A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools. *Cogent Education*, *3*(1). https://doi.org/10.1080/2331186X.2016.1178441

Caruana, R., & Niculescu-Mizil, A. (2004). Data mining in metric space: An empirical analysis of supervised learning performance criteria. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 69–78). https://doi.org/10.1145/1014052.1014063

Dhilipan J., Vijayalakshmi N., Suriya S., & Arockiya, C. (2021). Prediction of students performance using machine learning. *IOP Conference Series: Materials Science and Engineering, 1055*, 012122. https://doi.org/10.1088/1757-899X/1055/1/012122

Ekubo, A. E. (2020). *Predictive system for characterizing low performance of undergraduate students using machine-learning techniques*. PhD thesis, North-West University, South Africa.

Farooq, M. S., Chaudhry, A. H., Shafiq, M., & Berhanu, G. (2011). Factors affecting students' quality of academic performance: A case of secondary school level. *Journal of Quality and Technology Management*, *7*(2), 1–14.

Flanagan, B., Majumdar, R., & Ogata, H. (2022). Early-warning prediction of student performance and engagement in open book assessment by reading behavior analysis. *International Journal of Educational Technology in Higher Education*, *19*(41), 1–23. https://doi.org/10.1186/s41239-022-00348-4

Frimpong, E. A., Agyeman, G. A., & Ofosu, F. F. (2016). Institutional factors affecting the academic performance of polytechnic students in Ghana. *International Journal of Humanities & Social Science Studies (IJHSSS)*, *2*(5), 102–109.

Glewwe, P., & Kremer, M. (2006). Schools, teachers, and education outcomes in developing countries. In E. Hanushek, & F. Welch (Eds.), *Handbook of the economics of education, vol. 2* (pp. 945–1017). https://doi.org/10.1016/S1574-0692(06)02016-2

Gullo, F. (2015). From patterns in data to knowledge discovery: What data mining can do. *Physics Procedia*, *62*, 18–22. https://doi.org/10.1016/j.phpro.2015.02.005

Hughes, G., & Dobbins, C. (2015). The utilization of data analysis techniques in predicting student performance in massive open online courses (MOOCs). *Research and Practice in Technology Enhanced Learning*, *10*(1), 1–18. https://doi.org/10.1186/s41039-015-0007-z

Li, X., Zhang, Y., Cheng, H., Li, M., & Yin, B. (2022). Student achievement prediction using deep neural network from multi-source campus data. *Complex and Intelligent Systems*, *8*, 5143–5156. https://doi.org/10.1007/s40747-022-00731-8

Longe, O. (2017). Graduate unemployment in Nigeria: Causes, consequences and remediable approaches. *American International Journal of Contemporary Research*, *7*(4), 63–73.

Mitra, D. (2011). Pennsylvania's best investment: The social and economic benefits of public education. Pennsylvania State University.

Mohammed, A. (2022). Data mining classification simplified: Key types, steps & 6 best classifiers. Hevo. https://hevodata.com/learn/data-mining-classification

Neumann, U., Riemenschneider, M., Sowa, J. P., Baars, T., Kälsch, J., Canbay, A., & Heider, D. (2016). Compensation of feature selection biases accompanied with improved predictive performance for binary classification by using a novel ensemble feature selection approach. *BioData Mining*, *9*(36), 1–14. https://doi.org/10.1186/s13040-016-0114-4

Nsiah, H. (2017). Fear of failure and the academic performance of students from low-income families. *International Journal of Education and Social Science, 4*(10), 19–26.

Nurmi, J. E., Aunola, K., Salmela-Aro, K., & Lindroos, M. (2003). The role of success expectation and task-avoidance in academic performance and satisfaction: Three studies on antecedents, consequences and correlates. *Contemporary Educational Psychology*, *28*(1), 59–90. https://doi.org/10.1016/S0361-476X(02)00014-0

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.

Ololube, N. P. (2013). The problems and approaches to educational planning in Nigeria: A theoretical observation. *Mediterranean Journal of Social Sciences*, *4*(12). https://doi.org/10.5901/mjss.2013.v4n12p37

Oyebade, S. A., & Dike, C. (2013). Restructuring Nigerian tertiary (university) education for better performance. Paper presented at the 11th Annual Meeting of the Bulgarian Comparative Education Society, Plovdiv, Bulgaria, 14-17 May.

Oyerinde, O. D., & Chia, P. A. (2017). Predicting students' academic performances – A learning analytics approach using multiple linear regression. *International Journal of Computer Applications*, *157*(4), 37–44. https://doi.org/10.5120/ijca2017912671

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

Romero, C. B., Ventura, S. N., Pechenizkiy, M., & Baker, R. S. (Eds.) (2010). *Handbook of educational data mining*. CRC Press. https://doi.org/10.1201/b10274

Serdyukov, P. (2017). Innovation in education: What works, what doesn't, and what to do about it? *Journal of Research in Innovative Teaching & Learning*, *10*(1), 4–33. https://doi.org/10.1108/JRIT-10-2016-0007

Ugar, A. A. (2018). ASUU strike: The Federal Government and Nigerian educational system. *International Journal of Education and Research*, 6(5), 19–32.

UN Department of Economic and Social Affairs (UN DESA). (2013). *World economic and social survey 2013: Sustainable development challenges*.

Vergaray, A. D., Guerra C., Cervera, N., & Burgos, E. (2022). Predicting academic performance using a multiclassification model: Case study. *International Journal of Advanced Computer Science and Applications (IJACSA)*, *13*(9), 881–889. https://doi.org/10.14569/IJACSA.2022.01309102

Wirth, R., & Hipp, J., (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining* (pp. 29–39). Springer-Verlag.

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 1–9. https://doi.org/10.1186/s40561-022-00192-z