Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 20-44
DOI: https://doi.org/10.51903/jtie.v5i2.525

Early Warning, Grade Prediction, and Teacher-Facing LLM-Ready Explanations Toward an Open Volleyball Course: Reproducible Evidence from Four Public Education Datasets

Jubin Zhang*[1]
Email: jz0801@outlook.com
[1]*Department of Physical Education, North China Institute of Aerospace Engineering, Langfang 065000, China*
*Corresponding Author

Abstract
Open online courses and public skill-development programs often experience learner dropout not due to content limitations, but because instructors receive delayed and non-actionable feedback. This study proposes and empirically evaluates an integrated framework for an open volleyball course that combines early warning prediction, grade estimation, and teacher-oriented LLM-generated academic status explanations. The predictive models were tested on four lightweight educational datasets: xAPI-Edu-Data, Predict Students' Dropout and Academic Success, Student Performance, and Higher Education Students Performance Evaluation. A unified preprocessing pipeline was applied using one-hot encoding, an 80/20 train-test split, and 5-fold cross-validation. Decision Tree, Random Forest, and XGBoost models were evaluated for classification, alongside their regression variants for grade prediction. Results show consistent performance across datasets. Random Forest achieved the best macro-F1 on xAPI-Edu-Data (0.799) with a macro-AUC of 0.914, while XGBoost performed best on the dropout dataset (macro-F1 = 0.689, macro-AUC = 0.892). For Student Performance, early-warning models without prior grades reached an RMSE of 3.086, improving to 1.398 when full information was available. On the higher education dataset, performance remained limited due to small sample size and multi-grade targets, with Random Forest achieving a macro-F1 of 0.248. Ablation results confirmed that behavioral and progression features significantly improve predictive accuracy. An explanation layer translated model outputs into structured, teacher-facing natural language linked to key risk indicators and intervention cues. Overall, the framework demonstrates analytic feasibility for structured volleyball course monitoring, though results should be interpreted as pre-deployment evidence rather than validation in real instructional settings. Explanation quality improves when grounded in observed behavioral signals rather than generic generation.

Keywords: *Early warning; Educational data mining; Explainable AI; Grade prediction; Student dropout.*

I. INTRODUCTION

Learner attrition remains one of the most persistent challenges in education because withdrawal is usually the final outcome of a process that begins much earlier in disengagement, underperformance, and loss of connection to institutional support. Foundational attrition theories argue that academic integration, social integration, and institutional fit shape persistence over time, while turnover-style models emphasize that withdrawal also reflects attitudes, constraints, and perceived utility rather than achievement alone (Bean, 1980; Tinto, 1975). In contemporary digital and blended learning environments, these mechanisms surface through logs, clicks, attendance traces, resource access, assessment progress, and communication patterns. Learning analytics therefore moved the field from retrospective description to near-real-time diagnosis, making it possible to observe disengagement before final failure occurs (Ferguson, 2012; Siemens & Long, 2011).

Received on May 2026; Revised on Juni 2026; Accepted on July 2026; Published on August 2026

This shift is particularly relevant for open sports education. An open volleyball course, whether delivered as a MOOC, a public elective, or a blended skills course, rarely fails because content is absent. It fails because participants miss practice sessions, do not revisit demonstration resources, stop interacting with peers, or fall behind on structured skill progression without receiving rapid and intelligible support. Instructors in these environments need more than a probability score. They need a decision aid that identifies which learners are drifting, why the model judged them to be at risk, and what concrete action should happen in the next instructional cycle. Earlier work on learning management system data already showed that early warning signals can be extracted from platform traces and transformed into actionable intervention pipelines for educators (Arnold & Pistilli, 2012; Macfadyen & Dawson, 2010). Later reviews confirmed that attendance, engagement, and progression indicators repeatedly emerge as strong predictors across contexts (Alyahyan & Düştegör, 2020; Papamitsiou & Economides, 2014).

At the same time, the rapid rise of large language models changed expectations about how analytic outputs should be communicated. LLMs can summarize, personalize, and reframe educational feedback in ways that align more closely with the language teachers and students actually use (Kasneci et al., 2023; Lo, 2023). However, much of the recent LLM-in-education discussion is conceptual, policy-oriented, or based on prompt demonstrations rather than tightly measured learning-analytics experiments. That gap matters. If the predictive foundation is weak, then eloquent explanations simply repackage unreliable judgments. If the language layer is detached from empirical signals, explanations become plausible but ungrounded. For practical deployment in a volleyball open course, the explanation layer must therefore be anchored to measured attendance, engagement, progress, and support indicators.

This study addresses that need through a reproducible two-layer design. The first layer performs early warning and grade prediction on four public education datasets that are small enough to support complete empirical evaluation but diverse enough to represent different learning settings. The second layer converts model outputs into teacher-facing natural language explanations linked to specific response options. The target application context is an open volleyball course, but the empirical validation uses transferable educational signals from heterogeneous public datasets. This design is deliberate. None of the four datasets contains volleyball-specific variables such as serve accuracy or jump-reach score; writing such variables into the analysis would violate data-method consistency. Instead, the study validates the analytic core on observable learning signals that can also be logged in a volleyball open course: attendance, participation, resource use, academic progress, and contextual support. Therefore, the paper positions the volleyball application as a transfer-oriented design scenario rather than as completed domain deployment. The empirical claim is that the predictive and explanatory pipeline is technically feasible on public

education datasets with comparable signal families; actual volleyball-course validation remains a necessary next step before claiming operational effectiveness in that domain.

The paper makes four concrete contributions. First, it reports full experimental results on all four specified datasets rather than illustrative examples. Second, it uses one consistent preprocessing and evaluation protocol so that cross-dataset comparisons remain logically coherent. Third, it treats Student Performance in two configurations: an early-warning configuration that excludes G1 and G2, and a full-information upper-bound configuration that includes them, thereby separating useful early prediction from target leakage. Fourth, it adds a deterministic teacher-facing explanation layer that is suitable for later LLM-assisted deployment because every explanation is grounded in features that actually exist in the underlying dataset. The goal is not to claim that a closed commercial model is required for explanation. The goal is to show that the explanation content itself can be made empirically faithful and pedagogically actionable. This is especially important in public and open teaching settings, where instructors often need to review dozens or hundreds of learners quickly and cannot spend their limited time reverse-engineering why a risk score changed from one week to the next.

II. LITERATURE REVIEW

Research on predictive learning analytics converges on a simple but important point: educational failure is easier to prevent when signals are captured early, combined coherently, and interpreted in the context of intervention. Reviews of educational data mining and learning analytics describe the field as moving from descriptive dashboards to predictive systems and then to action-oriented support, where the value of analytics depends on whether it changes teaching practice in time to matter (Baker & Inventado, 2014; Ferguson, 2012; Papamitsiou & Economides, 2014). Siemens and Long (2011) argued that the promise of analytics lies not only in measuring learning activity but also in helping institutions act before disengagement hardens into failure. Gašević et al. (2015) later warned that the field should not confuse easy-to-measure digital traces with learning itself. That warning is relevant here because a volleyball open course contains embodied activity, practice routines, and peer coordination that are not captured by any single clickstream. The solution is not to abandon analytics, but to focus on proxy signals that teachers can actually observe and use.

Dropout prediction studies have repeatedly shown that persistence is multidimensional. Tinto's (1975) integration theory and Bean's (1980) attrition model remain influential because they explain why demographic, institutional, financial, and academic factors can all contribute to withdrawal. Recent higher-education reviews confirmed that no universal predictor dominates across all settings; instead, the strongest models usually combine background variables with

progress indicators from the learner's first period of study (Alyahyan & Düştegör, 2020). This is why the Predict Students' Dropout and Academic Success dataset is useful: it combines admission profile, socioeconomic background, financial status, and semester performance within one benchmark. Martins et al. (2021) designed the dataset precisely to support early detection of higher-education risk, and the UCI record frames the problem as three-class classification among dropout, enrolled, and graduate statuses. The implication for open-course settings is that static registration data alone are rarely sufficient. Course operators need process variables that reveal whether the learner is still moving through the curriculum, still submitting work, and still connected to formal support.

