



Enhancing High School Learning Management Systems Through Explainable AI-Based Behavioral Analytics for Personalized Learning Support

Nurhasanah

(Universitas Mercu Buana,
West Jakarta, Indonesia)

Yaya Sudarya Triana✉

(Universitas Mercu Buana, West
Jakarta, Indonesia)

Hadi Santoso

(Universitas Mercu Buana,
West Jakarta, Indonesia)

OPEN ACCESS

ARTICLE HISTORY

Received: June 2, 2026

Revised: July 21, 2026

Accepted: August 19,
2026

KEYWORDS

Behavioral learning
analytics;

Explainable AI;

Learning management
system;

Personalized learning
support;
SHAP.

ABSTRACT

Purpose – This study develops an Explainable AI (XAI)-enhanced Learning Analytics framework for analyzing student learning behaviors in a senior high school Learning Management System (LMS) and supporting retrospective performance-risk identification using score-independent behavioral features.

Methods – The study analyzed 57,095 LMS quiz activity records from 111 students collected between July 2024 and October 2025. Seven behavioral variables were initially engineered, while Answer Accuracy, Learning Consistency, and Learning Efficiency were excluded from predictive modeling because they are derived from score-related information. The final models used Quiz Frequency, Learning Duration, Activity Level, and Dominant Study-Time Pattern. Random Forest, Decision Tree, and Logistic Regression were evaluated, with SHAP TreeExplainer used for interpretation.

Findings – Decision Tree achieved the highest accuracy (73.9%), macro F1-score (0.638), and macro ROC-AUC (0.771) on the 23-student test set. Random Forest achieved 69.6% accuracy, 0.631 macro F1-score, and 0.669 macro ROC-AUC, while Logistic Regression achieved 56.5%, 0.455, and 0.656, respectively. Learning Duration had the highest mean absolute SHAP value.

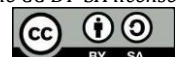
Research Implications – The results indicate that score-independent behavioral features provide a limited-to-moderate signal for retrospective performance classification. Prospective early-warning utility requires temporal validation.

Originality – The study integrates XAI with behavioral Learning Analytics while explicitly separating score-independent predictors from the score-derived target to support interpretable educational decision-making.

Correspondence Author: ✉ yaya.sudarya@mercubuana.ac.id

To cite this article : Nurhasanah, Triana, Y. S., & Santoso, H. (2026). Enhancing High School Learning Management Systems Through Explainable AI-Based Behavioral Analytics for Personalized Learning Support. *Journal of Deep Learning, Computer Vision and Digital Image Processing*, 4(3), 94–106. <https://doi.org/10.61255/decoding.v4i3.1676>

This is an open access article under the CC BY-SA license



INTRODUCTION

The rapid digitalization of education has catalyzed a paradigm shift in pedagogical practices. The integration of advanced information technologies has become a foundational pillar for creating flexible, student-centered, and sustainable learning environments in the era of Education 4.0.[1]. Central to this transformation is the widespread adoption of Learning Management Systems (LMS), which serve as the primary digital infrastructure for facilitating online learning activities and supporting interaction between teachers and students [2]. In the context of senior high schools, LMS platforms have become indispensable for streamlining academic administration, facilitating continuous assessments, and transcending the physical boundaries of traditional classrooms [3].

Beyond their administrative utility, LMS platforms possess a profound, yet frequently underutilized, capability: the automated generation of granular behavioral data. Every digital interaction—from login frequencies and resource navigation patterns to quiz response times and assignment submission punctuality is meticulously logged [4]. These digital footprints collectively construct a comprehensive profile of students' behavioral learning patterns, offering empirical insights into their cognitive engagement, learning consistency, and academic readiness [5]. Consequently, these behavioral indicators hold immense strategic value for educators seeking to understand the underlying mechanics of student performance beyond mere summative assessment scores [6].

Despite the proliferation of LMS deployments, the translation of this rich behavioral data into actionable pedagogical intelligence remains critically limited. In many senior high schools, LMS platforms are relegated to static repositories for content distribution and digital testing, devoid of analytical depth [7]. This underutilization highlights the potential value of Learning Analytics (LA), a discipline dedicated to measuring, collecting, and analyzing learner data to support educational processes [8]. By applying LA, educational institutions may be able to transition from reactive grading to more proactive, evidence-based decision-making, supporting the identification of students who may benefit from additional academic support [9]. Timely interventions, informed by such analytics, may help provide academic scaffolding before learning difficulties escalate [10].

