

Placement Model for Students into Appropriate Academic Class using Machine Learning

Anadi Stella Uju¹, Khalid Haruna^{2*}, Ibrahim A. Lawal³, Raliya Abubakar⁴, Aminu Usman Jibril⁵ and Rabi Mustapha⁶

^{1,2,4,5}Department of Computer Science, Bayero University, Kano, Nigeria.

³Department of Information Technology, Bayero University, Kano, Nigeria.

⁶Department of Computer Science, Kaduna State University, Kaduna Nigeria.

ujustella84@gmail.com¹, kharuna.cs@buk.edu.ng^{2*}, alawal.it@buk.edu.ng³, raliyatsafe@gmail.com⁴,
auj4real@icloud.com⁵, rabichubu@gmail.com⁶

*Corresponding Author

Abstract

Choosing the right academic major for junior secondary students into senior secondary school will assist both students and their teachers toward achieving the academic goal. Traditionally, students seeking admission into senior classes (Gambia, Sierra Leone, Ghana, Liberia and Nigeria) must have passed stipulated examinations like Basic Education Certificate Examination (BECE) and/or West Africa Junior Certificate Examination, which are done at the end of year three (at a sitting). They must pass the exam(s) satisfactorily with no emphasis on any of Science, Art or Commercial related subjects. Some schools use “Mock exam” or “Placement exam” as the basis for their placement of students but all are done at a sitting (end of year three). Though this method is to an extent valid but associated with some challenges (bias) as it does not carry along the student’s academic history in making decision for placement into appropriate class. In this research, we proposed a model that predicts appropriate academic class of Science, Art or Commercial for Junior students based on the progressive academic performances (history) of their predecessors on related subjects using ten supervised machine learning techniques. Two evaluation techniques were applied. The validity of the proposed model was tested by comparing the proposed model against a baseline, in which the proposed model showed a better performance in predicting appropriate academic class. The correlation coefficient between the proposed model and the baseline was 0.3. This research will benefit all stakeholders in education and students in particular because their academic performances over time stands a better chance for appropriate placement.

Keywords: Placement Model; Appropriate Academic Class; Machine Learning; Correlation coefficient

1. Introduction

The need for proper selection of academic major for students’ going into senior secondary school is of paramount importance towards achieving academic goals. It is a useful strategy to mitigate failure, promote the achievement of better results and to better manage resources in higher education. It is also established that some students discontinue their educational pursuit (dropout) due to class misplacement from their basic education level (Migueis et al., 2018). Researches have shown that students drop out of school for various reasons such as parental poverty (Mathias et al., 2021), sickness (Sagatun et al., 2014), subjects studied at university as well as the secondary school grades (Mohamed & Ayman, 2019), nature of the learning environment (Ekanem et al., 2021), teacher-student relationship (Eknash et al., 2020), insecurity (Ramaswami et al., 2022), economic recession (Okoji et al., 2021), accumulated failure (Amelec et al., 2019) and so on. Predictive models have been proven to be an important approach toward achieving remarkable improvements in both productivity, and proficiency in almost all human endeavors (Patricia et al., 2020).

The placement exams (a sitting exam) used by schools for placement of students into various classes suffer some challenges as students tends to be less productive when writing exams under ill health condition, unsecured environment, psychological emotions from family crisis and even exam phobia (Okoji et al., 2021). Using their performances from such record to do placement will be biased.

In this paper, we proposed a placement model for students into appropriate academic class using machine learning. This model will use the progressive academic performance of students for eight terms on selected subjects to make placement. The generated datasets can be used for further research in educational field and the model can be used to predict appropriate academic class for students using their progressive academic performance.

The remaining sections of this paper are as follows: a detailed literature review is presented in section 2. Section 3 presents our proposed methodology. Results and discussion is presented in section 4. Conclusion and recommendations for future work is presented in section 5.

2. Literature Review

Most of the researches on student academic performance are geared towards the management of at-risk students at the higher education level (Chawla et al., 2020) and, to a greater extent, in the distance education modality (Naser et al., 2020). Very little information has been found on the application in basic or secondary education, where just simple analyses of the information based on statistical methods have been carried out (Okubo et al., 2018). There are some differences and/or advantages in applying data mining over just using statistical models (Helal et al., 2018). Therefore, any predictive model designated to analyzing the students' academic performance must take cognizance of the complex and nonlinear relationships that exist among variables (Neema et al., 2019).

An adaptive recommendation system was proposed by Mohamed et al. (2018) for predicting a suitable education path(s) for students in college preparatory year. The adaptability was achieved by automatically applying different data mining techniques for extracting relevant features and building a tailor-made model for each education path. It recommends a suitable engineering department among seven engineering departments for each student based on his academic performance.

Vijay et al., (2017) mainly focuses on the prediction of a student's university result by making use of different attributes using KNN. These attributes might be of quantitative and qualitative type. The quantitative attributes used are Internal Assessments, Attendance percentage, Number of On-Duties taken and Overall Assignments completed. The qualitative attributes include Subject feedback, Faculty feedback, and whether the student is a Day Scholar/Hosteller. KNN was used against the historical data of students for a more accurate prediction of results. This algorithm makes use of the Euclidean distance which is used to find the nearest record and predicts better results which help students maximize their academic output.