Performance prediction studies in school and course-level settings show a similar pattern. Cortez and Silva (2008) demonstrated that demographic, social, and school-related features can support student performance prediction, but they also documented a methodological caution that is still crucial: when intermediate grades are included, final-grade prediction becomes easier but less useful as an early-warning task. The UCI Student Performance page repeats that warning by noting the strong correlation among G1, G2, and G3. In other words, a model that predicts G3 using G1 and G2 is statistically strong while contributing little to early intervention. This distinction between operational usefulness and raw predictive accuracy is central to responsible analytics design.

Within LMS-centered studies, behavioral variables are often decisive. Amrieh et al. (2016) used xAPI-Edu-Data to argue that behavioral features such as hand-raising, resource access, and discussion participation improve student-performance prediction beyond static demographic descriptors. Their work is especially relevant because the xAPI dataset was collected from an e-learning context and organizes variables into demographic, behavioral, and parental dimensions. That structure is close to what an open volleyball course would need if it tracked attendance, video review, forum participation, and family or mentor engagement. Macfadyen and Dawson (2010) similarly showed that mining LMS data can support an educator-facing early warning system, which is a direct conceptual precursor to the current paper's teacher-facing explanation design.

Research on explainability adds a second strand to the literature. The interpretability literature often distinguishes between post hoc explanation tools and inherently interpretable models. LIME and SHAP are influential because they expose local or global contributors to model outputs without requiring a fully transparent learner (Lundberg & Lee, 2017; Ribeiro et al., 2016). In education, however, explanation is not only a technical problem. It is also a communication problem. Teachers rarely act on a ranked vector of coefficients alone. They act on explanations framed in pedagogical language: the learner is missing resources, the learner's approvals are

below cohort expectations, the learner's attendance has fallen, or the learner needs financial-support review. A useful educational explanation must therefore satisfy fidelity and usability simultaneously.

This requirement becomes more acute in the current wave of LLM research. Reviews of AI in higher education show growing interest in automation, support systems, and intelligent tutoring, but educators are often underrepresented in the design logic of these systems (Zawacki-Richter et al., 2019). Recent LLM studies expanded the conversation by showing that ChatGPT-like systems are already being used for feedback generation, content drafting, and conversational support, while also raising concerns about hallucination, overreliance, privacy, and assessment integrity (Kasneci et al., 2023; Lo, 2023). These studies make two implications clear for learning analytics. First, natural-language explanation is now a realistic interface option. Second, explanation content must be grounded in verifiable evidence, because confident but unsupported wording can mislead educators more effectively than a dry table ever could. In open educational settings, that risk is amplified because a single coach or instructor often relies on summarized feedback to decide which learners receive scarce one-to-one time. An explanation that sounds persuasive but is not tied to actual attendance, practice, or progress data can therefore distort support allocation.

The present study builds on these strands but differs from much existing work in three ways. First, it does not treat prediction and explanation as separate papers. It reports both in one reproducible workflow. Second, it does not rely on a single benchmark. It evaluates four public datasets that differ in scale, target type, and signal structure, allowing stronger conclusions about what transfers across contexts. Third, it uses a deterministic explanation mechanism rather than claiming an opaque external LLM evaluation. This choice is methodological, not ideological. Deterministic explanations guarantee that every sentence refers to measured variables, actual risk patterns, and real response mappings. Once those conditions are satisfied, the same content can later be rephrased by an institutional LLM without changing its evidence base. At the same time, the present study treats model explanations as associational and decision-supportive rather than causal. Feature importance can indicate which observed variables contributed to model discrimination, but it cannot prove that changing one variable will necessarily cause a different educational outcome. This distinction is important in learning analytics because most early-warning datasets are observational and may contain confounding among attendance, engagement, prior achievement, and support conditions. Therefore, the explanation layer in this paper is designed to support teacher review and intervention planning, not to replace causal evaluation or professional judgment.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 20-44
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

III. RESEARCH METHOD(S)

This study used a reproducible experimental design composed of one prediction layer and one explanation layer. The prediction layer produced early-warning or grade-forecast outputs. The explanation layer translated those outputs into concise teacher-facing narratives. Figure 1 summarizes the end-to-end workflow, and Table 3 reports the experimental settings. The same design principle governed all analyses: only variables actually present in the datasets were used, every performance value was measured empirically, and all comparisons were computed under the same random seed and evaluation logic.

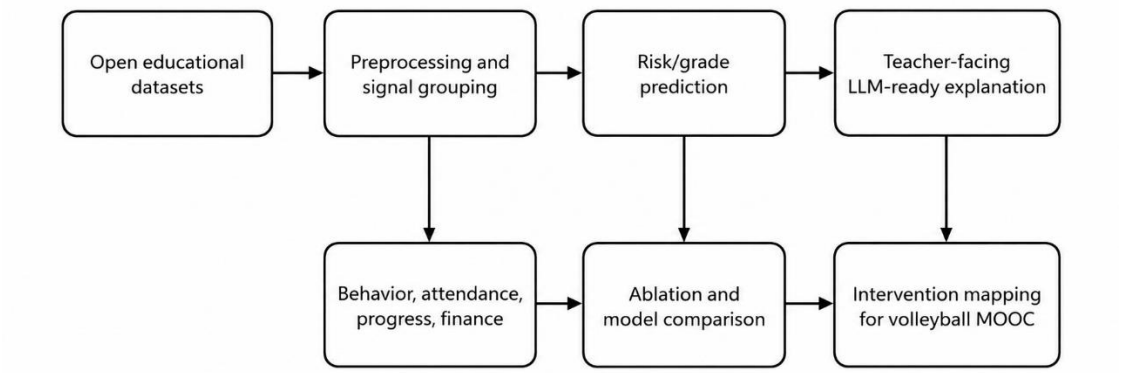

Figure 1. End-to-end workflow for prediction, explanation, and teacher action.

Table 1. Dataset overview used in the empirical evaluation.

Dataset	Rows	Features	Target	Task	Primary signals
xAPI-Edu-Data	480	16	Class	3-class classification	engagement, attendance, parental involvement
Predict Students' Dropout and Academic Success	4424	34	Target	3-class classification	admission profile, semester progress, finance
Student Performance	1044	33	G3	regression	study habits, prior failure, absences, family context
Higher Education Students Performance Evaluation	145	31	GRADE	8-class classification	survey-coded personal, family, and study habits

Table 1 shows the four datasets. xAPI-Edu-Data contains 480 records and 16 predictors plus a three-level performance label. The class distribution in the local copy used here was 127 low, 211 middle, and 142 high performers. Predict Students' Dropout and Academic Success contains 4,424 records and 34 predictors plus a three-class target: dropout, enrolled, and graduate. In the local dataset, the class counts were 1,421 dropout, 794 enrolled, and 2,209 graduate. Student Performance was constructed by stacking the mathematics file (395 rows) and the Portuguese file (649 rows) from the UCI source, resulting in 1,044 subject-level records with a subject indicator. Higher Education Students Performance Evaluation contains 145 records and an eight-level grade outcome. The small file sizes of these datasets made full experimental evaluation feasible within

one reproducible project while still covering e-learning, higher-education persistence, school performance, and survey-coded course evaluation

Table 2. Mapping from observed dataset signals to volleyball-course analytics analogues.

