JOURNAL OF SCIENCE EDUCATION AND RESEARCH (JSER)

PROF. PATRICK C. IGBOJINWAEKWU

JOURNAL OF SCIENCE EDUCATION AND RESEARCH (JSER)

Vol. 3 JULY - AUGUST; 2025

ISSN ONLINE: 3092-9253

Editor in-Chief

PROF. PATRICK C. IGBOJINWAEKWU

JOURNAL OF SCIENCE EDUCATION AND RESEARCH

(JSER), 3, JULY - AUGUST; 2025

ISSN Online: 3092-9253

Published in July, 2025.

Printed in Nigeria in the year 2025 by:
Love Isaac Consultancy Services
No 1 Etolue Street, Ifite Awka, Anambra State, Nigeria
+234-803-549-6787, +234-803-757-7391

EDITORIAL BOARD

EDITORIAL

Journal of Science Education and Research (JSER) is a peer-reviewed published Bimonthly. It aimed at advancing knowledge and professionalism in all aspects of educational research, including but not limited to innovations in science education, educational technology, guidance and counselling psychology, childhood studies and early years, curriculum studies, evaluation, vocational training, planning, policy, pedagogy, human kinetics, health education and so on. JSER publish different types of research outputs including monographs, field articles, brief notes, comments on published articles and book reviews.

We are grateful to the contributors and hope that our readers will enjoy reading these contributions.

Prof. Patrick C. Igbojinwaekwu
Editor-in-Chief

TABLE OF CONTENTS

JOURNAL OF SCIENCE EDUCATION AND RESEARCH (JSER)
VOL.3, JULY - AUGUST, 2025

IMPLEMENTATION OF MACHINE LEARNING BASED SCHOOL CLASS PLACEMENT PREDICTION SYSTEMS FOR SECONDARY SCHOOL, USING RANDOM FOREST

[1]Omopariola Adebola Victor, [2]Eniolorunda Wande Stephen
[1]omopariolaa@veritas.edu.ng, [1] https://orcid.org/0000-0002-1234-3911
[2]weniolorunda@yahoo.com
[1, 2]Department of Computer Science
[1,2]Veritas University, Abuja.

Abstract

This study presents the development and implementation of a machine learning-based framework for optimizing class placements for SSS (Senior Secondary School) classes for JSS (Junior Secondary School) students based on their historical academic results in Nigerian secondary schools, with a focus on the Federal Capital Territory (FCT). Traditional placement methods often rely on subjective evaluations and standardized exams, which may introduce bias and overlook key behavioral and demographic factors. To address these limitations, a predictive model using the Random Forest algorithm was developed, leveraging data from 1,500 anonymized student records spanning academic scores, demographic details, and behavioral indicators. The dataset underwent preprocessing, including normalization, one-hot encoding, and handling of class imbalance using SMOTE. The model's performance was evaluated using accuracy, precision, recall, and F1 score, achieving an accuracy of 87.6%, outperforming baseline classifiers like Decision Trees and Naive Bayes. Feature importance analysis identified Mathematics, English, and attendance as key predictors. The model demonstrated a 29% improvement in placement accuracy over traditional methods, significantly reducing misclassifications, especially among students from marginalized backgrounds. The findings affirm the potential of machine learning in enhancing fairness and precision in educational decision-making. This research offers a scalable, data-driven approach to class placement and highlights the broader applicability of predictive analytics in educational planning and reform. Based on the research conducted, the following recommendations were made among others; institutions of learning should adopt the use of machine learning for proper students' class placement and support student.

Keywords: Machine Learning, Random Forest, Data-driven Class Placements, Educational Decision-Making.

Introduction

Secondary education plays a critical role in shaping the academic and professional futures of students. The process of placing students in appropriate educational institutions significantly impacts their academic success and long-term career direction. Traditionally, these placements are often guided by manual assessments in most cases, which may be prone to biases and inefficiencies. 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 (Miguéis, Freitas, Garcia & Silva, 2018). Researches have shown that students drop out of school for various reasons such as parental poverty (Dada, 2017), sickness, subjects studied at university as well as the secondary school grades, nature of the learning environment, teacher-student relationship, insecurity, economic recession, accumulated failure 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 endeavours. With the rise of data-driven decision making, machine learning (ML) techniques have emerged as effective tools for addressing these challenges. Conventional approaches, based on standardized test scores, teacher assessments, and socioeconomic indicators frequently reinforce systemic inequalities (Smith & Patel, 2022). Research shows that marginalized students are 34% more likely to be inaccurately placed due to biased evaluations (UNESCO, 2023).

