Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.17, No.5, 2026

Predicting Student Academic Success Using Comparative Machine Learning and Explainable Artificial Intelligence

Isaiah Ifeanyi Nweze
Department of Computing
University of Greater Manchester
Bolton, United Kingdom
0009-0002-7054-783X

Ezekiel Nwibo Gabriel
Department of Applied Computer Science
University of Greater Manchester
Bolton, United Kingdom
0009-0004-4656-2946

Paul Maduabuchi Agu
Robotics, Autonomous Systems and Telecommunications Engineering
University of Greater Manchester Bolton, United Kingdom
pma1aes@bolton.ac.uk

Chukwuka Abraham Nwovu
Department of Computing
University of Greater Manchester
Bolton, United Kingdom
nwovuchukwuka2019@gmail.com

Charles Ugwute
Department of Computing
University of Greater Manchester
Bolton, United Kingdom
ogbonnacharles684@gmail.com

God's Favour Fred-Ibe
Department of Computer Science
Ebonyi State University
Abakaliki, Nigeria
gfredibe@gmail.com

Abstract

Student academic success prediction is a significant educational data mining task because it allows early intervention, enhances retention, and assists with academic planning. This paper investigates whether demographic, academic history, and socioeconomic factors can predict student outcomes in the Open University Learning Analytics Dataset (OULAD). The dataset (studentInfo.csv) consists of 32,593 student records with 12 variables. The final analytical dataset comprised 21,562 observations after excluding withdrawn students and recording the result as a binary target. This analysis involved a descriptive exploration and analysis of socioeconomic deprivation using the Index of Multiple Deprivation (IMD) band. A chi-square test showed that the IMD band and the final academic outcome are statistically associated ($\chi^2 = 472.42$, $p < 0.001$), indicating that performance depends on socioeconomic background. Three supervised machine learning models, Logistic Regression, Decision Tree, and Random Forest, were developed and compared. Logistic Regression achieved the highest discrimination performance (ROC-AUC = 0.672), while Random Forest achieved the highest recall (0.794) and F1-score (0.748). To increase interpretability, feature importance, permutation importance and SHAP analyses were used. Findings have shown that the most significant predictors of academic success are prior educational experience, frequency of prior attempts, credits completed, age, and selected socioeconomic variables. The implications of these results are that interpretable machine learning offers a useful structure to student risk identification and that structural and academic preparedness factors remain relevant in defining student success.

Keywords-*Educational Data Mining, Explainable Artifitial Intelligence, Machine Learning, Predictive Analytics, SHAP, Student Performance Prediction*

DOI: 10.7176/JEP/17-5-05
Publication date: May 30th 2026

I. Introduction

Predicting student academic performance has grown in interest due to its practical relevance in higher education. Universities are increasingly focusing on identifying students who may perform poorly, lose interest in their studies, or need additional academic support. Predictive analytics provides an opportunity to convert historical learner data into timely and actionable information. Rather than relying on end-of-term grades or withdrawal history to indicate academic challenges, predictive models can detect patterns associated with non-success earlier in the process [1], [2].

Student performance is influenced by a combination of academic, demographic, and contextual factors. The influence of academic variables, including prior educational attainment, course load, and prior attempts, is significant, while demographic variables, including age, disability, gender, and socioeconomic background, provide additional context. Academic achievements should thus not be viewed solely in terms of individual effort; they are also influenced by broader structural and institutional factors. It therefore follows that performance-prediction research must not simply focus on maximising predictive accuracy but also on the meaningful interpretation of the factors driving model results [3], [4].

Machine learning approaches are well-suited to this challenge, as they can identify relationships among numerous variables simultaneously. The main problem, though, is that as models become more predictive, they tend to be less interpretable [5]. In practice, academic staff and administrators might be unwilling to use black-box systems without understanding why a student is identified as at risk. Explainable artificial intelligence (XAI) is thus significant, as it relates model outputs to recognisable features and patterns [6], [7].

This paper examines student-level data from the Open University Learning Analytics Dataset (OULAD) to explore the prediction of academic performance using descriptive analysis, statistical testing, comparative machine learning, and explainability. The study does not focus on a single model; instead, it compares Logistic Regression, Decision Tree, and Random Forest to balance predictive power and interpretability. The objectives of this study are: to examine the distribution of academic outcomes within the dataset, to evaluate the relationship between socioeconomic deprivation and academic performance, to compare selected machine learning models for binary student-success prediction, and to identify key predictive variables using feature importance and SHAP-based analysis.

II. Background and Motivation

Educational data mining is an interdisciplinary field that focuses on discovering valuable patterns in information generated within learning settings. Academic outcome prediction is one of its largest applications, as it aims to forecast future outcomes based on historical data. Those predictions may aid institutional planning, individualised feedback, advice, and the design of interventions [8].

