Interdisciplinary Journal of Pedagogy and Research in Media Technology
2026, Vol. 02, No. 01, 9-27
https://doi.org/10.64268/inspire.v2i1.117

Early-warning analytics with LLM intervention rationales for student retention decisions: Classroom interaction modeling with xAPI-edu-data and dropout/success prediction

Qi Xin*
University of Pittsburgh, PA, USA
***Corresponding author:** qix29@pitt.edu

ABSTRACT
Background: Student-retention early-warning systems have substantially improved predictive performance, yet their outputs often remain limited to risk scores that provide little guidance for educational intervention.
Aims: This study proposes and empirically evaluates a reproducible two-stage framework that integrates classroom interaction modeling, institutional retention prediction, and structured LLM-ready intervention rationales for student-retention decision support.
Methods: The framework was evaluated using two complementary benchmark datasets. xAPI-Edu-Data was used for classroom interaction modeling, whereas Predict Students' Dropout and Academic Success was used for institutional retention prediction. The datasets were analyzed independently rather than merged. Stratified train-test splits, ensemble machine learning models, feature ablation, feature-importance analysis, probability calibration, and rationale-quality evaluation were employed.
Result: LightGBM achieved the highest Macro-F1 (0.7775) on the xAPI-Edu-Data benchmark, while XGBoost produced the best overall performance on the institutional retention dataset, achieving a multiclass accuracy of 0.7672, a Macro-F1 of 0.6964, and a ROC-AUC of 0.8889. In binary dropout prediction, XGBoost achieved a ROC-AUC of 0.9360 and an Average Precision of 0.9101. Behavioral engagement, attendance, academic progression, and tuition-related variables consistently emerged as the most informative predictors. The structured rationale layer achieved complete evidence alignment, actionability, and monitoring specificity while generating 392 unique intervention rationales.
Conclusion: The proposed framework demonstrates that early-warning analytics can move beyond risk prediction by integrating predictive analytics, explainable AI, and structured intervention rationale generation into a transparent, evidence-grounded decision-support workflow for improving student-retention decisions in higher education.

ARTICLE HISTORY
Submitted: March 08, 2026
Accepted: April 23, 2026
Published: June 29, 2026

KEYWORD
Early-Warning Analytics;
Educational Decision Support;
Learning Analytics;
Student Retention;
Xapi.

To cite this article:
Qin, Q. (2026). Early-warning analytics with LLM intervention rationales for student retention decisions: Classroom interaction modeling with xAPI-edu-data and dropout/success prediction. *Interdisciplinary Journal of Pedagogy and Research in Media Technology*, *2*(1), 9-27.

INTRODUCTION

Student retention remains a central institutional problem because withdrawal decisions usually develop through accumulated academic, social, financial, and organizational pressures rather than through a single event. Retention theory links persistence to academic and social integration over time (Tinto, 1975, 1993), while recent learning-analytics reviews show that educational data can support early identification, targeted feedback, and intervention, even though implementation and sustained impact remain difficult (Ifenthaler & Yau, 2020; Wong & Li, 2020; de Oliveira et al., 2021; Rahmani et al., 2024). As higher education increasingly adopts data-driven decision support, the ability not only to predict student risk but also to justify intervention decisions in an understandable and actionable manner has become an equally important institutional challenge.

Early-warning research has shown that learning-management-system traces, course signals, and open analytics pipelines can identify at-risk students early enough for human action (Macfadyen & Dawson, 2010; Arnold & Pistilli, 2012; Jayaprakash et al., 2014). These studies collectively demonstrate substantial progress in predictive analytics for identifying students at risk before academic failure becomes irreversible. However, many systems still communicate risk as a score, alert color, or ranked list. Such outputs are useful for screening, but they do not by themselves tell teachers, tutors, or advisors why a student was flagged, what action should be taken, or how the case should be monitored after the first intervention. Consequently, improvements in predictive performance have not always translated into equally meaningful support for educational decision-making.

This translation problem is especially important in educational decision-making. General explanation techniques such as LIME and SHAP can identify influential features, but their raw outputs do not automatically become pedagogically meaningful support plans (Ribeiro et al., 2016; Lundberg & Lee, 2017). Recent work on explainable AI in education emphasizes that explanations should be designed around educational stakeholders, human-centred interfaces, and the risks of misinterpreting automated recommendations (Khosravi et al., 2022). Collectively, these studies suggest that explanation quality should not be judged solely by technical fidelity or feature attribution accuracy. Rather, explanations should enable educators to understand behavioral evidence, justify intervention priorities, and support subsequent monitoring decisions. This perspective shifts explainability from a purely algorithmic property toward an operational component of educational decision support.

Although explainable AI improves the transparency of predictive models, it rarely transforms quantitative explanations into coherent intervention narratives that educators can immediately interpret and apply. Large language models add a further opportunity because they can generate fluent, contextualized educational text. Recent studies show that LLMs can support feedback generation, teaching assistance, content generation, and educational recommendation, while also raising concerns about transparency, replicability, privacy, and ethical deployment in real learning contexts (Brown et al., 2020; OpenAI, 2023; Kasneci et al., 2023; Yan et al., 2024). A retention-support system therefore requires a design in which LLMs function as a communication and reasoning layer rather than as the predictive engine itself, thereby preserving a transparent and auditable relationship between behavioral evidence, prediction outputs, explanatory reasoning, and recommended interventions.

Public educational datasets also differ substantially in the aspects of student behavior they capture. Interaction-oriented datasets such as xAPI-Edu-Data primarily describe students' learning activities within digital classrooms, whereas institutional datasets such as Predict Students' Dropout and Academic Success incorporate demographic, academic progression, and enrollment characteristics. Rather than forcing heterogeneous datasets into an artificial record-level integration, these complementary sources provide an opportunity to examine whether evidence-grounded intervention rationales can remain consistent across different representations of student learning and institutional performance.

Existing work has established the feasibility of predicting student risk and has developed explanation methods for machine-learning models. However, these research streams have largely evolved independently. Prediction studies predominantly evaluate classification performance, while explainability studies emphasize feature attribution without operationalizing how explanations become actionable educational interventions. Fewer studies provide a reproducible pipeline that connects classroom interaction signals, institutional dropout/success prediction, and teacher-facing intervention rationales without relying on unsupported dataset fusion or uncontrolled free-form text generation. The literature therefore still lacks an operational framework that systematically converts predictive evidence into structured intervention rationales whose evidence alignment, actionability, auditability, and follow-up monitoring potential can be empirically evaluated.

The objective of this study is to develop and empirically evaluate a two-stage early-warning framework that links classroom interaction modeling and institutional dropout/success prediction with structured, evidence-grounded, LLM-ready intervention rationales. Specifically, the proposed framework treats predictive analytics as the evidence-generation layer and LLM-supported rationale construction as the communication layer for educational decision support. The study makes three contributions: (1) it evaluates xAPI-Edu-Data and Predict Students' Dropout and Academic Success as complementary public benchmarks for early-warning analytics; (2) it identifies behaviorally and institutionally meaningful risk signals without forcing an artificial record-level merge between heterogeneous datasets; and (3) it formalizes intervention rationales as a structured and measurable output layer whose evidence alignment, actionability, and monitoring specificity connect predictive analytics, explainable AI, and LLM-supported educational decision making.

METHOD

This study used a staged benchmark design rather than an artificial record-level merge. The two datasets came from different educational contexts, populations, feature schemas, and outcome definitions, so merging them would have produced synthetic student records without empirical meaning. xAPI-Edu-Data was used to model classroom interaction behavior, while Predict Students' Dropout and Academic Success was used to model institutional dropout and academic success. Figure 1 summarizes this design.

