JASMINE
Journal of Intelligent Systems and Machine Learning

Articles in Press

Decision Tree-Based Early Warning System for Academic Failure: Comparative Analysis with Random Forest and Logistic Regression

Virzan Pasa Nugraha[1,*], Fidi Supriadi[1], David Setiadi[1]
[1]Department of Informatics, Faculty of Information Technology, Sebelas April University, Sumedang, Indonesia

*Author to whom any correspondence should be addressed.
E-mail: 220660121054@student.unsap.ac.id

Received: November 12, 2025
Accepted for publication: December 16, 2025

ABSTRACT
A Decision Tree–based early warning system for academic failure was evaluated using a dataset of 649 student observations and compared with Random Forest and Logistic Regression models through nested ten-fold cross-validation. The mean accuracy of the Decision Tree model was 0.9169, compared to 0.9260 for the Random Forest model and 0.8922 for the Logistic Regression model. Although the Random Forest model achieved the highest raw accuracy, paired statistical testing indicates that its performance difference with the Decision Tree model is not statistically significant (paired t-test, $p = 0.104950$). The difference between the Decision Tree model and the Logistic Regression model is statistically significant before Bonferroni correction ($p = 0.01734$) but becomes non-significant after adjustment (Bonferroni-adjusted $p = 0.05201$). The Decision Tree model was therefore selected to balance competitive predictive performance with interpretability through explicit and readable decision rules. In out-of-fold evaluation, the Random Forest model achieved the strongest results, with an accuracy of 0.9245, high receiver operating characteristic and precision–recall performance, a balanced accuracy of 0.8409, a minority-class recall of 0.7200, and twenty-eight false-negative predictions. Feature-importance analysis showed that G2 was the most influential variable with a relative importance of 0.2417, followed by G1 with 0.1992. Shapley value analysis and an ablation study further demonstrated that removing G1 and G2 substantially reduced overall accuracy and minority-class recall. These findings support the use of the Decision Tree model in educational contexts, where transparent rule-based decisions can guide early academic interventions while maintaining performance comparable to more complex ensemble methods.

Keywords: Decision Tree, Early Warning System, Academic Failure Prediction, Model Interpretability, Shapley Value Analysis

I. Introduction

Student achievement in secondary education plays a crucial role in shaping future academic and professional opportunities. However, academic failure at this stage remains a persistent concern for many educational institutions, affecting long-term student outcomes and placing additional demands on organizational resources [1]. These challenges have encouraged extensive research on predictive modelling for educational purposes, particularly the development of systems capable of identifying students at risk of academic failure early enough for meaningful intervention [2]. An effective early warning mechanism enables teachers, homeroom advisors, and guidance counselors to anticipate academic difficulties and respond proactively. As noted by Khan et al., predictive approaches must be presented in a transparent and interpretable form to ensure that educators can understand each prediction and act upon it [3].

Previous studies show that early academic indicators, including demographic profiles, attendance patterns, and first-period grades, are strong predictors of final outcomes. Contemporary research has employed diverse computational models to operationalize these predictors. Some studies attempt to predict performance using only pre-enrollment variables [4], [5], [6]. While others rely on mid-semester data. Within the broader domain of educational outcome modelling, many high-performing predictive techniques, particularly ensemble methods and deep learning models rely on complex internal structures that are not inherently transparent, making their decision processes difficult to operationalize and communicate to non-technical educational stakeholders without additional explainability mechanisms [7], [8], [9]. Despite extensive research, two substantial gaps remain. First, relatively few studies assess whether transparent and inherently interpretable models can perform competitively when evaluated through rigorous statistical testing rather than mere descriptive comparison. Second, although interest in interpretable artificial intelligence is growing, existing work seldom provides operational interpretability in the form of explicit, human-readable rules that teachers can directly use for classroom-level intervention.

To address these limitations, the present study investigates the Decision Tree model, an inherently transparent method widely recognized for its ease of interpretation [10]. Although commonly assumed to underperform compared with more complex ensemble or statistical approaches, the Decision Tree model is evaluated here through a statistically validated comparison with the Random Forest model, and Logistic Regression [4],[11],[12]. The evaluation incorporates nested ten-fold cross-validation, systematic hyperparameter optimization, synthetic oversampling to address class imbalance, and paired statistical testing to determine whether observed differences in predictive accuracy are statistically meaningful.

In addition to comparing models, this study examines the timing of prediction by evaluating two scenarios: one using only first-period grades (G1) and another combining first-period and second-period grades (G2). This comparison contributes to ongoing discussions regarding the balance between early,but potentially less stable,predictions and later,but more reliable,assessments [13]. The objectives of this study are threefold: (1) to identify the predictive value of early-term and mid-term indicators; to determine whether the Decision Tree model offers competitive performance relative to more complex methods; and (3) to analyze the transparent rule structures generated by the Decision Tree model and their implications for early educational intervention. Complementary interpretability techniques, including Shapley value analysis and permutation importance, are used to validate the internal reasoning of the models. The resulting rule structures, driven primarily by G2, G1, and prior academic failures, align with established patterns in the educational literature and underscore their practical relevance.