One of the driving forces of this research is the desire to go beyond accurate reporting. A model that predicts the majority class well in an educational setting might still be of little use if it does not identify students who require the most support. On the other hand, a model with moderate performance can also shed light on the variables that influence academic performance. Deeper interpretation should thus be supplemented by predictive evaluation [9].

The second motivation is the socioeconomic inequality in education. Despite theories of ability, effort, or engagement in the context of academic success, social and economic factors dictate learners' chances. The IMD band offered by OULAD enables the study of this problem empirically [10]. The analysis of IMD, combined with predictive modelling, is used to assess the descriptive and predictive relevance of socioeconomic disadvantages in the study.

The last reason is the growing significance of explainability in deployed machine-learning systems. Access to support, advice and opportunities can be influenced by educational decisions. It is thus not sufficient to say that a model makes a prediction, but to be able to comprehend the variables that affect that prediction. This is particularly necessary when the model's outputs can guide institutional policy or intervention strategies [11].

III. Related Work

Previous studies in educational data mining have shown that it is possible to predict student performance based on a combination of demographic, academic history, and behavioural variables. Systematic reviews recognise logistic regression, decision trees, random forests, support vector machines, gradient boosting, and neural networks as the most used methods. However, no single model is consistently more effective across various datasets and educational settings [8], [9], [13]. These papers also emphasise that the effectiveness of models is strongly determined by the quality of their features, class imbalance, and the target prediction task, i.e., whether it is to prioritise recall for at-risk students or balanced classification performance.

Within this body of literature, student background variables, including prior educational attainment, number of previous attempts, age, and socioeconomic context, consistently emerge as important predictors of academic outcomes [1], [4], [8]. At the same time, recent studies emphasise that predictive accuracy alone is insufficient for practical deployment in educational settings. Institutions require models that not only perform well but also provide interpretable, actionable insights to support fair, informed decision-making [2], [7], [10]. This has led to growing interest in explainable artificial intelligence methods that complement conventional performance metrics with interpretable explanations.

Among these sources, variables related to student background, such as prior education level, number of prior attempts, age, and socioeconomic background, are repeatedly identified as significant predictors of academic outcomes [1], [4], [8]. Simultaneously, recent research indicates that predictive accuracy alone is insufficient for practical use in educational contexts. Institutions need models that are not only effective but also provide interpretable, actionable insights to aid in making fair, informed decisions [2], [7], [10]. This has contributed to the increased interest in explicable artificial intelligence techniques that do not replace traditional performance indicators with interpretable explanations.

The study of interpretable machine learning and explainable AI has thus become increasingly relevant in the student analytics domain. In this field, surveys indicate that post hoc explanation methods, including permutation importance and SHAP, may help understand the impact of individual features on model predictions, thereby enhancing transparency and trust [6], [14], [19]. This is of particular concern in the field of education, where predictive outputs can influence advising decisions, support resource allocation, and inform institutional policy.

Regardless of these achievements, most studies are either focused on predictive performance without adequate interpretability or on explanation without rigorous statistical validation. To fill this gap, the current study will integrate descriptive analysis, statistical testing, comparative machine learning, and explainability into a single framework based on the OULAD dataset.

IV. Dataset Description

A. Source and Structure

The dataset utilised in this research is based on the Open University Learning Analytics Dataset, which is a popular open-source educational dataset that is aimed at research in the fields of learning analytics and prediction of student performance. The dataset is publicly available and can be accessed via:
https://archive.ics.uci.edu/dataset/320/open+university+learning+analytics+dataset

The dataset is the studentInfo.csv component of OULAD [12]. The raw file consists of 32,593 student records and 12 variables, including course information, demographics, educational background, academic history, socioeconomic status, and outcome. The variables used in the analysis are summarised in Table I.

TABLE I *Variables Contained in studentInfo.csv and Their Descriptions*

Variable	Description
code_module	Module code
code_presentation	Module presentation period
id_student	Unique student identifier
gender	Student gender
region	Geographic region
highest_education	Highest prior education level
imd_band	Index of Multiple Deprivation band
age_band	Student age category
num_of_prev_attempts	Number of previous attempts
studied_credits	Number of studied credits
disability	Disability status
final_result	Final academic outcome

B. Original Outcome Distribution

The initial final_result variable had four categories: Pass, Withdrawn, Fail, and Distinction. The numerical results for these outcomes are given in Table II. Pass was the most frequent, followed by Withdrawn, Fail, and Distinction. Since the modelling aim was to differentiate between academic success and failure among completed results, withdrawn records were excluded from the primary predictive analysis.