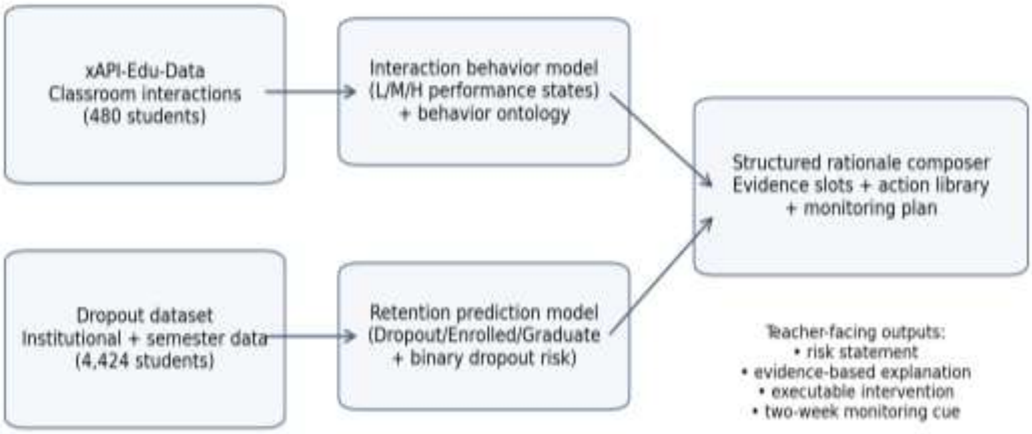

Figure 1. Two-stage early-warning framework linking behavior modeling, retention prediction, and structured intervention rationales

Table 1 summarizes the two benchmark datasets. xAPI-Edu-Data contained 480 records and 16 predictors after excluding the target column, with a three-class performance label coded as L, M, and H. Predict Students' Dropout and Academic Success contained 4,424 records and 36 predictors after excluding the target column, with institutional outcomes coded as Dropout, Enrolled, or Graduate. The two datasets were treated as complementary benchmarks rather than merged in Table 1.

Table 1. Overview of the two specified datasets used in the study.

Dataset	Instances	Predictors	Outcome	Primary signal
xAPI-Edu-Data	480	16	Class (L/M/H)	Classroom interaction and engagement traces
Predict Students' Dropout and Academic Success	4424	36	Dropout / Enrolled / Graduate	Enrollment, socioeconomic, and semester performance

Table 2 reports the full-sample class distributions. The xAPI outcome was moderately balanced, while the dropout/success benchmark was imbalanced toward Graduate and had a smaller Enrolled class. This structure justified macro-averaged and class-wise metrics rather than relying only on accuracy.

Table 2. Full-sample class distributions for the xAPI and dropout/success benchmarks.

Dataset	Class	Count	Share
xAPI	H	142	0.2958
xAPI	L	127	0.2646
xAPI	M	211	0.4396

Dataset	Class	Count	Share
Dropout	Dropout	1421	0.3212
Dropout	Enrolled	794	0.1795
Dropout	Graduate	2209	0.4993

All evaluations used stratified 80/20 train-test splits with random state 42. Preprocessing was fitted only on the training data and included imputation, one-hot encoding for categorical variables, and numeric standardization for linear and SVM models. The xAPI branch compared a dummy baseline, logistic regression, RBF-SVM, random forest, XGBoost, and LightGBM. The dropout/success branch compared the main linear and tree-based models for multiclass status prediction and binary dropout warning. Table 3 summarizes the fixed model families and hyperparameters used for reproducibility.

Table 3. Fixed model families and hyperparameters used in the benchmark comparisons.

Model	Key hyperparameters
Logistic Regression	max_iter=2000/3000; class_weight=balanced; solver=lbfgs
RBF-SVM	C=2.0; kernel=rbf; probability=True; class_weight=balanced
Random Forest	n_estimators=300; class_weight=balanced(_subsample)
XGBoost	n_estimators=220/260; max_depth=4/5; learning_rate=0.05; subsample=0.9; colsample_bytree=0.9
LightGBM	n_estimators=220/260; num_leaves=31; learning_rate=0.05; subsample=0.9; colsample_bytree=0.9; class_weight=balanced
Calibration model	Best binary model + isotonic calibration (cv=3)

Model performance was evaluated with accuracy, macro-F1, weighted-F1, one-vs-rest ROC-AUC, and log loss. The binary dropout task additionally used average precision, Brier score, expected calibration error, and recall@10%, recall@20%, and recall@30% to reflect early-warning triage settings. The xAPI branch also included an ablation test comparing the full feature set, a behavior/attendance/parent-engagement subset, and a background-only subset.

For explanation and rationale generation, encoded feature contributions were grouped back to original variables so that outputs referred to interpretable educational concepts rather than one-hot encoded fragments. The best binary dropout model supplied local evidence for the structured rationale layer. Each rationale contained four slots: estimated risk, model-derived evidence, teacher/advisor action, and monitoring cue. Rationale quality was evaluated by evidence alignment, actionability coverage, monitoring specificity, text length, and unique-text count.

RESULTS AND DISCUSSION

Result

The descriptive analysis was conducted to examine behavioral differences among students across the three observed performance classes before predictive modeling. The analysis focused on classroom interaction, attendance, parental involvement, and school satisfaction variables derived from the xAPI-Edu-Data dataset. Understanding these behavioral patterns provides an initial empirical basis for identifying variables that may contribute to subsequent dropout and academic success prediction. Table 4 summarizes the descriptive profile of these behavioral variables across the observed performance groups.

Table 4. Descriptive xAPI behavior profile by observed performance class.

Class	n	Raised hands mean	Visited resources mean	Announcements viewed mean	Discussion means	Above-7 absence rate	Parent survey yes rate	School satisfaction good rate
L	127	16.8898	18.3228	15.5748	30.8346	0.9134	0.2205	0.3386
M	211	48.9384	60.6351	40.9621	43.7915	0.3365	0.6066	0.6209
H	142	70.2887	78.7465	53.3803	53.6620	0.0282	0.8028	0.8310

Table 4 shows the xAPI descriptive profile by performance class. Low-performing students had substantially lower levels of raised-hand participation, resource visits, announcement views, parent survey responses, and school satisfaction, together with a markedly higher rate of students having more than seven absences. High-performing students exhibited the opposite pattern, suggesting that classroom interaction, attendance, and school engagement variables captured meaningful behavioral differences across performance groups. While these descriptive statistics demonstrate overall behavioral variation, they do not reveal how students are distributed within the multidimensional behavioral feature space. To complement the descriptive analysis, a dimensionality-reduction visualization was performed to examine the distribution of student behavioral patterns across the three performance classes. Figure 2 presents the PCA projection of the xAPI behavioral variables by performance class.

