IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 28, Issue 4, Ser. 2 (July. – August. 2026), PP 51-54
www.iosrjournals.org

Machine Learning Framework for Predicting Students' Academic Placement in Areas of Study in Senior Secondary Schools in Abakaliki

Egwu Nkemjika Samuel[1], J.O Ugah[2], Azemobor Daniel[3], Ugboaja, S.G.[4]

[1]*Department of Computer Science, Ebonyi State University, Nigeria*
[2]*Department of Computer Science, Ebonyi State University, Nigeria*
[3]*Department of Cybersecurity, Evangel University Akaeze, Nigeria*
[4]*Department of Computer Science, Michael Okpara University of Agriculture Umudike, Nigeria*

Abstract:
Academic placement into Science, Arts, or Commercial classes is an important decision in Nigerian senior secondary schools. In many schools, the decision still depends mainly on examination scores, teacher judgement, and school rules. This can ignore important student characteristics and may lead to unsuitable placement. Recent studies show that machine learning can analyse educational data and support more objective decisions. This study developed a machine learning framework using student records collected from selected secondary schools in Abakaliki, Ebonyi State. Cognitive, affective, and psychomotor characteristics were considered. Five supervised learning algorithms were evaluated: Decision Tree, Multinomial Logistic Regression, Support Vector Machine, Gradient Boosting Machine, and Random Forest. The models were assessed using accuracy, precision, recall, F1-score, and confusion matrix. The final model was integrated into a Streamlit web application developed with Python and scikit-learn. Random Forest produced the best overall result, with 94.80% accuracy, 0.95 precision, 0.94 recall, and 0.94 F1-score. The confusion matrix also showed a high number of correct classifications across the three placement classes. Conclusion: The developed framework provides a practical data-driven decision-support tool for academic placement. It reduces dependence on subjective judgement and considers a broader range of student characteristics. The system is intended to support, rather than replace, teachers, guidance counsellors, and school management.

Key Word: *Academic placement; Educational data mining; Machine learning; Random Forest; Secondary education; Decision support.*

Date of Submission: 07-08-2026 Date of Acceptance: 20-08-2026

I. Introduction

Academic placement is an important stage in secondary education because it influences the subjects students study, their later educational choices, and possible career directions. In the Nigerian senior secondary school system, students are commonly placed into Science, Arts, or Commercial areas after junior secondary education. The full project used for this article identified that placement in the selected schools was mainly based on examination results, teacher recommendations, and school policies. Such a process may not fully consider students' interests, behaviour, attendance, practical abilities, discipline, and other relevant characteristics.

Machine learning offers a useful approach for improving educational decisions because it can learn patterns from historical data and apply the learned patterns to new cases. A 2024 review by Ersozlu, Taheri, and Koch examined 77 publications and reported that supervised machine learning dominated recent educational data studies, with Random Forest, Decision Tree, Support Vector Machine, and Logistic Regression among the frequently used methods [1]. Recent research has also shown that educational data mining can identify patterns in student performance and provide useful information for academic support [2].

Other recent studies have moved beyond the use of academic scores alone. Hemdanou et al. compared feature selection and extraction approaches for student performance prediction and emphasized the importance of individual, family, and social factors [3]. Al-Tameemi et al. combined unsupervised and supervised learning and reported that clustered Random Forest, Decision Tree, and neural network models can improve prediction in multiclass educational datasets [4]. These findings support the use of multiple student characteristics when building intelligent educational systems.

However, much of the recent literature focuses on predicting grades, pass/fail status, dropout, or general academic performance rather than assigning students to appropriate academic streams. The full study reviewed for this article similarly identified a gap in practical machine learning systems designed specifically for Science, Arts, and Commercial placement in Nigerian senior secondary schools. Therefore, this study developed and evaluated a machine learning framework that combines cognitive, affective, and psychomotor information to support student academic placement.

II. Material And Methods

Study Design and Location

The study adopted a system-development and machine-learning experimental approach. Data were obtained from selected secondary schools in Abakaliki, Ebonyi State, Nigeria. The schools identified in the full work included Holy Ghost Foundation Secondary School, Assemblies of God Secondary School, Urban Model Secondary School, and Girls' High School. The target task was multiclass classification into Science, Arts, or Commercial placement.