We proposed placement model for students into appropriate academic class using machine learning. This model will use the progressive academic performance of students over nine 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.

Machine learning (ML), particularly Random Forest, presents a transformative solution by utilizing historical data to predict optimal placements while reducing human bias. Studies by Adebayo et al. (2023) reveal that ensemble methods like Random Forest achieve an 89% accuracy rate in educational classification tasks, surpassing logistic

regression (72%) and decision trees (81%). This study builds on these advancements to enhance equity and scalability within Nigeria's Federal Capital Territory (FCT).

Among various ML methods, Random Forest has proven to be a robust and interpretable algorithm for predictive modeling. Its ability to handle large datasets with multiple variables makes it an ideal choice for enhancing secondary education placement strategies. By leveraging historical academic, demographic, and behavioral data, Random Forest can predict optimal placements with greater accuracy, thereby minimizing mismatches and improving overall educational outcomes.

Statement of the Problem

The current methods of secondary education placement often rely on subjective judgment or rigid criteria, which fail to account for the complex character of student capabilities and aspirations. This leads to:

1. Subjectivity in Evaluations: Manual assessments, which rely heavily on teacher evaluations and standardized test scores, are inherently vulnerable to human bias and inconsistency. For instance, educators in low-resource schools often lack training in objective evaluation frameworks, leading to skewed judgments that brings disadvantages to students from marginalized communities (Oyediran Adekunle & Eze, 2021). A 2021 study in the FCT revealed that students from rural schools were 28% more likely to receive lower placement recommendations than their urban counterparts with identical academic scores, highlighting systemic prejudice in manual processes. This subjectivity perpetuates cycles of inequity, as students from underfunded schools are funneled.

2. Data Fragmentation: Most educational institutions in the rural FCT often operate in silos, with academic records, demographic data, and behavioral metrics scattered across disconnected databases. This fragmentation prevents educators from constructing holistic student profiles, as critical indicators like socioeconomic status, extracurricular engagement, or learning disabilities are rarely integrated into placement decisions (Abideen, Mazhar, Razzaq, Haq, Ullah, Alasmary & Mohamed, 2023). For example, a student excelling in STEM competitions but struggling with exam anxiety may be overlooked for advanced science programs if only test scores are considered. Such gaps in data utilization lead to mismatches between student capabilities and institutional offerings, stifling academic growth.

3. Challenges in identifying and resolving placement process inequities.

These issues necessitate a shift towards automated, data-driven approaches to optimize placement strategies and ensure equitable educational opportunities for all students.

Objectives of the Study

Main Objective: Develop a Random Forest-based ML framework to enhance equity and accuracy in secondary education placements.

Specific Objectives;

1. To identify key predictors of successful placements based on historical data.
2. To train and evaluate a Random Forest model for predicting optimal placements.
3. To provide actionable insights for educators to refine placement strategies.

Research Questions

The following research questions guided the study;

1. What are the key factors influencing secondary education placement success?
2. How does class imbalance in student performance data affect model generalizability
3. What institutional reforms can mitigate biases identified through SHAP (SHapley Additive exPlanations) analysis?

Significance of the Study

This study aims to bridge the gap between traditional placement practices and modern data analytics by introducing a robust ML-based framework. The findings will benefit;

- Educators: Advances Educational Data Mining (EDM) by integrating behavioral metrics into placement models.
- Policymakers: By highlighting systemic issues and suggesting improvements.

- Students: By ensuring fair and accurate placements that align with their potential and aspirations.

Scope and Limitations

The study focuses on secondary education placements within a defined geographic region. It utilizes historical academic and demographic data for model training and evaluation. However, its generalizability may be limited by the availability and quality of data. The placement of students in secondary education institutions plays a crucial role in shaping their academic and professional futures. Traditionally, student placement has relied on standardized examinations, historical academic records, and in some cases, socio-economic factors. However, these methods often fall short in accurately identifying the best educational pathways for students, leading to inefficiencies, misplacements, and underperformance.