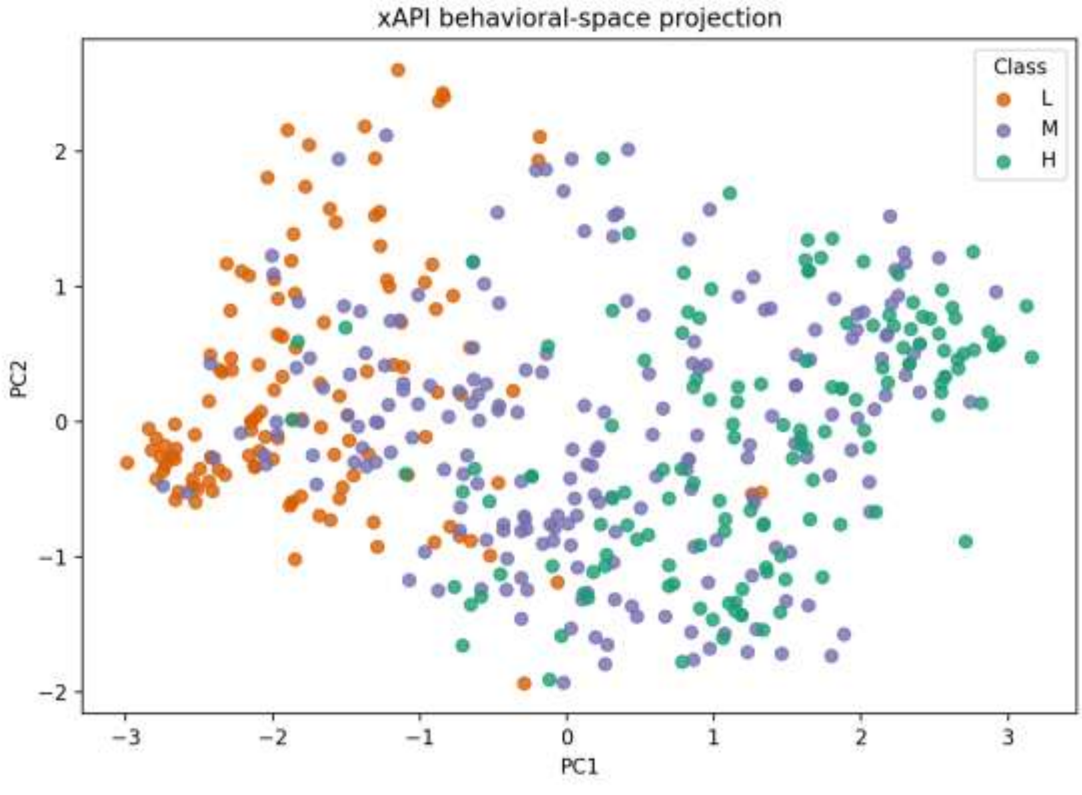

Figure 2. Principal-component projection of xAPI behavioral variables by performance class.

Figure 2 demonstrates that the three performance groups occupy distinguishable regions within the behavioral feature space after projection onto the first two principal components. Low-performing students are predominantly concentrated on the negative side of the first principal component (PC1), whereas high-performing students are more frequently distributed on the positive side, with medium-performing students occupying an intermediate region and partially overlapping both groups. This distribution indicates that classroom interaction variables capture meaningful behavioral variation associated with academic performance, despite the expected overlap arising from the complexity of student learning behaviors. The visualization is intended as an exploratory representation rather than a classification result, providing qualitative evidence that the selected behavioral indicators contain discriminative information suitable for subsequent predictive modeling. These findings complement the descriptive statistics in Table 4 by visually confirming that engagement and attendance patterns contribute to the separation of student performance profiles and provide empirical support for the development of the proposed early-warning prediction framework.

While the PCA visualization indicates that the behavioral variables contain meaningful discriminative information, predictive modeling is required to quantify how effectively these patterns distinguish student performance categories. Accordingly, multiple machine learning algorithms were evaluated using the xAPI-Edu-Data dataset. Table 5 reports the hold-out benchmark comparison on the xAPI-Edu-Data dataset.

Table 5. Held-out benchmark comparison on xAPI-Edu-Data.

Model	CV Acc	CV Macro-F1	Test Acc	Test Macro-F1	Test W-F1	Test AUC	LogLoss
LightGBM	0.7632	0.7705	0.7708	0.7775	0.7685	0.89	0.7333
Random Forest	0.7943	0.7983	0.7604	0.7675	0.7602	0.9181	0.5279
Logistic Regression	0.7371	0.7447	0.7292	0.7373	0.7262	0.8934	0.6093
XGBoost	0.7657	0.7728	0.7292	0.7348	0.7261	0.8856	0.627
RBF-SVM	0.7527	0.7584	0.6979	0.7132	0.697	0.884	0.5903
Dummy	0.4401	0.2037	0.4375	0.2029	0.2663	0.5	20.2746

Table 5 shows that all machine learning models substantially outperformed the dummy baseline, confirming that student performance classification was driven by meaningful behavioral patterns rather than class distribution alone. LightGBM achieved the highest test Macro-F1 score (0.7775), indicating the best overall balance in multiclass prediction, whereas Random Forest obtained the highest ROC-AUC (0.9181), demonstrating superior discrimination across performance classes. Logistic Regression and XGBoost produced competitive results with comparable accuracy and weighted F1 scores, while RBF-SVM showed relatively lower predictive performance, particularly in test accuracy and Macro-F1. The large performance gap between the evaluated models and the dummy classifier further suggests that classroom interaction variables extracted from the xAPI-Edu-Data dataset provide informative features for early-warning prediction, with ensemble tree-based methods offering the most robust overall performance.

Although Table 5 summarizes the numerical performance of the evaluated models, a visual comparison provides a clearer illustration of the relative differences in predictive performance across algorithms. Figure 3 presents the hold-out Macro-F1 comparison for the evaluated xAPI benchmark models.

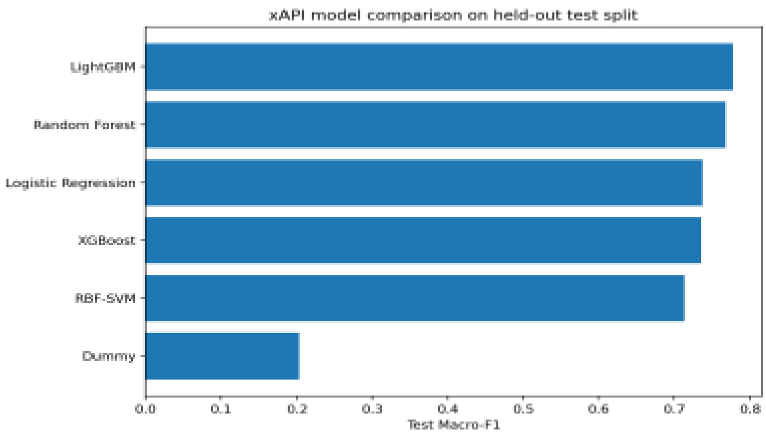

Figure 3. Held-out macro-F1 comparison for the xAPI benchmark models.

Figure 3 visually confirms the superiority of the tree-based ensemble models over the remaining algorithms in terms of hold-out Macro-F1 performance. LightGBM achieved the highest Macro-F1 score, followed closely by Random Forest, while Logistic Regression and XGBoost produced comparable intermediate performance. RBF-SVM yielded slightly lower predictive accuracy, whereas the dummy classifier performed substantially worse than all trained models. The relatively small performance differences among the leading machine learning algorithms indicate that the behavioral variables extracted from the xAPI-Edu-Data dataset consistently provide informative signals for student performance prediction, with ensemble methods offering the most balanced multiclass classification capability.

While the benchmark comparison identifies the most effective prediction algorithms, it does not indicate which categories of behavioral information contribute most to model performance. To address this question, an ablation study was conducted by progressively reducing the available feature sets. Table 6 reports the feature ablation study on the xAPI-Edu-Data dataset.

Table 6. xAPI ablation study comparing full, behavior-focused, and background-only feature sets.