Delna et al., (2018) used KNN to predict the category to which a student fails based on the skills, ability and behavior of the student after 4 weeks of 12 weeks course, similar to the work of Tuomas et al., (2019) who predicted at-risk students in online course on touch-typing skill after the very first lesson.

Iam-On et al., (2017) applied ensemble ML techniques for predicting dropout students in the Thai University based on their first-year academic performance. 124 students were used and the performance of the model was appreciative. Migueis et al., (2018) used the information available at the end of the first-degree year of students' academic career to segment students based on evidence of failure or high performance.

Amalec et al., (2020) applied ML in school dropout prediction based on Columbia University students' academic performance. They focus their interest in determining the multiple factors that most affect the performance of students (dropout and failure) at the different educational levels (basic, middle and higher education) through the use of the large amount of information that current computer equipment allows to store in databases. Ten classifiers were used on both demographic and academic features. Random forest emerged the highest accuracy of 85%.

Farouk et al., (2019); Jie Xu et al., (2017); Fares et al., (2021); Arindam et al., (2022); Amoo et al., (2017) predicted students' academic performance by tracking the academic performance of students for a short term. This yielded a better result when compared with existing methods (one sitting exam).

Table 1: Summary of Related works

Year	Title	Technique	Findings	Limitations	Author(s)
2019	Adaptive recommendation system using machine learning algorithms for predicting student's best academic program	Predictive approach.	There is no specific learning algorithm that best predicts all Engineering students of Al-Azharuni. Cairo, into various departments.	Used one-year preparatory course for the higher institution.	Mohammed Ezz and Ayman Elshenawy
2019	Hybrid approach for selection of academic major using naïve bayes (2019)	Naïve bayes technique	Naïve Bayes is a perfect classifier for Predicting students best academic major using performance history from JSS1-JSS3 for a small dataset.	Used few datasets.	Farouk lawan Gambo, Adamu sani yahaya, Abdullahi Abubakar, Khalid Haruna and Aminu. A. Aliyu
2018	Early segmentation of students according to their academic performance	Predictive approach (DT, SVM, NB, boosted trees).	Early classification reduces the failure rate, promote academic achievement	Used for year one-degree students.	Migueis V., Freitas A., Paulo J.V Garciab P.
2017	Improved student dropout prediction in Thai university using an ensemble of mixed-type data clustering	Predictive approach.	Using an ensemble of mixed-type data clustering generate a better prediction of dropout in higher institutions	Used for year one-degree students.	Lam-On, N. Boongoen (2017)
2017	Result Prediction Using K-Nearest Neighbor Algorithm for Student Performance improvement	KNN Classification Algorithm	Prediction of a student's university result by making use of different attributes.	Performed on only 124 students.	Vijay Muralidharan, Pravien M & Janani Balaji
2020	Data mining applied in school dropout prediction	NB, DT, RF, SVM.	Naïve Bayes Produced the highest accuracy in the prediction of at-risk students.	Used year one-degree students to predict at risk students.	Amelec Vitoria et al (2020)
2018	Prediction of Students' Performance at University Level	KNN Classification Algorithm	Predict the category to which a student falls based on their skills, ability and behavior of the students.	Performed on university students	Delna Sebastian and Saravanan K. N.

2017	Progressive prediction of student performance in college programs	Recurrent Neural Network	Predicted student final grade by progressively adjusting weight.	Used few datasets.	JieXu, Yuli Han, Daniel Marcu, and Mihaela Van Der Schaar.
2018	A Neural Network Approach for Students' Performance Prediction.	Artificial Neural Network	Prediction of students' final grade by updating weekly performance for 15 weeks of online course.	Performed on university students.	F. Okubo, T. Yamashita, A. Shimada & H. Ogata
2022	An approach to Predict a Student's Academic Performance using Recurrent Neural Network	Recurrent Neural Network	Used first and second term academic performance to predict third term grades.	Used very few datasets.	Arindam Mondal & Joydeep Mukherjee
2017	Predictive modelling and analysis of the academic performance of secondary school students:	Artificial Neural Network approach	WAEC grades of science students can serve as an important feature to predict their PUTME result.	ANN model does not specify an explicit mathematical model for the relationship.	Amoo M. Adewale, Alaba O. Bamidele and Usman O. Lateef
2022	A Novel Hybrid Classification Approach for Predicting Performance of Student in E-learning.	Hybrid approach	Predicting students' performance in virtual classes	For higher institution students	Ali, H., Mohamed, C., Abdelhamid, B., Ourdani, N., Alami, T. E.

From the reviewed literature, it was observed that most of the data used for analysis were obtained at a very early stage regardless of the improvement that may occur later. Predicting student performance is usually not a one-time task (Jie Xu et al., 2017). You cannot accurately predict a student's academic potential only through his early grades or even through his static demographic traits (Ali et al., 2022). Later prediction yields better accuracy and is preferred when intervention is not needed (Menno & Alfa, 2018). This requires continuous tracking of the student's performance over a longer period of time, to reveal his/her true potentials. It is also clear from the literature that many researchers have worked so much on the prediction of student's performance mainly in higher institutions based on first year or preparatory year performance. But little attention has been given to basic education levels (primary and Junior Secondary schools) which is the foundation class where all students offer the same subjects.

