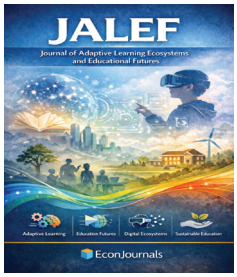


Journal of Adaptive Learning Ecosystems and Educational Futures

ISSN: 2146-4138

available at http: www.econjournals.com

Journal of Adaptive Learning Ecosystems and Educational Futures, 2025, 1(2), 1-8.

Behavioral Learning Analytics for Academic Risk Stratification in Smart Learning Platforms: A Reproducible Study Using the Public xAPI-Edu-Data Dataset

Mahmoud A. Zaher*

Data Science Department, Faculty of Artificial Intelligence, Horus University (HUE), Egypt. *Email: mzaher@horus.edu.eg

Received: 18 August 2025 **Accepted:** 05 November 2025 **DOI:** https://doi.org/10.32479/jalef.24053

ABSTRACT

Digital learning platforms generate rich behavioral traces that enable institutions to identify students who face academic challenges. Yet most institutional decisions continue to lack strong ties with research that can be tested and verified. The research investigates whether learning analytics createable through public xAPI- Edu-Data dataset access can assist in determining academic risk levels for students who study through digital platforms. The dataset contains 480 student records which include 17 variables that describe their demographic and behavioral and parental and attendance characteristics. The researchers tested three classification models which included logistic regression and random forest and extra trees. The team developed a Python-based analytical pipeline which used one-hot encoding and executed five-fold cross-validation and conducted hold-out testing. The extra trees model achieved the strongest cross-validated performance (accuracy = 0.7917, macro-F1 = 0.7959) and the best hold-out results (accuracy = 0.7986, macro-F1 = 0.8049, macro ROC-AUC = 0.9215). The feature-importance analysis revealed that student absence and resource visits and hand raises and announcement views and discussion participation and parental survey completion served as the main performance class predictors. Educational technology data provides an interpretable foundation for developing early warning systems which enable targeted academic support while conducting evidence-based student monitoring. The research presents three elements which include a reproducible analytical workflow and a comparative model assessment and an empirically grounded interpretation of behavioral indicators associated with academic performance.

Keywords: Learning Analytics, Educational Data Mining, Student Performance Prediction, Educational Technology, Academic Risk, Explainable Analytics
JEL Classification: I23

1. INTRODUCTION

Through digital transformation educational institutions now provide services and track student engagement while helping students achieve their academic goals. The combined use of learning management systems, student portals, communication dashboards, online assessments, and resource repositories results in the creation of extensive behavioral data which institutions can use to assess academic risk and distribute their support resources more effectively. Many institutions continue to treat their digital traces as operational exhaust instead of using them to build strategic evidence bases which support managerial decision-making.

Educational technology research shows that institutions can use digital platform behavioral data to predict student performance and engagement while assessing dropout risks (Hu et al., 2014; Jokhan et al., 2018; Flanagan et al., 2022; Dai et al., 2025). The field has developed from prediction accuracy assessment into research about explainability and ethics and trust and feedback and intervention design (Lim et al., 2021; Slade et al., 2019; Tzimas and Demetriades, 2024; Escolano-Perez and Losada, 2024). The gap continues to exist between method-driven research and research results which provide direct guidance to institutions about student assistance and attendance tracking and digital service development.

This research study demonstrates its findings through an analytic study which allows researchers to reproduce their results by using digital learning data to create managerial performance metrics. The study shows which activities on the educational platform help to detect academic risk while the study predicts future outcomes through its technical functions. The analysis uses the public xAPI-Edu-Data dataset introduced through prior work on xAPI-based educational mining (Amrieh et al., 2015; 2016) and compares three supervised learning models on the same data pipeline.

The study makes three contributions. First, it provides a reproducible empirical design based on a fully public dataset, thereby supporting transparency and independent verification. Second, it identifies an interpretable set of indicators associated with academic risk, including absence patterns, resource visitation, and participation intensity. The study shows how its research findings affect early warning systems, digital service design, and targeted intervention in educational settings.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature and summarizes related studies. Section 3 presents the research gap and conceptual framing. Section 4 describes the dataset and the methodology. Section 5 reports the empirical findings. Section 6 discusses implications for educational management and digital service strategy. The seventh section of the paper presents its conclusion together with its research limitations and research directions for future studies.

2. RELATED WORK

Educational data mining together with learning analytics has developed into an advanced system which includes methods for making predictions and conducting diagnostics and implementing interventions (Papadogiannis et al., 2024; Batool et al., 2023; Maulidiya et al., 2024). The initial research showed that online behavior patterns could forecast results for virtual and blended learning environments (Hu et al., 2014; Jokhan et al., 2018), while the latest studies have shown that behavior patterns together with explainable models and feedback systems and platform decisions create better results for academic research.