Feature set	k	Accuracy	Macro-F1	AUC	LogLoss
Full feature set	16	0.7604	0.7675	0.9181	0.5279
Behavior/attendance/parent only	7	0.7292	0.7421	0.8763	0.6197
Background only	9	0.5625	0.5612	0.7413	1.3063

Table 6 demonstrates that predictive performance remained relatively high when the model was trained exclusively on behavioral, attendance, and parental engagement variables, achieving a Macro-F1 score of 0.7421 compared with 0.7675 for the complete feature set. In contrast, using only background variables resulted in a substantial decline in performance, reducing the Macro-F1 score to 0.5612 and the AUC to 0.7413. These findings indicate that observable classroom behaviors account for most of the predictive information required for student performance classification, whereas demographic and static institutional characteristics provide comparatively limited discriminative value. The ablation results therefore reinforce the central role of behavioral learning analytics in the proposed early-warning framework and demonstrate that accurate prediction can be achieved without relying heavily on background descriptors alone.

Having established that behavioral variables contribute most substantially to predictive performance, the next analysis identifies which individual features exert the greatest influence on the classification model. Table 7 reports the permutation importance of the xAPI variables obtained from the Random Forest model.

Table 7. Top xAPI variables by permutation importance in the Random Forest model.

Feature	Perm. importance	SD
StudentAbsenceDays	0.2291	0.0364
VisITedResources	0.0292	0.0159
Relation	0.0275	0.0185
gender	0.0197	0.0168
raisedhands	0.0195	0.0204
ParentAnsweringSurvey	0.019	0.0215
Topic	0.0148	0.008
AnnouncementsView	0.004	0.0151
GradeID	0.0026	0.01
PlaceofBirth	0.0015	0.0138

Table 7 indicates that StudentAbsenceDays was by far the most influential predictor, with a permutation importance (0.2291) substantially exceeding that of all other variables. The next most important predictors included VisitedResources, Relation, gender, raisedhands, and

ParentAnsweringSurvey, while variables such as AnnouncementsView, GradeID, and PlaceofBirth contributed only marginally to predictive performance. This distribution suggests that regular attendance and active classroom engagement provide stronger evidence of student performance than relatively static demographic characteristics. The findings are consistent with the earlier ablation analysis, reinforcing that observable learning behaviors constitute the primary source of predictive information within the proposed early-warning framework. While Table 7 summarizes the numerical importance of individual predictors, a graphical representation provides a clearer comparison of their relative contributions to the classification model. Figure 4 presents the permutation importance ranking of the ten most influential xAPI variables.

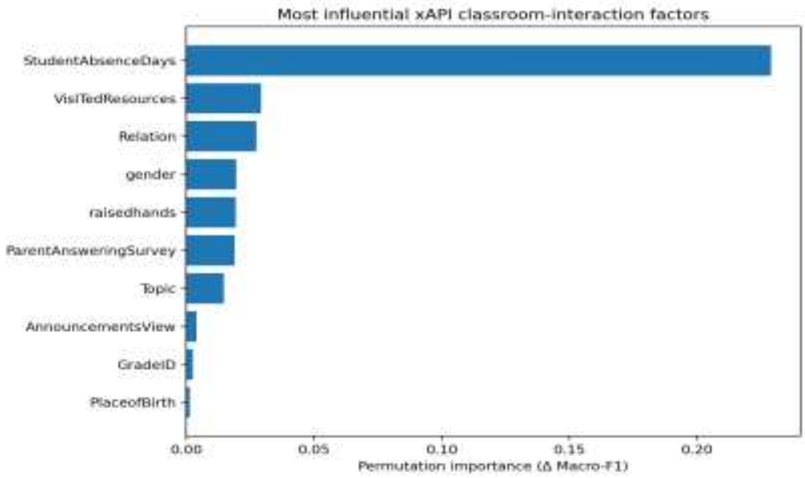

Figure 4. Ten most influential xAPI variables by permutation importance.

Figure 4 visually reinforces the dominance of StudentAbsenceDays over all other predictors, highlighting attendance as the most critical behavioral indicator for distinguishing student performance. A noticeable decline in importance is observed after the leading feature, with VisitedResources, Relation, gender, raisedhands, and ParentAnsweringSurvey contributing moderate predictive value, whereas the remaining variables exhibit comparatively small effects. The steep importance gradient indicates that a limited number of behavioral indicators account for most of the model's predictive capability, suggesting that effective early-warning systems may be developed using a compact yet highly informative subset of classroom interaction features. This ranking also provides an evidence-based foundation for constructing targeted intervention rationales in the subsequent LLM-supported decision-support stage.

While the preceding analyses focused on behavioral interaction data from the xAPI-Edu-Data benchmark, the proposed framework was further evaluated using an institutional dataset to examine its applicability across a broader student retention context. Table 8 reports the benchmark comparison on the *Predict Students' Dropout and Academic Success* dataset.

Table 8. Held-out multiclass comparison on Predict Students' Dropout and Academic Success.

Model	CV Acc	CV Macro-F1	Test Acc	Test Macro-F1	Test W-F1	Test AUC	LogLoss
XGBoost	0.7796	0.7068	0.7672	0.6964	0.7578	0.8889	0.5644
LightGBM	0.7703	0.7128	0.7424	0.6931	0.7479	0.8839	0.6007
Logistic Regression	0.7564	0.7131	0.7254	0.6885	0.7384	0.8787	0.6433
Random Forest	0.772	0.6822	0.7616	0.6681	0.7411	0.8868	0.598
Dummy	0.4993	0.222	0.4994	0.2221	0.3327	0.5	18.0422

Table 8 demonstrates that all evaluated machine learning models substantially outperformed the dummy baseline, confirming that institutional student records contain meaningful predictive information for multiclass dropout and academic success classification. XGBoost achieved the highest overall performance, with a test accuracy of 0.7672, a Macro-F1 score of 0.6964, and the highest ROC-AUC of

0.8889, indicating the most balanced discrimination across the three outcome classes. LightGBM produced comparable results, whereas Logistic Regression and Random Forest showed slightly lower but still competitive predictive performance. The consistent superiority of the ensemble tree-based models across both benchmark datasets suggests that they effectively capture the complex relationships between institutional characteristics and student academic outcomes.

Although the overall classification performance demonstrates the effectiveness of the prediction models, class-level evaluation is required to determine how accurately each student outcome category is identified. Table 9 presents the class-wise precision, recall, and F1 scores for the best-performing multiclass prediction model.

Table 9. Class-wise precision, recall, and F1 for the best multiclass dropout/success model.

Class	Precision	Recall	F1	Support
Dropout	0.8016	0.7254	0.7616	284.0
Enrolled	0.5366	0.4151	0.4681	159.0
Graduate	0.8059	0.9208	0.8596	442.0

Table 9 indicates considerable variation in predictive performance across the three outcome classes. The Graduate class achieved the highest recall (0.9208) and F1 score (0.8596), indicating that successful students were identified with high reliability. The Dropout class also exhibited strong classification performance, with an F1 score of 0.7616 and balanced precision and recall values. In contrast, the Enrolled class produced substantially lower precision (0.5366), recall (0.4151), and F1 score (0.4681), suggesting that students who remained enrolled represented the most difficult group to distinguish. This result is expected because the enrolled category reflects an intermediate academic status that shares characteristics with both dropout and graduate outcomes, thereby increasing classification ambiguity. Nevertheless, the model maintained robust performance for the two institutionally most critical outcomes, namely identifying students at risk of dropping out and those likely to graduate successfully.