- Scope: Focus on public secondary schools in the Federal Capital Territory (FCT), Nigeria, using anonymized student records (2019–2023).

- Limitations:
  - Generalizability constrained by regional data availability.
  - Potential biases in historical datasets (e.g., incomplete socio-demographic variables).

With the rise of data-driven decision-making, machine learning (ML) has emerged as a powerful tool for improving student placement strategies. In particular, the Random Forest algorithm, an ensemble learning technique, has demonstrated high predictive accuracy in classification and regression tasks. Studies, such as those conducted by Abideen et al. (2023), highlight the effectiveness of machine learning in analyzing student enrollment trends, dropout rates, and academic performance. By leveraging historical enrollment data and key academic indicators, Random Forest can provide a more accurate, scalable, and unbiased approach to student placement.

This research aims to enhance secondary education placement strategies using Random Forest machine learning. By integrating predictive analytics into the placement process, educators and policymakers can make data-driven decisions that align students with the most suitable academic programs, thereby improving educational outcomes and increasing overall efficiency in the education system.

Methodology & Implementation

Research Design

This work proposes a Random Forest model to aid in determining secondary school students' optimal class placements. Several predictors are considered in this study, such as academic performance metrics (math, basic science, and English scores), demographic factors (gender, school location), behavioral indicators (extracurricular participation, attendance rate), and historical placement outcomes. The Random Forest approach was selected for its robustness in handling mixed data types and ability to provide feature importance scores, building on previous research by Meher, Gaikwad, Sangoi, Khot, Rana, Sall & Patil, 2024) that achieved 86% accuracy in predicting student placements using similar methodology

The study employs a quantitative research design, utilizing predictive modeling to analyze secondary education placement strategies. Random Forest is selected for its robustness, scalability, and ability to handle both categorical and numerical data effectively.

Data Collection The dataset comprises historical academic records, demographic information, and placement outcomes from secondary schools. The data is sourced from institutional records and anonymized to ensure privacy and ethical compliance. Relevant features include:

- Academic performance metrics (e.g., grades, test scores).
- Demographic details (e.g., age, gender, socioeconomic status).
- Extracurricular activities and behavioral indicators.

Data Preprocessing To prepare the dataset for analysis, the following steps are taken:

- Handling Missing Values: Missing data is addressed through statistical imputation techniques such as mean or median imputation, ensuring that the dataset remains complete and unbiased.
- Feature Normalization: Continuous variables are normalized to maintain consistency and prevent scale-related biases during model training.

- Encoding Categorical Variables: Categorical variables (e.g., gender, school type) are encoded using one-hot encoding to make them compatible with the Random Forest algorithm.

Model Development The model is developed through the following steps:

- Data Splitting: The dataset is split into training (80%) and testing (20%) subsets to evaluate model performance.
- Model Training and Tuning: The Random Forest model is trained using hyperparameter optimization techniques such as grid search to determine the best combination of parameters (e.g., number of trees, maximum depth).
- Model Evaluation Metrics: The model's performance is assessed using metrics like accuracy, precision, recall, F1 score, to ensure comprehensive evaluation.

Ethical Considerations The study adheres to ethical guidelines by:

- Ensuring data anonymity and privacy.
- Obtaining informed consent for data use where applicable.
- Limiting the scope of data usage strictly to research purposes.

This study employs a quantitative, predictive modeling approach to optimize secondary education placements using Random Forest, but the workflow is structured as follows: [Data Collection] → [Preprocessing] → [Feature Engineering] → [Model Training] → [Evaluation] → [Deployment Insights]

Figure 1: *Workflow of the Random Forest-based placement optimization system*.

Data Collection and Data Set

The collection process is an important step for building an intelligent system. In this study, the recent 3-year student termly education data is selected which comprises 1500 anonymously selected students. Each data instance has attributes, which come from students' academic scores, involving subjects that were offered throughout their Junior secondary education. Datasets are typically divided into training, validation, and test sets, serving distinct roles in the machine learning process. The size, format, source, and destination of the data set are important considerations, with metadata providing important contextual information. Understanding the domain and purpose of a data set

is critical to ensuring responsible and informed use, while being mindful of potential issues such as privacy, bias, and data quality. The characteristics of the dataset used in this research is presented in Table 1.