Student Variables and Data Preparation

The framework considered student information from three major educational domains. Cognitive information included JS1, JS2, and JS3 academic averages, cumulative average, and academic trend. Affective information included behaviour, punctuality, diligence, classroom participation, discipline, responsibility, attendance, teacher recommendation, and parent recommendation. Psychomotor information included sports, cultural activities, music, drama, chess, and creative skills. The implemented system also computed cumulative average and academic trend from the junior secondary results.

Before model development, the records were cleaned and organised. Numeric fields were converted into appropriate numeric values, missing or inconsistent entries were handled, and the placement class was treated as the target variable. The prepared data were then divided for model training and testing. This process follows the general educational data-mining practice of preparing data, selecting relevant features, training models, and evaluating their predictive performance [1,3].

Machine Learning Models

Five supervised learning algorithms were evaluated: Decision Tree Classifier, Multinomial Logistic Regression, Support Vector Machine (SVM), Gradient Boosting Machine (GBM), and Random Forest. The use of several models allowed the study to compare different learning approaches before selecting the most suitable model. Random Forest was finally selected because it produced the strongest overall evaluation result and can combine many decision trees to improve stability and reduce the effect of overfitting.

System Development

The application was developed in Python. Pandas and NumPy supported data handling, scikit-learn supported model development, Joblib was used to save the trained model, and Streamlit was used to build the web interface. The system allows authorised users to enter a student's information, generate a recommended placement, view prediction confidence and evaluation information, process students in batches, and download placement reports.

Evaluation Metrics

The models were evaluated with accuracy, precision, recall, F1-score, and a confusion matrix. Accuracy measured the overall proportion of correct predictions. Precision measured the correctness of predicted classes, recall measured the ability to identify the correct classes, and F1-score provided a balance between precision and recall. The confusion matrix was used to show correct and incorrect predictions for Science, Arts, and Commercial classes.

III. Result

The five algorithms produced different levels of predictive performance. Random Forest achieved the best overall result. The complete evaluation reported 94.80% accuracy, 0.95 precision, 0.94 recall, and 0.94 F1-score. Gradient Boosting Machine was the next best model, while Decision Tree recorded the lowest accuracy among the five models.

Table 1: Performance of the evaluated machine learning algorithms.

Algorithm	Accuracy (%)	Precision	Recall	F1-Score
Decision Tree	84.10	0.83	0.82	0.82
Multinomial Logistic Regression	86.50	0.85	0.85	0.85
Support Vector Machine	88.20	0.88	0.87	0.87
Gradient Boosting Machine	90.40	0.90	0.90	0.90
Random Forest	94.80	0.95	0.94	0.94

The confusion matrix for Random Forest showed 85 Science students correctly classified as Science, 76 Arts students correctly classified as Arts, and 78 Commercial students correctly classified as Commercial. The remaining values represented a relatively small number of misclassifications. This pattern supports the conclusion that the selected model was able to distinguish the three academic placement groups effectively.

Table 2: Confusion matrix of the selected Random Forest model.

Actual Class	Predicted Science	Predicted Arts	Predicted Commercial
Science	85	3	2
Arts	4	76	5
Commercial	3	4	78

The developed interface also supported individual and batch prediction. For individual prediction, the system displayed the recommended academic placement, class probabilities, prediction confidence, academic trend, reasons for the recommendation, strengths, areas for review, and model evaluation metrics. For batch prediction, an uploaded spreadsheet could be processed and the predicted placements exported. The implementation displayed a success notification after a batch file was processed successfully and an error message when processing failed.

IV. Discussion

The results show that ensemble learning is suitable for academic placement when the input contains different types of student information. Random Forest performed better than the other four models used in this study. This is consistent with the broader literature showing that tree-based and ensemble models are widely used for educational prediction [5] also reported strong results from an educational predictive-analytics architecture that combined Random Forest with other learning methods, showing the value of ensemble approaches when student data contain complex feature interactions [5].

A major strength of the present framework is that it does not rely only on examination scores. It combines academic records with affective and psychomotor factors. This is important because recent research has shown that student outcomes are influenced by several interacting characteristics and that feature selection can significantly affect predictive performance [3]. The system therefore provides a broader basis for placement decisions than a manual process that uses only a few indicators.