TABLE II *Original Academic Outcome Distribution*

Outcome	Count
Pass	12,361
Withdrawn	10,156
Fail	7,052
Distinction	3,024

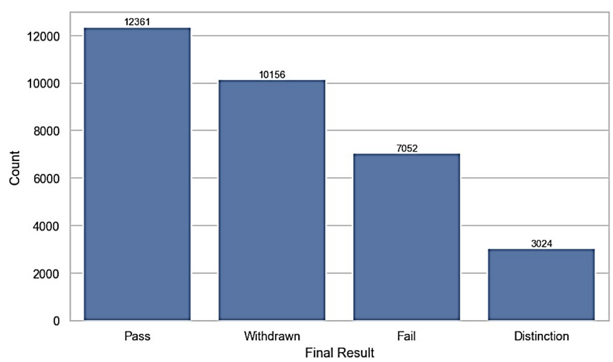

Fig. 1. Distribution of original academic outcomes in the dataset

C. Binary Target Construction

The withdrawn category was removed, and the target variable was recoded as a binary outcome: Success (Pass or Distinction) and Non-success (Fail). This gave 15,385 successful students and 7,052 unsuccessful students. The distribution was cleaned to 68.57% success and 31.43% non-success. The recoded target is summarised in Table III and Fig. 2. The class imbalance is moderate and can influence classifier behaviour, particularly the trade-off between precision and recall.

TABLE III *Binary Target Distribution After Removal of Withdrawn Students*

Target Class	Count	Percentage
Success (Pass/Distinction)	15,385	68.57%
Non-success (Fail)	7,052	31.43%

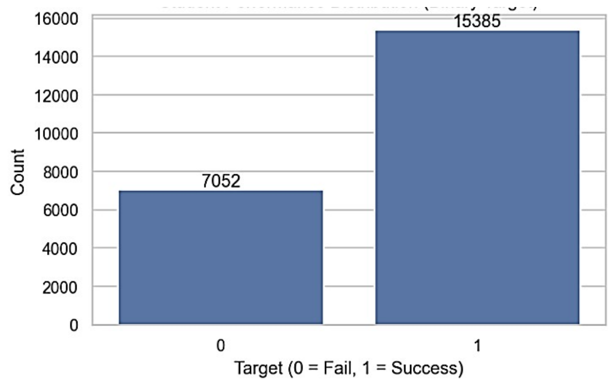

Fig. 2. Student performance distribution after target recoding and cleaning

V. Research Methodology

A. Data Preprocessing

1. Cleaning Procedure

The preprocessing was made in a structured pipeline. First, the records that had Withdrawn as the final_result were deleted. Second, the binary target variable was created. Third, id_student was excluded because it is an identifier rather than a meaningful predictor and including it would introduce spurious patterns. Fourth, rows with missing values were eliminated. Lastly, categorical variables were one-hot encoded to create a machine-readable feature space [13]. The modelling dataset had 21,562 rows and 41 columns, comprising all other variables and the target variable after preprocessing. The main preprocessing results are summarised in Table IV. There were 40 features in the predictor matrix.

TABLE IV *Summary of Preprocessing Outcomes*

Item	Value
Original number of records	32,593
Records after removing Withdrawn	22,437
Final rows after dropping missing values	21,562
Final number of columns	41
Predictor variables used for modelling	40
Target variable	Binary (1 = Pass/Distinction, 0 = Fail)

B. Encoded Feature Space

One-hot encoding has been used to encode categorical attributes as binary indicators. The resulting characteristics were module and presentation dummies, region indicators, educational-background categories, IMD bands, age bands, gender, and disability status. Such representation can be used with standard classification algorithms while maintaining the interpretability of categorical effects [14].

C. Train-Test Split

The data were divided into training and test groups in an 80/20 split, with random_state=42 and stratified by the target category. This generated 17,249 training observations and 4,313 test observations. Stratification was used to ensure that the class ratio was the same in both partitions.

VI. Exploratory Data Analysis

The first exploratory analysis revealed that Pass was the most common academic outcome, followed by Withdrawn, Fail, and Distinction. Following recoding, successful outcomes were more prevalent than unsuccessful outcomes, confirming a moderate class imbalance that must be considered when assessing the model.

One of the objectives of the research was to investigate the correlation between academic outcomes and socioeconomic background. The count-based perspective on academic outcomes by IMD band indicated that passing was higher than failing in each deprivation band, whereas raw counts are not always sufficient because group sizes vary. A more informative analysis was then conducted using percentage analysis.

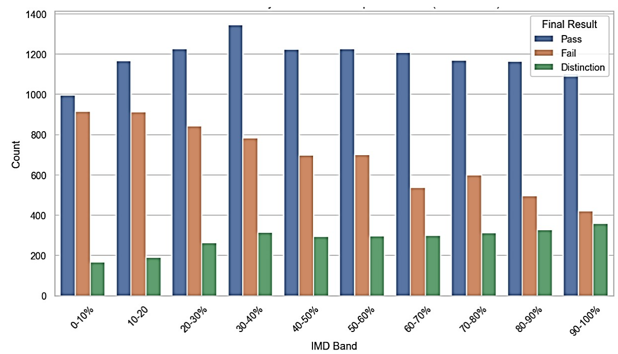

Fig. 3. Academic outcome by socioeconomic deprivation band (count-based view)