While the class-wise evaluation summarizes predictive performance for each outcome category, it does not reveal the specific misclassification patterns made by the model. To better understand these prediction errors, the confusion matrix of the best-performing multiclass model was examined. Figure 5 presents the confusion matrix of the best-performing multiclass dropout/success prediction model.

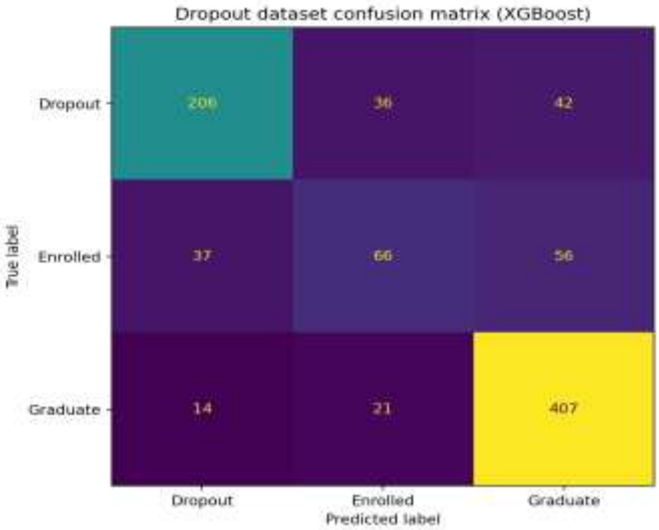

Figure 5. Confusion matrix for the best multiclass dropout/success model.

Figure 5 sows that the model classified the Graduate and Dropout classes with relatively high accuracy, as indicated by the large number of correctly predicted instances along the main diagonal. In contrast, the Enrolled class exhibited substantially more misclassifications, with many cases being predicted as either Dropout or Graduate. This pattern is consistent with the class-wise evaluation presented in Table 9 and reflects the transitional nature of enrolled students, whose academic profiles often overlap with both successful and at-risk learners. The confusion matrix therefore indicates that the proposed model effectively distinguishes the two institutionally most critical outcomes while identifying the intermediate enrolled status as the primary source of classification uncertainty. Because many institutional early-warning systems ultimately focus on identifying students at risk of dropping out, an additional binary classification analysis was conducted to evaluate the proposed framework under a practical early-warning scenario. Table 10 reports the binary dropout early-warning performance across the evaluated machine learning models.

Table 10. Binary dropout early-warning comparison across model families.

Model	Accuracy	F1	ROC-AUC	AP	Brier	R@10%	R@20%	R@30%	ECE
XGBoost	0.8983	0.8295	0.936	0.9101	0.0823	0.3134	0.5986	0.7887	0.0257
LightGBM	0.8791	0.812	0.933	0.9036	0.0884	0.3134	0.588	0.7887	0.0372
Logistic Regression	0.8802	0.8185	0.9317	0.902	0.0927	0.3099	0.588	0.7887	0.0698
Random Forest	0.8814	0.7977	0.9257	0.8948	0.0933	0.3134	0.588	0.7746	0.0541

Table 10 demonstrates that XGBoost achieved the strongest overall performance across the binary early-warning evaluation, obtaining the highest accuracy (0.8983), F1 score (0.8295), ROC-AUC (0.9360), and Average Precision (0.9101), while also maintaining the lowest Expected Calibration Error (0.0257). LightGBM produced highly competitive results, whereas Logistic Regression and Random Forest showed slightly lower predictive and calibration performance. Importantly, the recall values indicate that approximately 78.9% of at-risk students could be identified within the highest-risk 30% of predictions, highlighting the practical usefulness of the proposed framework for prioritizing early intervention. These findings demonstrate that ensemble tree-based models not only provide strong predictive discrimination but also produce well-calibrated risk estimates suitable for operational early-warning decision support.

Although the binary evaluation demonstrates strong predictive performance, effective early-warning systems also require reliable probability estimates to support risk-based intervention decisions. Therefore, the calibration characteristics of the best-performing binary model were further examined. Figure 6 presents the ROC and precision–recall curves for the best binary dropout prediction model before and after isotonic calibration.

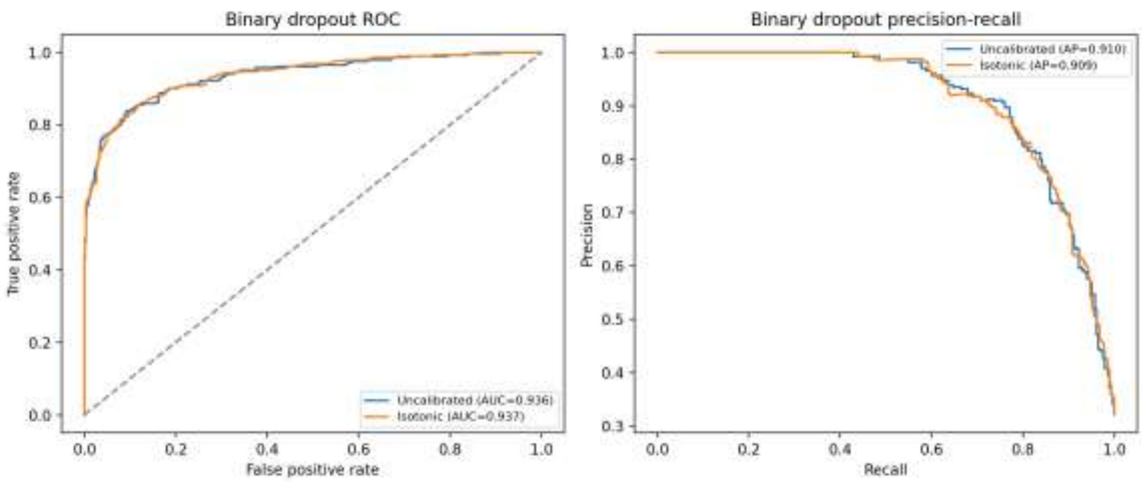

Figure 6. ROC and precision-recall curves for the best binary dropout model before and after isotonic calibration.

Figure 6 shows that isotonic calibration produced only minor changes in both the ROC and precision–recall curves, indicating that the model maintained strong discriminative performance after probability calibration. The nearly overlapping curves suggest that calibration improved the reliability of predicted probabilities without substantially altering the model's ability to distinguish between dropout and non-dropout students. This result demonstrates that the predictive performance remained stable while producing probability estimates that are more appropriate for risk-based decision making, thereby strengthening the suitability of the proposed framework for operational early-warning applications. To quantify the impact of probability calibration on predictive performance and early-warning effectiveness, the calibrated and uncalibrated models were compared across multiple evaluation metrics. Table 11 reports the calibration comparison for the best binary dropout prediction model.

Table 11. Calibration comparison for the best binary dropout model.

Version	Accuracy	F1	ROC-AUC	AP	Brier	ECE	R@10%	R@20%	R@30%
Uncalibrated	0.8983	0.8295	0.936	0.9101	0.0823	0.0257	0.3134	0.5986	0.7887
Isotonic	0.8859	0.8069	0.9371	0.9093	0.0835	0.0276	0.3134	0.5986	0.7923

Table 11 indicates that isotonic calibration preserved the overall predictive performance of the model while producing only marginal changes in discrimination metrics. Accuracy, F1 score, ROC-AUC, and Average Precision remained highly comparable between the calibrated and uncalibrated versions, whereas the Expected Calibration Error (ECE) increased only slightly from 0.0257 to 0.0276. Importantly, the recall values at the top 10% and 20% highest-risk predictions remained unchanged, and the recall at the 30% threshold increased slightly from 0.7887 to 0.7923, indicating that calibration did not compromise the model's operational ability to prioritize at-risk students. These findings suggest that isotonic calibration enhances the interpretability of predicted probabilities while maintaining the strong ranking performance required for practical early-warning intervention systems.