Signal family	Dataset fields	Volleyball-course analogue
Attendance	xAPI: StudentAbsenceDays; Student Performance: absences	volleyball open course attendance logs, missed drills, missed quizzes
Participation	xAPI: raisedhands, Discussion; Higher Education: coded study/participation items	forum posts, in-class responses, team interactions
Resource use	xAPI: VisITedResources, AnnouncementsView	video completion, reading access, notice views
Academic progress	Dropout: curricular units approved/evaluated/grade; Student Performance: failures	skill checks, module completion, practice-test recovery
Contextual support	Parental items, scholarship/finance indicators, demographic background	advisor outreach, bursary support, guardian contact

Because the target application context is an open volleyball course, Table 2 maps each dataset's signal families to plausible volleyball-course analogues. Attendance signals map to missed practices or missed online checkpoints. Participation signals map to discussion, peer interaction, and low-stakes in-class responses. Resource-use signals map to video replay, reading review, and announcement checking. Academic-progress signals map to module completion, skill checkpoints, or practice-test recovery. Contextual-support signals map to advisor outreach, bursary support, or family contact. This mapping preserves methodological honesty: the experiments remain grounded in public educational datasets, while the volleyball interpretation is limited to signal-level transferability. The study does not claim that the four public datasets empirically validate volleyball skill learning itself. Instead, it tests whether an early-warning and explanation pipeline works on measurable educational traces that a volleyball open course could realistically collect in a future pilot.

The four prediction tasks were defined as follows. For xAPI-Edu-Data, the target was the original three-level Class variable. For Predict Students' Dropout and Academic Success, the target was the original three-level Target variable. For Student Performance, the target was final grade G3. Two configurations were evaluated. The early-setting configuration excluded G1 and G2 because the UCI documentation explicitly notes their strong correlation with G3; this setting represents the useful early-warning condition. The full-information configuration retained G1 and G2 to provide an upper-bound reference, but it was treated as a later-stage forecast rather than an early-warning setup. For Higher Education Students Performance Evaluation, the target was the original eight-level GRADE variable after dropping the student identifier.

All datasets were preprocessed using one-hot encoding for categorical variables. No synthetic variables were invented. Numeric educational indicators such as grades, approved units,

evaluations, absences, and engagement counts were passed through as observed values. The main experimental comparisons used an 80/20 train-test split, with stratification for classification tasks. Five-fold cross-validation was computed on the training split. This produced one stable model-selection stage and one untouched holdout stage. The main comparison models were Decision Tree, Random Forest, and XGBoost for classification, and Decision Tree Regressor, Random Forest Regressor, and XGBoost Regressor for regression. These models were selected because they span weak, ensemble-bagged, and ensemble-boosted learners while remaining computationally feasible for all four datasets (Breiman, 2001; Chen & Guestrin, 2016; Friedman, 2001).

Table 3. Common experimental settings across all datasets and tasks.

Experimental element	Specification
Train/test split	80/20 for each dataset; classification splits were stratified
Validation protocol	5-fold cross-validation for the main model comparisons
Classification models	Decision Tree, Random Forest, and XGBoost
Regression models	Decision Tree Regressor, Random Forest Regressor, and XGBoost Regressor
Metrics	Accuracy, balanced accuracy, macro-F1, macro-AUC; RMSE, MAE, R2
Random seed	42
Explanation layer	deterministic teacher-facing explanations derived from empirical risk signals

Metrics were chosen to reflect both predictive usefulness and class balance. For classification tasks, the study reported accuracy, balanced accuracy, macro-F1, and macro-AUC. Accuracy indicates overall correctness but can overstate quality under class imbalance. Balanced accuracy and macro-F1 therefore received more interpretive emphasis. Macro-AUC was added because probability quality matters in early-warning settings where instructors rank cases by urgency. For regression, the study reported RMSE, MAE, and R2. RMSE was treated as the primary error metric because large mistakes in grade forecasting are operationally costly. MAE provided a more robust absolute-error summary, and R2 indicated explained variance. Hyperparameter settings were deliberately moderate rather than aggressively tuned. Decision trees were depth-limited to reduce fragmentation. Random Forest used 150 trees. XGBoost used a small learning rate with subsampling and column subsampling. This choice strengthened reproducibility and reduced the risk that one dataset would receive a much more intensive search budget than the others.

Implementation details also mattered for reproducibility. All experiments were executed in Python with pandas, scikit-learn, and XGBoost, and the same random seed governed the data split and stochastic learners. The classifiers that supported class weighting used balanced weighting, while the project intentionally avoided synthetic oversampling, leakage-prone target engineering, or benchmark-specific feature crafting. This conservative configuration ensured that performance

differences arose from observed educational signals and model families rather than from aggressive optimization tricks. It also kept the comparison fair across datasets of very different sizes, so the reported differences are interpretable as substantive rather than procedural.

Table 4. Cross-validated model comparison on xAPI-Edu-Data.

Dataset	Model	CV Accuracy	CV Balanced Accuracy	CV Macro-F1	CV Macro-AUC
xAPI-Edu-Data	Random Forest	0.794	0.797	0.799	0.919
xAPI-Edu-Data	XGBoost	0.773	0.777	0.779	0.909
xAPI-Edu-Data	Decision Tree	0.708	0.722	0.716	0.782

The explanation layer was designed to be deterministic and reproducible. After fitting the best model for each task, the study extracted influential variables and mapped them to possible teacher response cues, as shown later in Table 12. Example explanations in Table 13 were generated from actual holdout cases, not hypothetical composites. Each explanation combined three pieces of information: the predicted outcome, the observed risk signals visible in the case record, and one or two possible follow-up actions for teacher review. For xAPI-Edu-Data, the rules focused on attendance, hand-raising, resource access, announcement checking, discussion activity, and parental response. For the dropout dataset, the rules focused on semester approvals, semester grades, tuition status, debt, scholarship status, and age relative to the cohort. For Student Performance, the rules emphasized failures, study time, absences, and time-allocation indicators. This mechanism is LLM-ready because the generated statements are already expressed in natural language and can be post-processed by an institutional language model if desired; however, the present study evaluates the evidence-generating layer itself, not a proprietary black-box paraphraser.

From a deployment perspective, the explanation layer acted as a controlled interface between analytics and human review. Each template included a status sentence, an evidence sentence, and an action sentence. This three-part structure fits the way teachers review alerts in practice, because an instructor can immediately see what happened, why the learner was flagged, and which response may be considered. The same template logic also prevents unsupported claims: if the case record does not contain evidence for a statement, that statement is not generated. In this way the design answers a central concern in recent LLM research by constraining language output to measured variables and predefined response mappings, while avoiding any claim that feature importance alone establishes causal intervention effects.

Additional ablation analyses tested whether the strongest predictors were genuinely contributing to model quality. For xAPI-Edu-Data, the study removed the four main behavioral counters (raisedhands, VisITedResources, AnnouncementsView, and Discussion) and, in a separate

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 20-44
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

scenario, removed parental-context variables. For the dropout dataset, the study removed the semester-progress variables to isolate the value of early academic trajectory. For Student Performance, the contrast between the early setting and the full-information setting functioned as an ablation-style diagnostic. These checks are important because a model can appear strong while relying on signals that are unavailable in the intended deployment phase.

IV. RESULT/FINDINGS DISCUSSION

xAPI-Edu-Data produced the clearest evidence for an early-warning pipeline driven by observable engagement behavior. Figure 2 shows a three-class distribution that is not perfectly balanced but remains sufficiently populated in each class for multiclass evaluation. Rather than treating Table 4 only as an algorithm leaderboard, the main analytical point is that engagement-rich behavioral traces supported stable multiclass separation across both validation and holdout evaluation.