3. Methodology

This section shows the methodology used in carrying out the research. Three different datasets were used. The first dataset was generated from students' performance whose placement were based by choice (no defined pattern). The second and third datasets were generated from the same students (progressive academic records from JSS1 –JSS3 and Mock performance in one sitting exam respectively) from the five schools.

3.1 Dataset

The characteristics of the dataset used in this research is presented in Table 2.

Table 2: Dataset

Dataset	Source	Description
Subject category	WAECeducurriculum, (2019)	Subjects are categorized into three classes of science, art and commercial in the following subject combinations; Science: Biology, Chemistry, Physics, Basic science and Basic technology. Art: Economics, Literature, Government, Social studies and Cultural and creative art. Commercial: Account. Commerce, Entrepreneur, Civics and Business studies.
Terms Students	Federal Airport Authority of Nigeria (FAAN); Prince of Wales Secondary School, Sierra Leone; Nusrat Senior Secondary School, Gambia; Akosombo International School, Ghana and Our Lady of Grace International school, Liberia.	Examination scores of eight terms (three sessions) for junior students. 1521 student records were collected. Twenty-one (1.4%) were dropped during preprocessing. 1500 instances were used for the analysis.
Dataset	Source	Description
Subject category	WAECeducurriculum, (2019)	Subjects are categorized into three classes of science, art and commercial in the following subject combinations; Science: Biology, Chemistry, Physics, Basic science and Basic technology. Art: Economics, Literature, Government, Social studies and Cultural and creative art. Commercial: Account. Commerce, Entrepreneur, Civics and Business studies.
Terms Students	Federal Airport Authority of Nigeria (FAAN); Prince of Wales Secondary School, Sierra Leone; Nusrat Senior Secondary School, Gambia; Akosombo International School, Ghana and Our Lady of Grace International school, Liberia.	Examination scores of eight terms (three sessions) for junior students. 1521 student records were collected. Twenty-one (1.4%) were dropped during preprocessing. 1500 instances were used for the analysis.
Dataset	Source	Description
Subject category	WAECeducurriculum, (2019)	Subjects are categorized into three classes of science, art and commercial in the following subject combinations; Science: Biology, Chemistry, Physics, Basic science and Basic technology. Art: Economics, Literature, Government, Social studies and Cultural and creative art. Commercial: Account. Commerce, Entrepreneur, Civics and Business studies.

<p>Terms Students</p>	<p>Federal Airport Authority of Nigeria (FAAN); Prince of Wales Secondary School, Sierra Leone; Nusrat Senior Secondary School, Gambia; Akosombo International School, Ghana and Our Lady of Grace International school, Liberia.</p>	<p>Examination scores of eight terms (three sessions) for junior students. 1521 student records were collected. Twenty-one (1.4%) were dropped during preprocessing. 1500 instances were used for the analysis.</p>
---------------------------	---	--

3.2 The Existing Method

The existing manual method as used by the sampled schools, subject students to a placement examination (at a sitting) at the end of year three (3). The average score of all the subjects per student were taken and students were categorized based on their overall average performance. The classes are categorized based on the following scale:

- i) 65 marks and above __ science class
- ii) 64 – 55 marks __ commercial class
- iii) 54 and below __ art class

Here, science class is prioritized (higher precedence) over other classes. That is, sciences are for outstanding students and art class for weak students.

3.3 Algorithm of the proposed model

Stage 1: Data is collected from students from five schools (Q_i).

Stage 2 (i) Do preprocessing by sorting student scores

(ii) Cleaning by dropping records less than eight terms ($t < 8$)

(iii) Apply the equation $\hat{Y} = (\sum_{i=1}^t Q_i) \frac{1}{8}$. Where \hat{Y} = preprocessed record, t = number of terms, i = scores per term and Q = students. Then drop $\hat{Y} < 50$.

Stage 3: Do classification by applying ten classifiers on Junior Mock Record (JMR), Junior Progressive Record (JPR) then Base Record (BR).