Overall, the findings of this study show that the Decision Tree model achieves predictive performance that is statistically comparable to that of the Random Forest model and the Logistic Regression model, while providing explicit and interpretable rule structures suitable for real-world early intervention. These results demonstrate the practical value of transparent predictive systems in education and reinforce the study's contribution to advancing interpretable early warning methodologies.

II. Related work

Research on predictive modelling in education has expanded significantly in recent years, with numerous studies applying advanced computational techniques to forecast student performance. A dominant pattern in contemporary literature is the prioritization of maximum predictive accuracy, which has led many studies to emphasize complex ensemble architectures. Recent systematic reviews report that methods such as Gradient Boosting, CatBoost, and Random Forest frequently outperform simpler models across diverse student datasets [1]. Comparative evaluations also show a consistent preference for high-capacity classifiers,including Support Vector Machines, Naive Bayes, and various ensemble combinations, over inherently interpretable models such as the Decision Tree [4], [7]. Findings from related domains further reinforce this trend, demonstrated that Random Forest outperformed the Decision Tree model in a separate classification context, strengthening the widespread assumption that model complexity directly correlates with predictive strength [14].

Despite these advancements, recent studies highlight a growing methodological imbalance in the field. As Gunasekara and Saarela observe, much of the existing work "compares performance, but rarely evaluates explainability," resulting in models that may achieve high accuracy yet remain unsuitable for operational use in educational environments [9]. For stakeholders such as teachers and counselors, model transparency is essential not merely for interpretability but for ethical justification of interventions, especially when predictions influence academic or behavioral decisions. Several authors note that the opacity of complex models limits accountability and trust, undermining their practical utility within early-warning contexts [2], [8]. This challenge persists even with the adoption of post hoc interpretability tools such as SHapley value methods or local surrogate explainers, which provide approximations but do not yield the explicit, rule-level reasoning required for classroom-level decision-making.

Recent research on explainable artificial intelligence in education further illustrates the disconnect between technical innovation and practical deployment. Although post hoc explanation techniques offer insights into feature contributions, they often function as add-on layers rather than inherent components of the prediction process. Systematic reviews consistently note that such methods remain fragmented, difficult for non-experts to interpret, and insufficient for guiding actionable interventions [5], [9]. In contrast, rule-based models such as the Decision Tree offer an intrinsically transparent structure, yet they are underrepresented in rigorous comparative studies that evaluate whether their interpretability comes at a statistically significant cost to predictive performance.

Recent studies have demonstrated that model-agnostic interpretability techniques can augment the transparency of machine learning models in educational contexts. For example, Miranda et al. employed SHAP-based explanations to interpret predictions of student academic performance and showed that such approaches allow educators to understand the relative contribution of behavioral and engagement features [15]. Their findings reinforce the importance of incorporating explainability into predictive systems, particularly when complex models are used in real-world early-warning scenarios.

The limitations identified above highlight two critical gaps in the current literature: (1) a lack of studies that evaluate interpretable models and complex models under equivalent, statistically validated experimental conditions; and (2) a shortage of research demonstrating how interpretable models can yield operational, human-readable rules that directly support early academic interventions.

This study addresses these gaps by conducting a comparative analysis of the Decision Tree model against two widely adopted complex models, the Random Forest model and the Logistic Regression model, using a rigorous nested cross-validation design and paired statistical tests. Unlike prior work that reports accuracy descriptively, this study evaluates whether observed performance differences are statistically meaningful. Furthermore, by focusing on early academic indicators such as G1, G2, absences, and prior failures, the study emphasizes features that are both actionable and pedagogically relevant. Through the extraction and analysis of explicit IF–THEN decision rules, this study advances an interpretable modelling approach that aligns with the practical needs of school-level stakeholders while maintaining predictive performance competitive with more complex models.

III. Material and Methods

This study adopted a structured and reproducible machine-learning workflow to develop an interpretable early-warning model for academic failure. The methodological pipeline comprises data acquisition and specification, data preprocessing, model development, nested cross-validation, hyperparameter optimization, class-imbalance handling, model evaluation, statistical significance testing, interpretability analysis, ablation experiments, and misclassification examination. Figure 1 illustrates the methodological framework applied in this research.

A. Data Source and Specification

The Student Performance dataset [16] was utilized, consisting of 649 instances and 33 attributes. A binary target variable was constructed from the final grade (G3) using:

- Fail = 1 for G3 $\leq$ 9
- Pass = 0 for G3 $\geq$ 10

The resulting class distribution was imbalanced (Fail = 100; Pass = 549). Predictor variables included academic (G1, G2, failures), behavioural (absences, studytime), contextual (higher, romantic, schoolsup), and target features as shown in Table 1. Two prediction scenarios were evaluated: (1) Early Warning Model (EWM): pre-enrollment features and G1; (2) Mid-Term Model (MTM): pre-enrollment features, G1, and G2.