Although calibration confirms that the predicted probabilities are reliable for operational use, effective early-warning systems also require transparent explanations of the factors driving individual predictions. Accordingly, the contribution of each predictor was examined using SHAP-based model interpretation. Figure 7 presents the reliability curves for the best binary dropout prediction model before and after isotonic calibration.

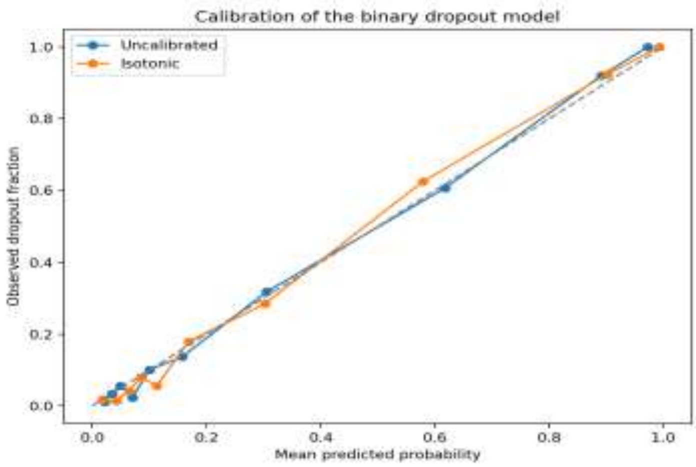

Figure 7. Reliability curve for the best binary dropout model before and after isotonic calibration.

Figure 7 demonstrates that both the calibrated and uncalibrated models closely follow the diagonal reference line, indicating good agreement between predicted probabilities and observed dropout frequencies. The isotonic-calibrated model exhibits slightly improved alignment across the probability range, particularly in the mid-probability region where accurate risk estimation is most critical for intervention planning. The close correspondence between the two curves confirms that probability calibration enhances the reliability of risk estimates without introducing substantial distortion to the model's predictive behavior. This provides additional support for using calibrated probabilities as the basis for communicating individualized dropout risk in practical early-warning systems. Having established that the predicted probabilities are both accurate and well calibrated, the next analysis identifies the individual predictors that contribute most strongly to dropout risk within the best-performing model. Table 12 reports the most influential predictors of binary dropout risk based on mean absolute SHAP values.

Table 12. Top grouped predictors of binary dropout risk by mean absolute SHAP value.

Feature	Mean |SHAP|
Curricular units 2nd sem (approved)	1.2023
Tuition fees up to date	0.4405
Course	0.3318
Curricular units 1st sem (approved)	0.29
Age at enrollment	0.2643
Curricular units 2nd sem (grade)	0.1665
Scholarship holder	0.1461
Mother's qualification	0.144
Mother's occupation	0.1345
Curricular units 2nd sem (enrolled)	0.1324
Father's occupation	0.1244
Unemployment rate	0.1226

Table 12 indicates that academic progression variables were the dominant determinants of dropout risk. The number of approved curricular units in the second semester emerged as the most influential predictor by a substantial margin, followed by tuition fees up to date, course, approved curricular units in the first semester, and age at enrollment. Additional factors, including second-semester grades, scholarship status, parental education and occupation, and regional unemployment rate, contributed comparatively smaller but still meaningful effects. Overall, the SHAP ranking suggests that dropout risk is driven primarily by students' academic progress and institutional engagement rather than demographic background alone. These findings provide an interpretable evidence base for constructing individualized intervention rationales in the subsequent LLM-supported decision-support component of the proposed early-warning framework. While the SHAP analysis identifies the most influential predictors of dropout risk, numerical feature importance alone does not provide educators with actionable guidance. Figure 8 presents the global SHAP importance of the most influential predictors in the best binary dropout prediction model.

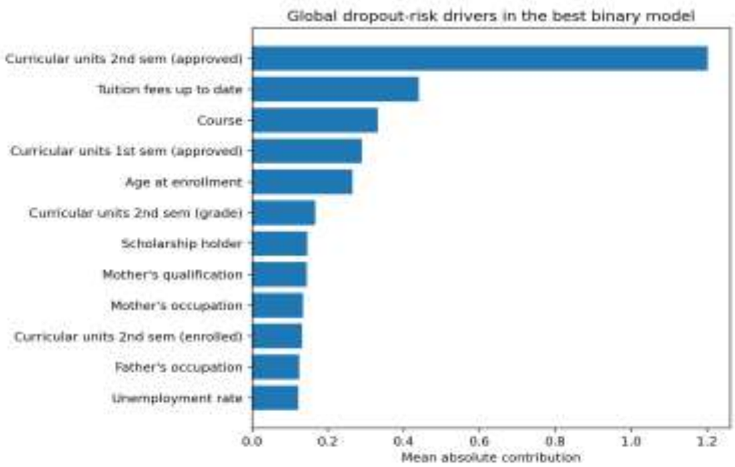

Figure 8. Global dropout-risk drivers in the best binary model.

Figure 8 visually confirms that academic progression variables dominate the global prediction of dropout risk. Approved curricular units in the second semester contributed substantially more than any other feature, followed by tuition payment status, course, approved first-semester curricular units, and age at enrollment. The remaining predictors exhibited progressively smaller contributions, indicating that the model relies primarily on students' academic progress and institutional engagement rather than demographic characteristics. This global importance pattern complements the SHAP ranking presented in Table 12 and establishes an interpretable evidence layer that can be transformed into structured intervention rationales for educators. Although these predictors explain why the model identifies students as being at risk, practical early-warning systems also require explanations that are understandable, actionable, and sufficiently specific to guide follow-up interventions. Table 13 reports the comparative evaluation of feature-list, generic, and structured intervention rationales.

Table 13. Comparative evaluation of feature-list, generic, and structured intervention rationales.

Method	Evidence alignment	Actionability coverage	Monitoring specificity	Average words	Unique texts
Feature list only	1.0	0.0	0.0294	20.0724	224
Generic teacher advice	0.0	1.0	1.0	12.0	1
Structured rationale	1.0	1.0	1.0	81.6855	392

Table 13 demonstrates the clear advantage of the proposed structured rationale approach over the two baseline explanation strategies. Although the feature-list baseline achieved perfect evidence alignment, it failed to provide actionable recommendations or monitoring guidance, limiting its practical value for educational intervention. Conversely, the generic teacher advice baseline produced actionable recommendations but lacked evidence alignment and generated only a single generic response for all students, preventing individualized decision support. The proposed structured rationale simultaneously achieved perfect evidence alignment, complete actionability, and full monitoring specificity while producing 392 unique intervention texts for 392 high-risk students, demonstrating its ability to generate personalized, evidence-grounded recommendations. These findings indicate that the proposed rationale layer effectively bridges predictive analytics and educational decision making by transforming model explanations into interventions that are both interpretable and operationally useful.

Having established the effectiveness of the structured rationale generation approach, the next analysis examines the distribution of intervention recommendations assigned to high-risk students. This provides additional insight into the types of educational support most frequently suggested by the proposed decision-support framework. Figure 9 presents the distribution of primary intervention categories generated for high-risk students.