TABLE V *Percentage Distribution of Academic Outcomes Across IMD Bands*

IMD Band	Distinction (%)	Fail (%)	Pass (%)
0–10%	8.08	44.04	47.88
10–20%	10.08	39.14	50.77
20–30%	10.92	36.10	52.98
30–40%	11.53	35.60	52.87
40–50%	13.27	31.08	55.65
50–60%	14.42	30.31	55.28
60–70%	15.73	27.48	56.79
70–80%	16.79	26.37	56.84
80–90%	17.81	24.95	57.24
90–100%	19.04	22.39	58.56

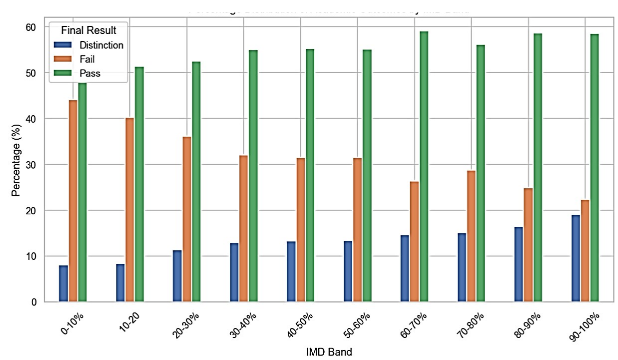

Fig. 4. Percentage distribution of academic outcomes by IMD band

A. Statistical Validation of the IMD Association

A chi-square test of independence was used to statistically test the descriptive IMD pattern. The statistics presented in Table VI were the result of the contingency table of imd_band and final_result. Since the p-value is less than 0.05, the relationships between socioeconomic deprivation and academic outcomes are statistically significant. The observed differences in IMD are therefore not random fluctuations but a meaningful pattern.

TABLE VI *Chi-Square test of association between IMD band and academic outcome*

Statistic	Value
Chi-square (χ^2)	472.4199
p-value	6.4728×10^{-89}
Degrees of freedom	18
Interpretation	Statistically significant association between IMD band and academic outcome

B. Outcome by Highest Education, Gender, and Age Band

Additional exploratory plots revealed a significant difference in terms of previous schooling. The results of students with A-level or equivalent qualifications were better than those with qualifications below A-level. The differences by gender were also visible, but not as pronounced as those for educational background or IMD band. Age-band comparisons indicated that the 0-35 group was the highest, whereas the 35-55 and 55+ groups were lower. The following are the exploratory views, as shown in Figs. 5–7.

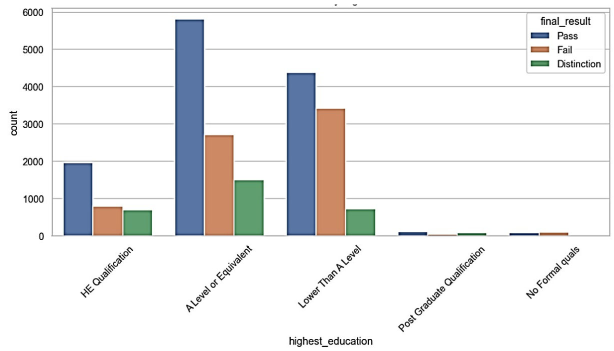

Fig. 5. Academic outcome by highest level of education

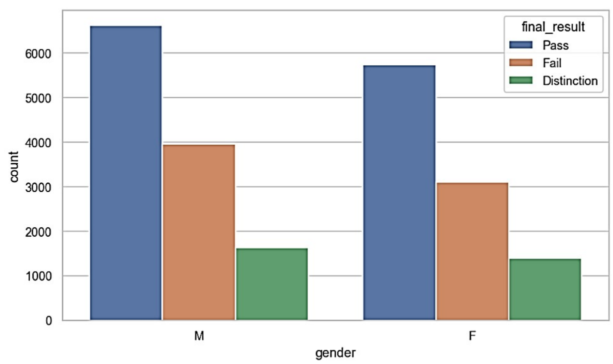

Fig. 6. Academic outcome by gender

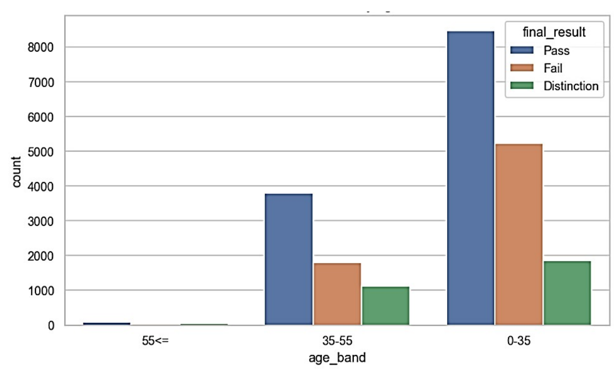

Fig. 7. Academic outcome by age band

VII. Model Development

A. Algorithms Used

Three supervised classification models were trained: Logistic Regression, Decision Tree, and Random Forest. Logistic Regression was used as the baseline linear classifier because it is easy to understand and performs well on structured tabular data. The decision tree provided clear classification using rules, but this was likely to be more volatile and prone to overfitting. Random Forest was also added as an ensemble method capable of capturing nonlinear patterns and enhancing robustness. To address class imbalance, all models were trained with class_weight='balanced'. Random Forests had 200 estimators and a fixed random state to ensure reproducibility [15].