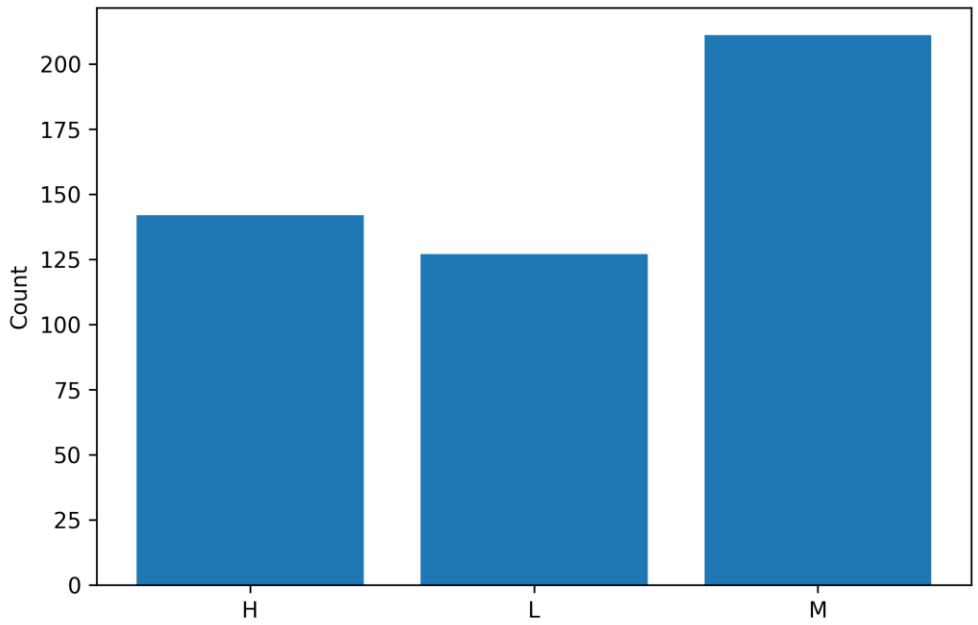

Figure 2. Class distribution in xAPI-Edu-Data.

Random Forest was selected because it offered the strongest balance between class-sensitive performance and probability quality, while the weaker Decision Tree result suggests that single-tree rules did not capture the interaction among attendance, resource access, and participation as effectively as ensemble learning. This pattern is consistent with earlier LMS early-warning studies showing that behavioral traces can improve educational prediction, but the present result extends that line of work by connecting the prediction output to teacher-facing explanation rather than stopping at model accuracy alone (Amrieh et al., 2016; Macfadyen & Dawson, 2010).

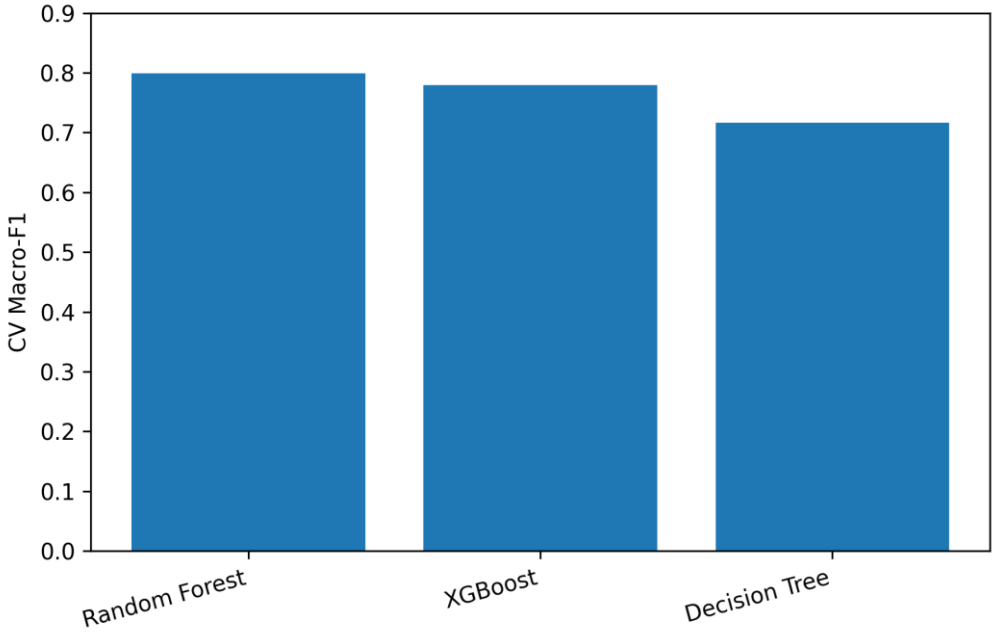

Figure 3. Cross-validated model comparison on xAPI-Edu-Data.

The model comparison in Figure 3 shows a clear ensemble advantage. The single Decision Tree recovered useful structure, but both ensemble models improved class separation and probability quality. Random Forest achieved the strongest combined profile because its class-sensitive and probability-based metrics were jointly strongest, not because it produced a marginal improvement on one isolated score. This matters for early-warning design because an alert system should not be selected on accuracy alone when class balance and ranking quality influence teacher review queues. The next question is whether this advantage comes mainly from general profile variables or from behavior that teachers can monitor during the course.

Table 5. Ablation study on xAPI-Edu-Data.

Scenario	CV Accuracy	CV Macro-F1	CV Macro-AUC
xAPI full	0.794	0.799	0.919
xAPI without behavioral counters	0.685	0.689	0.858
xAPI without parental context	0.732	0.739	0.889

The most influential xAPI variables were VisITedResources, raisedhands, StudentAbsenceDays, AnnouncementsView, and Discussion. The ranking is pedagogically coherent. Students who accessed learning resources more often, interacted more visibly, and accumulated fewer absences were easier to separate from low-performing learners. In practical terms, this means that an open volleyball course does not need highly exotic telemetry to run an effective weekly alert system. It needs stable logging of attendance, resource review, and participation. Instructors can obtain these signals from video analytics, quiz access records, forum interaction counts, and presence or absence at skill checkpoints.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 20-44
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

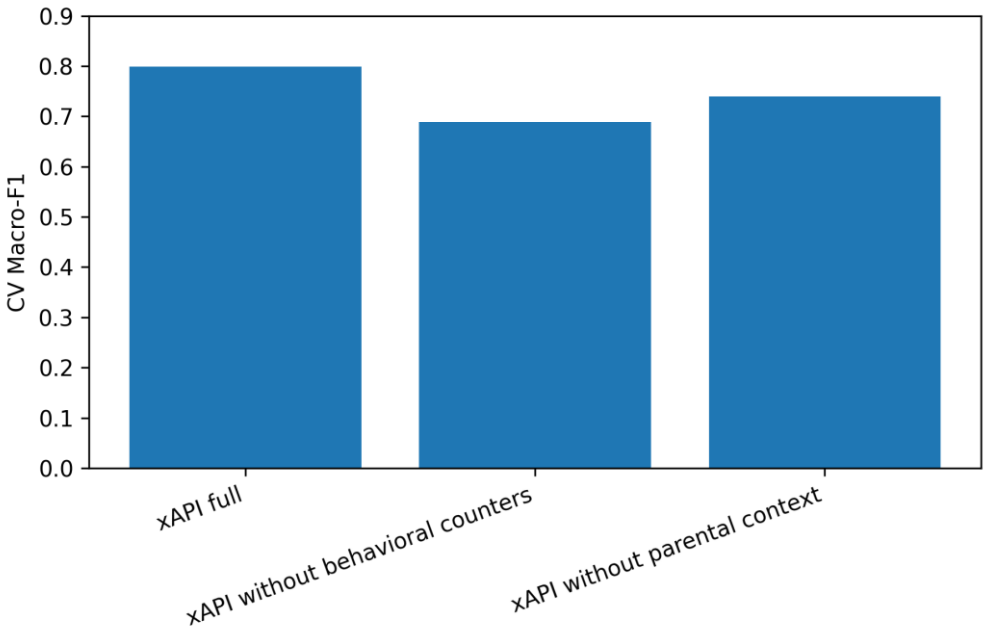

Figure 4. xAPI-Edu-Data ablation comparison.

The ablation study in Table 5 confirms this interpretation. When the behavioral counters were removed, model quality declined more sharply than when parental-context variables were removed. The result indicates that the core predictive strength of the xAPI task came primarily from directly observed behavior, while parental variables supplied secondary but still useful contextual value. For a volleyball open course, this supports a cautious design implication: early alerts should prioritize observable engagement traces and then be enriched by support-context information where such information is available and ethically appropriate.