However, the current landscape of Learning Analytics is fraught with a significant methodological bottleneck: the "black-box" nature of advanced machine learning models. While contemporary predictive algorithms achieve high accuracy, they often lack transparency, generating forecasts without elucidating the specific behavioral variables driving those outcomes [11]. In secondary education settings, where teachers require interpretable insights to inform instructional design, this opacity severely diminishes trust and practical utility [12]. Analytical systems that fail to explain their reasoning hinder pedagogical reflection and limit the integration of data-driven strategies in daily classroom practices [13].

To dismantle this barrier, the integration of Explainable Artificial Intelligence (XAI) into Learning Analytics emerges as a transformative solution. XAI methodologies, such as SHAP (SHapley Additive exPlanations), are engineered to demystify algorithmic decision-making by quantifying the contribution of individual features to a model's prediction [14]. In an educational context, XAI empowers teachers to comprehend precisely why a student is flagged as at-risk—whether due to declining quiz participation, erratic login patterns, or prolonged inactivity on specific modules [15]. This transparency bridges the gap between complex data science and practical pedagogy, fostering trust and supporting targeted instructional interventions. [16].

The synergy of Behavioral Patterns Learning Analytics and XAI is particularly potent for driving personalized learning and continuous examination enrichment in senior high schools. Personalized education necessitates adaptive systems that tailor content, pacing, and feedback to individual learner profiles [17]. By leveraging explainable behavioral analytics, LMS platforms can assist teachers in identifying students who may benefit from personalized learning support and more informed instructional planning based on observable behavioral patterns [18].

Nevertheless, a critical review of recent literature reveals prominent empirical lacunae. First, the majority of LA and XAI research is heavily skewed toward higher education, neglecting the distinct developmental and behavioral characteristics of senior high school students [19]. Second, existing

studies disproportionately prioritize predictive metrics over interpretability and pedagogical usability [20]. Third, there is a distinct scarcity of frameworks that integrate XAI specifically for personalized examination enrichment and continuous assessment preparation in secondary education [21]. Most current models primarily focus on final grade prediction rather than providing interpretable insights that support teachers in understanding students' learning behaviors and identifying those who may require additional academic support [22].

Addressing these gaps, this study develops and implements a Behavioral Patterns Learning Analytics framework augmented by Explainable Artificial Intelligence within a senior high school LMS environment. The primary objective is to analyze students' digital behavioral patterns for retrospective identification of performance risk and to provide interpretable information that may inform learning support. Specifically, this research aims to develop an interpretable predictive model that identifies students at potential academic risk and explains the behavioral factors associated with those predictions, thereby supporting evidence-based instructional decision-making [23].

Theoretically, this study advances the Learning Analytics discourse by embedding XAI into the behavioral analysis of secondary education, thereby enriching the literature on human-centered educational AI [24]. Practically, this study provides teachers with an interpretable, data-driven decision-support tool to facilitate the identification of students who may require additional academic support and to inform evidence-based instructional decision-making in Indonesian senior high schools [25].

METHOD

Research Design

This study employed a retrospective machine learning development and dashboard prototyping approach. The dashboard development was informed by selected Design-Based Research (DBR) principles concerning the development of practical educational solutions in authentic contexts [26]. However, the study does not claim to implement a full DBR cycle involving systematic co-design, repeated classroom implementation, stakeholder evaluation, and theory refinement. Instead, the analytical workflow comprised four major stages: data collection, preprocessing, model development, and dashboard implementation. In parallel, data-driven Learning Analytics emphasizes the systematic identification and analysis of patterns derived from educational data to support evidence-based pedagogical decision-making [27].

Participants and Data Collection

The participants in this study consisted of 111 senior high school students who actively utilized the school's Learning Management System (LMS) during the observation period between July 2024 and October 2025. A total sampling approach was employed, whereby all students with recorded learning activities in the LMS were included in the study. During this period, the LMS generated 57,095 raw quiz activity records, representing individual student interactions such as quiz participation, learning duration, answer accuracy, and learning outcomes [28]. These event-level records were subsequently aggregated through the feature engineering process to construct student-level behavioral profiles. Consequently, the final machine-learning dataset consisted of 111 student instances, with each instance representing one student's aggregated behavioral characteristics derived from the raw LMS activity records. Student identities were anonymized prior to data analysis to ensure confidentiality and protect participants' privacy throughout the research process [29].