B. Data Preprocessing

Categorical variables were transformed using One-Hot Encoding. Numerical features retained their original scale except when standardized within model-specific pipelines (e.g., Logistic Regression). The dataset contained no missing values. To mitigate class imbalance (Fail $\approx$ 0.1541), SMOTE was applied only inside the inner cross-validation training folds to prevent information leakage. SMOTE was implemented with sampling_strategy = 'auto' and k_neighbors = 5.

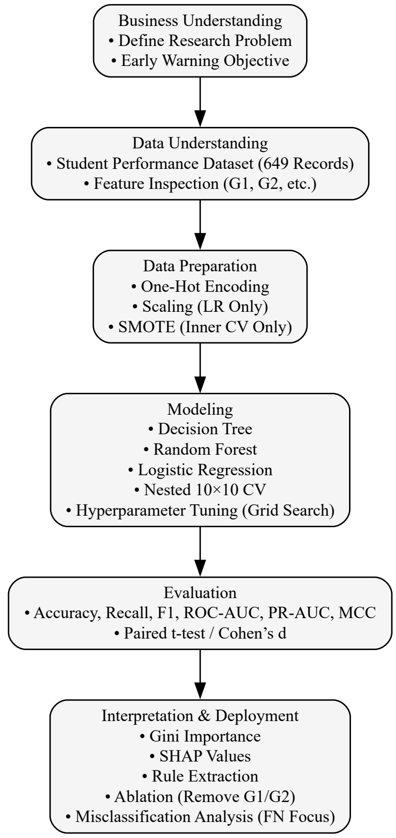

Figure 1. Methodological framework adapted from the CRISP-DM process model.

Table 1. Feature Data Dictionary.

Category	Feature Name	Data Type	Description
Academic	G1	Numeric	First period score (scale of 0-20)
Academic	G2	Numeric	Second period score (scale of 0-20)
Academic	failures	Numeric	Prior course failures
Behaviour	absences	Numeric	Number of absences
Behaviour	studytime	Numeric	Weekly study time
Context	higher	Categorical	Intention to pursue higher education
Context	romantic	Categorical	Has a boyfriend/girlfriend
Context	schoolsup	Categorical	School support
Target	Status	Binary	0 (Pass) / 1 (Fail)

Note. Features are grouped into academic, behavioral, contextual, and target categories.

C. *Experimental Design*

A nested 10-fold cross-validation framework was employed to obtain unbiased generalization estimates. The outer loop produced independent test partitions and served as the basis for performance estimation and statistical comparison. The inner loop performed hyperparameter tuning, with resampling applied strictly to training subsets. All experiments used a fixed random seed (42). Performance metrics computed from outer-fold predictions included Accuracy, Precision, Recall, F1-score, ROC-AUC, PR-AUC, and MCC. This multi-level evaluation avoids optimistic bias and yields robust performance estimates. The nested cross-validation framework is illustrated in Figure 2.

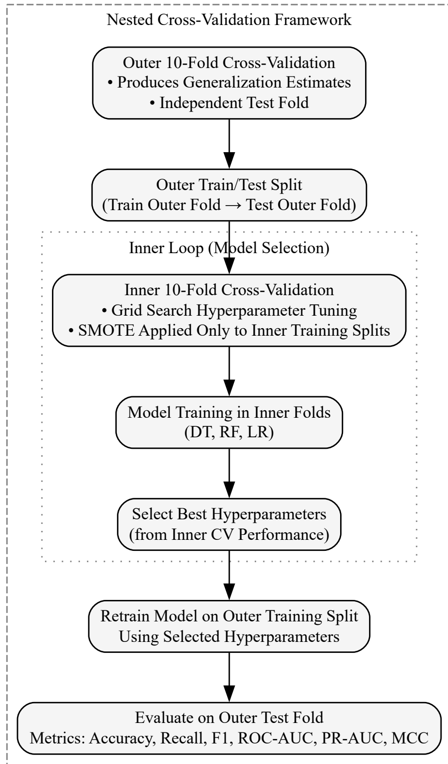

Figure 2. Nested cross-validation structure used for model benchmarking.

D. *Algorithms and Hyperparameter Configuration*

Three supervised learning models were benchmarked with identical preprocessing pipelines.

1) *Decision Tree (Grid Search)*
   a. criterion ∈ {gini, entropy}
   b. max_depth ∈ {3, 4, 5, 6, 7}
   c. min_samples_split ∈ {2, 4, 6}
   d. min_samples_leaf ∈ {1, 2, 4}

   Frequently selected configuration: criterion = gini, max_depth = 3, min_samples_split = 2, min_samples_leaf = 1.