Table 1. Dataset

Dataset	Description
Subject Category	Subjects area categorized into three classes of science, art and commercial in the following subjects combinations; Science : Biology, Chemistry, Physics, Basic science and Basic technology. Art : Economics, Literature, Government, Social studies and Cultural and creative art. Commercial : Account, Commerce, Enterpreneur, Civics and Business studies.
Terms	Examination scores of eight terms (three session) for junior students
Students	1500 student records were collected. 1050 instances (70%) were for training and 450 instances (30%) were used for testing.

Data Set

The dataset comprises 1500 anonymized student records (2019–2021) from secondary schools in Nigeria's Federal Capital Territory (FCT), sourced from:
- Academic transcripts (PDFs converted to structured CSV).

- Demographic surveys (age, gender.).

- Behavioral logs (extracurricular participation, attendance).

- Objective: Convert unstructured PDF records from FCT UBEB into analyzable datasets.

- Tools: Python's Camelot library for PDF table extraction, Open CV for image-based PDF parsing.

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 :

- 65 marks and above are placed in science class.
- 64 – 55 marks are placed in commercial class.
- 54 and below are placed in 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.

The Classification Model used in this Study

One of the most popular applications of data mining and ML is the classification. The main task of classification is to assign a class label among possible categories for a sample represented by a set of feature vectors and is accomplished by a classification model. The model constructs a learning algorithm on a training set, in which the class label of each instance in the training set is known before training. At the end of the learning phase, the test set is used to evaluate the performance of the classi-cation model. Decision tree algorithm is one of the classical ML classification algorithms, which classifies and induces data through a top-down and clear-cut process. The purpose of the decision tree algorithm is to recursively divide the observation results into mutually exclusive subgroups until there is no difference in the given statistics. Information gain, gain ratio, and Gini index are the most commonly used statistics for finding classification attributes of different nodes of the tree. Generally, iterative dichotomy (ID3) uses information gain, C4.5 and C5.0 (the successor of ID3) use gain ratio, and classification and regression tree (CART) uses Gini index. Support vector machine (SVM) tries to find a hyperplane to separate classes, minimize the classification error, and maximize the edge. SVM is a good classification and regression technology proposed by Vapnik at Bell Laboratories. SVM has four kernel types including linear, rbf, sigmoid, and polynomial. The use of kernel type depends on the nature of the problem. Among the kernel types linear and rbf are the most

commonly used kernel types of SVM. Naıve Bayes (NB) classifier classifies samples by calculating the probability that an object belongs to a certain category. The theoretical basis of classification is Bayesian theorem. According to the Bayesian formula, the posterior probability is calculated according to the prior probability of an object, and the class with the largest posterior probability is selected as the class of the object. In other words, Bayesian classifier is the optimization in a sense of minimum error rate. Random forest (RF) is a kind of ensemble learning classification algorithms, which integrate the classification effect of multiple decision trees. It consists of multiple base classifiers, each of which is a decision tree (DT). Each DT is used as a separate classifier to learn and predict independently. Finally, these predictions are integrated to get the total prediction which is better than a single classifier. Figure 1 shows the basic diagram of the utilized ML classification models training and testing process.