All student identities were anonymized prior to data extraction and analysis, with each participant assigned a unique numerical identifier (`user_id`) to ensure confidentiality. The study was conducted with the knowledge and approval of the school administration, and data access was granted exclusively for academic research purposes. No personally identifiable information, including names, addresses, or contact details, was retained in the analytical dataset. The research adhered to ethical principles of data minimization, purpose limitation, and participant confidentiality, consistent with established guidelines for educational data research.

Instrumentation and Feature Engineering

The primary instrumentation relied on authentic LMS-generated behavioral log data rather than self-reported questionnaires. The raw data included user identification, quiz metadata, timestamps, quiz participation records, response accuracy, and learning duration. Through feature engineering, event-level records were aggregated into student-level behavioral profiles representing each student's learning behavior [28]. Seven engineered variables were initially defined: (1) Quiz Frequency total count of quiz sessions per student (unit: count); (2) Learning Duration mean duration per quiz session in minutes, computed as $\text{mean}((\text{time_end} - \text{time_start}) / 60)$ (unit: minutes); (3) Answer Accuracy percentage of correct responses out of all answered items, computed as $(\Sigma \text{ correct}) / (\Sigma \text{ correct} + \Sigma \text{ wrong}) \times 100\%$ (unit: %); (4) Learning Consistency stability of quiz scores computed as $100 - \sigma(\text{score})$, where σ denotes the standard deviation of quiz scores across sessions (unit: %); (5) Learning Efficiency ratio of mean quiz score to mean duration, computed as $\text{mean}(\text{score}) / \text{mean}(\text{duration})$ (unit: score per minute); (6) Activity Level average number of quiz sessions per week across the observation period (unit: sessions/week); and (7) Dominant Study-Time Pattern modal time-of-day category of quiz engagement, classified as Morning (00:00–11:59), Afternoon (12:00–16:59), or Evening (17:00–23:59).

Student performance was classified into three categories based on the average quiz score: Very Good (score ≥ 80), Moderate ($60 \leq \text{score} < 80$), and Needs Improvement (score < 60). Because Answer Accuracy, Learning Consistency, and Learning Efficiency are derived from quiz-response or quiz-score information that overlaps with the information used to define the target, these three variables were excluded from predictive modeling. The final predictive feature set therefore consisted of four score-independent behavioral variables: Quiz Frequency, Learning Duration, Activity Level, and Dominant Study-Time Pattern. The three excluded variables were not used as model inputs or SHAP predictors. Prior to model development, preprocessing procedures included duplicate removal, anomaly correction (score values exceeding 100 were clipped to 100), mean imputation for missing values, and IQR-based outlier treatment to improve data quality [28]. Preprocessing parameters were estimated using the training data only and subsequently applied unchanged to the held-out test data to prevent preprocessing leakage.

Machine Learning Models and Explainable AI

To predict student performance categories, three classification algorithms were implemented using Python 3.8+ and the Scikit-learn library [30]. All predictive features were derived exclusively from the four score-independent behavioral variables (Quiz Frequency, Learning Duration, Activity Level, and Dominant Study-Time Pattern). Random Forest was configured with $n_estimators=100$, $max_depth=5$, $class_weight="balanced"$, and $random_state=42$; Decision Tree with $max_depth=5$, $class_weight="balanced"$, and $random_state=42$; and Logistic Regression with $max_iter=1000$, $class_weight="balanced"$, and $random_state=42$. No hyperparameter optimization was performed; the reported parameters were specified a priori. Model selection was based on macro F1-Score because of the imbalanced multi-class target. The Decision Tree achieved the highest macro F1-Score (0.638) and was therefore selected as the primary predictive model. SHAP was applied using TreeExplainer to explain the Decision Tree predictions.

To improve the interpretability of the predictive models, SHapley Additive exPlanations (SHAP) was incorporated as the Explainable Artificial Intelligence (XAI) component. SHAP can support global and local interpretation; however, this study reports global feature importance using mean absolute SHAP values for the Decision Tree model, as presented in Figure 6. This analysis provides transparent information about the relative contribution of the behavioral features to the model's predictions [16], [17].