Stage 4: Perform evaluation and validation using 10-fold cross validation

3.4 Activity chart of the proposed model

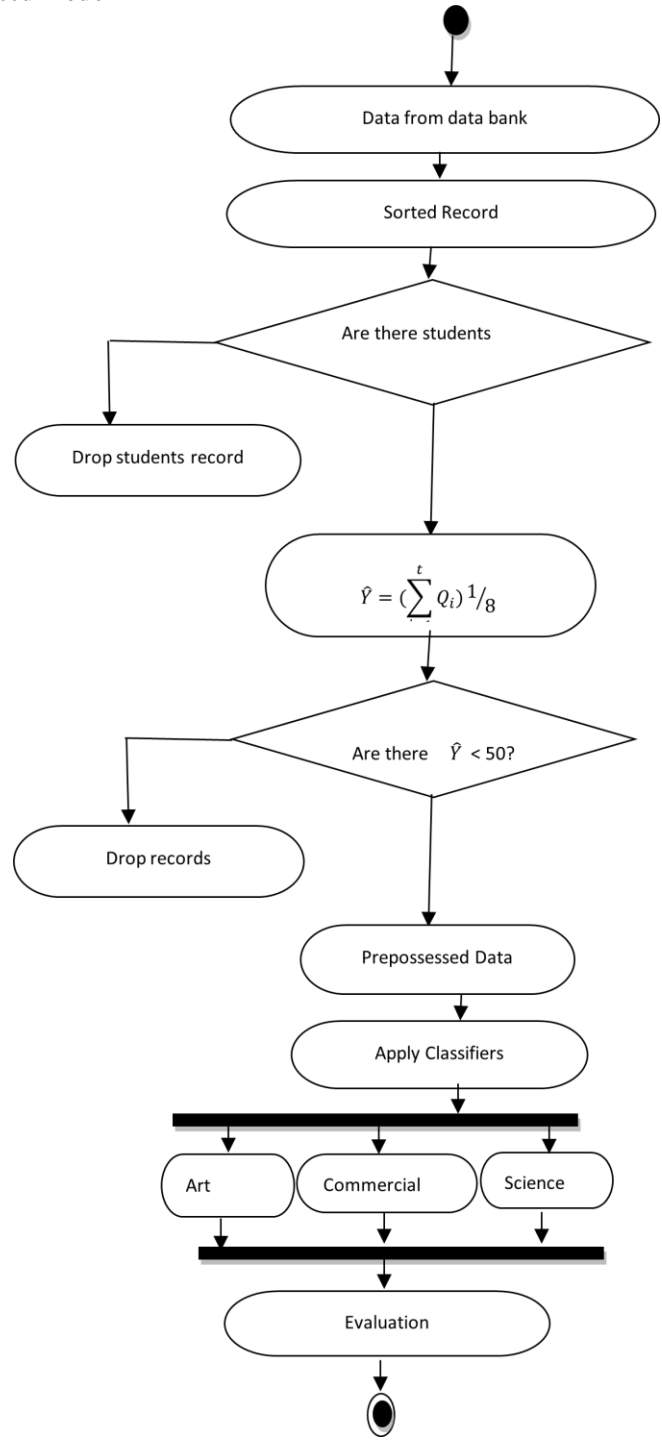


Figure 1: Activity Chart of the Proposed Model

3.5 Proposed Placement Model for Students into Appropriate Academic Class using Machine Learning

Our proposed model is divided into four stages as represented in figure 2. It is made up of data resource and collection, preprocessing, classification and evaluation stages.

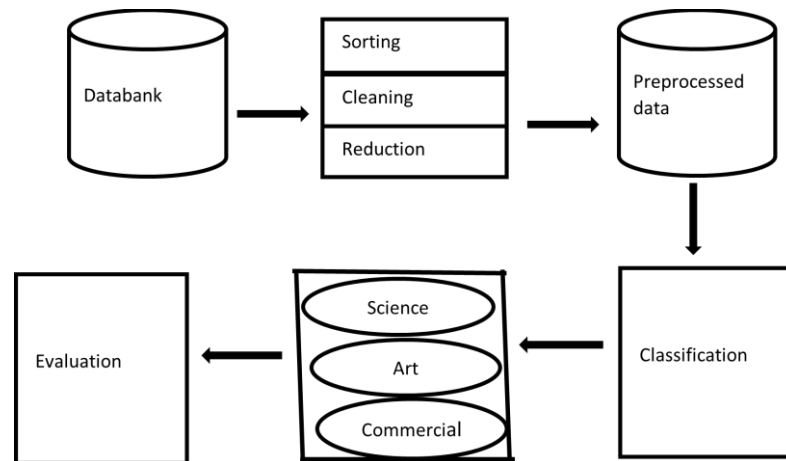


Figure 2: Architecture of the proposed model

Stage 1: Data Resource and Collection

The source of the data used in this research was obtained from five private schools (Federal Airport Authority of Nigeria (FAAN); Prince of Wales Secondary School, Sierra Leone; Nusrat Senior Secondary School, Gambia; Akosombo International School, Ghana and Our Lady of Grace International school, Liberia) from the five West Africa Examination Council, (WAEC) member countries. Only academic record of students who enrolled into JSS1 class in 2015 were used for experiments two and three while year 2001 students record were used for experiment one (base record) with emphasis on the following subjects: Basic science, Basic technology, Social studies, Civic education, Cultural and creative art and Business studies. Every given student in each class has to offer three combinational subjects either for science, art or commercial as stipulated by the WAEC policy, (2013) and shown in Figure 3. Eight academic terms must be completed in JSS (Junior secondary school). 1521 students’ records were collected for the analysis.