B. Evaluation Criteria

Accuracy, precision, recall, F1-score, and ROC-AUC were used to evaluate model performance. These measures represent different dimensions of classification behaviour. ROC-AUC was used as the primary model selection criterion because it evaluates a classifier's ability to discriminate between classes across thresholds [16].

VIII. Comparative Model Performance

TABLE VII *Comparative Performance of Logistic Regression, Random Forest, and Decision Tree*

Model	Accuracy	Precision	Recall	F1 Score	ROC-AUC
Logistic Regression	0.633898	0.779339	0.643466	0.704915	0.671581
Random Forest	0.636680	0.707295	0.793927	0.748111	0.585999
Decision Tree	0.582425	0.708641	0.654725	0.680617	0.542361

The results of the three models are compared in Table VII and Fig. 8. Logistic Regression achieved the highest ROC-AUC (0.672), indicating the best overall discrimination between success and non-success. It also had the best precision. Random Forest achieved the best recall (0.794) and F1-score (0.748), so it is more sensitive to positive cases but has lower discriminative power. The decision tree scored the lowest, particularly in ROC-AUC, indicating worse generalisation [17].

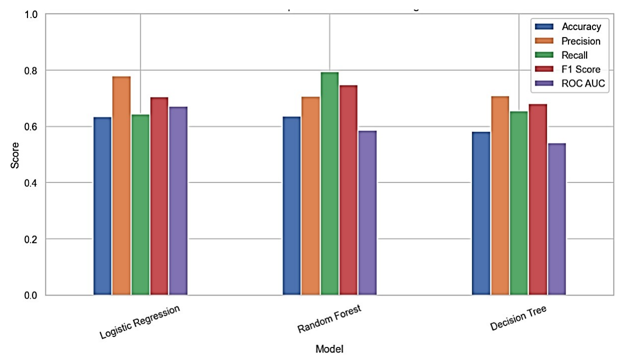

Fig. 8. Performance comparison of machine learning models across evaluation metrics

IX. ROC Analysis

The comparison of the ROC curves among the models revealed a distinct ranking: Logistic Regression first, then Random Forest, and finally Decision Tree. Fig. 9 indicates that the Logistic Regression achieved the optimal AUC, with the Decision tree curve being closest to the diagonal. The importance of this result is that greater algorithmic complexity did not necessarily yield better discrimination in this case.

Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.17, No.5, 2026

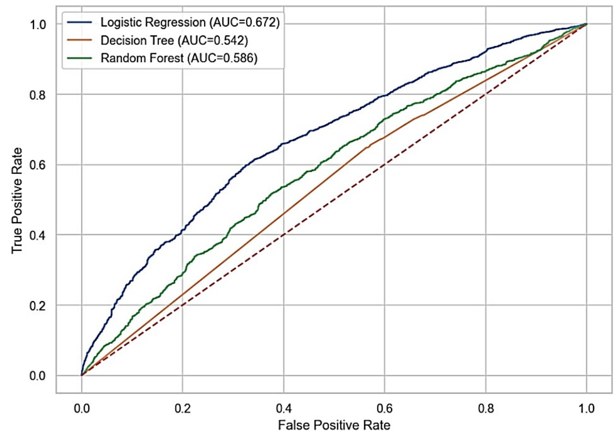

Fig. 9. ROC curve comparison across classification models

X. Confusion Matrix analysis

A. Logistic Regression

Logistic Regression was chosen as the primary model for detailed interpretation because it achieved the highest ROC-AUC. The classification report is summarised in Table VIII. The normalised confusion matrix in Fig. 10 indicates that 61 per cent of actual non-success and 64 per cent of actual success were correctly classified. Logistic Regression was the most balanced in handling the two classes compared to the other models [18].