Random Forest Classifier

Random Forest was evaluated as an ensemble classifier capable of modeling non-linear relationships among behavioral variables. The method constructs multiple decision trees and aggregates their predictions using majority voting.

$$\hat{y} = \underset{c}{\operatorname{argmax}} \sum_{k=1}^k \mathbf{1}[h_k(x) = c] \quad (1)$$

In this study, the Random Forest model was implemented using 100 decision trees ($n_{\text{estimators}} = 100$) to classify students into three performance categories: Needs Improvement, Moderate, and Very Good. To address class imbalance, the `class_weight="balanced"` parameter was applied, which increases the relative training weight of minority classes, and a fixed `random_state=42` was set to ensure reproducibility.

Decision Tree Classifier

$$Gini(t) = 1 - \sum_{i=1}^c p_i^2 \quad (2)$$

Decision Tree was employed as a baseline classification model to evaluate the effectiveness of single-tree learning compared with ensemble learning approaches. The model was configured with `max_depth=5` to prevent overfitting, `class_weight="balanced"` to address class imbalance, and `random_state=42` for reproducibility.

Logistic Regression

$$P(y = c|x) = \frac{e^{z_c}}{\sum_{j=1}^c e^{z_j}} \quad (3)$$

Logistic Regression was utilized as a baseline linear classification model. Since the target variable consisted of three performance categories, the Softmax function was applied to estimate class probabilities. The model was configured with `max_iter=1000` to ensure convergence, `class_weight="balanced"` to address class imbalance, and `random_state=42` for reproducibility.

Model Evaluation Metrics

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

$$AUC = \int_0^1 TPR(FPR^{-1}(x)) dx \quad (6)$$

$$Macro AUC = (1/C) \times \sum AUC_c, \quad (7)$$

To identify the most suitable predictive model for behavioral learning analytics, the performance of Random Forest, Decision Tree, and Logistic Regression was evaluated using Accuracy, macro Precision, macro Recall, macro F1-Score, and macro ROC-AUC.

Procedures and Evaluation Metrics

For multiclass ROC-AUC, a one-vs-rest strategy was used. Class-specific AUC values were computed from the test-set probabilities, and macro ROC-AUC was obtained as their arithmetic mean: $Macro AUC = (1/C) \times \sum AUC_c$, for $c = 1, \dots, C$. The reported macro AUC therefore summarizes the three class-specific one-vs-rest AUC values equally.

RESULTS AND DISCUSSION

Student Performance Distribution and Class Imbalance

The analysis of student performance distribution revealed a significant concentration of students in the lower performance tier. Among the 111 participants, 63 students (56.8%) were categorized as Needs Improvement (average score < 60), 36 students (32.4%) as Moderate, and only 12 students (10.8%) as Very Good. This skewed distribution highlights the class imbalance in the dataset. To reduce potential bias toward the majority class, the `class_weight="balanced"` parameter was applied during model training.

Figure 1: Student Performance Distribution
(n=111 students, average quiz score threshold)

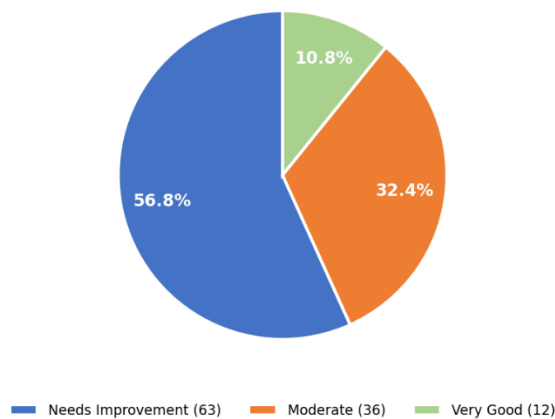


Figure 1. Student Performance Distribution

Figure 2 depicts the relationship between average learning duration per quiz session and average quiz scores across 111 students. A distinct behavioral trend emerged: students in the Needs Improvement category tended to exhibit longer average quiz durations despite lower scores, while students in the Very Good category generally completed quiz sessions in shorter time with higher scores. This negative association between duration and performance is consistent with the quantitative finding ($r = -0.231$, Figure 3), and may reflect that lower-performing students require more time per session due to greater difficulty with quiz content, rather than indicating greater effort or engagement. This visualization was used only for exploratory descriptive analysis and was not used as a predictive model input or to construct the target. Although average quiz score is displayed in Figure 2, it was not used to compute any of the four score-independent predictive features.

