Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

From Prediction to Support: Explainable Machine Learning and Rule-Based Interventions for Identifying Students at Academic Risk

Aminat Salami ADEBAYO;[1] Chika Judith ABOLLE-OKOYEAGU;[2] Jo Ann TAIT [3]

[1,2,3] School of Computing, Engineering and Technology, Robert Gordon University, Aberdeen, United Kingdom
[1] a.adebayo11@rgu.ac.uk; https://orcid.org/0009-0008-0729-2954
[2] j.abolle-okoyegu@rgu.ac.uk; https://orcid.org/0000-0001-6046-6986
[3] j.e.tait@rgu.ac.uk; https://orcid.org/0000-0002-0795-3918

Correspondence: Aminat Salami ADEBAYO

Abstract

Early identification of students at risk of academic failure may enable institutions to offer timely and proportionate support. This study developed and compared three supervised classification models, XGBoost, Random Forest and Decision Tree, and linked their predictions to a prototype rule-based intervention dashboard. A public synthetic student-performance dataset containing 6,607 records was preprocessed by imputing missing values and encoding categorical variables. Academic risk was defined by an examination-score threshold, with scores below 65 classified as high risk and scores of 65 or above as low risk. Data were divided into 80% for training and 20% for testing. Models were assessed using accuracy, precision, recall, F1-score, confusion matrices and area under the receiver operating characteristic curve. SHAP values were used to examine the direction and relative importance of predictors. XGBoost achieved the strongest test performance, with 0.92 accuracy, 0.92 weighted precision, 0.92 weighted recall, 0.92 weighted F1-score and an AUC-ROC of 0.976. Random Forest achieved 0.91 accuracy and an AUC-ROC of 0.965, while Decision Tree achieved 0.83 accuracy and an AUC-ROC of 0.809. For the high-risk class, XGBoost produced a precision of 0.89 and recall of 0.86. Across the models, hours studied, attendance, and previous scores were the most influential

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

features. These informed a Power BI prototype that paired risk classifications with suggested support. Findings show technical feasibility of combining explainable prediction with recommendations. However, because the dataset is synthetic, results do not establish effectiveness in a university population. External validation, fairness assessment, and human oversight are required before operational use.

Keywords: ***Learning analytics; Student success; Academic risk; XGBoost; Explainable artificial intelligence; Prescriptive analytics; Early warning systems***

1. Introduction

Universities increasingly use learning analytics to understand engagement, monitor progression and direct student support. One important application is the early identification of students whose academic records indicate a heightened risk of failure or withdrawal. Timely identification may allow teaching and support teams to offer tutoring, mentoring, study-planning assistance or wellbeing referrals before difficulties become entrenched. Predictive models can help with this task by detecting patterns across academic, behavioural and contextual variables that would be difficult to assess consistently at scale (Herodotou et al., 2019; Wong & Li, 2020).

Prediction alone has limited practical value. A risk score does not explain why a student has been flagged, which action is appropriate, or whether the proposed action respects the student's circumstances. Black-box predictions may also weaken trust among educators and students. Explainable artificial intelligence offers one response by showing how individual variables contribute to a model's output (Adadi & Berrada, 2018; Alamri & Alharbi, 2021; Akinloye, 2026). In educational settings, explanation is especially important because model outputs may affect access to support and may reproduce inequalities contained in the data.

A second gap concerns the connection between predictive and prescriptive analytics. Much of the literature focuses on forecasting grades, failure or dropout, while fewer studies translate predictions into transparent and reviewable support

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

recommendations (Ramaswami et al., 2022). Dashboards can make model outputs accessible, but their effects depend on design, interpretation and the surrounding support process. Visual feedback can aid monitoring, although it may also influence motivation and anxiety in unintended ways (Valle et al., 2021). A useful early-warning system must therefore connect prediction, explanation and intervention while keeping academic judgement and student participation at the centre.

This study develops a prototype workflow that compares three tree-based classifiers and uses SHAP explanations to inform rule-based intervention suggestions. It addresses three questions: (1) How accurately do XGBoost, Random Forest, and Decision Tree classify high- and low-risk cases in a public synthetic student-performance dataset? (2) Which features contribute most strongly to the classifications? (3) How can predictions and explanations be translated into a dashboard that supports structured academic intervention? The contribution is methodological. It demonstrates an end-to-end architecture and reports its internal test performance, while treating institutional deployment and educational impact as questions for later validation.

---

2. Literature Review

Predictive learning analytics

Machine learning allows for more predictive models to be applied to student data through the detection of patterns that may go unnoticed when utilising traditional statistical methods. Algorithms such as Logistic Regression (LR), Decision Trees (DT), Naïve Bayes (NB), Support Vector Machines (SVM), Random Forest (RF), Gradient Boosting Decision Trees (GBDT) and Extreme Gradient Boosting (XGBoost) have been employed in supervised learning tasks to predict student performance, retention and dropout probability (Adeniran et al., 2024; Okafor, 2025). However, there is discussion on if the benefit of moving towards larger and more complex models carries educational significance past predictive performance.