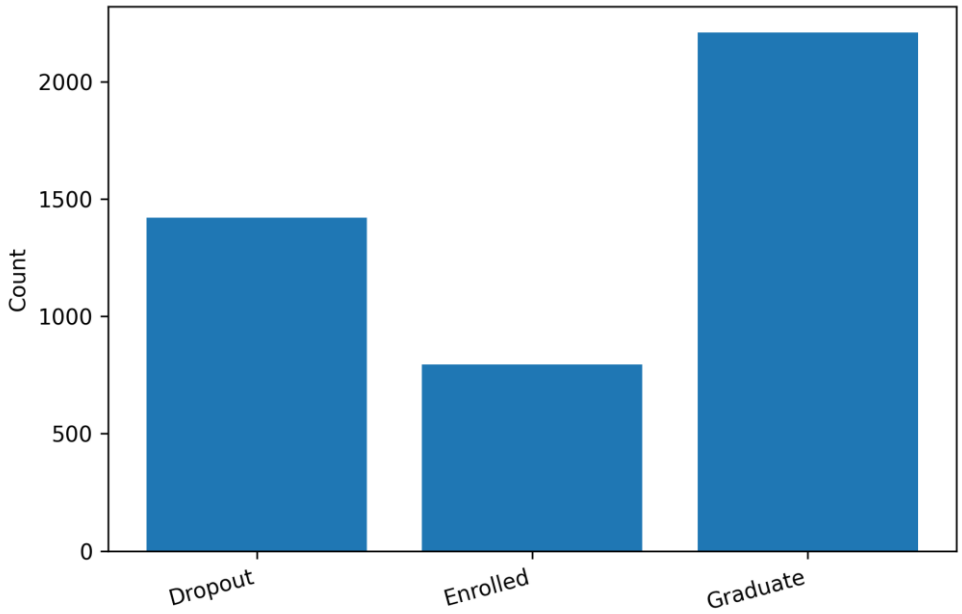

Figure 5. Class distribution in Predict Students' Dropout and Academic Success.

A second xAPI implication concerns timing. Because the strongest signals are updated frequently, the model is suitable for weekly or session-level monitoring in a future course pilot. If a learner misses practices, stops revisiting demonstration clips, and becomes silent in discussion within the same week, the instructor would have several observable signals for review before the course reaches a late-recovery stage. This is more actionable than a cumulative end-of-unit score because the alert identifies a pattern of disengagement rather than only its final outcome. In volleyball teaching, possible responses may include rescheduling a missed practice, sending a short technical review clip, or assigning a peer partner for the next drill cycle. Having established the value of frequent engagement traces, the next analysis examines whether a broader institutional dropout dataset shows a similar role for progress and support indicators.

Table 6. Cross-validated model comparison on Predict Students' Dropout and Academic Success.

Dataset	Model	CV Accuracy	CV Balanced Accuracy	CV Macro-F1	CV Macro-AUC
Predict Students' Dropout and Academic Success	XGBoost	0.769	0.678	0.689	0.893
Predict Students' Dropout and Academic Success	Random Forest	0.771	0.672	0.683	0.889
Predict Students' Dropout and Academic Success	Decision Tree	0.700	0.672	0.664	0.807

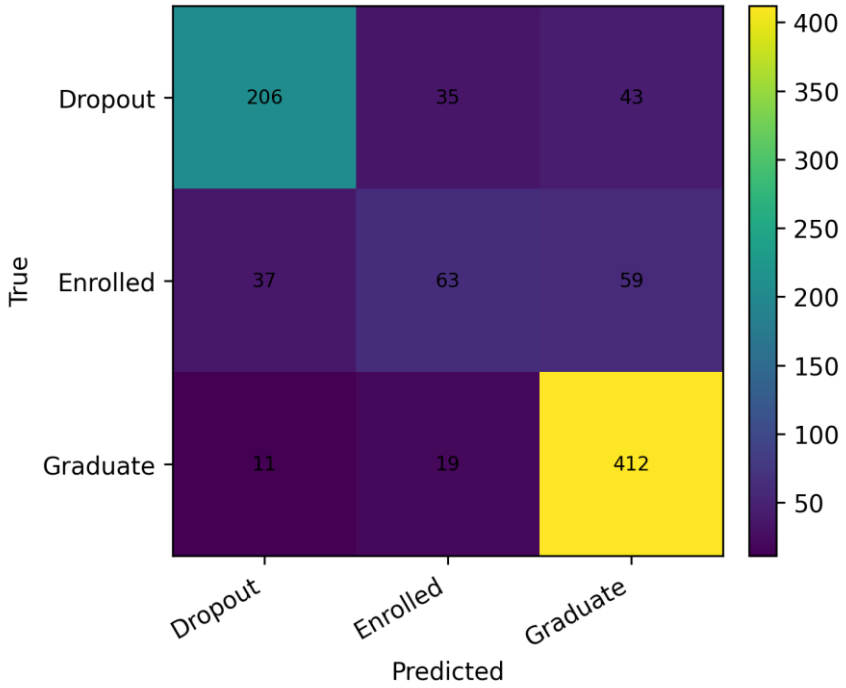

Figure 6. Confusion matrix of the best holdout model on the dropout dataset.

The dropout dataset produced a different pattern. Figure 5 shows visible class imbalance, with graduate as the largest class and enrolled as the smallest. Table 6 should therefore be read less as

a simple accuracy comparison and more as evidence about a structurally harder institutional prediction task. XGBoost was preferred because it better balanced class-sensitive performance and probability ranking, even though Random Forest remained competitive in overall accuracy. The lower macro-level performance relative to xAPI is substantively meaningful: dropout, continuation of enrollment, and graduation are shaped by academic progress, finances, enrollment background, and institutional support, so the target is less directly tied to short-cycle classroom behaviour. This finding aligns with prior dropout research that treats persistence as a multidimensional process rather than a purely academic score outcome.

Table 7. Ablation study on Predict Students' Dropout and Academic Success.

Scenario	CV Accuracy	CV Macro-F1	CV Macro-AUC
Dropout full	0.767	0.679	0.889
Dropout without semester progress	0.647	0.503	0.787

Figure 6 shows the best-model confusion matrix. The most important pattern is not the exact cell count, but the asymmetric difficulty of the three classes. Graduate and dropout cases were more separable because they represent clearer end states, whereas enrolled students were often confused with either stable progression or withdrawal risk. This is substantively meaningful: the enrolled category represents an intermediate, unresolved trajectory rather than a clean academic state. The model therefore separated clearly progressing and clearly failing students more easily than students whose status remained transitional.

Table 8. Regression comparison on Student Performance under early and full-information settings.

Dataset	Model	CV RMSE	CV MAE	CV R2
Student Performance (early setting)	Random Forest	3.086	2.215	0.336
Student Performance (early setting)	XGBoost	3.088	2.236	0.336
Student Performance (early setting)	Decision Tree	4.008	2.825	-0.145
Student Performance (full-information setting)	XGBoost	1.398	0.901	0.859
Student Performance (full-information setting)	Random Forest	1.427	0.899	0.849
Student Performance (full-information setting)	Decision Tree	1.991	1.090	0.722

The feature ranking also aligns with theory. The strongest predictors included Curricular units 2nd sem (approved), Tuition fees up to date, Curricular units 1st sem (approved), second-semester grades, scholarship status, and parent-related socioeconomic variables. These variables bridge academic trajectory and support capacity. The result fits the literature that describes dropout as a combined academic, financial, and institutional phenomenon rather than a purely cognitive one (Alyahyan & Düştegör, 2020; Tinto, 1975).