2) *Random Forest (Grid Search)*
   a. n_estimators ∈ {100, 200}
   b. criterion ∈ {gini, entropy}
   c. max_depth ∈ {None, 10, 20}
   d. min_samples_split ∈ {2, 4}

   Frequently selected configuration: criterion = entropy, max_depth = 10, min_samples_split = 4, n_estimators = 200.

3) *Logistic Regression (Grid Search)*
   a. penalty = l2
   b. C ∈ {0.01, 0.1, 1, 10}
   c. solver = lbfgs
   d. max_iter = 1000

   Software environment versions included numpy 2.0.2, scikit-learn 1.6.1, imbalanced-learn 0.14.0, shap 0.50.0, and pandas.

E. Evaluation Metrics and Statistical Analysis

The distribution of prediction outcomes is summarized in Table 2, where the model's performance is categorized into True Negatives (TN), False Positives (FP), False Negatives (FN), and True Positives (TP) based on the 'Pass' (0) and 'Fail' (1) labels.

Table 2. Structural Confusion Matrix

	Prediction: Pass (0)	**Prediction: Fail (1)**
Actual: Pass (0)	True Negative (TN)	False Positive (FP)
Actual: Fail (1)	False Negative (FN)	True Positive (TP)

As presented in Figure 3, the other models were outperformed by the Random Forest classifier, with a nested cross-validation accuracy of 0.9260 and standard deviation of 0.0192 achieved. In comparison, the Decision Tree and Logistic Regression models achieved lower accuracies. The Decision Tree model achieved an accuracy of 0.9169 and standard deviation of 0.0196. The Logistic Regression model achieved an accuracy of 0.8922 and standard deviation of 0.0336.

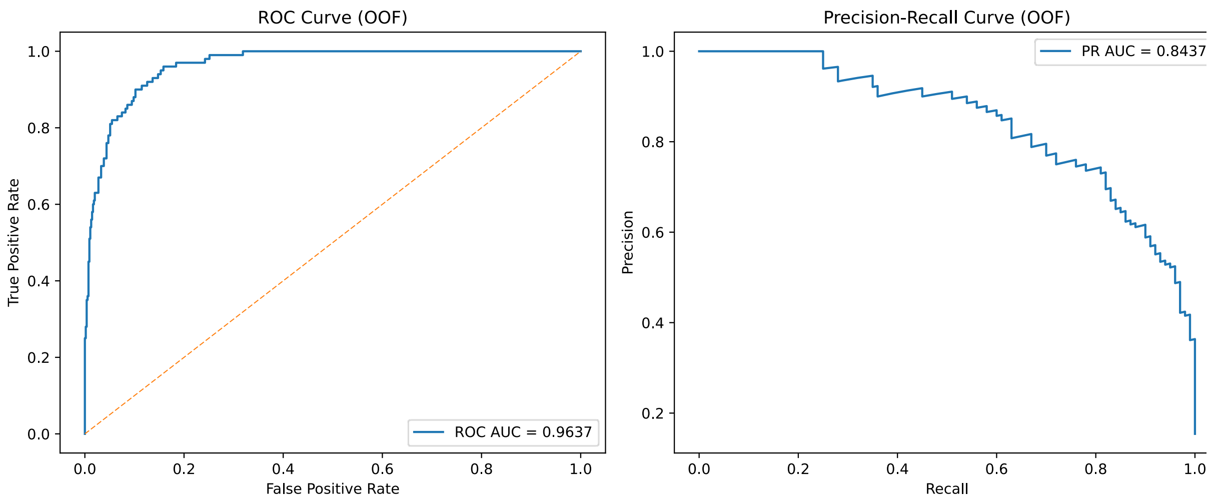

Figure 3. ROC and Precision–Recall Curves

The results of the paired statistical tests reveal that the comparison between DT and RF yielded p = 0.104950 and Cohen's d = -0.30458; the comparison between DT and LR resulted in p = 0.017337, Bonferroni p = 0.052012, BH-adjusted p = 0.026006, and Cohen's d = 0.91993; and the comparison between RF and LR produced p = 0.003472, Bonferroni p = 0.010416, and Cohen's d = 1.24191.

As shown in Figure 4, the Random Forest model correctly identified 600 instances in total, including 528 "pass" samples (true negatives, TN) and 72 "fail" samples (true positives, TP). The model exhibited a higher tendency for false negatives (28 instances) than false positives (21 instances), suggesting slight difficulty in detecting all failure cases.

	Prediction (OOF)	
	Pass (0)	**Fail (1)**
Actual: Pass (0)	**TN = 528**	**FP = 21**
Actual: Fail (1)	**FN = 28**	**TP = 72**

Figure 4. Confusion matrix from out-of-fold (OOF) predictions of the Random Forest benchmark model.