Traditional methods are still used to allow institutions to find a balance between accuracy, efficiency, and model interpretability. Ahmed et al. (2021) showed that LR, NB, SVM, and GBDT were sufficient in identifying students at risk prior to failing

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

a course. Zhang et al. (2022) employed Decision Trees and Random Forest algorithms and were able to consistently predict academic performance by understanding nonlinear relationships between academic and behavioural predictors. Machine learning allows institutions to utilise a flexible modelling approach when addressing educational data that consists of numerous variables.

Lately, there has been a focus on ensemble methods such as Random Forests and boosting models. Villar and De Andrade (2024) analysed Decision Trees, SVM, Random Forest, XGBoost, CatBoost, and LightGBM. The authors reported that the boosting models performed the best in dropout prediction with AUC's over 0.90. High model performance is not always beneficial if the results cannot be understood or carried through into educational use. Many of these models sacrifice their model interpretability as their complexity grows.

Metrics other than model accuracy should be considered when analysing the performance of a model. Precision, recall, sensitivity, and true positive rates can allow us to understand how well a model minimises false positives and false negatives (Li et al., 2023; Omigbodun & Akinloye, 2026). False negatives can cause students to not receive the help they need, while false positives can cause students to be labelled as needing help when they do not. As stated before, ensemble models such as Random Forests and Gradient Boosting models tend to show great performance. There is still some discussion on whether obtaining higher degrees of accuracy should be the goal.

Performance is highlighted in many studies, but there are fewer studies focused on how teachers can use these predictions. This has created a rise in explainable artificial intelligence.

Explainability and intervention

While more complex machine learning algorithms have enhanced the predictive performance of learning analytics models, these improvements have often come at the cost of transparency. Ensemble-based algorithms such as Random Forest, XGBoost and LightGBM often perform best on educational datasets by learning complex nonlinear representations of the data. While these algorithms can make accurate predictions, they often do so without providing insight into how decisions are made on an individual level, thus becoming a "black box" (Adadi & Berrada, 2018). Predictions made by learning analytics algorithms have educational

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

consequences, meaning they should be able to be understood, assessed, and validated by practitioners.

The necessity for explainability is becoming widely accepted when using machine learning to make decisions about student support. Baneres et al. (2021) recently wrote that interpretability is a significant issue because limiting the understandability of ML models decreases user trust and prevents the implementation of such systems. When a student is identified as being at risk, teachers should have enough information to understand whether the student is facing actual hardships, if it is a false positive, or if there is a fault in the model. For this reason, there should be more emphasis on why students are being identified as they are rather than if they are being correctly identified.

Explainable AI (XAI) is a collection of methods that allow users to interpret the predictions made by machine learning models. XAI helps determine how much weight each feature had in the prediction process. SHapley Additive exPlanations (SHAP) is one method of explainable AI that has received growing popularity. SHAP creates explanations using Shapley values, allowing users to see both global and local importance of features (Lundberg et al., 2020). SHAP has previously been applied in learning analytics to discover what features impact predictions of students' final grades and risk of dropping a course (Villar & De Andrade, 2024). Alamri and Alharbi (2021) and Melo et al. (2022) have also found explainable models can help increase transparency in early warning systems by showing what features are considered important rather than a binary at-risk prediction. Despite this, there is still discussion about the validity of these post hoc explainable models.

While post hoc explanations increase the interpretability of machine learning models, they do not confirm causation. XAI methods like SHAP allow users to understand how the model is coming to its decisions, but this does not mean that the language used by the AI is causal. Melo et al. (2022) found that SHAP had the highest Explainability Index at 78% compared to LIME at 57% and Shapley Values at 35%. While SHAP was able to provide more accurate explanations than other XAI techniques, it still did not allow users to know the exact reasoning behind the model's decisions. Therefore, while explainability can help increase transparency in early warning models, it does not replace the educator's knowledge when advising students.

While there has been an increase in research focused on explainable predictions, there is still a debate on the trade-off between accuracy and model simplicity. While

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

using a black-box model may provide the best accuracy, it is often difficult for teachers to understand how students were grouped. Using explainable models can allow users to understand how students are systematically grouped, creating more trust in the product. However, while many studies have gone into making accurate and explainable predictions, less literature discusses how to translate explanations to provide teachers with actionable support. The current study aimed to address this gap in literature by using SHAP-based explanations along with predictions to allow for transparent and informed prescriptive support.

Predictors of academic risk

The predictors used in estimating student risk vary between studies. While many sources concur that previous academic achievement along with attendance and learning engagement data make up some of the best predictors of risk, there is mixed consensus as to the predictive power of demographic and contextual factors. Attendance and grades earned in the first semester have been frequently used due to their ability to capture student adaptation to learning at the university during the start of the term (Gonzalez-Nucamendi et al., 2023; Osborne & Lang, 2023; Omigbodun et al. 2026). Furthermore, Tucker and McKnight (2017) found that high school GPA and university GPA from the first term were better predictors of future academic performance than standardised admission testing such as the American College Test (ACT). Thus previous research has supported the use of quantitative measures of learning behaviour to predict student success.