A major area of research is predicting performance. Hu et al. (2014) created early warning systems for online learning performance, demonstrating that learning-portfolio data can pinpoint students at risk during course delivery. In blended higher education, Jokhan et al. (2018) built on this idea by showing that an early warning system can give useful hints about how well students will do in a course. Simultaneously, Amrieh et al. (2016) illustrated that ensemble methods can enhance performance classification when behavioral features from an educational platform are incorporated. Recent studies persist in demonstrating the predictive significance of digital or admission-related attributes (Kaensar and Wongnin, 2023; Zahruddin et al., 2023).

A second stream focuses on the importance of learning behavior and feedback. Lim (2021) looked at how learning analytics feedback helps students learn on their own by combining behavioral traces with what they remember. Flanagan et al. (2022) employed reading-behavior analytics to facilitate early warning predictions regarding both performance and engagement. Tzimas and Demetriades (2024) stated that guidance derived from learning analytics can influence self-regulated learning, satisfaction, and performance outcomes.

A third stream is about trust and governance. According to Slade et al. (2019), disclosure, trust, and perceived benefit are still very important when it comes to using learning analytics. This is important for institutional deployment because predictive systems are not just technical systems; they are social and organizational arrangements that are built into policy, communication, and accountability.

A fourth stream suggests that AI-enabled analytics will become more common. Escolano-Perez and Losada (2024) utilized decision trees to discern significant profiles in secondary education. Borna et al. (2024) employed click data and AI-driven analysis to investigate performance prediction and learning assessment. Dai et al. (2025) advanced institutional action by amalgamating at-risk identification with feedback intervention. Foundational reviews and mapping studies also demonstrate the swift proliferation of educational data mining and intelligent learning environments (Batool et al., 2023; Papadogiannis et al., 2024; Maulidiya et al., 2024).

Table 1 summarizes representative studies relevant to this manuscript.

3. RESEARCH BACKGROUND AND ANALYTICAL FRAMEWORK

The empirical and methodological literature reviewed in Section 2 indicates that digitally mediated learning environments produce several classes of signals that are relevant to academic monitoring: participation intensity, resource-consumption behavior, attendance regularity, and selected indicators of family-school interaction. However, prior studies often treat these signals either as isolated correlates of performance or as inputs to prediction models without a sufficiently explicit analytical structure linking raw behavioral traces, statistical learning, and actionable institutional interpretation. The present study therefore formulates academic risk stratification as a reproducible multiclass learning problem in which educational activity records are transformed into a structured feature space and then mapped to performance categories representing high, medium, and low achievement.

From an analytical perspective, the problem can be expressed as follows. Let the dataset contain N student records, with the i^{th} record represented by a feature vector $x_i \in \mathbb{R}^p$ after preprocessing and encoding, and let the associated observed outcome be $y_i \in \{L, M, H\}$, where L, M, and H denote low, medium, and high performance, respectively. The objective is to estimate a predictive mapping,