TABLE VIII Classification Report for Logistic Regression

Class	Precision	Recall	F1-score
0 (Non-success)	0.45	0.61	0.52
1 (Success)	0.78	0.64	0.70
Accuracy			0.63

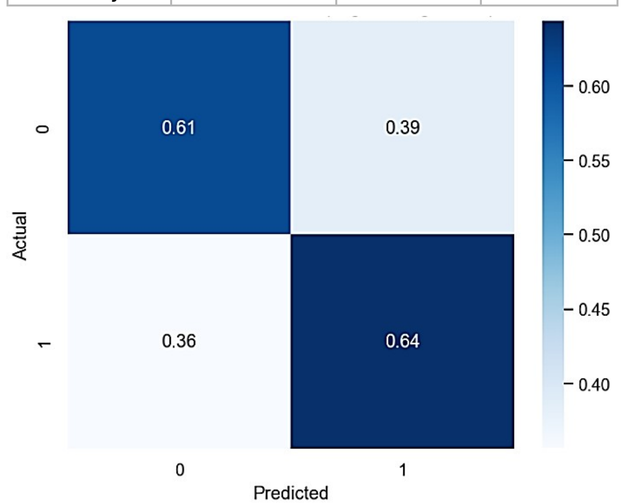

Fig. 10. Normalised confusion matrix for Logistic Regression

B. Random Forest

Random Forest generated another confusion matrix pattern. Indeed, as in Fig. 11, only 30 per cent of real non-success cases were correctly identified, whereas 79 per cent of real success cases were. The model, therefore, favoured the majority success class. This illustrates that strong recall for one class does not necessarily imply balanced predictive utility.

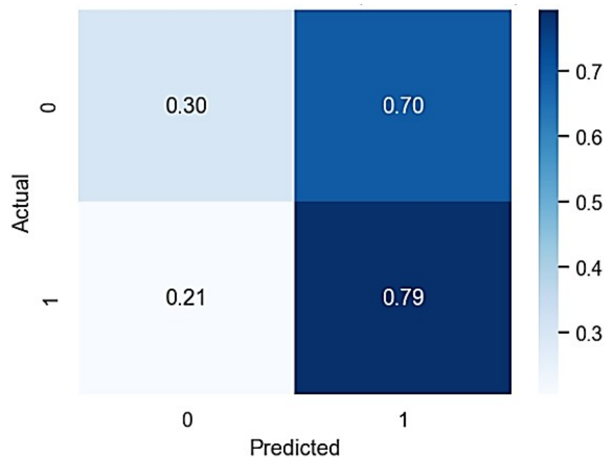

Fig. 11. Normalised confusion matrix for Random Forest

XI. Feature Importance and Explainability

A. Random Forest Built-in Importance

The strongest predictor, according to the Random Forest feature importance ranking, was studied_credits, followed by gender_M, num_of_prev_attempts, age_band_35-55, and, lastly, highest education: Lower Than A Level. These rankings are reported in Table IX and visualised in Fig. 12. Overall, study load, prior attempts, age, educational background, disability, and selected contextual categories influenced the Random Forest predictions.

TABLE IX Top Ten Random Forest Feature Importances

Rank	Feature	Importance
1	studied_credits	0.131567
2	gender_M	0.047585
3	num_of_prev_attempts	0.037470
4	age_band_35-55	0.035219
5	highest_education_Lower Than A Level	0.035211
6	code_presentation_2013J	0.032186
7	disability_Y	0.029984
8	code_module_FFF	0.028351
9	imd_band_30-40%	0.028193
10	imd_band_20-30%	0.027276

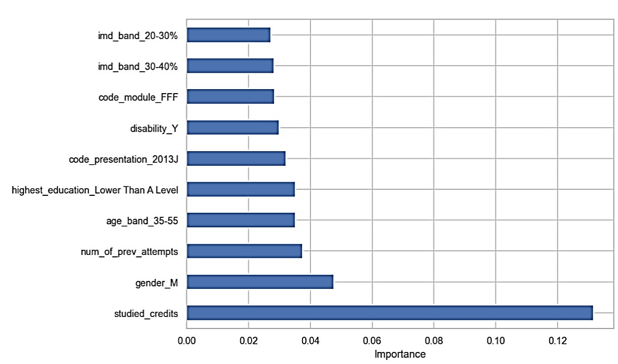

Fig. 12. Random Forest feature importance

B. Permutation Importance for Logistic Regression

To interpret the best-performing model more directly, permutation importance was computed for the Logistic Regression model. Table X ranking and Fig. 13 plot once more illustrate the significance of background in education, module setting, and socioeconomic bands. These findings support the conclusion that even the most discriminative models are affected by educational background and socioeconomic context.

TABLE X *Top Ten Permutation Importance Features for Logistic Regression*

Rank	Feature	Importance
1	highest_education_Lower Than A Level	0.035660
2	code_presentation_2014J	0.020914
3	code_module_DDD	0.019198
4	imd_band_80-90%	0.014607
5	num_of_prev_attempts	0.012196
6	code_module_BBB	0.012010
7	imd_band_90-100%	0.011871
8	imd_band_70-80%	0.011500
9	code_module_FFF	0.011268
10	imd_band_60-70%	0.011176