In addition to learning behaviour, demographic, socioeconomic, and other variables have been explored as potential supplemental predictors. Ahmed et al. (2021) used a total of twenty variables to develop their prediction model. Some of these variables were numeric like age, family size, and hours spent on the internet, while others were categorical like gender, department, parent's job, and university. Another frequently used type of behavioural predictor is the students' engagement with the school's Learning Management System (LMS). Research has shown that students who are more active on the LMS generally perform better academically than their less active peers (Helal et al., 2018). Supplemental variables have been shown to consistently improve prediction performance; however, they do not always have an obvious connection to education like previous academics or course attendance. Demographic and socioeconomic factors run the risk of unfairly targeting certain groups of students and invading privacy (Gonzalez-Nucamendi et al., 2023). Although these variables can improve the accuracy of predictions, they should be used with caution.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

Machine learning models have been used to predict at-risk students with success. Decision Trees, Random Forests, Support Vector Machines (SVM), Logistic Regression, Naïve Bayes, and Gradient Boosting Decision Trees (GBDT) have all been used to identify patterns in data to better predict students who may be at risk of academic failure (Ahmed et al., 2021; Zhang et al., 2022). Studies have found that when comparing multiple machine learning models, the model with the highest overall accuracy is not always the best model (Li et al., 2023; Dare, 2024). Other factors such as sensitivity, true positive rate, and classification accuracy should also be taken into consideration when comparing models. This further indicates that the selected predictors should allow a model to meet the specific needs of the task rather than solely maximize overall accuracy.

Existing literature has examined a variety of factors that can be used to predict students who may be at risk of academic failure. However, one question has not been thoroughly addressed. Can the students who have been identified as at risk from these prediction models receive impactful support? Addressing this question is necessary to ensure that these models have a positive impact on students. We want to not only predict which students may need support but also provide them with that support. This research aims to use IF- based prescriptive analytics to translate the results of machine learning models into actionable support recommendations.

From prediction to educational action

Questions about how much automation should be involved in learning analytics have been raised with the recent increase in AI implementation. Automated predictive models can process large quantities of information and easily detect relationships between factors and academic risk. Running multiple predictors of student success, such as grades and attendance, simultaneously could help institutions efficiently provide preventative support to students at the start of their college career.

Machine learning models can help institutions make these determinations quickly and consistently. While automation has the potential to support larger scales of students, there are multiple reasons that automated decision-making may not lead to adaptive support. Predictive analysis does not consider all factors that may affect a student's life. Automated identification of risk groups has also been critiqued for determining too much information about students. Being placed in a risk category is just showing the likelihood of a student fitting a pattern. For example, missing classes can put students at risk of dropping out, but missing classes can be caused

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

by working, being sick, having a disability, caring for someone, or many other circumstances.

A bad midterm grade can mean that a student is struggling, but it could also be that they had an unexpected doctor's appointment. Making decisions without factoring in the human element could lead to suggested interventions that do not fit the student's actual needs. Automated decisions have the potential to be harmful if the institution trusts the analysis to make decisions with no human input. Learning analytics has shifted in recent years from making decisions for people to informing decisions made by people.

Visualisations of student data, like dashboards and recommender systems, can help make decisions by allowing humans to review risk factors, understand potential causes, and choose how to intervene. Decisions should be left to the professionals who can bring their knowledge about the student into the equation. Information not available to analytics could make all the difference in how to approach a student about their performance. How students are talked to about their performance through learning analytics can cause increases in motivation or stress (Valle et al., 2021). The learning interventions themselves are heavily dependent on the process they are used in (Wong & Li, 2020; Akinloye 2025). Although it makes supporting students more efficient, it is unknown how much analytics should weigh on academic decision-making. Studies have shown the benefits of predictive analysis but fail to show how automated decisions can work with human decision-makers. There needs to be a system in place that encourages analytics to be used as a support tool.

---

3. Methodology

Study Design and Data Source

The study used a quantitative model-development design based on the publicly available Student Performance Factors dataset hosted on Kaggle. The dataset contains 6,607 synthetic records and a mixture of academic, behavioural, demographic, psychological and socioeconomic variables. Examples include hours studied, attendance, previous scores, tutoring sessions, motivation, peer influence,

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

access to resources, parental involvement, school type and gender. No records from Robert Gordon University or another identifiable institution were used.

The outcome was derived from examination scores. Records with an examination score below 65 were labelled high risk (class 0), and records with a score of 65 or above were labelled low risk (class 1), following the operational threshold used in the source project. This is a modelling definition rather than a universal definition of academic failure. The threshold should be adapted to programme regulations and validated locally before any institutional application.

Preprocessing

Data were loaded into a Pandas data frame and inspected for structure, descriptive statistics and missing values. Missing numerical observations were imputed using a measure of central tendency, while missing categorical observations were replaced by the mode. Ordinal categorical variables were label-encoded, and non-ordinal categories were converted into binary indicator variables. Irrelevant columns were removed. The processed data were separated into predictors and the binary risk outcome and divided into training and test subsets in an 80:20 ratio. The held-out test set contained 1,322 observations, comprising 426 high-risk and 896 low-risk cases.

Because the outcome was constructed from examination score, the examination score itself must be excluded from the predictor matrix. The project report indicates that predictors and target were separated, but the reproducible code and processed feature list should be checked before submission to confirm that neither examination score nor any direct transformation of it remained among the inputs. This check is essential to rule out target leakage.