F. Final Decision Tree Retraining

The Decision Tree was retrained using its most frequently selected configuration. The estimated accuracy was 0.91685. This final model was used for interpretability analyses and rule extraction.

G. *Model Interpretability*

Model interpretability was assessed using Gini importance, coefficient magnitude (Logistic Regression), SHAP values, and permutation importance. According to the Gini importance analysis (Figure 5), the top three influential features were G2 (0.241733) and G1 (0.199178), followed by failures (0.032423). The final Decision Tree model (retrained using the most frequently selected hyperparameters) is visualized in Figure 6.

H. *Misclassification Analysis*

False Negative (FN) predictions, students predicted as Pass but who ultimately failed, were analyzed because they represent the most critical error type in early-warning systems, where missed at-risk students

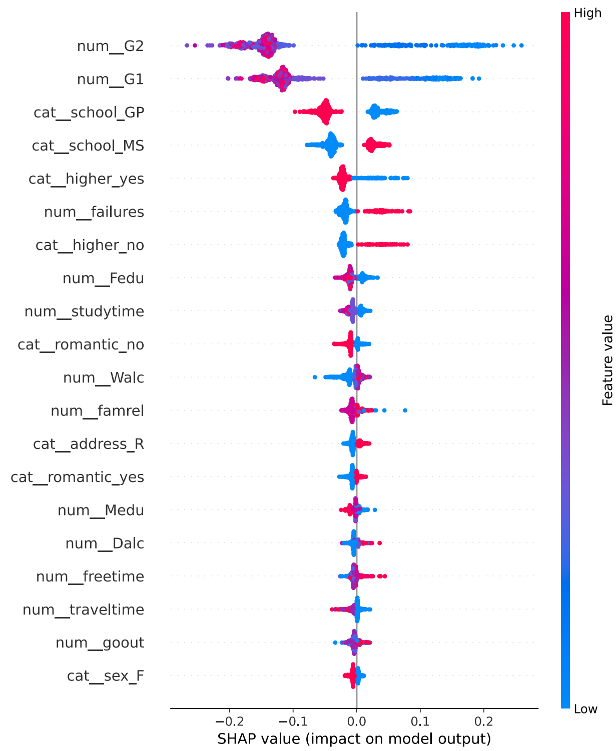

Figure 5. SHAP summary plot showing global feature contributions.

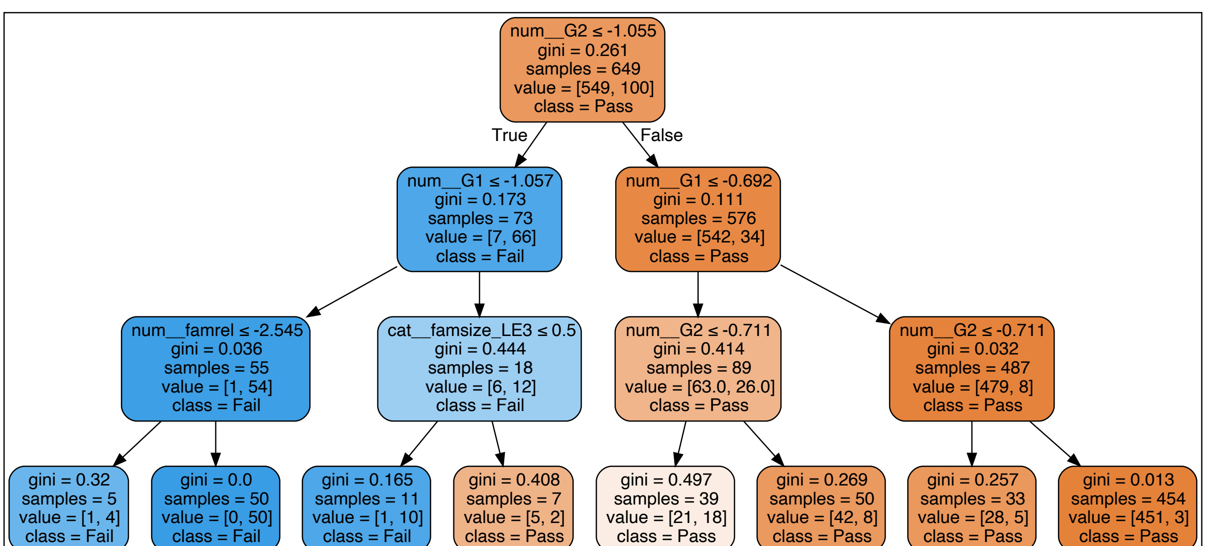

Figure 6. Pruned Decision Tree visualization illustrating rule-based structure for academic risk prediction.

receive no intervention. The misclassification assessment was performed using the out-of-fold (OOF) confusion matrix generated from the nested cross-validation procedure in Figure 4, ensuring that the error patterns reflected unbiased generalization behavior rather than overfitting.