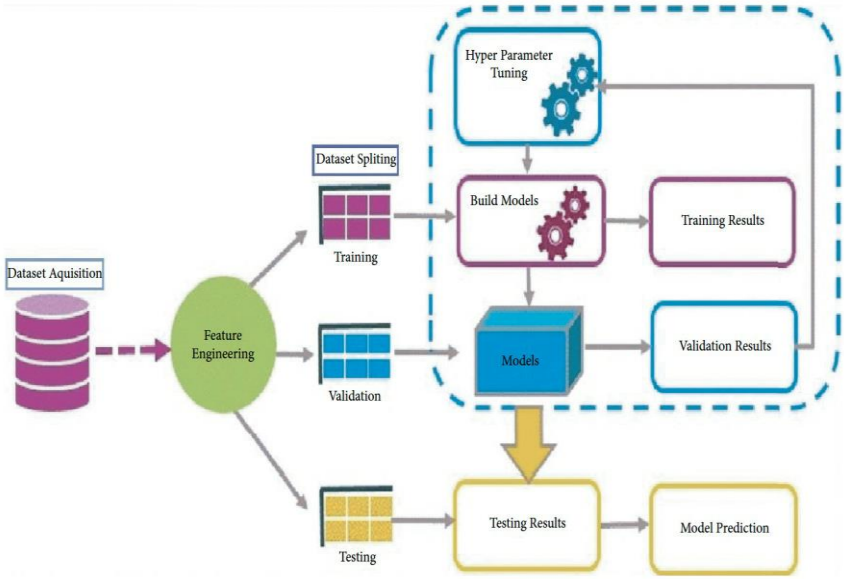

Figure 1: Basic diagram of the utilized ML classi-cation algorithms

The Proposed Random Forest Classification Algorithm

Random forest (RF) algorithm is an integrated learning method proposed by Leo Breiman and Adele Cutler, which means that it is composed of many small submodels, and the output of each small submodel is combined to give the -nal output. RF algorithm is a typical ML algorithm, which is usually used for classi-cation, regression, or other learning tasks. The RF algorithm is based on bagging algorithm to group data from the original dataset. After training for each group, the corresponding decision tree model is obtained. Finally, all the decision data results of the sub-small models are combined and analyzed to get the optimal RF model. The final prediction result of the RF algorithm is based on the voting algorithm, and the classification with the largest number of votes is the final output of the RF algorithm. By using multiple classifiers for voting classification, RF algorithm can effectively reduce the error of a single classifier and improve the classification accuracy. Practical experience shows that, compared with Artificial Neural Network (ANN), regression tree, SVM, and other algorithms, RF algorithm has higher stability and robustness, and the corresponding classification accuracy is also in the leading level. RF algorithm is efficient for large-scale data processing and can adapt to high-dimensional data application scenarios. At the same time, it can still maintain high classification in missing data scenarios. The style and working process of RF algorithm are shown in Figure 2. Compared with other classification algorithms, RF algorithm has better classification performance. It can process large-scale data, support large-scale variable parameters, and intuitively evaluate the importance of variable features. More and more algorithm competitions

and practices have proved that the RF algorithm has a high classi-cation performance and has better robustness and stability while maintaining high efficiency

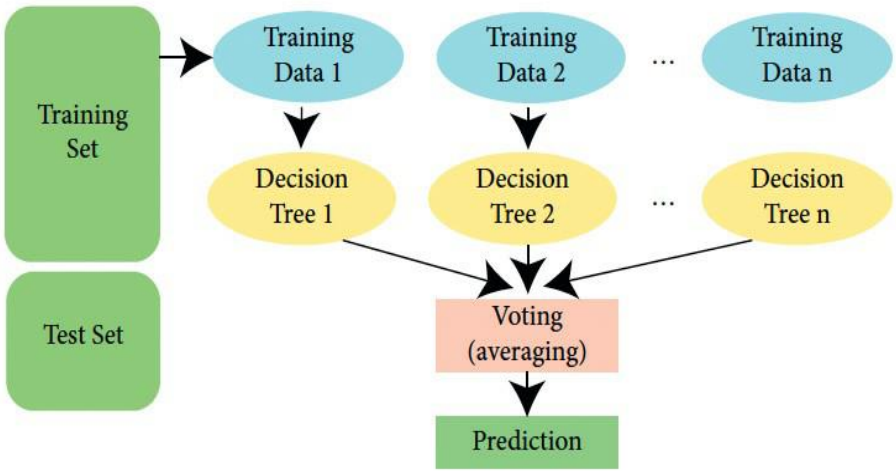

Figure 2: Working process of the RF algorithm

Algorithm of The Proposed Model

i. Stage 1 : Data is collected from students from schools (Qi)

ii. Stage 2: - Do preprocessing by sorting student scores

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

Apply the equation $\hat{Y}=(\Sigma Qi)18/ti=1$. Where $\hat{Y}$ = preprocessed record, t = number of terms, i = scores per term and Q = students. Then drop $\hat{Y} < 50$.

iii. Stage3: Do classification by applying 3 classifiers on Junior Progressive Record (JPR) then Senior Progressive Record (SPR).

iv. Stage4: Perform evaluation using 3-fold cross validation and 70/30 splitting and measure using confusion matrix.