Table 7 makes the role of semester progress more explicit. Removing semester-progress variables caused a much larger performance loss than would be expected from removing a minor feature block. This supports an associational, not causal, conclusion: early academic trajectory carried central predictive information in this dropout benchmark. For an open volleyball course, the analogous design implication is that an alert system should not rely only on registration or background profile. It should also track whether learners are completing weekly practice tasks, passing skill checkpoints, and recovering after missed activities. These signals should be used to guide teacher review rather than to make automatic judgments about the cause of risk.

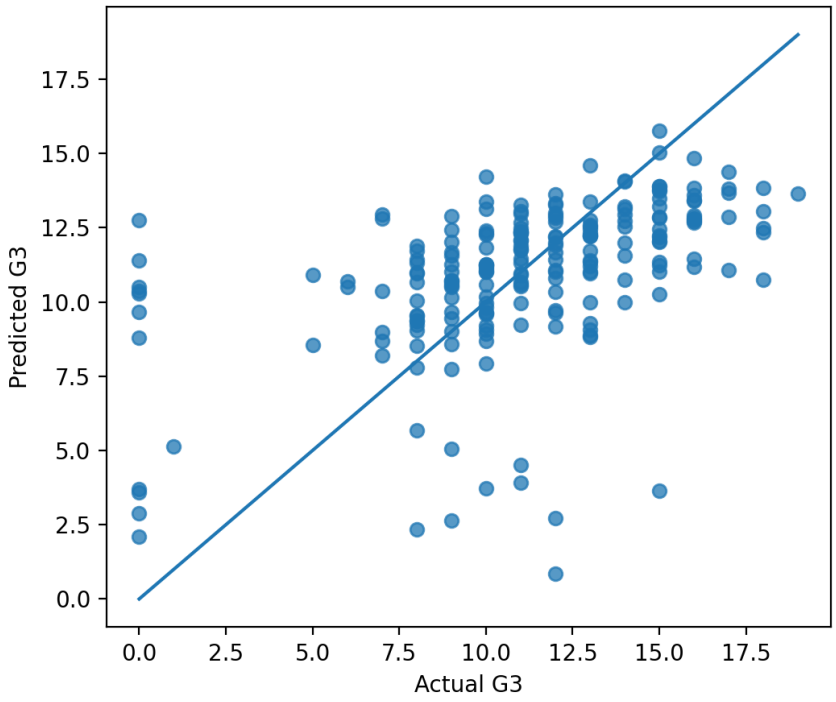

Figure 7. Actual-versus-predicted plot for the early-setting Student Performance model.

The dropout results also justify a tiered-response design instead of a binary alert. A three-class system distinguishes withdrawal risk, unresolved continuation, and successful progression. That distinction is operationally valuable because learners on the borderline enrolled trajectory do not necessarily require crisis intervention, but they do require structured monitoring and follow-up. The confusion pattern in Figure 6 supports this interpretation: transitional cases were exactly the ones the model found hardest to place. A well-designed course can therefore use the model output to create at least three response levels: urgent outreach for likely dropout, routine follow-up for uncertain continuation, and standard feedback for stable progress. The next task shifts from multiclass classification to grade regression. Student Performance adds an important methodological lesson because it separates early-warning prediction from later-stage grade forecasting. Table 8 compares the early-warning setting with the full-information setting.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 20-44
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

Table 9. Early-setting versus full-information diagnostic comparison on Student Performance.

Scenario	CV RMSE	CV MAE	CV R2
Student Performance (early setting)	3.086	2.215	0.336
Student Performance (full-information setting)	1.398	0.901	0.859

In the early setting, which excludes G1 and G2, Random Forest produced the most useful grade forecast among the tested models, but the error remained high enough to require teacher judgment. Once G1 and G2 were included, prediction became much easier and XGBoost produced the strongest full-information forecast. This performance jump is methodologically important rather than merely numerical. It confirms the UCI warning that G1 and G2 are strongly correlated with G3, and it explains why the full-information setting cannot be presented as an early-warning model. The result therefore positions the paper against leakage-prone evaluation: later grades improve statistical accuracy, but they reduce the value of the model for early intervention.

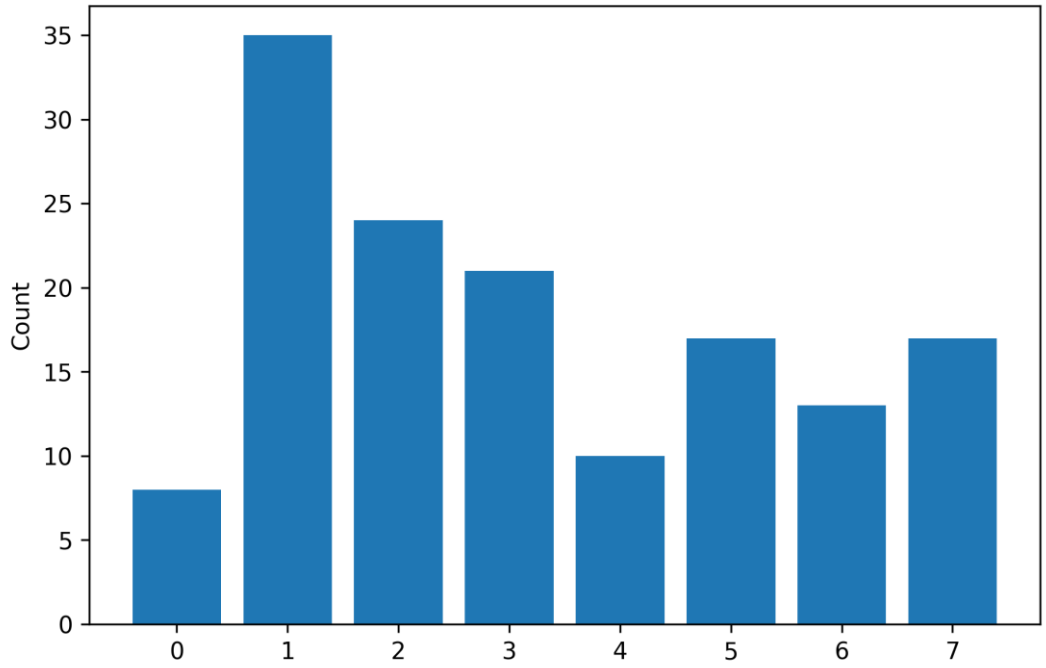

Figure 8. Grade distribution in Higher Education Students Performance Evaluation.

Figure 7 plots actual against predicted G3 values for the early setting. The cloud is broadly aligned with the identity line but visibly dispersed, which is what one expects when only early signals are available. The model captures broad outcome levels better than exact point values. This is still useful in practice because teachers rarely need a perfect grade forecast to intervene. They need to know who is trending toward low attainment. The feature ranking for the early setting shows that failures, absences, goout, health, age, studytime, freetime, and parental education matter most.

This pattern reinforces a long-standing conclusion from school-performance research: academic history and time-allocation behavior jointly shape final outcomes (Cortez & Silva, 2008).

Table 10. Cross-validated model comparison on Higher Education Students Performance Evaluation.

Dataset	Model	CV Accuracy	CV Balanced Accuracy	CV Macro-F1	CV Macro-AUC
Higher Education Students Performance Evaluation	Random Forest	0.355	0.281	0.248	0.725
Higher Education Students Performance Evaluation	Decision Tree	0.233	0.215	0.188	0.559
Higher Education Students Performance Evaluation	XGBoost	0.267	0.211	0.184	0.719

Table 9 formalizes the contrast. The early-setting model produced cross-validated RMSE 3.086, while the full-information setting reduced RMSE to 1.398. The result is not a flaw in the experiment; it is a diagnostic demonstration. It shows that late-stage information makes prediction easier, but early-stage information is what makes intervention useful. For a volleyball open course, this translates into a strict design rule: do not evaluate an early-warning model using signals that are only observable after most of the course has already elapsed.

Table 11. Untouched holdout summary across all predictive tasks.