$$\hat{f} : xi \rightarrow \hat{y}i \quad (1)$$

Table 1: Summary of representative related studies

Study	Context	Data/Platform	Method	Main contribution
Amrieh et al. (2015)	LMS-based school data	xAPI educational traces	Preprocessing+ predictive modeling	Introduced the xAPI-based educational dataset and demonstrated the value of behavioral variables for student-performance analysis.
Amrieh et al. (2016)	Educational platform data	xAPI-Edu-Data	Ensemble methods	Showed that behavioral features can improve student-performance prediction relative to simpler feature sets.
Hu et al. (2014)	Online higher education	Online learning portfolios	early warning analytics	Established that online behavioral traces can be used for early warning systems during course delivery.
Kuzilek et al. (2017)	Distance higher education	OULAD	Open dataset publication	Provided one of the most widely used open learning analytics datasets, strengthening reproducibility in the field.
Jokhan et al. (2018)	Blended higher education	Course warning signals	Predictive analytics	Demonstrated the usefulness of early warning systems for blended-course performance.
Slade et al. (2019)	Learning analytics governance	Institutional analytics context	Conceptual/ empirical conference study	Highlighted student trust, disclosure, and perceived benefit as crucial to the legitimacy of learning analytics systems.
Lim et al. (2021)	Learning analytics feedback	Behavioral logs+learner recall	Triangulated analysis	Connected analytics feedback to self-regulated learning processes rather than prediction alone.
Flanagan et al. (2022)	Open-book assessment	Digital reading behavior logs	Early warning models	PREDICTED both performance and engagement from learning-behavior signals.
Batool et al. (2023)	Broad EDM literature	Survey of prediction studies	Survey/review	Synthesized the student-performance prediction literature and clarified common methods, variables, and gaps.
Zahruddin et al. (2023)	Student demographics	Demographic attributes	Data mining case study	Showed that even limited structured attributes can support classification-oriented student-performance prediction.
Kaensar and Wongnin (2023)	University admissions	Admission data	Machine learning+tuning	Demonstrated early prediction with admission features and hyper parameter tuning in a university setting.
Tzimas and Demetriades (2024)	Higher education	Learning analytics guidance	Mixed methods	Linked learning analytics guidance to self-regulated learning, satisfaction, and performance.
Papadogiannis et al. (2024)	EDM field overview	Literature synthesis	overview/review	Positioned educational data mining as a foundational interdisciplinary area connecting AI, data mining, and learning support.
Escolano-Perez and Losada (2024)	Secondary education	Executive-function data	Decision trees	Illustrated the interpretive value of decision trees for explaining learning result profiles.
Maulidiya et al. (2024)	Smart learning environments	Bibliometric dataset	Bibliometric analysis	Showed the thematic evolution of smart learning environments and the growing role of analytics-driven personalization.
Borna et al. (2024)	Clickstream-based assessment	Click data	AI/ML analytics	Demonstrated how click-level digital traces can support performance prediction and assessment insights.
Dai et al. (2025)	At-risk student support	Institutional LA pipeline	Risk identification+ feedback intervention	Extended prediction toward intervention by linking at-risk identification with feedback mechanisms.

that minimizes classification error while preserving interpretability at the level of student-support policy and digital service intervention. In probabilistic terms, the task is to estimate the posterior class distribution,

$$P(Y = c \mid X = x_i),\ c \in \{L, M, H\} \tag{2}$$

and to assign each observation to the class with the largest posterior probability,

$$\hat{y}_i = \arg \max_{c \in \{L,M,H\}} P(Y = c \mid x_i) \tag{3}$$

This formulation is suitable for academic risk analytics because it yields both a discrete class decision and a probabilistic representation of uncertainty, which can be used to support escalation thresholds in student monitoring systems.

3.1. Problem Formulation and Analytical Logic

The predictor space is organized around four substantive domains. The first domain comprises behavioral engagement variables, including raised hands, visited resources, viewed announcements, and discussion activity. The second domain captures attendance regularity, represented by the recorded absence category. The third domain contains contextual educational descriptors, such as stage, grade level, topic, and semester. The fourth domain reflects family-related interaction indicators, including parental survey response and parental satisfaction with the school. Collectively, these domains form a behavioral-operational representation of the student learning context.

The analytical assumption is that the observed outcome Y is not driven by a single attribute but by a joint configuration of behavioral intensity, attendance discipline, contextual learning conditions, and family-engagement signals. This can be expressed through an unknown data-generating relationship,

$$Y = f(B, A, C, F) + \varepsilon \tag{4}$$

where B denotes the vector of behavioral engagement measures, A denotes attendance status, C denotes contextual academic

descriptors, F denotes family-related indicators, and e captures unobserved variation. Because the functional form of $f(\cdot)$ is not known a priori, the study compares a linear probabilistic benchmark with non-linear ensemble learners. This design allows the empirical analysis to distinguish between approximately additive relationships and higher-order interactions that may be present in student behavior data.

3.2. Mathematical Specification of the Predictive Models

Three complementary model families are used. The first is multinomial logistic regression, which provides an interpretable baseline by modeling the log-odds of class membership relative to a reference class. For class c, the model can be written as,

$$\log \frac{P(Y=c \mid x_i)}{P(Y=r \mid x_i)} = \beta_{0C} + x_i^{\top} \beta_c \quad (5)$$

where r is the reference class, β_{0c} is an intercept term, and β_c is the parameter vector for class c. The corresponding Softmax probability is,

$$P(Y=c \mid x_i) = \frac{\exp(\beta_{0c} + x_i^{\top} \beta_c)}{\sum_{k\in\{L,M,H\}} \exp(\beta_{0k} + x_i^{\top} \beta_k)} \quad (6)$$

This specification is useful for identifying directional linear effects after encoding categorical variables, but it may underrepresent interaction-rich structures in educational behavior data.

The second and third models are tree-based ensembles. Let $T_b(x_i)$ denote the class prediction from the b^{th} decision tree in an ensemble of size B. The ensemble decision is obtained by aggregating across trees,

$$\hat{y}_i = \text{mode}\{T_1(x_i), T_2(x_i), \ldots, T_B(x_i)\} \quad (7)$$

Random forest introduces stochasticity through bootstrap resampling and random feature selection at each split, thereby reducing variance and improving generalization. Extra trees increases randomization further by drawing split thresholds more aggressively, which often improves robustness when the predictor space contains mixed data types and potentially non-linear boundary structures. In operational terms, these ensemble models are attractive because they can absorb threshold effects such as abrupt changes associated with absenteeism categories while still preserving post hoc interpretability through feature-importance analysis.