The 65-point cut-off created two classes with different frequencies. In the held-out test set, 426 observations were labelled high risk, and 896 were labelled low risk. The high-risk group therefore represented 32.2% of the test cases. The imbalance was moderate rather than extreme, but it justified reporting class-specific measures. No resampling procedure was reported. Retaining the observed class distribution has the advantage of evaluating the models against the dataset as constructed, though future work should test whether class weighting or threshold adjustment improves identification of the high-risk group.

Model Development

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

Three supervised classifiers were trained: Decision Tree, Random Forest and XGBoost. The ensemble models were instantiated with 100 estimators. Random Forest tuning was used after an initial training accuracy of 1.00 suggested overfitting. The source report does not retain a complete, publication-ready record of the random seed, all hyperparameters, cross-validation procedure or software environment. These details should be recovered from the original notebook and added before journal submission to permit replication.

Performance was assessed on the held-out test set using accuracy, weighted precision, weighted recall, weighted F1-score and AUC-ROC. Class-specific precision, recall and F1-score were examined because the high-risk class was smaller and false negatives have practical importance. Confusion matrices were used to show the number and direction of classification errors.

Accuracy expresses the proportion of all test cases classified correctly. Precision for the high-risk class expresses the proportion of high-risk alerts that corresponded to high-risk labels. Recall expresses the proportion of all labelled high-risk cases found by the model. F1-score provides a harmonic balance between precision and recall. AUC-ROC summarises discrimination across classification thresholds. These measures answer different questions and should be read together. For a support system, high-risk recall is particularly important because it quantifies missed cases, while precision indicates the likely workload created by false alerts.

Model Explanation

TreeExplainer from the SHAP package was applied to each fitted classifier. SHAP distributes the difference between a model's prediction and its baseline prediction across the input variables. A positive or negative SHAP value shows the direction in which a feature moved a particular model output, while the magnitude shows the strength of that contribution within the fitted model. Summary plots combine these local explanations to provide a global view of features, importance and direction. The interpretation remains conditional on the model, dataset and encoding choices. Correlated predictors may share or redistribute importance, and a high SHAP value is not evidence of a causal effect.

Prescriptive Rules and Dashboard Design

The three features that appeared consistently among the most influential predictors were used to organise the recommendation logic. Where a case was classified as high risk and showed low attendance, the prototype suggested a staff conversation

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

to identify barriers and agree an attendance or access plan. Where limited study time was prominent, suggestions included study-planning support, tutoring and structured learning sessions. Where previous attainment was low, the proposed response focused on diagnostic academic support and targeted review of prerequisite knowledge. More than one recommendation could apply to a case.

Recommendations for low-risk cases were developmental rather than remedial. They included voluntary peer mentoring, skills workshops and continued monitoring where appropriate. The inclusion of low-risk recommendations was intended to avoid treating the dashboard solely as a failure-detection mechanism. Nevertheless, low-risk students should not be burdened with peer-support responsibilities simply because the model predicts success. Participation should be voluntary, supported and consistent with institutional safeguarding arrangements.

Table 1. Prototype mapping from explanatory signals to staff-reviewed support options

Model signal or context	Interpretive question for staff	Possible support response
Low attendance contributes to a high-risk prediction	Are there academic, health, disability, caring, work or access barriers?	Supportive contact; attendance or access plan; referral to relevant service
Limited study time contributes to a high-risk prediction	Does the student need help with planning, workload or study methods?	Study-planning session; tutoring; structured study group; learning-skills support
Low previous score contributes to a high-risk prediction	Is there a specific prerequisite or assessment skill gap?	Diagnostic review; targeted academic tutoring; formative practice and feedback
Several risk signals occur together	Which need is most immediate, and what support is acceptable to the student?	Coordinated case review with one named contact and an agreed follow-up date

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

Model signal or context	Interpretive question for staff	Possible support response
Low-risk prediction	Would the student value continued development or voluntary peer activity?	Optional skills development; voluntary peer mentoring; routine monitoring

Technical Implementation

TreeExplainer from the SHAP package was applied to the three fitted models. SHAP summary plots were used to rank features and inspect whether high or low feature values pushed predictions toward high or low risk. The consistently prominent features were then used in a rule-based recommendation layer. High-risk cases could prompt staff review for mentoring, tutoring, attendance support or structured study planning. Low-risk cases could receive general development opportunities or be invited to voluntary peer-mentoring activity. Recommendations were presented in a Power BI prototype alongside counts and proportions by predicted risk, student-level filters and a table connecting each classification to suggested action.

The analysis was undertaken in Python using Pandas and NumPy for data preparation, scikit-learn for Decision Tree and Random Forest modelling and performance metrics, XGBoost for gradient boosting, and SHAP for explanation. Matplotlib and Seaborn supported exploratory visualisation. Trained models and categorical encoders were saved for reuse, and the processed output was exported for Power BI. Power Query and a Python connector were used to bring predictions and recommendations into the dashboard. A generated index served as a non-identifying record key within the prototype.