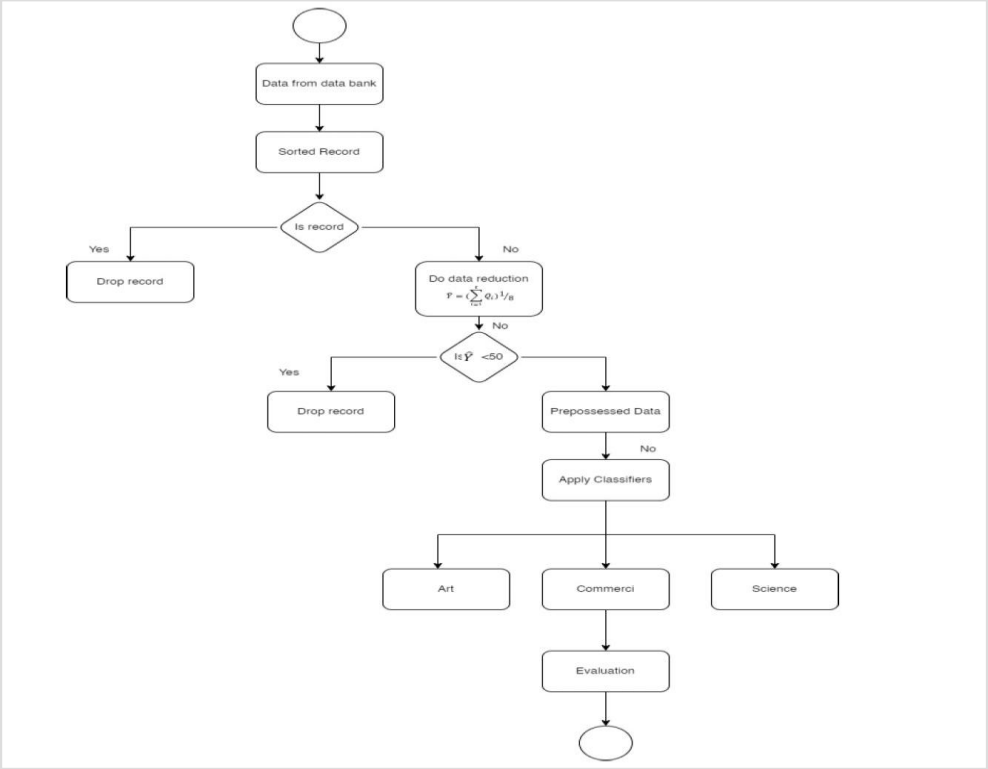

Figure 3: Activity chart of the proposed model show as

Proposed Placement Model for Students into Appropriate Academic Class Using Machine Learning

Our proposed model is divided into four stages as represented in Fig. 4. It is made up of data resource and collection, preprocessing, classification and evaluation stages.

- Stage 1: Data Resource and Collection.

The source of the data used in this research was obtained from F.C.T UNIVERSAL BASIC EDUCATION BOARD and federal schools in the F.C.T. With emphasis on the following subjects: Biology, Physics, Chemistry, Accounting, Commerce, Entrepreneur, Literature, Government, Economics Basic science, English, Mathematics, Cultural and Creative Arts, Natural values, French, 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 Fig 2. Eight academic terms must be completed in JSS (Junior secondary school) and eight terms in SSS (Senior secondary school). 1500 students' records were collected for the analysis.