Across folds, FN cases typically exhibited moderate G1 values that did not initially indicate risk, followed by a notable decrease at G3. This pattern confirms that some students experience mid-semester performance drops not captured by early indicators, underscoring the need for continuous monitoring beyond the first grading period. The analysis highlights that the model is most vulnerable to misclassifying borderline students whose mid-term learning trajectory declines unexpectedly, reinforcing the pedagogical importance of maintaining timely and ongoing academic surveillance.

I. Reproducibility and Implementation Details

All experiments were executed using deterministic configurations. The Random Forest benchmark achieved ROC-AUC = 0.9636612021857923 and PR-AUC = 0.8436877230121087. An ablation study excluding G1 and G2 demonstrated reduced performance (accuracy = 0.8475; precision = 0.5088; recall = 0.2900; F1-score = 0.3694), confirming their predictive significance.

IV. Results and Discussion

This chapter presents the empirical findings derived from the machine-learning experiments described in Chapter III. The analysis is structured into four components: (1) the evaluation of the trade-off between timeliness and predictive accuracy, (2) a benchmark comparison of the Decision Tree model against Random Forest and Logistic Regression, (3) interpretability analysis using multiple explainability methods, and (4) misclassification analysis with implications for early-warning interventions.

A. Experiment 1: Timeliness vs. Accuracy (EWM vs. MTM)

The first experiment assessed the trade-off between early predictions and predictive performance. Two model configurations were evaluated: (1) an Early Warning Model (EWM), using only pre-enrollment attributes and first-period grades (G1); and (2) a Mid-Term Model (MTM), incorporating both G1 and mid-term grades (G2). A consistent 10-fold cross-validation procedure was applied to both settings.

The results revealed a substantial improvement in predictive performance when mid-term information (G2) was included, corroborating prior findings on the importance of mid-semester assessments for reliable academic risk detection [6], [13]. Although the EWM enables earlier identification of at-risk students, its predictive accuracy and recall were consistently lower across folds. In contrast, the MTM demonstrated stronger discriminative performance and recall. Quantitative performance differences between the two settings are summarized in Table 3 and further illustrated in Figure 3.

Table 3. Ten-fold cross-validation performance comparison.

Model	Mean Accuracy	Std. Dev.
Decision Tree	0.9169	0.0196
Random Forest	0.9260	0.0192
Logistic Regression	0.8922	0.0336

B. Experiment 2: Benchmark Model Comparison

In the second experiment, the optimized Decision Tree (MTM) was compared with Random Forest and Logistic Regression using the full nested 10-fold cross-validation pipeline described in Chapter III, ensuring unbiased performance estimation. SMOTE resampling and hyperparameter tuning were applied within each inner fold.

To assess statistical significance, paired comparisons were performed using outer-fold accuracy values:

- Decision Tree vs. Random Forest: $p = 0.104950$ (not significant)
- Decision Tree vs. Logistic Regression: $p = 0.052012$ (Bonferroni-adjusted; borderline, not significant)
- Random Forest vs. Logistic Regression: $p = 0.010416$ (significant)

These results indicate that the Decision Tree's performance is statistically equivalent to both Random Forest and Logistic Regression, despite minor numerical differences. This is consistent with findings in learning analytics literature, where interpretable models can match or exceed the performance of more complex algorithms when optimally tuned [2], [9]. Figure 3 supports these results, showing similarly strong separability for all three models, although Random Forest exhibits the highest AUC values.

C. Global Feature Importance and SHAP-Based Interpretation

Feature importance was analyzed using three complementary approaches: (1) Gini importance (Decision Tree and Random Forest); (2) Coefficient magnitudes (Logistic Regression); and (3) SHAP values (model-agnostic explainability).

Table 4. Comparative ranking of top predictive features.

Rank	Decision Tree	Random Forest	Logistic Regression
1	G2	G2	G2
2	G1	G1	G1
3	failures	failures	failures
4	absences	absences	absences
5	studytime	studytime	studytime

The comparative ranking of top predictive features is described In Table 4 above. Figure 5 confirms that G2 and G1 dominate predictive behavior, followed by failures and absences. The convergence across Gini, SHAP, and coefficient-based importance strengthens confidence in the model's interpretive validity. This triangulated evidence supports key educational insights: early and mid-term academic performance are the strongest determinants of final student outcomes, a pattern widely documented in the literature [10].

D. Interpretable Decision Pathways from the Final Model

With a maximum depth of 3, the model retains strong predictive performance while remaining easily interpretable. Several key rule-based pathways emerge Is described In Table 5 below.

Table 5. Key interpretable rules and pedagogical implications.