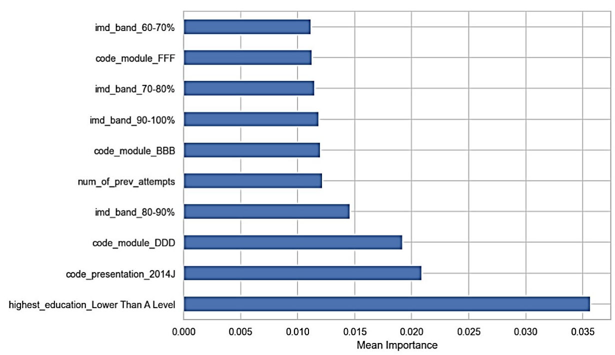

Fig. 13. Permutation importance of features for Logistic Regression

C. SHAP Analysis

The random forest model was subjected to SHAP analysis to better understand the effects of features [19]. The best SHAP features are reported in Table XI. The SHAP summary and bar plots in Figs. 14 and 15 show that increasing the values of highest_education_Lower Than A Level and num_of_prev_attempts shifts the predictions towards a non-successful result but studied_credits has a positive influence on success. The effects of age band, gender, disability, and chosen IMD bands are smaller but significant. These explainability findings, when combined, suggest that prior preparedness and academic history are better predictors of outcome than most demographic variables.

TABLE XI *Top Features Identified by SHAP Analysis*

Rank	Feature
1	highest_education_Lower Than A Level
2	num_of_prev_attempts
3	age_band_35-55
4	studied_credits
5	gender_M
6	code_presentation_2013J
7	disability_Y
8	code_module_FFF
9	imd_band_20-30%
10	imd_band_30-40%

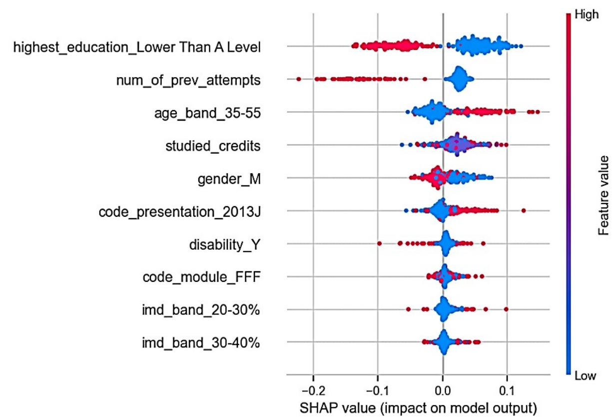

Fig. 14. SHAP summary plot for the Random Forest model

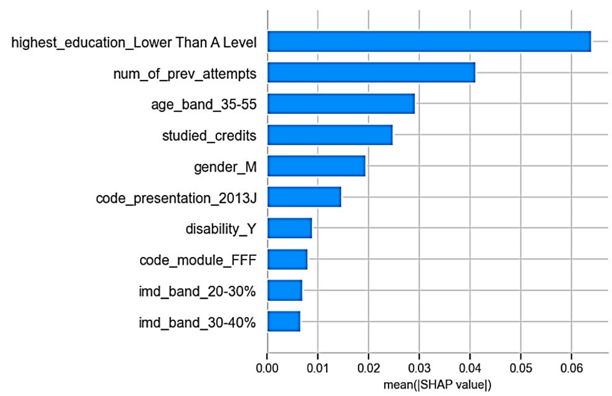

Fig. 15. SHAP feature-importance bar plot

IV. Discussion of Findings

This study combined descriptive analysis, statistical testing, comparative machine learning, and explainability to examine the prediction of student academic success. There are several findings of note. First, the exploratory and statistical analyses clearly show that socioeconomic deprivation matters. The highest failure rates and lowest distinction rates were seen in the most deprived IMD bands, and the lowest in the least deprived. The chi-square outcome justified that this correlation is statistically significant. The dataset, however, cannot be one in which academic performance is independent of socioeconomic context.

Second, the findings indicate that the best classifier depends on the purpose of the evaluation. The strongest overall discriminator was Logistic Regression, which had the best ROC-AUC and precision. Random Forest had the highest recall and F1-score, indicating it has a greater capacity to recognise successful students. However, its inability to detect both non-success and success makes it less appealing when the main objective is early risk detection. The decision tree was consistently the least competitive model.

Third, academic history and prior educational background were among the most effective predictors of outcome. This pattern was replicated in Random Forest importance, permutation importance, and SHAP analysis. Highest_education_Lower Than A Level and num_of_previous_attempt were consistently negative, indicating that students with lower prior preparation or a history of poor performance on previous attempts might need extra institutional support.

Fourth, it was also studied_credits, particularly in Random Forest. This implies that academic outcomes are associated with study intensity or workload, but the underlying mechanism cannot be fully identified from the studentInfo.csv file alone. Lastly, demographic variables, including gender and age group, were present but less influential than academic-history variables.

The research also shows the usefulness of explainable AI in learning. In addition to using performance measures as the sole source of information, the analysis provides feature-level descriptions that make the results more comprehensible and practical for educators and policymakers. At the same time, studying has limitations. It used only studentInfo.csv; more specific behavioural information, including virtual learning environment interactions and assessment records, was not included in the final models. The optimal ROC-AUC value of 0.672 thus shows moderate, not high, discrimination. The models are to be considered decision-support tools, not fully automated decision systems.