Model evaluation follows a multiclass setting. For each class c, precision and recall are defined as,

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c} \qquad \text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \quad (8)$$

with the class-specific F1 score given by,

$$\text{F1}_c = \frac{2\text{Precision}_c \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \quad (9)$$

To avoid dominance by the majority class, the study emphasizes macro-averaged metrics,

$$\text{Macro-}F1 = \frac{1}{3} \sum_{c\in\{L,M,H\}} F1_c \quad (10)$$

with analogous averaging for precision and recall. This choice is appropriate because institutional monitoring systems must identify all performance groups reliably, including the low-performance group that is most relevant for support intervention.

3.3. Operational Framework for Academic Risk Stratification

Figure 1 summarizes the analytical framework used in the study. The framework begins with the raw educational records, proceeds through preprocessing and encoded feature construction, and then branches into comparative model estimation. The selected model is subsequently interpreted through performance diagnostics and feature-importance analysis, after which the results are translated into operational categories for attendance support, engagement reinforcement, and family-oriented intervention.

The framework is intentionally sequential. First, raw platform and contextual records are standardized into a machine-readable analytical matrix. Second, model comparison identifies the classifier with the most stable multiclass performance. Third, explanatory outputs are used to distinguish variables that primarily reflect engagement intensity from those that signal structural academic risk, such as recurrent absenteeism. Finally, these findings are mapped to institution-level decision processes. In that sense, the framework does not end with prediction; it ends with an interpretable evidence layer that can support student advising, digital communication escalation, and targeted academic monitoring.

4. MATERIALS AND METHODS

4.1. Dataset

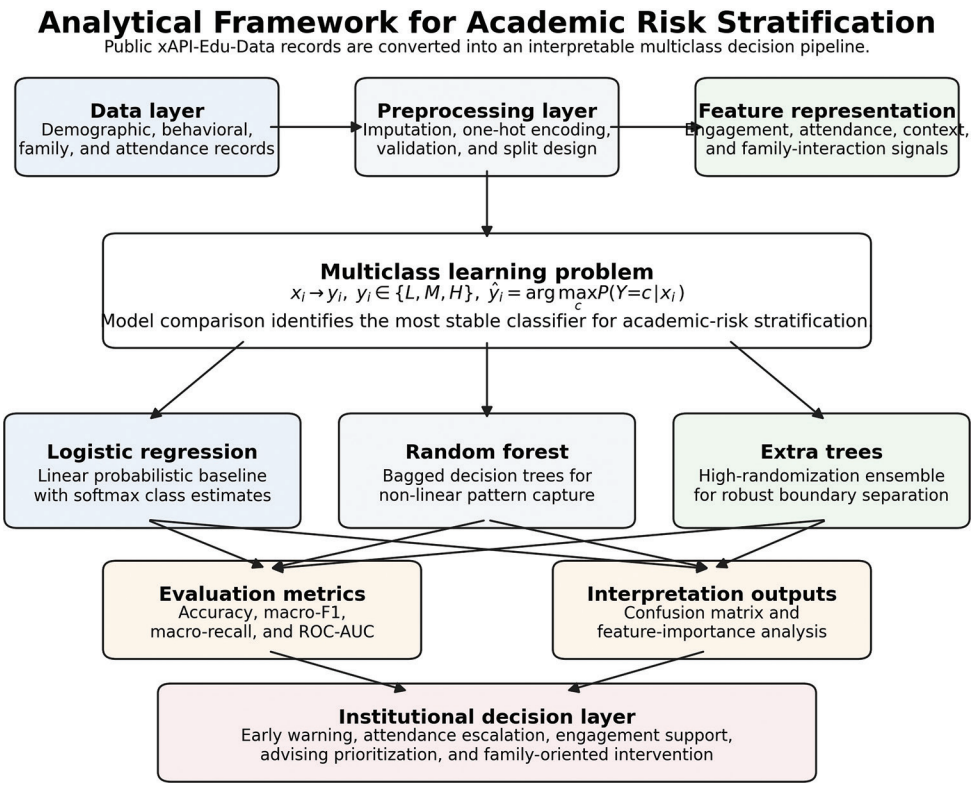
The empirical analysis uses the public *xAPI-Edu-Data* dataset, originally associated with xAPI-based educational mining work (Amrieh et al., 2015; 2016). The dataset contains 480 student observations and 17 variables. It combines demographic and school descriptors (e.g., gender, stage, grade, topic, semester), family-related variables (e.g., relation, parent answering survey, school satisfaction), behavioral variables (raised hands, visited resources, announcements viewed, discussion participation), attendance information, and a three-class performance target variable (H, M, L).