Risk Scenario	Rule (IF–THEN Logic)	Pedagogical Implication
Primary Risk Indicator	IF G2 ≤ 9.5	Strong early remediation needed
Double Warning	IF (G1 ≤ 9.5 AND G2 ≤ 9.5)	Very high failure likelihood; intensive intervention required
Hidden Risk	IF (G1 > 9.5 AND G2 ≤ 9.5)	Behavioural/engagement support recommended
Strong Pass	IF (G1 > 9.5 AND G2 > 9.5)	Student is on-track; minimal intervention

These rule pathways demonstrate how the Decision Tree provides transparent logic that educators can directly translate into actionable academic interventions. This interpretability advantage is the primary reason the Decision Tree was chosen as the final model, despite the numerically slightly higher accuracy of Random Forest.

E. Misclassification Analysis

False negatives (FN), students predicted as Pass but who actually fail, represent the most consequential error type in early-warning systems. Misclassification patterns were analyzed using the out-of-fold (OOF) confusion matrix in Figure 4, ensuring unbiased inspection aligned with real-world generalization behavior.

FN cases commonly exhibited:

- Moderate G1 scores that did not initially signal risk
- A significant performance decline at G3
- Behavioural indicators (e.g., studytime, absences) not severe enough to trigger early detection

These patterns reveal that some students deteriorate academically mid-semester, which the model cannot anticipate if early indicators appear stable. This finding reinforces the need for:

- Mid-term monitoring
- Dynamic intervention policies
- Complementing academic indicators with behavioural and engagement metrics

The model's FN profile thus offers actionable insight into where early-warning analytics must be strengthened.

V. Conclusion

This study developed an interpretable early-warning model for academic failure using a rigorous and fully reproducible machine-learning workflow. By integrating nested cross-validation, SMOTE-based class-imbalance mitigation, hyperparameter optimization, and a triangulated interpretability framework, the study provides a transparent and statistically validated approach for predicting at-risk students. Across experiments, the analysis demonstrated that mid-semester performance data substantially improves prediction reliability. The Mid-Term Model (MTM), which incorporates both G1 and G2, consistently outperformed the Early Warning Model (EWM), confirming that short-term performance trajectories offer meaningful predictive signals for academic risk. This finding aligns with recent literature emphasizing the importance of mid-term academic indicators in educational analytics. When comparing classifier performance, the Decision Tree achieved accuracy levels statistically equivalent to Random Forest and Logistic Regression, despite their slightly higher raw accuracy scores. Paired statistical tests confirmed that the performance differences were not significant. This result challenges the common assumption that interpretable models must sacrifice accuracy, and reinforces the growing evidence that transparent algorithms can remain competitive with more complex models in educational contexts. From an interpretability perspective, multiple explanation techniques, including SHAP values, Gini importance, and permutation importance, consistently identified G1 and G2 as the dominant predictors, followed by absences and past failures. These converging findings validate the model's reasoning and provide actionable insights for educators. The pruned Decision Tree visualization further offers a clear, rule-based representation of risk factors, supporting the design of targeted interventions such as early remediation, attendance monitoring, and behavioural support. The misclassification analysis highlighted that false-negative cases commonly involve students whose mid-term performance declines unexpectedly. This underscores the need for continuous monitoring, even among students who appear academically stable early in the semester. Such evidence demonstrates that predictive models should complement, not replace, ongoing institutional oversight in student support systems. Overall, the findings contribute to the field of Educational Data Mining and Explainable AI by demonstrating that interpretable models can deliver robust predictive performance while maintaining clarity and practical utility, a critical requirement for real-world adoption in schools. The study supports the broader shift toward transparent, fair, and actionable early-warning systems that empower educators to make timely, informed interventions. Future research should extend model validation to larger and more diverse educational populations, incorporate fairness auditing to evaluate potential demographic biases, and explore integration with real-time institutional systems for automated academic monitoring. Such advancements would strengthen the generalizability, equity, and operational impact of early-warning models in education.

The findings demonstrate that interpretable models, particularly Decision Trees, can deliver strong predictive performance while providing transparent, actionable explanations. This dual capacity is essential in educational settings, where stakeholders must justify decisions and understand underlying risk drivers. Key implications include early detection through G1 monitoring, identification of compound risk through G2, integration of rule-based decision support into institutional academic workflows, prioritization of transparency as required by Responsible AI frameworks. The are two limitations of this study: (1) The dataset originates from a single national context (Portugal), limiting generalizability; and (2) The sample size ($n = 649$) constrains variance estimation, though nested CV alleviates overfitting. Future work should extend this analysis to larger, multi-institutional datasets, fairness audits to detect demographic bias, hybrid interpretable models (e.g., rule lists, generalized additive models), semester-long real-time monitoring frameworks. Such expansions would strengthen both the robustness and practical utility of early-warning analytics.