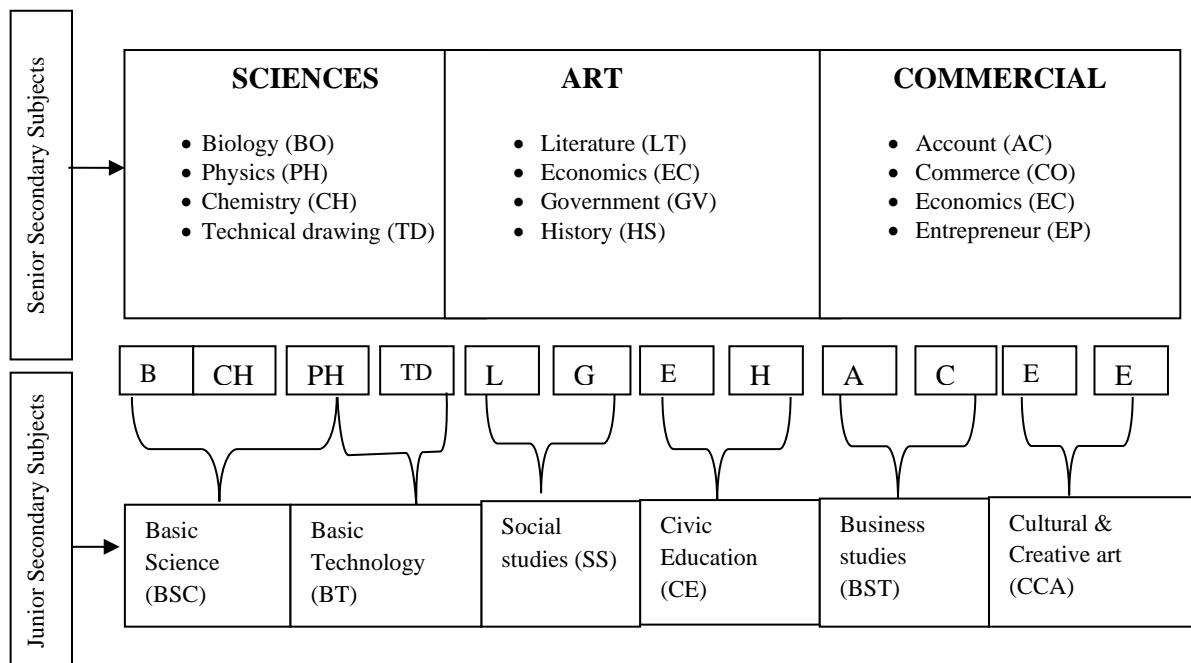


Figure 3: Summary of relationship between subjects
 Source: WAECeducurriculum, (2019)

Stage 2: Data Preprocessing

Data preprocessing is the preparation of data into a usable form. To achieve this, the collected datasets have to undergo certain stages which include:

- Data Sorting: during data collection, records of students were randomly collected irrespective of class (Science, Art or Commercial). These records will then be sorted based on the actual class which they are presently in and labelled.
- Data Cleaning: Is the process of detecting and correcting corrupt or inaccurate record from a record set. Nine students with incomplete records (less than 8 terms) were extracted from the datasets as they represented only 0.6% of the sample size.
- Data Reduction: In ML, data reduction is done to bring records into a single representation. In this model, we applied Equation 1)

$$\hat{y}_i = \left(\sum_{i=1}^t Q_i \right)^{1/8} \tag{1}$$

where i represent the scores obtained by student Q and t number of terms. Then, Q_i^t denote student i 's academic scores at term (t). Our independent variable is the scores obtained by the students (predictors) and the dependent variable is the targeted output (Science, Art or Commercial). The \hat{y}_i for all the students will be filtered by setting a threshold of $\hat{y}_i \geq 50$. Twelve students (0.8%) whose scores fall below this range were extracted. This threshold was set in conformity to the set criteria by West Africa Examination Council law, (1995). Out of the 1521 students' records collected, 21 students record were dropped in this stage leaving 1500 samples to be used as input in the next stage.

Stage 3: Classification and prediction

Classification is a technique used to assign a class of unseen records as correctly as possible by using a collection of records called a training dataset, where each tuple in the training set comprises of a set of attributes, and then one of the attributes is called a class. Because is supervised learning, the class label (art:0, commercial:1 and science:2) was assigned to all the datasets for python compatibility. The preprocessed data were divided into 70:30 for training and testing respectively. Then, SMOTE (Synthetic Minority Oversampling Technique) was applied on the training datasets to balance the minority class. Ten classifiers [Linear Regression (LiR), Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), Logistic Regression (LoR), Recurrent Neural Network (RNN), Gaussian Process Classifier (GPC) Adaboost, K-nearest neighbor (KNN) and Support Vector Machine (SVM)] were passed on the datasets for classification.

Stage 4: Performance Evaluation

In this phase, the result from the existing model and that of the proposed model was evaluated to ascertain the differences in their performance. Confusion matrix was used to measure accuracy, precision, recall and F1 score of all datasets. The validity of the proposed model was also tested. The general formula for these measures as stated by Chawla et al., (2020) are presented in Equations (2)-(5)

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of data points}} \tag{2}$$

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})} \tag{3}$$

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Negative})} \tag{4}$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

4. Results and Discussion

The results of the three experiments are presented in Table 3 and graphically represented in Figure 4.

Table 3: Result of Experiment

Evaluation Metric	Classifier	Ensemble		Regression			Non-Linear	Neural Network	Gaussian		
		RF	Adaboost	LiR	LoR	SVM	DT	RNN	NB	GPC	KNN
Accuracy1		53/95/87	52/71/60	44/74/57	44/65/60	46/65/54	64/91/73	61/72/65	48/75/60	73/93/90	59/94/81
	Accuracy2	72/93/80	71/83/80	58/76/74	46/58/54	68/70/69	73/92/88	50/57/49	64/80/49	71/90/88	71/85/92
Precision		82/98/90	74/72/71	42/83/81	66/42/39	46/66/62	80/84/82	57/50/49	43/83/80	93/82/80	95/82/80
Recall		81/92/88	71/69/67	44/67/65	66/32/43	55/69/65	81/99/93	72/68/65	49/69/65	93/82/80	94/81/80
F1 Score		81/94/89	72/70/68	42/74/72	66/36/40	50/67/65	80/90/87	63/57/55	45/75/71	93/82/80	94/81/80

In the first experiment using Base Record (BR), considering the state-of-the-art performance evaluation technique (cross validation-accuracy2), decision tree attained the highest accuracy of 73% and logistic regression lowest accuracy of 46%. GPC attained the highest accuracy of 73%, logistic and linear regression the least performed (70/30 splitting-accuracy1) with 44%.