Ethics and Governance

The study analysed a public synthetic dataset and did not involve identifiable human participants, contact with students or access to institutional records. Formal operational use would require institutional ethical review, data-protection assessment, role-based access controls, security testing, bias and fairness analysis, clear retention rules, and a process through which students can understand and challenge a classification. Predictions should supplement

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

professional judgement and conversation with the student. They should not determine progression, discipline, funding or access to services.

---

4. Results

Overall Model Performance

XGBoost produced the strongest overall results, followed closely by Random Forest. Decision Tree was materially weaker across all summary measures. XGBoost achieved test accuracy and weighted F1-score of 0.92, with an AUC-ROC of 0.976. Random Forest achieved accuracy of 0.91 and an AUC-ROC of 0.965. Decision Tree achieved accuracy of 0.83 and an AUC-ROC of 0.809.

Table 2. Overall test-set performance

Model	Accuracy	Weighted precision	Weighted recall	Weighted F1	AUC-ROC
XGBoost	0.92	0.92	0.92	0.92	0.976
Random Forest	0.91	0.90	0.91	0.90	0.965
Decision Tree	0.83	0.83	0.83	0.83	0.809

Performance for the High-risk Class

For high-risk students, XGBoost attained precision of 0.89, recall of 0.86 and F1-score of 0.87. It correctly identified 366 of 426 high-risk cases and missed 60. Random Forest correctly identified 349 high-risk cases and missed 77, yielding recall of 0.82. Decision Tree correctly identified 316 and missed 110, yielding recall of 0.74. XGBoost therefore reduced false negatives by 17 compared with Random Forest and by 50 compared with Decision Tree on this test split.

Table 3. Class-specific performance

Model	Class	Precision	Recall	F1-score	Support

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

Model	Class	Precision	Recall	F1-score	Support
XGBoost	High risk	0.89	0.86	0.87	426
XGBoost	Low risk	0.93	0.95	0.94	896
Random Forest	High risk	0.88	0.82	0.85	426
Random Forest	Low risk	0.92	0.95	0.93	896
Decision Tree	High risk	0.74	0.74	0.74	426
Decision Tree	Low risk	0.88	0.88	0.88	896

Table 4. Confusion matrices on the held-out test set

Model	Actual high predicted high	Actual high predicted low	Actual low predicted high	Actual low predicted low
XGBoost	366	60	47	849
Random Forest	349	77	48	848
Decision Tree	316	110	111	785

Explanations and Intervention Prototype

SHAP analysis identified hours studied, attendance and previous scores as the most influential features across the three models. Higher attendance generally shifted XGBoost predictions toward the low-risk class, while lower attendance shifted them toward high risk. Fewer hours studied and lower previous scores were also associated with high-risk predictions. These patterns are educationally plausible and align with earlier research on engagement and prior attainment, although the analysis does not establish causality.

The XGBoost predictions were incorporated into a Power BI prototype. The dashboard displayed the number and percentage of cases assigned to each risk group, permitted filtering by a generated student index and risk category, and paired classifications with suggested interventions. The prototype demonstrates a possible

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

decision-support interface. It was not connected to live university systems, and the recommended interventions were not delivered or evaluated.

Comparative Error Analysis

The error distributions clarify the practical difference between the models. XGBoost generated 107 errors, consisting of 60 false negatives for the high-risk class and 47 false positives. Random Forest generated 125 errors, with 77 false negatives and 48 false positives. Its main disadvantage relative to XGBoost was therefore a larger number of missed high-risk cases rather than a substantial increase in unnecessary alerts. Decision Tree generated 221 errors, divided almost equally between 110 false negatives and 111 false positives. This pattern indicates weaker separation of both classes.

Among the 413 cases predicted as high risk by XGBoost, 366 were labelled high risk, producing the reported precision of 0.89. Among the 426 cases labelled high risk, 366 were found, producing recall of 0.86. These values illustrate the operational trade-off. If support teams acted on every high-risk XGBoost alert in this test set, 47 contacts would concern cases labelled low risk, while 60 labelled high-risk cases would receive no alert. Whether this balance is acceptable depends on the nature, cost and potential harm of the intervention. A low-cost invitation to optional study support can tolerate more false positives than a consequential or stigmatising action.

The small difference in overall accuracy between XGBoost and Random Forest should not be treated as proof that XGBoost will always be superior. Results are based on one dataset and one held-out split, without reported confidence intervals or repeated resampling. The AUC-ROC values nevertheless suggest that both ensemble methods separated the two classes effectively within these data. Decision Tree retained an advantage in simplicity and direct interpretability, but its lower recall and greater number of false alerts would limit its use as the principal classifier in this form.

Discussion

This study examined whether an explainable, tree-based classification workflow could move from academic-risk prediction to structured support recommendations. Within the held-out portion of the synthetic dataset, XGBoost produced the best discrimination and the strongest balance of precision and recall. Random Forest performed almost as well, while a single Decision Tree made substantially more

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

errors. The pattern is consistent with the capacity of ensemble methods to capture nonlinear relations and interactions in structured educational data (Badal and Sungkur, 2023; Villar and De Andrade, 2024).

The high-risk class deserves separate consideration. Overall accuracy can appear reassuring in an imbalanced dataset because the majority class dominates the measure. XGBoost still missed 60 of 426 high-risk cases, which corresponds to 14% of the high-risk test cases. In a real support system, those false negatives could represent students who receive no alert despite needing assistance. The appropriate threshold would therefore depend on institutional capacity and the relative cost of false negatives and false positives. A lower classification threshold might improve sensitivity but would generate more alerts for staff to review.