Figure 2 depicts the relationship between average learning duration per quiz session and average quiz scores across 111 students. A distinct behavioral trend emerged: students in the Needs Improvement category tended to exhibit longer average quiz durations despite lower scores, while students in the Very Good category generally completed quiz sessions in shorter time with higher scores. This negative association between duration and performance is consistent with the quantitative finding ($r = -0.231$, Figure 3), and may reflect that lower-performing students require more time per session due to greater difficulty with quiz content, rather than indicating greater effort or engagement. This underscores that behavioral time-on-task metrics must be interpreted cautiously in educational analytics contexts. This visualization was used only for exploratory descriptive analysis and was not used as a predictive model input or for target construction.

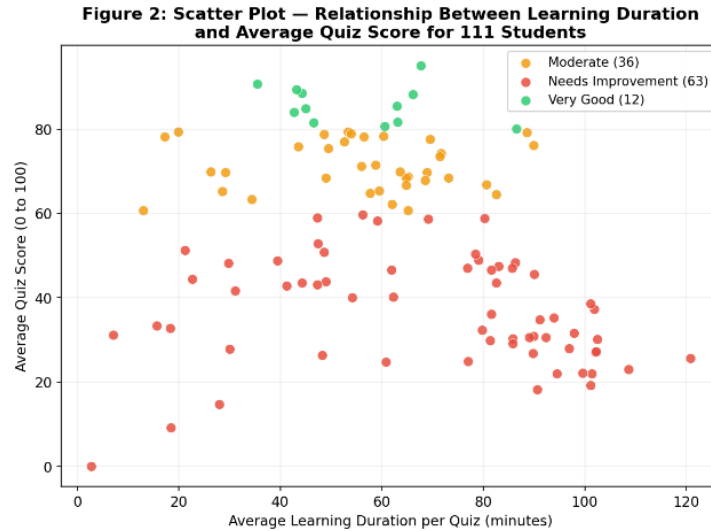


Figure 2. Relationship Between Learning Duration and Average Quiz Score

An exploratory Pearson correlation analysis was conducted between the four score-independent behavioral features and the ordinally encoded performance labels (Needs Improvement = 0, Moderate = 1, Very Good = 2) to assess direction and approximate association. This analysis was descriptive only and was not used for model training, feature selection, or causal inference. Because Dominant Study-Time Pattern is categorical, its encoded association should be interpreted as exploratory rather than as a continuous linear relationship.

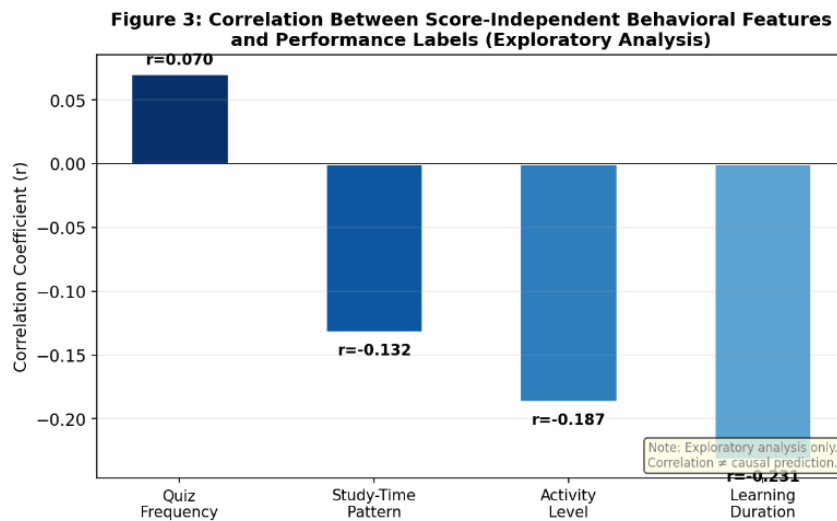


Figure 3. Correlation Between Behavioral Features and Performance Labels

Predictive Model Evaluation

Three machine learning algorithms Random Forest (RF), Decision Tree (DT), and Logistic Regression (LR) were evaluated using four score-independent behavioral features to identify the optimal predictive model. All features were derived from behavioral engagement data independent of the quiz score variable that defines the target label. Table 1 summarizes the comparative performance metrics on the test set ($n = 23$ students).