In the second experiment (JPR), Random Forest emerged with the highest accuracy of 93% and least with RNN 57% (accuracy2). Random forest obtained 95% against logistic regression and Support vector machine with least score of 65% accuracy (with 70/30 splitting-accuracy1).

The third experiment (JMR) recorded 90% accuracy which is the highest accuracy with 70/30 splitting and 54% as lowest with logistic SVM (accuracy1). KNN scored the highest accuracy of 92% with cross-validation while NB and RNN as the least performed with 49% (accuracy2).

In some experiments, 70/30 splitting evaluation metric out performed cross validation performance metric and vice versa. This research preferred the results of cross validation because it is more efficient and accurate and gives model the opportunity to train on multiple train-test split (Ali et al., 2022). It was also observed that there is a general higher accuracy in the results using JPR (progressive record) against the results of JMR (one sitting exam record) and BR (base record). Our model used subjects from junior secondary which have a direct link to subjects offered in senior secondary. Eight terms (longer time) were considered as students tend to progress with time. We also measured the correlation coefficient between the three datasets used for analysis and validated the performance of the proposed model by comparing the performances of the existing method, the proposed model and the base record. Results are presented in Table 4, figure 5 and figure 6 respectively.

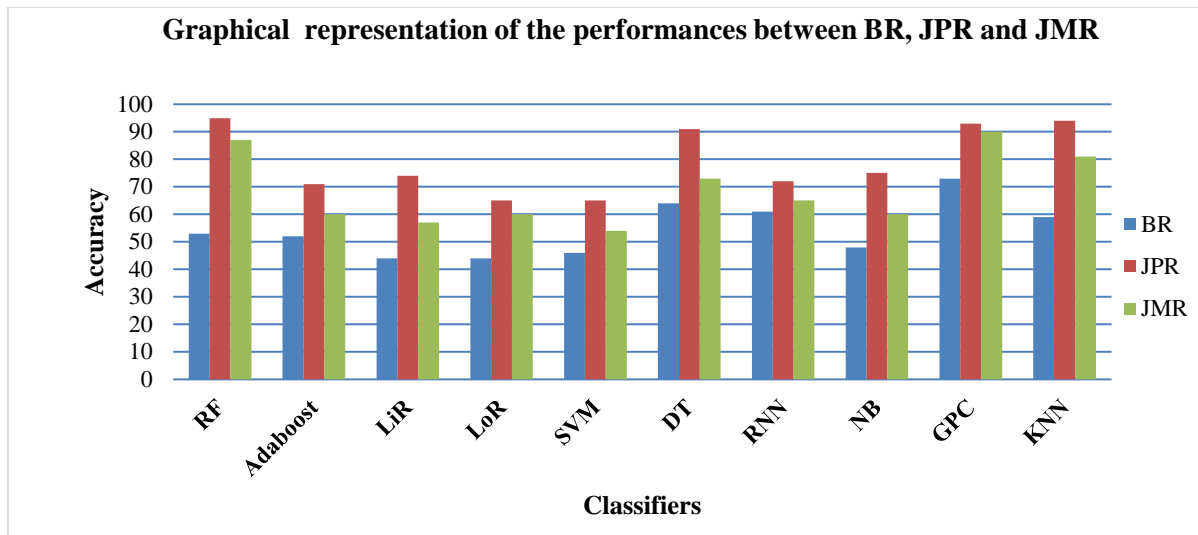


Figure 4: Graphical Representation of the Classifiers Performances between BR, JPR and JMR

Correlation coefficient is a valuable numerical measure of association between two variables which is a value between -1 and 1 indicating the strength of the association of the observed data for the two variables (Mohamed et al., 2019).

Table 4: Correlation coefficient of datasets

JMR/BR	0.736714
JPR/BR	0.419412

From Table 4, it can be seen that the correlation coefficient between JMR/BR is higher when compared with that of JPR/BR. This is because the records obtained from one sitting examination (JMR) suffers some challenges (bias) familiar with base records (BR). In this context, the lower the correlation coefficient value, the better the datasets and vice versa.