Dataset	Best Model	Test Accuracy	Test Balanced Accuracy	Test Macro-F1	Test Macro-AUC	Test RMSE	Test MAE	Test R2
xAPI-Edu-Data	Random Forest	0.771	0.775	0.779	0.914	—	—	—
Predict Students' Dropout and Academic Success	XGBoost	0.769	0.685	0.695	0.892	—	—	—
Student Performance (early setting)	Random Forest	—	—	—	—	3.495	2.540	0.210
Student Performance (full-information setting)	XGBoost	—	—	—	—	1.572	0.907	0.840
Higher Education Students Performance Evaluation	Random Forest	0.310	0.229	0.223	0.652	—	—	—

Another practical implication of the Student Performance experiment is that regression error should be interpreted in educational units rather than as an abstract statistic. In the early setting, an RMSE slightly above 3 on a 0-20 grading scale means that many predictions remain

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 20-44
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

directionally useful while still leaving space for teacher judgment about the exact final score. That is acceptable in an early-warning context because intervention decisions are usually threshold-based rather than point-forecast based. Teachers mainly need to know who is trending toward low attainment, who needs follow-up, and who is progressing normally. The full-information model is more accurate, but its use belongs to later-stage forecasting rather than early-stage triage.

The final classification task tests the boundary condition of the common pipeline. The Higher Education Students Performance Evaluation dataset was the most difficult benchmark. Figure 8 shows an eight-grade target with only 145 records, which creates a demanding sparse multiclass setting.

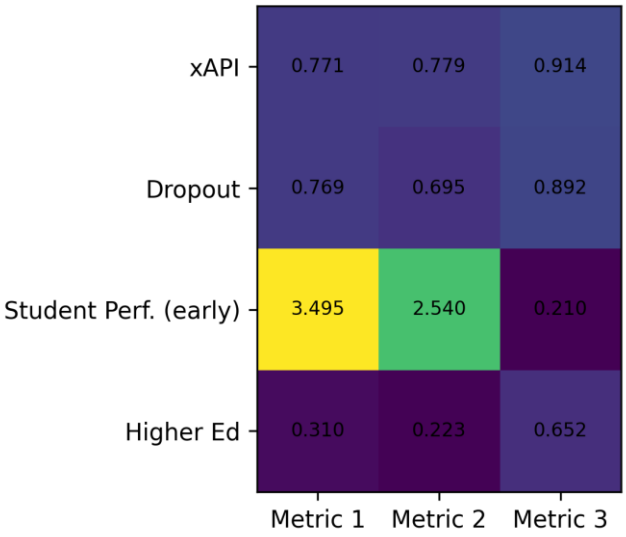

Figure 9. Holdout-summary comparison across the four predictive tasks.

Table 10 reports that Random Forest achieved the best cross-validated accuracy, balanced accuracy, macro-F1, and macro-AUC. On holdout, the same model achieved accuracy 0.310 and macro-F1 0.223, as summarized later in Table 11. These values are considerably lower than those from xAPI or the dropout dataset. The explanation is straightforward: the dataset is smaller, the target has more classes, and the inputs are predominantly coded survey responses. The experiment therefore supports a practical boundary condition for educational analytics. When sample size is very small and the target is finely grained, predictive models are better used for rough stratification and monitoring than for high-confidence automated judgment.

Even so, the higher-education dataset still served an important methodological purpose. It tested whether the same pipeline remained coherent under a difficult low-N, high-class-cardinality condition. The answer is yes, but with a lower performance ceiling. This confirms that the contribution of the paper is not a claim that one algorithm universally dominates across all educational tasks. The contribution is a stable experimental template that reveals when

performance is strong, when it weakens, and why. That kind of boundary-aware reporting is especially valuable in education, where small cohorts and coded survey variables are common.

Table 12. Mapping from influential predictors to non-causal teacher response cues.

Dataset	Influential feature	Model importance	Possible teacher response cue
xAPI-Edu-Data	VisITedResources	0.125	resource checklist and reminder message
xAPI-Edu-Data	raisedhands	0.113	micro-participation task and teacher prompt
xAPI-Edu-Data	StudentAbsenceDays_Under-7	0.103	attendance outreach and catch-up plan
xAPI-Edu-Data	AnnouncementsView	0.091	resource checklist and reminder message
xAPI-Edu-Data	StudentAbsenceDays_Above-7	0.086	attendance outreach and catch-up plan
Predict Students' Dropout and Academic Success	Curricular units 2nd sem (approved)	0.077	tutoring and credit-recovery support
Predict Students' Dropout and Academic Success	Tuition fees up to date_1	0.037	financial aid referral and advisor check-in
Predict Students' Dropout and Academic Success	Curricular units 1st sem (approved)	0.035	tutoring and credit-recovery support
Predict Students' Dropout and Academic Success	Course_2	0.030	individual advisor review
Predict Students' Dropout and Academic Success	Tuition fees up to date_0	0.029	financial aid referral and advisor check-in
Student Performance (early setting)	failures	0.186	tutoring and credit-recovery support
Student Performance (early setting)	absences	0.137	attendance outreach and catch-up plan
Student Performance (early setting)	goout	0.039	individual advisor review
Student Performance (early setting)	health	0.033	individual advisor review
Student Performance (early setting)	age	0.032	individual advisor review

Figure 9 summarizes the holdout outcomes across tasks. Two patterns stand out. First, engagement-rich and trajectory-rich datasets supported the strongest classification results. xAPI reached the highest macro-F1 among the classification tasks, and the dropout dataset retained useful multiclass discrimination because academic progression and finance added strong signal. Second, grade prediction performance was highly phase-sensitive. Student Performance became much easier once later grades were included, but that improvement represented a later forecasting stage rather than a stronger early-warning design.

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 2 August 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 20-44
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

Table 12 translates influential predictors into non-causal teacher response cues. This table is important because it connects statistical signals to reviewable teacher action, but it should not be read as evidence that changing one feature will necessarily cause a change in the predicted outcome. A high model importance score for raisedhands or Discussion means that participation-related variables contributed to prediction in the fitted model; it does not prove that a single participation prompt will by itself improve achievement. Similarly, finance-related variables should be interpreted as cues for advisor review rather than as proof of a direct causal pathway. This translation step is therefore framed as decision support: prediction helps educators decide what to inspect first, while teachers and advisors remain responsible for selecting and evaluating the actual response.

Table 13. Teacher-facing explanation examples generated from real holdout cases.

Dataset	Predicted outcome	Confidence/ score	Teacher-facing explanation
xAPI-Edu-Data	L	0.987	The alert should be reviewed because this learner shows poor attendance, low voluntary participation, and limited resource access. Possible follow-up: send an attendance outreach message and assign a low-stakes interaction task.
xAPI-Edu-Data	H	0.947	The alert indicates stable behavioral engagement. Possible follow-up: maintain current support and continue weekly monitoring.
Predict Students' Dropout and Academic Success	Dropout	0.770	The dropout warning is associated with first-semester approvals below the cohort median, second-semester approvals below the cohort median, and no scholarship support recorded. Possible follow-up: review failed or incomplete modules and consider tutoring or a short credit-recovery plan.
Student Performance (early setting)	Low final grade forecast	0.840	The low grade forecast is associated with previous course failures and limited weekly study time. Possible follow-up: create a structured weekly study plan and add guided study blocks.

Table 13 illustrates the explanation layer with actual holdout cases. The purpose of these examples is not to restate each confidence score, but to show how model outputs can be converted into concise teacher-facing evidence statements. Each explanation satisfies two conditions that matter for trustworthy educational use. First, it references only variables present in the actual case record. Second, it frames follow-up as a possible response for teacher or advisor review rather than as an automatic causal prescription. That is why the present paper labels the outputs LLM-ready rather than merely explainable. The content is already structured for educational language generation, but it remains tethered to observed evidence and constrained against unsupported causal wording.