The dataset is well suited to reproducible analysis because it is compact enough for direct inspection while still containing a meaningful combination of demographic, behavioral, and school-related features. This structure enables examination of how observable learning behaviors and attendance patterns relate to academic performance.

4.2. Research Design

The study uses a supervised classification design. The dependent variable is academic performance class. The explanatory variables are all non-target fields in the dataset.

Figure 1: Analytical framework for reproducible academic risk stratification in smart learning platforms



Four numerical engagement variables were retained in numeric form: raisedhands, VisITedResources, AnnouncementsView, and Discussion. All remaining predictors were treated as categorical and encoded through one-hot encoding. Three models were compared:

- Logistic regression as a transparent linear baseline;
- Random forest as a strong non-linear ensemble baseline;
- Extra trees as a high-variance, ensemble-based benchmark.

4.3. Preprocessing and Evaluation

A reproducible scikit-learn pipeline was built with median imputation for numeric variables, mode imputation for categorical variables, and one-hot encoding for categorical predictors. Model comparison relied on stratified five-fold cross-validation using accuracy, macro-F1, macro-precision, and macro-recall. After model selection, the best-performing model was trained on a 70/30 train-test split to obtain hold-out metrics, a confusion matrix, class-specific performance values, one-versus-rest ROC curves, and feature-importance estimates.

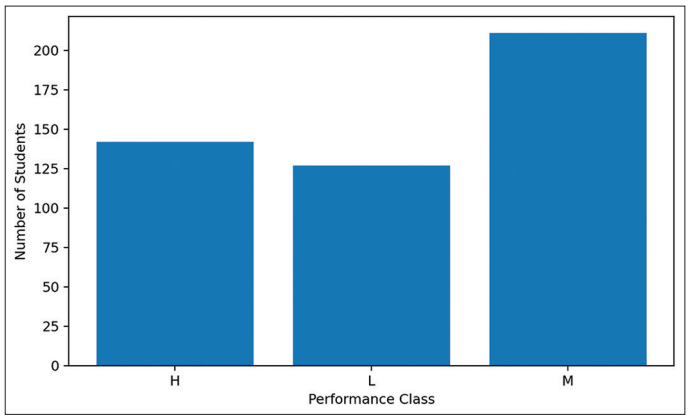
In addition to predictive modeling, the study reports descriptive means by class, chi-square tests for categorical attributes, and Kruskal-Wallis tests for numerical engagement indicators. This design allows the analysis to report not only classification performance but also inter-pretable group-level evidence.

5. RESULTS AND DISCUSSION

5.1. Descriptive Characteristics of the Sample

The dataset contains 480 observations. The class distribution is reasonably balanced for a three-class educational dataset: 211

Figure 2: Distribution of academic performance classes

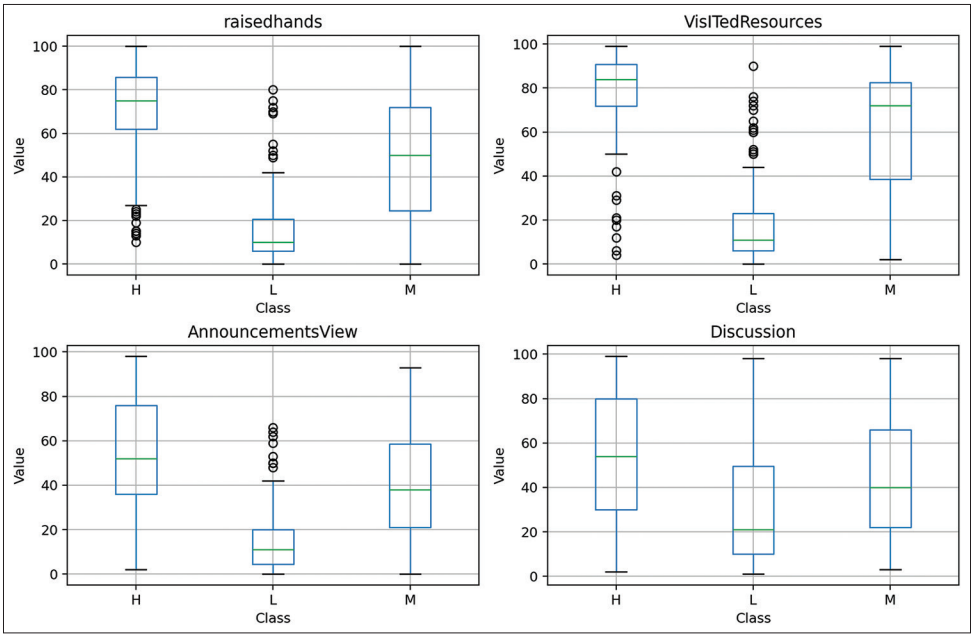


medium-performing students, 142 high-performing students, and 127 low-performing students. Figure 2 presents the class distribution.