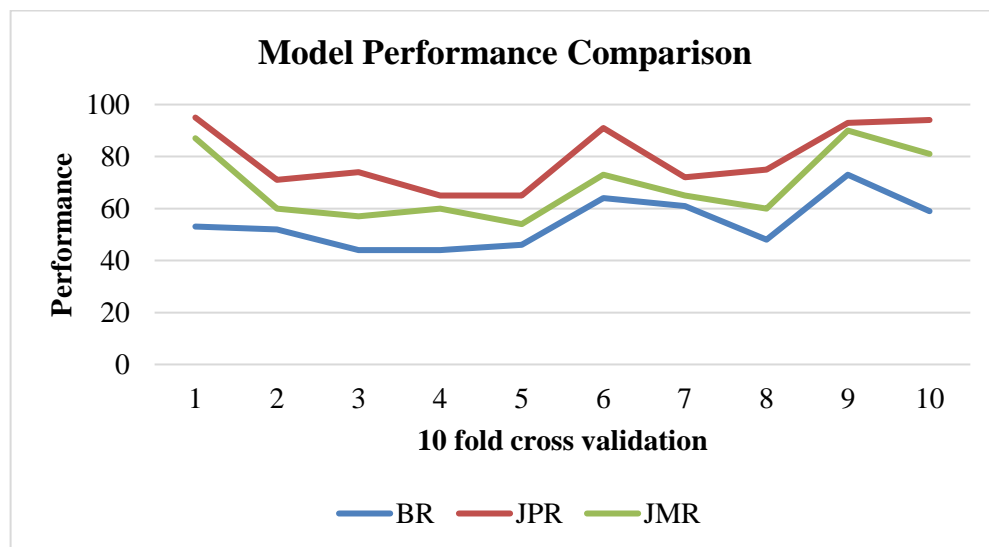


Figure 5: Validation of Model Performance

Figure 5 shows the model performances on the three datasets when ten classifiers were passed. 10-fold cross validation was used to test the validity of the models. JPR clearly outperformed the other models in all iteration of the validation process. That means, students will be appropriately predicted to the classes where they have higher academic potentials using the proposed model. The JMR (existing model) is valid to an extent (as stated in the abstract) but are biased due to factors that affect records generated from one sitting examination. The BR model recorded lowest model performance because the records used were not refined (preprocessed). This validation test justified the proposed model preferable over the other models within the context.

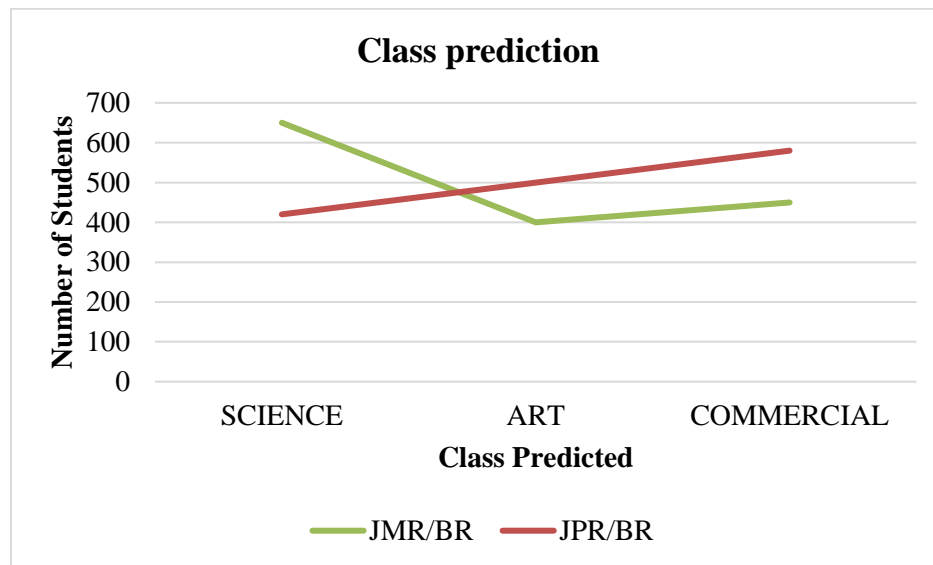


Figure 6: Class Prediction using JMR/BR and JPR/BR Datasets

In figure 6, out of the 1500 records used, 650, 400 and 450 students were placed into science, art and commercial classes respectively using JMR/BR as against 420, 500 and 580 students placed into science art and commercial classes respectively using JPR/BR. Their senior secondary performances were compared to the results of the placement and it showed that the proposed model performs better than the existing model.

5. Conclusion and Future Research Direction

The results obtained after the experiments showed that our proposed model can efficiently and accurately predict student’s appropriate academic class by taking their progressive academic performance on selected subjects for longer time (JPR) compared to the result obtained using one sitting exam (JMR) on all the thirteen subjects and the results obtained using base record (BR). The generated datasets can be used for further research in educational field. The modality for handling students below the threshold score ($\hat{y}_i < 50$) is a direction for future research.

References

- Ali, H., Mohamed, C., Abdelhamid, B., Ourdani, N., Alami, T. (2022). A Novel Hybrid Classification Approach for Predicting Performance of Student in E-learning. *Engineering and Communications Technologies*, vol 147. Springer, Cham. https://doi.org/10.1007/978-3-031-15191-0_5
- Amelec, V., Jesús, G., Gulian, W., Niebles, N., &Hugo, H. (2020). Data Mining Applied in School Dropout Prediction. *A case study. Procedia Computer Science*, 86, 56-70. doi:10.1088/1742-6596/1432/1/012092
- Amoo, M. Adewale, A., & Bamidele, U. (2017). Predictive modeling and analysis of academic performance of secondary school students: Artificial Neural Network approach. *International Journal of Science and Technology Education*. 9(1), 1-8. DOI: 10.5897/IJSTER2017.0415