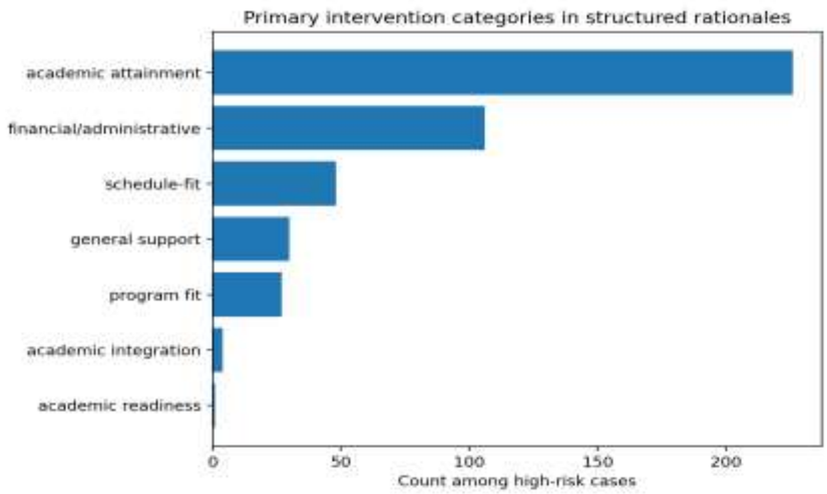

Figure 9. Distribution of primary intervention categories among high-risk cases.

Figure 9 shows that academic attainment was the dominant intervention category, accounting for the majority of recommendations generated for high-risk students. This was followed by financial or administrative support, schedule-fit adjustments, general academic support, and program-fit guidance, whereas recommendations related to academic integration and academic readiness occurred only infrequently. The distribution indicates that the proposed framework primarily associates dropout risk with deficiencies in academic progression while recognizing that financial constraints, scheduling issues, and program suitability also contribute to student persistence. Rather than producing generic advice, the structured rationale layer prioritizes intervention categories according to each student's evidence profile, thereby generating recommendations that are both evidence-grounded and contextually relevant for educational decision making. Table 14 provides representative examples of teacher-facing intervention rationales generated for high-risk students.

Table 14. Example teacher-facing rationales for high-risk students from the binary dropout test set.

Student id	Estimated dropout probability	Observed dropout	Top evidence features	Teacher-facing rationale
860	1.000	1	Curricular unit 2nd sem (approved); Tuition fees up to date; Age at enrollment	The student is in a high-risk segment (estimated dropout probability 1.000) because few second-semester units were approved, tuition status was not up to date, and age at enrollment was higher than most peers. The teacher/advisor should set a two-week completion plan and refer the student to billing or financial-aid support. Monitor approved units, completed evaluations, and support follow-through over the next two weeks.
90	1.000	1	Tuition fees up to date; Curricular units 2nd sem (approved); Age at enrollment	The student is in a high-risk segment (estimated dropout probability 1.000) because tuition status was not up to date, few second-semester units were approved, and age at enrollment was higher than most peers. The teacher/advisor should confirm a payment or waiver plan and require a two-week completion plan for pending assessments. Monitor support follow-through and new approved units.

Figure 9 shows that academic attainment was the dominant intervention category, accounting for the majority of recommendations generated for high-risk students. This was followed by financial or administrative support, schedule-fit adjustments, general academic support, and program-fit guidance, whereas recommendations related to academic integration and academic readiness occurred only infrequently. The distribution indicates that the proposed framework primarily associates dropout risk with deficiencies in academic progression while recognizing that financial constraints, scheduling issues, and program suitability also contribute to student persistence. Rather than producing generic advice, the structured rationale layer prioritizes intervention categories according to each student's evidence profile, thereby generating recommendations that are both evidence-grounded and contextually relevant for educational decision making. Table 14 provides representative examples of teacher-facing intervention rationales generated for high-risk students.

Discussion

The findings demonstrate that the proposed framework successfully integrates classroom interaction modeling, institutional retention prediction, and structured intervention rationale generation into a unified early-warning analytics framework. Rather than treating behavioral analysis, predictive modeling, explainability, and intervention planning as independent analytical processes, the proposed approach connects these components into a coherent workflow that supports educational decision making. The results further show that classroom interaction variables derived from the xAPI-Edu-Data benchmark capture immediate behavioral signals associated with student engagement, whereas the Predict Students' Dropout and Academic Success dataset contributes complementary institutional evidence reflecting longer-term academic progression and retention risk. This layered interpretation expands conventional early-warning analytics by combining short-cycle behavioral indicators with longer-cycle institutional characteristics within a single analytical framework. Such findings are consistent with retention theory, which argues that student persistence develops through continuous academic and social integration rather than isolated events (Tinto, 1975, 1993). They also reinforce recent learning-analytics research emphasizing that effective retention systems should support not only early prediction but also timely and actionable educational interventions (Ifenthaler & Yau, 2020; Wong & Li, 2020; de Oliveira et al., 2021; Rahmani et al., 2024).

The behavioral evidence identified across both benchmark datasets highlights the central role of observable learning engagement in explaining student success and retention. Attendance, classroom participation, learning-resource utilization, and curricular-unit completion consistently emerged as the strongest predictors of student outcomes, whereas demographic variables contributed comparatively little to predictive performance. This finding suggests that student persistence is more accurately reflected through continuously evolving learning behaviors than through relatively static background characteristics. From an educational perspective, behavioral engagement functions as an early manifestation of academic integration, enabling institutions to detect emerging learning difficulties before they develop into long-term retention problems. The consistency of these behavioral patterns across descriptive analysis, feature ablation, and feature-importance evaluation further indicates that the identified predictors represent stable educational signals rather than artifacts specific to a single dataset. Consequently, the proposed framework demonstrates that meaningful behavioral analytics can provide a reliable empirical foundation for evidence-based student retention strategies.

The comparative evaluation of machine learning models also provides important methodological insights into educational prediction. XGBoost achieved the strongest performance on the institutional retention dataset, whereas LightGBM and Random Forest produced highly competitive results on the xAPI-Edu-Data benchmark. These differences should not be interpreted as evidence of universal model superiority but rather as an indication that different educational datasets exhibit distinct structural characteristics requiring different modeling capabilities. Ensemble tree-based algorithms are particularly effective for heterogeneous educational tabular data because they naturally capture nonlinear relationships, interaction effects, and threshold-based behavioral patterns without extensive manual feature engineering. The consistent performance of these models across two independent benchmark datasets therefore suggests that predictive success depends not only on algorithm selection but also on the educational relevance of the behavioral and institutional variables being modeled. This finding shifts the emphasis of early-warning analytics from identifying a universally superior classifier toward developing prediction systems that are closely aligned with meaningful educational evidence.

Beyond predictive accuracy, the findings also emphasize the importance of trustworthy model interpretation in educational decision support. The calibration analysis demonstrated that reliable probability estimates can be maintained without sacrificing predictive discrimination, thereby improving the credibility of risk estimates used for intervention prioritization. At the same time, SHAP analysis consistently identified academic progression, tuition-fee status, classroom engagement, and attendance as the principal contributors to model predictions, providing transparent evidence regarding how the prediction models reached their decisions. These results indicate that explainability should not be viewed merely as a technical property of machine learning models but as an essential mechanism for increasing educators' confidence in AI-assisted decision making. By combining calibrated probability estimation with interpretable feature attribution, the proposed framework strengthens the transparency of early-warning analytics while preserving the predictive performance required for operational student-retention systems.