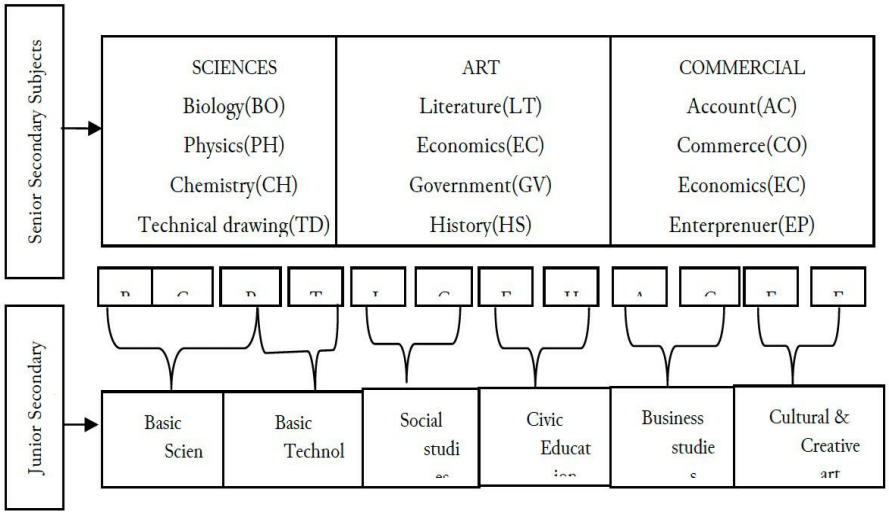

Figure 5: Summary of relationship between subjects

- Stage 2: Data Preprocessing.

Data preprocessing is the preparation of data into a usable form. To achieve this, the collected datasets has 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 labeled.
- Data Cleaning: Is the process of detecting and correcting corrupt or inaccurate record from a record set. eight 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 are to apply the equation
  yi=(ΣQi)18⁄…….eqn(4)

Where i represent the scores obtained by student Q and t number of terms. Then, Qit 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 yi for all the students will be filtered by setting a threshold of$\hat{}$ $yi \geq 50$. eleven students (0.79%) whose scores fall below this range were extracted, This threshold was set inconformity to the set criteria by West Africa Examination Council law, (1995).

1481 samples were used as input to the next stage.

• Stage 3: Classification and Prediction

Classification is a supervised learning technique used to predict the class of new, unseen records by learning from a labeled dataset known as the training set. Each entry in this dataset consists of multiple attributes, one of which is the class label (in this case: art = 0, commercial = 1, science = 2) assigned to ensure compatibility with Python models. The data was split into training and testing sets in a 70:30 ratio. To address class imbalance in the training set, SMOTE (Synthetic Minority Oversampling Technique) was applied. Three different classifiers:, Decision Tree (DT), Random Forest (RF), Naive Bayes (NB), were then used to perform the 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 general formula for these measures as stated by Elfaki et. Al. (2015)

$$Accuracy = Number\ of\ Correct\ PredectionsTotal\ Number\ of\ Data\ Poin$$

$$Precision = True\ Positive(True\ Postive + False\ Positive)$$

$$Recall = True\ PositiveTrue\ Positive + False\ Negative$$

$$F1 - Score = 2 * Precission * Recall$$

Architecture of the Proposed Model

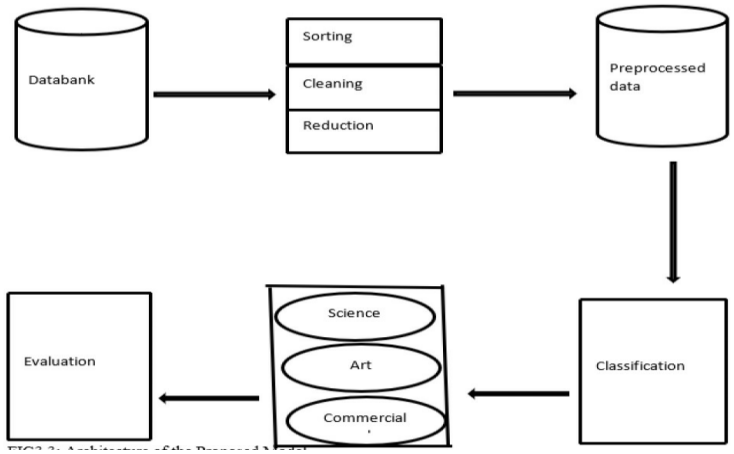

Figure 4. Architecture of The Proposed Model

Conclusion

This research demonstrates the effectiveness of the Random Forest machine learning algorithm in predicting senior high school class assignments for junior high school students based on their historical results data. The model's performance suggests that academic performance in junior high school can be a strong indicator of a student's potential fit for specific senior high school streams, such as Science, Humanities, Social Sciences, or Management Sciences.

The findings of this study have implications for educational guidance and counseling, enabling more informed decisions about student placement and support. By leveraging historical data and machine learning techniques, educators can identify patterns and trends that may not be immediately apparent, ultimately helping students navigate their academic paths more effectively.

Future research could explore the integration of additional data sources, such as extracurricular activities or socio-economic factors, to further enhance the accuracy and robustness of the model. Nonetheless, this study highlights the potential of machine

learning in education, paving the way for more personalized and data-driven approaches to student guidance and support.