- Arindam, M., & Joydeep, M. (2022). An approach to Predict a Student's Academic Performance using Recurrent Neural Network (RNN). *International Journal of Computer Application, Springer, Berlin, Heidelberg*, 2(1), 84–89. Doi:10.5120/ijca2022917352
- Chawla, N., Bowyer, K., Hall L., & Kegelmeyer, W. (2020) "Supporting students' learning- pathway choices by providing rule-based recommendation system". *International Journal of Education and Information Technologies*, 9, 81–94
- Delna, S., & Saravanan, K. (2018). Prediction of Students' Performance in University Level using KNN Classification Algorithm. *International Journal for Research in Technological Studies*. 5(6), 45 -56.
- Ekanem, S., Apebende, E., & Ekefere, E. (2021). Learning environment and pupils' academic performance: Implications for counseling Association of Nigeria. *The Counsellor Official Journal of Counselling Association of Nigeria (CASSON)*, 29(1): 32-42.
- Ekansh, M., Chandrima, R., & Siddharth S. (2020). Prediction of factors associated with the dropout rates of primary to high school students in India using data mining tools. *Advances in Intelligent Systems and computing* 1013, 242-251. https://doi.org/10.1007/978-981-32-9186-7_26
- Fares, B., & Khalid, M. (2018). Tracking and predicting student performance in university using RNN. *Neural computation*, 9(8):1735-1780.
- Farouk, L., Adamu, S., Abdullahi, A., Khalid, H., & Aminu, A. (2019). Hybrid Approach for Selection of Academic Major Using Naïve Bayes.
- Helal, S., Li, J., Liu, L., Ebrahimiea, E., Dawsonb, S., & Murrayc, D. (2018). Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, 161, 134–146.
- Iam-On, T., & Boongoen, T. (2017). Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings. *International Journal of Machine Learning and Cybernetics*, 8, 497–512. doi 10.1007/s13042-015-0341-x
- JieXu, K., Yeong H., & Mihaela, V. (2017). A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing*, 11(5):742-753.
- Mathias, U., & Dada, T. (2021). Sociological investigation of the determination factors and the effects of child street hawking in Nigeria: Agege Lagos, under survey. *International Journal of Asian Social Science*
- Mduma, N., Kalegele, K. & Machuve, D., 2019. A Survey of Machine Learning Approaches and Techniques for Student Dropout Prediction. *Data Science Journal*, 18(1), p.14. DOI: <http://doi.org/10.5334/dsj-2019-014>
- Menno V., & Alfa, A. (2018). Early prediction of students' performance in a distance learning university. *IOP Conference Series: Materials Science and Engineering*, 200, 1–10. *Education and Information Technologies*. doi: 10.5815/ijisa.2018.12.04
- Migueis, V., Freitas, A., Garciab, P., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, 115, 36–51. doi:10.1016/j.dss.2018.09.001
- Mohamed, E., & Ayman, E. (2019). Adaptive recommendation system using machine learning algorithms for predicting student's best academic program. *Education and Information Technologies, Springer Science+Business Media, LLC*. <https://doi.org/10.1007/s10639-019-10049-7>
- Naser, S., Zaqout, I., Atallah, R., Alajrami, E., & Abu Ghosh, M. (2020). Predicting student performance using artificial neural network: In the faculty of engineering and information technology. *International Journal of Hybrid Information Technology*, 8(2), 221–228
- Okoji, O. (2021). Effects of socio-economic status on educational attainment of secondary school students in Rivers State of Nigeria. *Gender and Behaviour*, 11(1): 5162-5166
- Okubo, F., Yamashita T., Shimada, A., & Ogata, H. (2018). A Neural Network Approach for Students' Performance Prediction. *Proc. ICCE2018*, p 739-744. <http://dx.doi.org/10.1145/3027385.3029479>
- Patricia, N., & Norbert, N. (2020). Challenges Encounter Using Cbt by 2015 Utme Candidate in Owerri Zone One Nigeria: Test Validity Implications. *IOSR Journal of Research and Method in Education (IOSR-JRME)* 5, 5 (2020), 28–35.
- Ramaswami, G., Susnjak, T., Mathrani, A., Lim, J. and Garcia, P. (2022), "Using educational data mining techniques to increase the prediction accuracy of student academic performance", *Information and Learning Sciences*, Vol. 120 No. 7/8, pp. 451-467. <https://doi.org/10.1108/ILS-03-2019-0017>

- .Sagatun, Å., Heyerdahl, S., Wentzel-Larsen, T., & Lien, L. (2014). Mental health problems in the 10th grade and non-completion of upper secondary school: The mediating role of grades in a population-based longitudinal study. *BMC Public Health* 14(1). <http://www.biomedcentral.com/content/pdf/1471-2458-14-16.pdf>
- Tuomas, T., & Hannu, T. (2019). Predicting and preventing student failure – using the k -nearest neighbour method to predict student performance in an online course environment. Typing Master Finland Oy, Eerikinkatu 4 A 16, FI-00100 Helsinki, Finland
- Vijay, M., Pravien, M., & JananiBalaji, (2017). Result Prediction Using K-Nearest Neighbor Algorithm For Student Performance Improvement. *International Research Journal of Electronics & Computer Engineering*. 1, 7-10. DOI: 10.24178/irjece.2017.3.1.07
- West Africa Examination Council policy, educriculum (2019). www.waec.org