Group means for the behavioral variables suggest a clear monotonic pattern. High performers record substantially higher values on raisedhands (70.29), VisITedResources (78.75), AnnouncementsView (53.38), and Discussion (53.66), while low performers record markedly lower values on the same dimensions (16.89, 18.32, 15.57, and 30.83, respectively). Figure 3 visualizes these differences.

Kruskal-Wallis testing confirms statistically significant class differences for all four engagement variables, with very small P-values. Chi-square testing also shows strong dependence between performance class and multiple categorical variables,

Figure 3: Engagement indicators by performance class



especially StudentAbsenceDays, ParentAnsweringSurvey, ParentschoolSatisfaction, and Relation.

5.2. Comparative Model Performance

Table 2 presents the five-fold cross-validation results. Extra trees achieved the highest macro-F1 and overall accuracy, followed closely by random forest. Logistic regression remained competitive but consistently weaker than the ensemble models (Figure 4).

5.3. Hold-out Performance of the Selected Model

The extra trees model was retained for final reporting. On the 30% hold-out set, it achieved an accuracy of 0.7986, macro-F1 of 0.8049, macro-precision of 0.8130, macro-recall of 0.7986, and macro one-versus-rest ROC-AUC of 0.9215. These values indicate that the model is not only accurate overall but also balanced across the three classes.

The confusion matrix in Figure 5 indicates that the model distinguishes the low-performing group especially well. Misclassifications occur primarily between the high and medium groups, which is common when class boundaries are educationally adjacent rather than sharply separated.

Class-level hold-out performance is also satisfactory. For class L, precision reaches 0.8889, recall 0.8421, and F1-score 0.8649. For class H, the model records precision of 0.8000, recall of 0.7442, and F1-score of 0.7711. For class M, precision is 0.7500, recall 0.8095, and F1-score 0.7786.

5.4. Predictor Importance and Interpretability

Figure 6 shows the most important predictors in the extra trees model. The top variables are StudentAbsenceDays Under-7, StudentAbsenceDays Above-7, VisITedResources, raisedhands, AnnouncementsView, Discussion, ParentAnsweringSurvey Yes, and family-related relation indicators.

Figure 4: Cross-validated model comparison

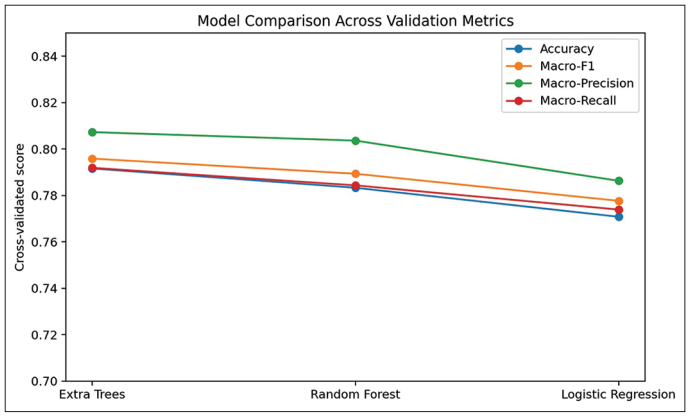


Table 2: Cross-validation performance by model

Model	Accuracy	Macro-F1	Macro-Precision	Macro-Recall
Extra Trees	0.7917	0.7959	0.8073	0.7919
Random Forest	0.7833	0.7894	0.8036	0.7843
Logistic Regression	0.7708	0.7777	0.7863	0.7739

The substantive meaning is important. Absence behavior is the single strongest signal, suggesting that attendance-related risk remains central even in digitally rich environments. Resource visitation and participatory signals such as raised hands and discussion activity further indicate that platform engagement carries meaningful information about academic outcomes. Finally, parent survey participation and school satisfaction suggest that family-school connection remains a relevant explanatory layer, even in a platform-centered dataset.