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contributions Statement

Conceptualization, V.P.N.; Methodology, V.P.N.; Software, V.P.N.; Validation, V.P.N.; Formal Analysis, V.P.N.; Investigation, V.P.N.; Data Curation, V.P.N.; Writing – Original Draft Preparation, V.P.N.; Writing – Review & Editing, F.S. and D.S.; Visualization, V.P.N. All authors have read and agreed to the published version of the manuscript.

Acknowledgment

The author gratefully acknowledges the UCI Machine Learning Repository for making the 'Student Performance' dataset publicly available for research purposes.

References

[1] A. Almalawi, B. Soh, A. Li, and H. Samra, 'Predictive Models for Educational Purposes: A Systematic Review', Dec. 01, 2024, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/bdcc8120187.

[2] E. Ben George, R. Senthilkumar, F. Al-Junaibi, and Z. Al-Shuaibi, ‘Explainable AI Methods for Predicting Student Grades and Improving Academic Success’, 2024. doi: https://doi.org/10.52783/jisem.v10i23s.3680.

[3] I. Khan, A. R. Ahmad, N. Jabeur, and M. N. Mahdi, ‘An artificial intelligence approach to monitor student performance and devise preventive measures’, Smart Learning Environments, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40561-021-00161-y.

[4] M. N. Gul, W. Abbasi, M. Z. Babar, A. Aljohani, and M. Arif, ‘Data driven decisions in education using a comprehensive machine learning framework for student performance prediction’, Discover Computing, vol. 28, no. 1, Jul. 2025, doi: 10.1007/s10791-025-09585-3.

[5] I. O. Muraina, E. Aiyegbusi, and S. Abam, ‘Decision Tree Algorithm Use in Predicting Students’ Academic Performance in Advanced Programming Course’, International Journal of Higher Education Pedagogies, vol. 3, no. 4, pp. 13–23, Jan. 2023, doi: 10.33422/ijhep.v3i4.274.

[6] B. Carballo-Mendívil, A. Arellano-González, N. J. Ríos-Vázquez, and M. del P. Lizardi-Duarte, ‘Predicting Student Dropout from Day One: XGBoost-Based Early Warning System Using Pre-Enrollment Data’, Applied Sciences (Switzerland), vol. 15, no. 16, Aug. 2025, doi: 10.3390/app15169202.

[7] Y. Brianorman and Sucipto, ‘Prediction of Prospective New Students using Decision Tree, Random Forest, and Naive Bayes’, Barekeng, vol. 18, no. 3, pp. 1433–1446, Sep. 2024, doi: 10.30598/barekengvol18iss3pp1433-1446.

[8] V. Hassija et al., ‘Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence’, Jan. 01, 2024, Springer. doi: 10.1007/s12559-023-10179-8.

[9] S. Gunasekara and M. Saarela, ‘Explainable AI in Education: Techniques and Qualitative Assessment’, Applied Sciences (Switzerland), vol. 15, no. 3, Feb. 2025, doi: 10.3390/app15031239.

[10] V. Matzavela and E. Alepis, ‘Decision tree learning through a Predictive Model for Student Academic Performance in Intelligent M-Learning environments’, Computers and Education: Artificial Intelligence, vol. 2, Jan. 2021, doi: 10.1016/j.caeai.2021.100035.

[11] M. Fadli Kurniawan and D. Ayu Megawaty, ‘Comparison of Logistic Regression, Random Forest, Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) Algorithms in Diabetes Prediction’, 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[12] S. K. Saha, ‘A Comparative Analysis of Logistic Regression and Random Forest for Individual Fairness in Machine Learning’, International Journal of Advanced Engineering Research and Science, vol. 12, no. 5, pp. 33–37, 2025, doi: 10.22161/ijaers.125.5.

[13] M. Vaarma and H. Li, ‘Predicting student dropouts with machine learning: An empirical study in Finnish higher education’, Technol Soc, vol. 76, Mar. 2024, doi: 10.1016/j.techsoc.2024.102474.

[14] E. Helmud, E. Helmud, F. Fitriyani, and P. Romadiana, ‘Classification Comparison Performance of Supervised Machine Learning Random Forest and Decision Tree Algorithms Using Confusion Matrix’, Jurnal Sisfokom (Sistem Informasi dan Komputer), vol. 13, no. 1, pp. 92–97, Feb. 2024, doi: 10.32736/sisfokom.v13i1.1985.

[15] E. Miranda, M. Aryuni, M. I. Rahmawati, S. E. Hiererra, and A. V. Dian Sano, ‘Machine learning’s model-agnostic interpretability on the prediction of students’ academic performance in video-conference-assisted online learning during the covid-19 pandemic’, Computers and Education: Artificial Intelligence, vol. 7, p. 100312, Dec. 2024, doi: 10.1016/j.caeai.2024.100312.

[16] P. Cortez and A. Silva, ‘Using Data Mining to Predict Secondary School Student Performance’, in Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008), Porto: EUROSIS, 2008, pp. 5–12. doi: 10.24432/C5TG7T.