Hours studied, attendance and previous scores emerged as the leading features. Their prominence supports an intervention model focused on study planning, participation and targeted academic help. Yet these features should not be interpreted in isolation. Low attendance may reflect caring duties, disability, employment, illness, transport difficulty or disengagement. Limited study time may arise from financial pressure or inaccessible teaching materials. A responsible response begins with enquiry, not assumption. The dashboard should prompt a supportive conversation and allow staff and students to select an appropriate action together.

The use of SHAP improves transparency by revealing how the model uses available features. It does not prove that changing a feature will change the outcome. Nor does it resolve fairness concerns if the training data contain biased patterns. Institutions should test performance across relevant groups, examine calibration, document the purpose and limitations of each variable, and avoid using protected characteristics as convenient proxies for risk. Explanations should be intelligible to students and staff, with routes for correction where records are inaccurate.

The proposed prescriptive layer is deliberately simple. Rule-based recommendations are easier to inspect than a second opaque model and can be aligned with the services an institution can actually provide. However, their effectiveness remains an empirical question. Mentoring, tutoring and peer support have potential benefits (Collier, 2022), but an alert system only improves outcomes when support is accessible, timely, acceptable and adequately resourced. Prospective evaluation should measure uptake, student experience, attainment,

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

retention and unintended effects rather than judging success by predictive accuracy alone.

Relationship to Previous Studies

The results accord with a broad pattern in educational prediction studies in which ensemble tree methods outperform a single Decision Tree on structured datasets. Badal and Sungkur (2023) similarly report the value of comparing algorithms rather than assuming that one classifier will perform consistently across educational settings. Villar and De Andrade (2024) show that boosted and ensemble approaches can produce strong predictions of dropout and academic success. The present analysis adds an integrated explanation and recommendation layer, though its synthetic setting means that direct numerical comparison with studies using institutional records would be misleading.

The prominence of attendance, study time and prior performance is also consistent with research that treats engagement and academic history as important indicators. Sequential engagement models demonstrate that patterns of participation can improve academic-performance prediction (Song et al., 2021), while analyses of virtual learning environment data show the value of behavioural traces at scale (Waheed et al., 2020). The present dataset is less granular because it does not represent week-by-week change. Its results should therefore be viewed as a cross-sectional demonstration rather than a dynamic early-warning model.

The use of SHAP responds to concerns about opaque educational algorithms. Systematic reviews have noted growing interest in explainable student-performance models, alongside continuing variation in how explanations are produced and evaluated (Alamri and Alharbi, 2021). In this study, SHAP made the model's dependence on a small group of plausible features visible. A stronger follow-up would test whether students and staff understand those explanations, whether they regard them as fair, and whether explanations help them select better support.

The prototype also reflects the argument that prediction should lead to an educationally meaningful response. Ramaswami et al. (2022) connect explainable machine learning with automated prescriptive analytics, while Wong and Li (2020) show that learning-analytics interventions vary widely in design and evidence. The current paper adopts a cautious middle position. It demonstrates how recommendations can be generated consistently, but reserves the final decision for

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

informed human review and explicitly avoids claims that the rules identify the most effective treatment.

---

5. Conclusion

An explainable machine-learning workflow combining prediction, SHAP interpretation and rule-based recommendations can provide a coherent prototype for student-support decision making. XGBoost achieved the strongest internal test performance, with 0.92 accuracy and an AUC-ROC of 0.976, and identified more high-risk cases than Random Forest or Decision Tree. Hours studied, attendance and previous scores were the leading model features and informed proposed support actions. These findings demonstrate technical feasibility within a synthetic dataset. They do not establish readiness for operational use. The next phase should reproduce the analysis, exclude any possibility of target leakage, validate the model on real and temporally separated institutional data, assess fairness and calibration, and evaluate whether human-reviewed interventions improve outcomes without causing harm.

Practical Implications

A staged implementation is advisable. First, the model should be rebuilt with de-identified institutional data under an approved governance framework. Second, temporal validation should test whether a model trained on one cohort performs on a later cohort. Third, staff should review alerts in a silent pilot in which predictions are compared with existing support processes but do not trigger automated action. Fourth, a small prospective pilot should assess workload, student acceptability and intervention uptake. Only after these stages should wider deployment be considered.

The dashboard should display the date and source of each input, the main factors contributing to the classification, uncertainty, and a clear statement that the output is advisory. Access should be restricted to staff with a defined support role. Students should be told how their data are used and given a practical route to query or correct information. Automated messaging based solely on risk status should be avoided because it may cause anxiety or stigma.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

Institutional preparation extends beyond model development. Each suggested action should have a named service owner, eligibility criteria, expected response time and a process for recording whether contact occurred. Alert volumes should be compared with staff capacity before thresholds are selected. If a model produces more cases than services can support, apparent technical success may simply create an unmanaged queue. The institution should also monitor whether some programmes or student groups receive more alerts but less effective assistance.