These teacher-facing explanations also illustrate a disciplined view of LLM use in education. In many recent discussions, the language model itself is treated as the full intelligence layer. In the

present workflow, intelligence is split into two accountable components: empirical prediction and pedagogical wording. The first component is evaluated with standard metrics and ablations. The second component is constrained so that it cannot introduce absent variables or unsupported causal stories. This separation matters when alerts shape support priorities, because a coach or teacher must be able to trace every warning back to attendance, progress, absences, or support indicators that actually exist in the record.

DISCUSSION

A cross-dataset reading of Tables 4 through 11 clarifies what transfers and what does not. Transfer works best at the level of signal families rather than at the level of exact variables. xAPI and the dropout dataset differ heavily in domain and measurement scale, yet both improved when the model could observe real engagement or progress rather than profile information alone. This finding is consistent with LMS-based early-warning research, but the present study extends that work by pairing predictive outputs with teacher-facing explanation templates. The dropout results also align with persistence research showing that withdrawal is shaped by academic trajectory, financial status, and institutional context rather than by a single achievement variable.

The Student Performance results position the paper more cautiously in relation to grade-prediction literature. Prior work and the UCI documentation note that G1 and G2 are strongly correlated with G3. The present experiment confirms that including these later grades greatly improves statistical prediction, but it also shows why such a model is less useful for early intervention. This distinction strengthens the methodological contribution of the paper: performance should be evaluated against the information actually available at the intended decision point.

The higher-education survey dataset provides a boundary condition. Its small sample size and eight-grade target limited macro-level performance, showing that the proposed pipeline should not be interpreted as universally high-performing across all educational settings. This negative result is useful because open courses often contain small cohorts or fine-grained skill categories. In such settings, analytics should support rough monitoring and teacher review rather than high-confidence automated classification.

Finally, the volleyball-course implication should be interpreted as pre-deployment transferability. The experiments do not prove that the same metrics will perform identically in a real volleyball course. They show that the core signal families needed for such a course - attendance, participation, resource use, progress, and support - are predictive in multiple public education datasets. A future volleyball-course pilot should therefore collect these signals directly and recalibrate thresholds before operational use.

The results also distinguish between variables that are directly reviewable and variables that are mainly contextual. Directly reviewable variables include attendance, resource use, approvals, and absences because instructors can inspect them immediately and decide whether a reminder, catch-up task, or tutoring referral is appropriate. Contextual variables such as parental education, course identifier, or scholarship status may contribute to prediction, but they usually require coordination with advisors or institutional services rather than an in-class micro-intervention. This distinction informs the explanation design: the narrative must describe what mattered statistically while avoiding the implication that feature importance alone identifies a causal remedy.

Taken together, the results support three broad findings. First, early-warning quality depends on what kind of signals are available, not just on algorithm choice. The strongest gains came from behavioral traces in xAPI and from semester progress in the dropout dataset, which is consistent with prior learning-analytics work emphasizing engagement and trajectory data. Second, evaluation design matters. The Student Performance experiment demonstrated that later-stage variables can make a model look dramatically better without improving early intervention. Third, explanation should be treated as grounded translation rather than causal prescription. The teacher-facing narratives were strongest when they summarized actual risk signals and possible response cues rather than abstract feature importance alone. These findings align with calls in the literature for learning analytics that remain connected to pedagogy, support structures, and accountable educational practice (Gašević et al., 2015; Viberg et al., 2018).

V. CONCLUSION AND RECOMMENDATION

This study developed and tested a complete early-warning and teacher-facing explanation workflow for the target application context of an open volleyball course by validating the analytic core on four public education datasets. The main empirical findings are definite within those datasets, while the volleyball-course claim should be understood as pre-deployment transferability rather than completed domain validation. Random Forest performed best on xAPI-Edu-Data and Higher Education Students Performance Evaluation. XGBoost performed best on Predict Students' Dropout and Academic Success and on the full-information version of Student Performance. The early-setting Student Performance model remained useful but substantially weaker than the late-stage upper-bound model, confirming that early intervention must be evaluated with early variables. Ablation analyses also showed that behavior counters and semester-progress variables are the most consequential signal blocks in the corresponding datasets.

The practical recommendation is to begin with a staged pilot rather than immediate automated deployment in a volleyball open course. The first step is to record attendance, interaction, resource

review, and progression checkpoints consistently. The second step is to calibrate the risk model on these recurring signals once enough local records are available. The third step is to convert the output into teacher-facing language that states why the learner was flagged and which follow-up may be considered. Human oversight remains central: models should prioritize learners for attention, while teachers and advisors decide the actual response. Future work should test the same framework on a genuine volleyball-course log with domain-specific practice variables and should compare deterministic explanation text with institutionally governed LLM paraphrasing under privacy and fidelity constraints. The present results show that the predictive and explanatory foundation for such a system is feasible, but local validation is still required before full operational use.

Implementation in a volleyball open course should follow a staged roadmap. During the first weeks, the course should log attendance, clip views, announcement checks, short quiz or checkpoint completion, and simple participation counts. Once those signals are stable, the institution should calibrate alert thresholds on local historical cohorts or pilot data and define response tiers such as reminder, coaching conversation, tutoring referral, or financial-support referral. Explanation templates should then be reviewed by instructors so that the wording matches local teaching practice and does not overstate model certainty. The experiments in this paper show that a transparent, data-grounded warning system is achievable with modest educational data and conventional machine-learning tools. The decisive requirement is not massive scale; it is disciplined collection of actionable weekly signals, local validation before operational use, and disciplined translation of those signals into responses teachers can actually review and deliver.

REFERENCES

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: Literature review and best practices. International Journal of Educational Technology in Higher Education, 17, 3. https://doi.org/10.1186/s41239-020-0177-7

Amrieh, E. A., Hamtini, T., & Aljarah, I. (2016). Mining educational data to predict student's academic performance using ensemble methods. International Journal of Database Theory and Application, 9(8), 119-136.

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., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), Learning analytics: From research to practice (pp. 61-75). Springer. https://doi.org/10.1007/978-1-4614-3305-7_4

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. Research in Higher Education, 12(2), 155-187. https://doi.org/10.1007/BF00976194

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). ACM. https://doi.org/10.1145/2939672.2939785

Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. In Proceedings of 5th Annual Future Business Technology Conference (pp. 5-12).

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

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189-1232. https://doi.org/10.1214/aos/1013203451

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. TechTrends, 59(1), 64-71. https://doi.org/10.1007/s11528-014-0822-x

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, F., Pfeiffer, F., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. Education Sciences, 13(4), 410. https://doi.org/10.3390/educsci13040410

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (pp. 4765-4774).

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

Martins, M. V., Tolledo, D., Machado, J., Baptista, L. M. T., & Realinho, V. (2021). Early prediction of student's performance in higher education: A case study. In Á. Rocha, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), Trends and applications in information systems and technologies (Vol. 1368, pp. 166-175). Springer. https://doi.org/10.1007/978-3-030-72660-7_16

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Educational Technology & Society, 17(4), 49-64.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135-1144). ACM. https://doi.org/10.1145/2939672.2939778

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

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. Computers in Human Behavior, 89, 98-110. https://doi.org/10.1016/j.chb.2018.07.027

Weidlich, J., Gašević, D., & Drachsler, H. (2022). Causal inference and bias in learning analytics: A primer on pitfalls using directed acyclic graphs. Journal of Learning Analytics, 9(3), 183-199. https://doi.org/10.18608/jla.2022.7577

Yılmaz, N., & Şekeroğlu, B. (2020). Student performance classification using artificial intelligence techniques. In R. A. Aliev, J. Kacprzyk, W. Pedrycz, M. Jamshidi, M. Babanli, & F. Sadikoglu (Eds.), 10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions (ICSCCW 2019) (Advances in Intelligent Systems and Computing, Vol. 1095). Springer.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? International Journal of Educational Technology in Higher Education, 16, 39. https://doi.org/10.1186/s41239-019-0171-0