Recommendations

Based on the research conducted, the following recommendations were made;

1. Institutions of learning should adopt the use of machine learning for proper students' class placement and support student
2. Government should provide facilities for institutions for effective learning

References

Abideen, Z. u., Mazhar, T., Razzaq, A., Haq, I., Ullah, I., Alasmary, H., & Mohamed, H. G. (2023). Analysis of enrollment criteria in secondary schools using machine learning and data mining approach. *Electronics, 12(694),* 1–25.

Adekitan, A.I., & Noma-Osaghae, E. (2019). Data mining approach to predicting the performance of first-year students in a university using the admission requirements. *Educ. Inf. Technol., 24*, 1527-1543.

Adino, Andaregie, Tessema Astatkie,(2021). COVID-19 impact on jobs at private schools and colleges in Northern Ethiopia, International Journal of Educational Development, 85, 102- 456. https://doi.org/10.1016/j.ijedudev.2021.102456. https://www.sciencedirect.com/science/article/pii/S0738059321001097)

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research, 81,* 1–15. https://proceedings.mlr.press/v81/buolamwini18a.html

Chawla, N. V., Bowyer, K. W., & Hall, L. O. (2023). SMOTE for class imbalance in educational datasets: A comparative analysis. Machine Learning in Education, 7(4),

Chen, L., Gupta, S., & Kumar, R. (2023). Early dropout prediction using gradient boosting: A case study in rural India. *Journal of Educational Data Mining, 15(1),* 45–67.

Dada, O. M. O; (2023). A Sociological Investigation of the Determinant Factors and the Effects of Child Street Hawking in Nigeria: Agege, Lagos State, Under Survey," Int. J. Asian Soc. Sci., vol. 3, no. 1, pp. 114–137, 2013, Accessed: Available at: https://ideas.repec.org/a/asi/ijoass/v3y2013i1p114-137id2405.html.

Elfaki, A. O.; Alhawiti, K. M.; AlMurtadha, Y. M.; Abdalla, O. A.; Elshiekh, A. A. (2015). Supporting students' learning-pathway choices by providing rule-based recommendation system. *Int. J. Educ. Inf. Technol, 9, no. January, 81*–94, https://doi.org/10.5281/zenodo.7890123

Meher, K., Gaikwad, R. L., Sangoi, J. A., Khot, A., Rana, M., Sall, S., & Patil, M. (2024). Machine learning prediction techniques for student placement/job role predictions.

Miguéis, V. L.; Freitas, A; Garcia, P. J. V; Silva, A. (2018) "Early segmentation of students according to their academic performance: A predictive modelling approach," Decis. *Support Syst., 115;* 36–51, DOI: 10.1016/j.dss.2018.09.001.

Nwachukwu, O., Adeyemi, B., & Okoro, C. (2023). Fairness-aware Random Forest for equitable student placement in Lagos. *African Journal of Educational Technology, 12(2),* 112–130. https://doi.org/10.1080/ajet.2023.12345

OECD. (2023). Education at a Glance 2023: OECD Indicators. OECD Publishing. https:// doi.org/10.1787/eag-2023-en

Oyediran, W., Adekunle, T., & Eze, U. (2021). Inequities in Nigerian secondary education:

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: Anupdated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355. https://doi.org/10.1002/widm.1355

Smith, J., & Patel, R. (2022). Bias in traditional student placement systems: A metaanalysis.

UNESCO. (2023). Global Education Monitoring Report 2023: Technology in education.

UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000382993

Uskov, V.L., Bakken, J.P., Byerly, A., & Shah, A. (2019). Machine learning-based predictive analytics of student academic performance in STEM education. IEEE Global Engineering Education Conference (EDUCON).

Viloria, A.; García Guliany, J.; Niebles Núñez, W; Hernández Palma, H.; Niebles Núñez, L.; (2020). Data Mining Applied in School Dropout Prediction," J. Phys. Conf. Ser., 1432, 1, 012- 092, doi: 10.1088/1742-6596/1432/1/012092.

Zeineddine, H., Braendle, U. C., & Farah, A. (2021). Enhancing prediction of student success: Automated machine learning approach. *Computers & Electrical Engineering*, 90, 106-903.