The system should support, rather than displace, existing sources of knowledge. Tutors, advisers and support staff may know about circumstances that are absent from the dataset. Students themselves remain the most important interpreters of their needs. A useful workflow would allow staff to record that an alert was reviewed, note whether the student accepted support, select a reason where no action was appropriate, and schedule follow-up. These records would create evidence for improving the process while avoiding a presumption that every model output is correct.

Ethical and Governance Implications

Student-risk prediction involves an asymmetry of power because institutions control both the data and many of the decisions that affect progression. Transparency must therefore extend beyond a technical explanation of features. Students need to know the purpose of the system, the categories of data used, who can see an alert, how long records are retained, and whether declining an offered intervention has any consequence. Clear communication is necessary even when the model is used only to offer support.

Fairness evaluation should include performance measures disaggregated by relevant groups, subject to lawful and ethical data handling. Equal overall accuracy can conceal different false-negative rates. If one group is more likely to be missed, the system may widen existing gaps in access to help. Calibration should also be checked so that comparable predicted risks have comparable observed outcomes across groups. Where differences appear, institutions should review data quality, feature design, thresholds and the surrounding support process rather than treating model adjustment as the only remedy.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

Human oversight must be substantive. A staff member should be able to disagree with a prediction, document why a recommendation is unsuitable and choose a different response. Oversight becomes nominal if staff lack time, training or authority to challenge the dashboard. Training should cover interpretation of precision, recall, uncertainty and SHAP values, together with the limits of causal inference. Governance should include periodic review, incident reporting and a clear decision about when a model should be suspended or retired.

Future Studies

The first research priority is a fully reproducible analysis. The original notebook should be archived with a data dictionary, preprocessing pipeline, complete hyperparameters, software versions and random seeds. Nested or repeated cross-validation would provide a more stable estimate of model selection performance. Confidence intervals and statistical tests should be used to assess whether observed differences between XGBoost and Random Forest are likely to persist beyond a single split. Calibration plots and precision-recall curves would add information that is particularly useful when class frequencies change.

The second priority is external and temporal validation. A model trained on one cohort should be tested on a later cohort without refitting, and preferably across programmes with different assessment structures. Predictor availability should be aligned with explicit intervention points during the semester. Performance drift should be monitored when curricula, assessment policies or patterns of engagement change. The outcome definition should also expand beyond a single examination threshold to include module completion, progression, withdrawal and student-defined measures of success.

The third priority is intervention evaluation. A silent pilot can assess alert validity and workload without contacting students. A subsequent prospective study should compare usual support with a carefully governed alert-assisted process. Outcomes should include academic performance, retention, speed of contact, service uptake, staff workload, student trust and possible distress. Qualitative interviews would help explain why some students accept or decline support and whether dashboard explanations feel accurate. Only this combination of technical, educational and experiential evidence can establish practical value.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

Limitations

The principal limitation is the synthetic dataset. It permits technical development without exposing personal records, but its distributions and relationships may not represent any real university. Reported performance is therefore internal to this dataset and test split. The risk outcome was created from a single examination-score threshold and does not capture programme-specific pass rules, progression, withdrawal or multidimensional student success. The dataset was split randomly rather than temporally or by institution, and the source report does not document repeated cross-validation, confidence intervals or statistical comparison between models.

A further concern is reproducibility. Full hyperparameters, random seeds, software versions, the exact processed feature set and the complete intervention rules should be recovered from the project files. care is needed to confirm that the examination score, from which the target was derived, was excluded from all predictor inputs. Without this confirmation, target leakage cannot be ruled out. The SHAP results describe model behaviour rather than causal mechanisms. Finally, the dashboard and interventions were prototypes and were not evaluated with students or staff.

Funding

This research received no external funding.

Conflicts of Interest

The author declare no conflict of interest.

Data Availability

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

The Student Performance Factors dataset used in the study is publicly available from Kaggle at https://www.kaggle.com/datasets/lainguyn123/student-performance-factors

---

References

Adadi, A. and Berrada, M. (2018) Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Adekitan, A.I. and Salau, O. (2019) The impact of engineering students' performance in the first three years on their graduation result using educational data mining. Heliyon, 5(2), e01250. https://doi.org/10.1016/j.heliyon.2019.e01250

Adeniran, A. A, Adeniran, A. O, Familusi, O. B. & Adedayo, O. (2024). The Outlook of Cyber Security in African Businesses: Issues and Way-out. *Management Analytics and Social Insights*, *1* (2), 260-271.

Akinloye, D. N. (2024). Grandparent-Grandchild Communication: Synthesis of Integrative Model. Psychology Nexus, 1(1), 1-15

Akinloye, D. N. (2025). Communication Pattern and the Influence of Social Media Contents on Young Adults' Wellbeing in Texas, United States of America. Journal of Communication, 6(2), 1-22.

Alamri, R. and Alharbi, B. (2021) Explainable student performance prediction models: A systematic review. IEEE Access, 9, 33132–33143. **DOI:** 10.1109/ACCESS.2021.3061368

Almalawi, A., Soh, B., Li, A. and Samra, H. (2024) Predictive models for educational purposes: A systematic review. Big Data and Cognitive Computing, 8, 187. https://doi.org/10.3390/bdcc8120187

Badal, Y.T. and Sungkur, R.K. (2023) Predictive modelling and analytics of students' grades using machine learning algorithms. Education and Information Technologies, 28, 3027–3057.