5.5. Discussion

The empirical findings demonstrate that behavioral learning analytics can function as an effective tool for academic risk

Figure 5: Confusion matrix and one-versus-rest ROC curves for the extra trees model

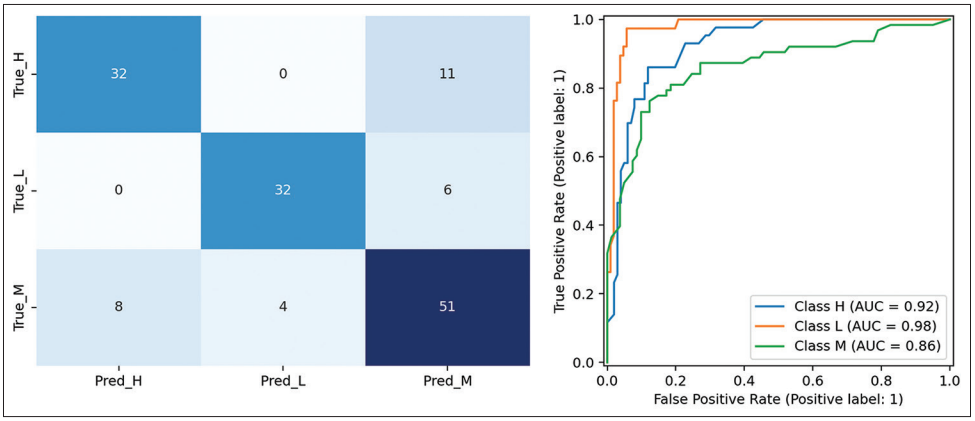
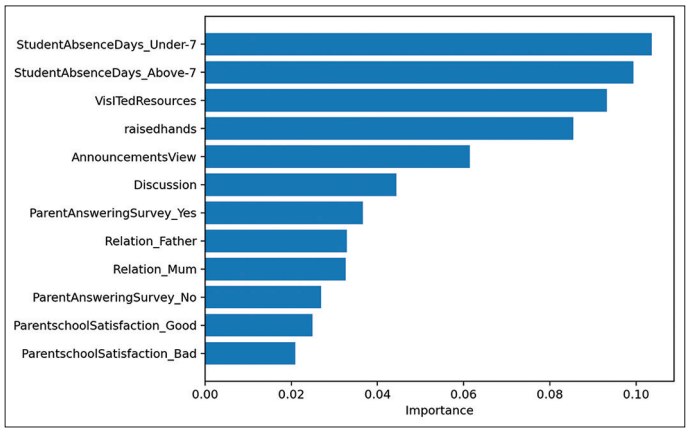


Figure 6: Top predictors of academic performance



stratification. The most reliable predictors are not obscure technical variables, but rather clear operational indicators such as absence, resource utilization, participation, announcements viewed, and parental response. This is significant because educational administrators are considerably more inclined to trust and implement a model when its indicators align with familiar institutional procedures.

The model's performance is also in line with what other researchers have found. The superiority of ensemble models over a simple linear baseline corresponds with research indicating that non-linear combinations of educational attributes frequently enhance predictive accuracy (Amrieh et al., 2016; Kaensar and Wongnin, 2023; Borna et al., 2024). The results do not indicate that increasingly intricate black-box models are invariably essential. The current extra trees model is effective and can be understood through feature ranking and confusion analysis.

Setting priorities is the most important thing for educational management to do right now. Institutions can't respond to all signals at once. The feature-importance profile gives a good order: First, watch how people act when they aren't there; second, get more people to use and interact with resources; and third, add communication that is directed at families when it makes sense. These results can help set dashboard thresholds, escalation workflows, advisor nudges, and redesign digital services when it comes to operations.

The findings further substantiate that learning analytics ought to be regarded as a socio-technical capability rather than merely a predictive tool. Previous research has focused on trust, feedback, self-regulation, and intervention design (Slade et al., 2019; Lim et al., 2021; Tzimas and Demetriades, 2024; Dai et al., 2025). The current findings enhance the existing literature by demonstrating that a public dataset can yield actionable management insights when the analysis is centred on institutional decisions rather than solely on algorithmic novelty.

6. CONCLUSION AND PRACTICAL IMPLICATIONS

This research analysed academic risk stratification utilising the public xAPI-Edu-Data dataset. By comparing logistic regression, random forest, and extra trees models, it showed that a small number of behavioural, attendance, and family-related variables can be used to accurately determine academic risk. The extra trees model gave the best overall results, with a hold-out macro-F1 of 0.8049 and a macro ROC-AUC of 0.9215.

The most significant predictors were days of absence, accessed resources, raised hands, viewed announcements, discussion engagement, and participation in parental surveys. These results show that data from educational technology can help with early intervention, setting service priorities, and supporting students digitally.

The limits of the research study become evident through its existing boundaries. The dataset contains insufficient data to provide complete coverage of a higher-education student information system. The target is a three-class performance label rather than a longitudinal retention outcome. Future research can extend this work by validating the same framework on larger open datasets such as OULAD, incorporating fairness analysis, and testing intervention- aware pipelines that move from prediction to action.