XIII. Conclusion

This study examined students' academic performance using descriptive statistics, statistical tests, comparative machine learning, and explainable AI. The findings indicate that socioeconomic deprivation is strongly associated with student outcomes; the Logistic Regression has the highest overall discrimination; the Random Forests have the highest recall and provide a useful nonlinear view; and experience, prior efforts, credits earned in school, and socioeconomic factors are among the most predictive factors.

The results show that student success is determined by structural context and academic preparedness. In practice, early-support strategies can be advantageous for educational institutions that serve students with weaker prior qualifications, a history of repeated failure, and less favourable socioeconomic conditions. The study should be extended in the future by using more OULAD tables, particularly clickstream and assessment data, to measure behavioural engagement and, perhaps, enhance predictive performance. Additional improvements can be achieved through resampling, threshold optimisation, and higher-capacity models such as gradient boosting.

References

[1] Z. Papamitsiou and A. A. Economides, “Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence,” IEEE Trans. Educ. Technol. Soc., vol. 17, no. 4, pp. 49-64, 2014.

[2] A. Almalawi, B. Soh, A. Li, and H. Samra, “Predictive models for educational purposes: A systematic review,” Big Data Cogn. Comput., vol. 8, no. 12, Art. no. 187, 2024.

[3] R. S. Baker, T. Martin, and L. M. Rossi, “Educational data mining and learning analytics,” in The Wiley Handbook of Cognition and Assessment: Frameworks, Methodologies, and Applications, 2016, pp. 379-396.

[4] I. Issah, O. Appiah, P. Appiahene, and F. Inusah, “A systematic review of the literature on machine learning application of determining the attributes influencing academic performance,” Decision Analytics J., vol. 7, Art. no. 100204, 2023.

[5] T. Freiesleben, G. König, C. Molnar, and A. Tejero-Cantero, “Scientific inference with interpretable machine learning: Analysing models to learn about real-world phenomena,” arXiv:2206.05487, 2022.

[6] A. Adadi and M. Berrada, “Peeking inside the black-box: A survey on explainable artificial intelligence (XAI),” IEEE Access, vol. 6, pp. 52138-52160, 2018.

[7] M. Saqr and S. López-Pernas, “Why explainable AI may not be enough: Predictions and mispredictions in decision making in education,” Smart Learning Environ., vol. 11, no. 1, Art. no. 52, 2024.

[8] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, “Educational data mining techniques for student performance prediction: Method review and comparison analysis,” Frontiers Psychol., vol. 12, Art. no. 698490, 2021.

[9] B. Albreiki, N. Zaki, and H. Alashwal, “A systematic literature review of student performance prediction using machine learning techniques,” Educ. Sci., vol. 11, no. 9, Art. no. 552, 2021.

[10] E. López-Meneses, P. C. Mellado-Moreno, C. Gallardo Herrerías, and N. Pelícano-Piris, “Educational data mining and predictive modelling in the age of artificial intelligence: An in-depth analysis of research dynamics,” Computers, vol. 14, no. 2, Art. no. 68, 2025.

[11] Y. Feldman-Maggor, M. Cukurova, C. Kent, and G. Alexandron, “The impact of explainable AI on teachers’ trust and acceptance of AI EdTech recommendations: The power of domain-specific explanations,” Int. J. Artif. Intell. Educ., pp. 1-34, 2025.

[12] J. Kuzilek, M. Hlosta, and Z. Zdrahal, “Open University Learning Analytics dataset,” Sci. Data, vol. 4, Art. no. 170171, 2017.

[13] B. Sekeroglu, R. Abiyev, A. Ilhan, M. Arslan, and J. B. Idoko, “Systematic literature review on machine learning and student performance prediction: Critical gaps and possible remedies,” Appl. Sci., vol. 11, no. 22, Art. no. 10907, 2021.

[14] H. O'Brien Quinn, M. Sedky, J. Francis, and M. Streeton, "Literature review of explainable tabular data analysis," Electronics, vol. 13, no. 19, Art. no. 3806, 2024.

[15] M. Aria, C. Cuccurullo, and A. Gnasso, "A comparison among interpretative proposals for Random Forests," Mach. Learn. Appl., vol. 6, Art. no. 100094, 2021.

[16] J. Li, "Area under the ROC Curve has the most consistent evaluation for binary classification," PLoS One, vol. 19, no. 12, Art. no. e0316019, 2024.

[17] B. Alnasyan, M. Basheri, and M. Alassafi, "The power of deep learning techniques for predicting student performance in virtual learning environments: A systematic literature review," Comput. Educ. Artif. Intell., vol. 6, Art. no. 100231, 2024.

[18] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," Inf. Sci., vol. 513, pp. 429-441, 2020.

[19] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Adv. Neural Inf. Process. Syst., vol. 30, 2017.