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

Baneres, D., Guerrero-Roldan, A.E., Rodriguez-Gonzalez, M.E. and Karadeniz, A. (2021) A predictive analytics infrastructure to support a trustworthy early warning system. Applied Sciences, 11, 5781. https://doi.org/10.3390/app11135781

Collier, P. (2022) How peer mentoring can help universities promote student success in a post-COVID-19 pandemic world. Metropolitan Universities, 33(1), 37–54. https://doi.org/10.18060/25222

Dare, A. N. (2024). Communication and Online Platforms: Pitfalls, Possibilities, and Pressures. Texas Tech University. Ttu-ir.tdl.org

Han, J., Kim, K.H., Rhee, W. and Cho, Y.H. (2021) Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation. Computers & Education, 163, 104041. https://doi.org/10.1016/j.compedu.2020.104041

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z. and Hlosta, M. (2019) A large-scale implementation of predictive learning analytics in higher education: The teachers' role and perspective. Educational Technology Research and Development, 67, 1273–1306.

Hlosta, M., Herodotou, C., Papathoma, T., Gillespie, A. and Bergamin, P. (2022) Predictive learning analytics in online education: A deeper understanding through explaining algorithmic errors. Computers and Education: Artificial Intelligence, 3, 100108.

Kokoc, M. and Altun, A. (2021) Effects of learner interaction with learning dashboards on academic performance in an e-learning environment. Behaviour & Information Technology. https://doi.org/10.1080/0144929X.2019.1680731

Maheshwari, A., Malhotra, B.S., Hada, M., Ranka, M. and Basha, M.S.A. (2024) Comparative analysis of machine learning models in predicting academic outcomes: Insights and implications for educational data analytics. Proceedings of the 2024 International Conference on Smart Systems for Applications in Electrical Sciences. https://doi.org/10.1109/ICSSES62373.2024.10561260

Melo, E., Silva, I., Costa, D., Viegas, C. and Barros, T. (2022) On the use of explainable artificial intelligence to evaluate school dropout. Education Sciences, 12, 845. https://doi.org/10.3390/educsci1212084

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

Okafor, C. J., Adeniran, A. A., & Adeniran, A. O. (2025). Machine Learning and AI for Predictive Maintenance and Grid Integration of Wind Farms. *Soft Computing Fusion with Applications*, *2*(2), 62-73.

Omigbodun, T. O., & Akinloye, D. N. (2026). Context and Application of Agnew's General Strain Theory to Explain Crime. Journal of International Review of Research Studies, 1(06), 1-11. https://doi.org/10.66104/s04mz37

Omigbodun, O. T., Familusi, O. B., Akinsola, S. A., Adeniran, A. A., & Adeniran, A. O. (2026). AI-Powered Workflow Automation in Small Businesses. *Open Journal of Economics and Commerce,* 7(1): 13-45. https://doi.org/10.22259/2638-549X.0701003

Ramaswami, G., Susnjak, T. and Mathrani, A. (2022) Supporting students' academic performance using explainable machine learning with automated prescriptive analytics. Big Data and Cognitive Computing, 6, 105. https://doi.org/10.3390/bdcc6040105

Song, X., Li, J., Sun, S., Yin, H., Dawson, P. and Doss, R.R.M. (2021) A sequential engagement-based academic performance prediction model. IEEE Intelligent Systems, 36, 46–53. **DOI:** 10.1109/MIS.2020.3006961

Valle, N., Antonenko, P., Valle, D., Sommer, M., Huggins-Manley, A.C., Dawson, K., Kim, D. and Baiser, B. (2021) Predict or describe? How learning analytics dashboard design influences motivation and statistics anxiety in an online statistics course. Educational Technology Research and Development, 69, 1405–1431. https://doi.org/10.1007/s11423-021-09998-z

Villar, A. and De Andrade, C.R. (2024) Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. Discover Artificial Intelligence, 4, 2. https://doi.org/10.1007/s44163-023-00079-z

Waheed, H., Hassan, S.U., Aljohani, N.R., Hardman, J., Alelyani, S. and Nawaz, R. (2020) Predicting academic performance of students from VLE big data using deep learning models. Computers in Human Behavior, 104, 106189. https://doi.org/10.1016/j.chb.2019.106189

Nexus Science Review
INTERNATIONAL JOURNAL OF INTERDISCIPLINARY RESEARCH

Received: 15/10/2026 - **Accepted:** 13/11/2026
Vol: 01.06
DOI: 10.66104/sxxfns69
Pages: 1-25

Wong, B.T.M. and Li, K.C. (2020) A review of learning analytics intervention in higher education (2011–2018). Journal of Computers in Education, 7(1), 7–28. https://doi.org/10.1007/s40692-019-00143-7

Author Responsibility Statement

ASA: Conceptualization, Writing - original draft, Literature, Methodology, Result, Conclusion; CJA: Introduction, Conclusion, Editing; JOT: Conceptualization, Writing - original draft, Editing; CRU: Introduction, Literature, Conclusion, Visualization. The authors read and approved the final manuscript.