The principal contribution of this study extends beyond prediction and explainability by introducing a structured rationale layer that directly connects analytical evidence with educational intervention. Previous studies on explainable artificial intelligence in education emphasize that explanations should remain human-centred, understandable, and sensitive to the needs of educational stakeholders (Khosravi et al., 2022). Similarly, recent research on large language models demonstrates considerable potential for supporting educational feedback, recommendation, and instructional assistance while simultaneously highlighting concerns regarding transparency, reproducibility, privacy, and ethical deployment (Kasneci et al., 2023; Yan et al., 2024). The proposed framework responds to these challenges by transforming predictive evidence into deterministic, evidence-grounded, and auditable intervention rationales rather than relying on unconstrained free-form text generation. Every recommendation remains explicitly linked to measurable behavioral or institutional evidence, allowing educators to understand not only the predicted level of risk but also the reasoning underlying each recommended intervention. This design extends current explainable learning analytics by positioning LLM-ready rationale generation as an evidence-communication layer that preserves both interpretability and accountability.

Taken together, the findings demonstrate that the proposed framework advances early-warning analytics from prediction-oriented modeling toward evidence-informed educational decision support. Instead of terminating the analytical process with risk probabilities or feature rankings, the framework integrates behavioral learning analytics, institutional prediction, explainable artificial intelligence, and structured rationale generation into a continuous workflow supporting educator-facing decision making. This integration establishes a transparent relationship between observed student behavior, predictive evidence, explanatory reasoning, and recommended interventions, thereby reducing the interpretive gap that commonly exists between machine learning outputs and educational practice. In doing so, the study contributes not only to improving predictive performance but also to strengthening the operational interpretability of AI-supported student retention systems within higher education.

IMPLICATIONS

The proposed framework has important practical implications for higher education institutions seeking to strengthen student-retention strategies through evidence-based decision support. Unlike conventional early-warning systems that primarily provide risk scores or prediction probabilities, the framework integrates behavioral analytics, predictive modeling, explainable artificial intelligence, and structured intervention rationales into a unified workflow that directly supports educators' decision-making processes. By explicitly linking risk predictions with interpretable evidence and actionable recommendations, the framework enables lecturers, academic advisors, and student-support offices to identify at-risk students, understand the factors contributing to their risk, and implement timely interventions based on transparent analytical evidence. This design also facilitates institutional accountability by ensuring that intervention recommendations remain traceable to measurable behavioral and academic indicators, thereby improving the operational value of AI-assisted retention systems in higher education.

LIMITATIONS

Several limitations should be acknowledged when interpreting the findings of this study. First, the empirical evaluation was conducted using two publicly available benchmark datasets that, although widely adopted in educational data mining research, may not fully capture the diversity of institutional contexts, learning environments, or student populations encountered in real-world higher education settings. Second, the predictive models were evaluated under fixed experimental configurations rather than within continuously evolving institutional data streams, where student behavior and educational conditions may change over time. Third, the structured intervention rationale was assessed primarily through evidence alignment, actionability, and monitoring specificity, without direct validation by lecturers, academic advisors, or student-support professionals. Consequently, while the proposed framework demonstrates methodological robustness and reproducibility, additional empirical evaluation is necessary before its effectiveness can be fully generalized to operational educational environments.

SUGGESTIONS

Future research should extend the proposed framework beyond benchmark evaluation toward authentic institutional deployment and human-centred validation. Longitudinal studies involving real learning-management systems and continuously updated student records would provide stronger evidence regarding the framework's effectiveness under dynamic educational conditions. In addition, future investigations should incorporate lecturers, academic advisors, and student-support practitioners in evaluating the usefulness, clarity, and trustworthiness of the generated intervention rationales during actual decision-making processes. Further research may also explore adaptive LLM-based rationale generation capable of incorporating institutional policies, disciplinary contexts, and multilingual educational settings while preserving transparency, auditability, and evidence-grounded reasoning. Such developments would contribute to the evolution of explainable and trustworthy AI systems that not only predict student risk but also provide personalized and operationally meaningful support for improving student retention.

CONCLUSION

This study proposed and empirically evaluated a two-stage early-warning framework that integrates classroom interaction modeling, institutional retention prediction, explainable artificial intelligence, and structured LLM-ready intervention rationale generation for student retention decision support. The findings demonstrate that behavioral learning analytics derived from the xAPI-Edu-Data dataset effectively capture short-term engagement patterns, while institutional variables from the Predict Students' Dropout and Academic Success dataset provide complementary evidence for longer-term retention prediction, with ensemble tree-based models achieving robust predictive performance across both benchmarks. Beyond predictive accuracy, the proposed framework transforms model outputs into structured, evidence-grounded intervention rationales that explicitly connect risk identification, explanatory evidence, recommended actions, and follow-up monitoring within a transparent and reproducible decision-support workflow. By integrating predictive analytics, explainable AI, and teacher-facing rationale generation into a unified framework, this study advances early-warning analytics beyond risk detection toward actionable educational decision support, providing a practical and interpretable approach for strengthening student retention strategies in higher education.

AUTHOR CONTRIBUTIONS STATEMENT

The sole author is responsible for all experiments and manuscript preparation.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

ADL Initiative. (2013). *Experience API (xAPI) specification* (Version 1.0.1). Advanced Distributed Learning Initiative.

Adadi, A., & 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

Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015). Preprocessing and analyzing educational data set using X-API for improving student's performance. In *2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*. IEEE. https://doi.org/10.1109/AEECT.2015.7360581

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). ACM. https://doi.org/10.1145/2330601.2330666

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 253–274). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.016

Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17. https://doi.org/10.5281/zenodo.3554657

Bowers, A. J. (2010). Grades and graduation: A longitudinal risk perspective to identify student dropouts. *The Journal of Educational Research, 103*(3), 191–207. https://doi.org/10.1080/00220670903382970

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review, 42*(4), 40–57.

de Oliveira, C. F., Sobral, S. R., Ferreira, M. J., & Moreira, F. (2021). How does learning analytics contribute to prevent students' dropout in higher education: A systematic literature review. *Big Data and Cognitive Computing, 5*(4), Article 64. https://doi.org/10.3390/bdcc5040064

Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). Predicting students dropout: A case study. In *Proceedings of the 2nd International Conference on Educational Data Mining* (pp. 41–50).

Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. arXiv. https://arxiv.org/abs/1702.08608

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5–6), 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development, 68*, 1961–1990. https://doi.org/10.1007/s11423-020-09788-z

Jayaprakash, S. M., Moody, E. W., Lauria, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics, 1*(1), 6–47. https://doi.org/10.18608/jla.2014.11.3

Kasneci, E., Sessler, K., Kuchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Gunnemann, S., Hullermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khosravi, H., Buckingham Shum, S., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gasevic, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence, 3*, Article 100074. https://doi.org/10.1016/j.caeai.2022.100074

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an early warning system for educators: A proof of concept. *Computers & Education, 54*(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

OpenAI. (2023). *GPT-4 technical report*. arXiv. https://arxiv.org/abs/2303.08774

Rahmani, A. M., Groot, W., & Rahmani, H. (2024). Dropout in online higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education, 21*, Article 19. https://doi.org/10.1186/s41239-024-00450-9

Realinho, V., Machado, J., Baptista, L., & Martins, M. V. (2022). Predicting student dropout and academic success. *Data, 7*(11), Article 146. https://doi.org/10.3390/data7110146

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations* (pp. 97–101). Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-3020

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40*(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 30–40.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89–125. https://doi.org/10.3102/00346543045001089

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). University of Chicago Press.

Wong, B. T.-M., & 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

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., Li, X., Jin, Y., & Gasevic, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90–112. https://doi.org/10.1111/bjet.13370