The study also has a practical decision-support focus. Islam et al. developed an explainable educational data-mining support system and showed the growing importance of giving users understandable information about machine learning predictions [6]. In a similar direction, the developed placement application presents reasons, strengths, confidence, and areas for review so that school personnel can understand and examine the recommendation. The final placement decision remains with the school, which is important because educational predictions should support professional judgement rather than automatically replace it.

Despite the strong result, the framework has limitations. The data came from selected schools in one geographical area, so performance may change when the model is applied to schools with different student populations or assessment practices. The system also depends on the quality and completeness of the supplied student records. Future work should test the framework with larger and more diverse datasets, evaluate fairness

across student groups, and examine the effect of the recommendations on later student performance. Recent work on fairness in educational algorithms also shows that predictive accuracy should be considered together with bias and responsible use [7].

V. Conclusion

This study developed a machine learning framework for predicting students' academic placement into Science, Arts, and Commercial classes in senior secondary schools. The framework analysed cognitive, affective, and psychomotor characteristics and compared five supervised machine learning algorithms. Random Forest produced the best performance with 94.80% accuracy, 0.95 precision, 0.94 recall, and 0.94 F1-score. The model was integrated into a web-based decision-support application that can process individual students and batches of student records.

The findings indicate that machine learning can improve the consistency, speed, and evidence base of academic placement decisions. The framework is especially useful because it considers more than examination results and provides information that teachers, guidance counsellors, and school administrators can review before making a final decision. The study therefore contributes a practical approach for applying educational data mining to academic placement in Nigerian secondary schools.

VI. Recommendations

Secondary schools should consider data-driven placement tools as support for teachers and guidance counsellors. Schools that adopt the framework should maintain accurate student records and train users to interpret predictions correctly. Future studies should validate the model with larger datasets from different states, examine fairness and generalisation, and investigate whether students placed with machine-learning support achieve better long-term academic outcomes.

References

[1] Z. Ersozlu, S. Taheri, and I. Koch, "A review of machine learning methods used for educational data," Education and Information Technologies, vol. 29, pp. 22125–22145, 2024. doi: 10.1007/s10639-024-12704-0.

[2] "Analyzing students' academic performance using educational data mining," Computers and Education: Artificial Intelligence, vol. 7, 100263, 2024. doi: 10.1016/j.caeai.2024.100263.

[3] A. Laakel Hemdanou, M. Lamarti Sefian, Y. Achtoun, and I. Tahiri, "Comparative analysis of feature selection and extraction methods for student performance prediction across different machine learning models," Computers and Education: Artificial Intelligence, vol. 7, 100301, 2024. doi: 10.1016/j.caeai.2024.100301.

[4] G. Al-Tameemi, J. Xue, I. Hadi, and S. Ajit, "A Hybrid Machine Learning Approach for Predicting Student Performance Using Multi-class Educational Datasets," Procedia Computer Science, vol. 238, pp. 888–895, 2024. doi: 10.1016/j.procs.2024.06.108.

[5] S. Khan, T. Mazhar, T. Shahzad, M. Amir Khan, W. Waheed, A. Waheed, and H. Hamam, "Predictive analytics in education—enhancing student achievement through machine learning," Social Sciences & Humanities Open, vol. 12, 101824, 2025. doi: 10.1016/j.ssaho.2025.101824.

[6] M. M. Islam, F. H. Sojib, M. F. H. Mihad, M. Hasan, et al., "The integration of explainable AI in Educational Data Mining for student academic performance prediction and support system," Telematics and Informatics Reports, vol. 18, 100203, 2025. doi: 10.1016/j.teler.2025.100203.

[7] J. A. Idowu, "Debiasing Education Algorithms," International Journal of Artificial Intelligence in Education, vol. 34, pp. 1510–1540, 2024. doi: 10.1007/s40593-023-00389-4.

[8] J. R and S. Selvakumari, "Design of a Prediction Model to Predict Students' Performance Using Educational Data Mining and Machine Learning," Engineering Proceedings, vol. 59, no. 1, 25, 2023. doi: 10.3390/engproc2023059025.