Several practical implications emerge from the analysis.

First, academic-risk monitoring can be simplified around a small number of high-yield indicators. Educational leaders do not necessarily need dozens of variables if absence and engagement signals already explain a substantial part of outcome variation.

Second, digital platform analytics can be integrated into service strategy. For example, low visitation of learning resources or weak announcement engagement can trigger a low-cost intervention before academic failure becomes visible through grades alone.

Third, family-related indicators in the dataset imply that communication systems matter. Even in digitally mediated learning environments, parent or guardian engagement may retain value as a support mechanism, especially in school or transition contexts.

Fourth, the reproducible nature of the analysis is itself useful for management research. Public data studies can serve as benchmarking exercises for institutions that are not yet ready to release internal data but still want an evidence-based starting point for policy or system design.

7. DATA AVAILABILITY

The analysis uses the public xAPI-Edu-Data dataset. A copy of the CSV file used in the analysis is included in the project package. The original dataset is associated with the xAPI educational mining work of Amrieh et al. (2015; 2016) and is publicly available through open repositories.

REFERENCES

Amrieh, E.A., Hamtini, T., Aljarah, I. (2015), Preprocessing and Analyzing Educational Data Set using X-API for Improving Student's Performance. In: Proceedings of the 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT). p1-5.

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.

Batool, S., Rashid, J., Nisar, M.W., Kim, J., Kwon, H.Y., Hussain, A. (2023), Educational data mining to predict students' academic performance: A survey study. Education and Information Technologies, 28(1), 905-971.

Borna, M.R., Saadat, H., Hojjati, A.T., Akbari, E. (2024), Analyzing click data with AI: Implications for student performance prediction and learning assessment. Frontiers in Education, 9, 1421479.

Dai, W., Lin, J., Jin, F.J.Y., Tsai, Y.S., Srivastava, N., Le Bodic, P., Gasevic, D., Chen, G. (2025), Learning analytics for early identification of at-risk students and feedback intervention. Journal of Learning Analytics, 12(3), 102-125.

Escolano-Perez, E., Losada, J.L. (2024), Using artificial intelligence in education: Decision tree learning results in secondary school students based on cold and hot executive functions. Humanities and Social Sciences Communications, 11(1), 1563.

Flanagan, B., Majumdar, R., Ogata, H. (2022), Early-warning prediction of student performance and engagement in open book assessment by reading behavior analysis. International Journal of Educational Technology in Higher Education, 19, 41.

Hu, Y.H., Lo, C.L., Shih, S.P. (2014), Developing early warning systems to predict students' online learning performance. Computers in Human Behavior, 36, 469-478.

Jokhan, A., Sharma, B., Singh, S. (2018), Early warning system as a predictor for student performance in higher education blended courses. Studies in Higher Education, 44(11), 1900-1911.

Kaensar, C., Wongnin, W. (2023), Predicting new student performances and identifying important attributes of admission data using machine learning techniques with hyperparameter tuning. EURASIA Journal of Mathematics, Science and Technology Education, 19(12), em2369.

Kuzilek, J., Hlosta, M., Zdrahal, Z. (2017), Open university learning Analytics dataset. Scientific Data, 4, 170171.

Lim, L.A., Gasevic, D., Matcha, W., Uzir, N.A., Dawson, S. (2021), Impact of Learning Analytics Feedback on Self-Regulated Learning: Triangulating Behavioural Logs with Students' Recall. In: Proceedings of the 11th International Conference on Learning Analytics and Knowledge (LAK '21). p364-374.

Maulidiya, D., Nugroho, B., Santoso, H.B., Hasibuan, Z.A. (2024), Thematic evolution of smart learning environments, insights and directions from a 20-year research milestones: A bibliometric analysis. Heliyon, 10 (5), e26191.

Papadogiannis, I., Wallace, M., Karountzou, G. (2024), Educational data mining: A foundational overview. Encyclopedia, 4(4), 1644-1664.

Slade, S., Prinsloo, P., Khalil, M. (2019), Learning Analytics at the Intersections of Student Trust, Disclosure and Benefit. In: Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19). p235-244.

Tzimas, D.E., Demetriadis, S.N. (2024), Impact of learning analytics guidance on student self-regulated learning skills, performance, and satisfaction: A mixed methods study. Education Sciences, 14(1), 92.

Zahruddin, N.A.B.M., Kamarudin, N.D., Mat Jusoh, R., Abdul Fataf, N.A., Hidayat, R. (2023), Case study: Using data mining to predict student performance based on demographic attributes. International Journal on Informatics Visualization, 7(4), 2460-2468.