Table 1. Comparative Performance of Machine Learning Models

Model	Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	ROC-AUC (Macro)
Logistic Regression	0.565	0.472	0.463	0.455	0.656
Decision Tree	0.739	0.657	0.631	0.638	0.771
Random Forest	0.696	0.775	0.606	0.631	0.669

The Decision Tree classifier achieved the best overall performance among the evaluated models, with an accuracy of 73.9%, macro F1-score of 0.638, and macro ROC-AUC of 0.771. Random Forest achieved 69.6% accuracy, 0.631 macro F1-score, and 0.669 macro ROC-AUC, while Logistic Regression achieved 56.5%, 0.455, and 0.656, respectively. The results indicate that the score-independent behavioral features provide a limited-to-moderate predictive signal for performance categories defined by quiz scores. The Decision Tree was therefore selected as the primary model for subsequent confusion-matrix, ROC, and SHAP analyses.

The classification performance of the Decision Tree model is detailed in the Confusion Matrix (Figure 4), evaluated on the test set ($n = 23$ students). The model correctly classified 11 of 13 Needs Improvement students, 5 of 7 Moderate students, and 1 of 3 Very Good students, yielding 17 correct predictions out of 23 (Accuracy = 73.9%). Misclassifications occurred primarily between adjacent performance categories: 2 Needs Improvement students were predicted as Moderate, 1 Moderate student as Needs Improvement, 1 Moderate student as Very Good, and 1 Very Good student as Needs Improvement and 1 as Moderate. No misclassification occurred between the Needs Improvement and Very Good categories. However, the model was not designed or optimized as an ordinal classifier, and formal ordinal metrics such as weighted kappa were not computed. Given the small test set ($n = 23$), these results should be interpreted with caution.

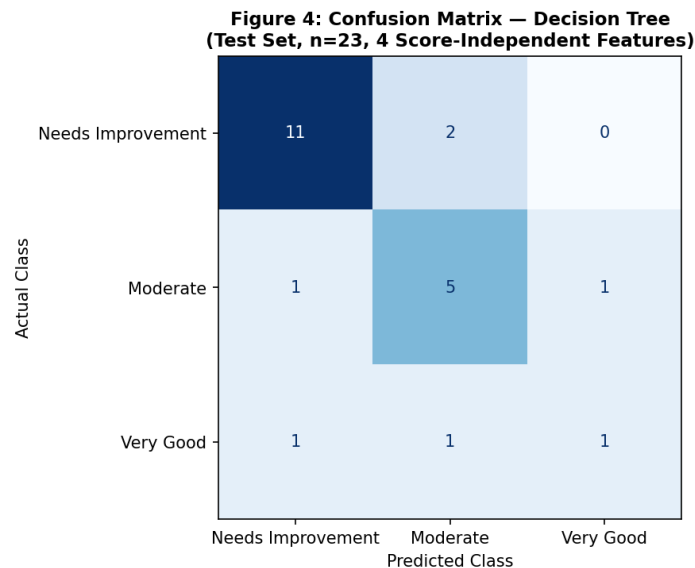


Figure 4. Confusion Matrix of Decision Tree Classifier (Test Set, $n=23$, 4 Score-Independent Features)

The discriminative power of the Decision Tree model across all three classes is corroborated by the ROC curves (Figure 5). The class-specific AUC values are $AUC(\text{Needs Improvement}) = 0.846$, $AUC(\text{Moderate}) = 0.826$, and $AUC(\text{Very Good}) = 0.642$. Their arithmetic mean is $(0.846 + 0.826 + 0.642) / 3 = 0.771$, consistent with the macro ROC-AUC reported in Table 1. The Needs Improvement and Moderate classes show moderate separation, whereas the Very Good class shows weaker separation ($AUC = 0.642$), likely influenced by the small number of Very Good students in the test set ($n = 3$). These results indicate limited-to-moderate discriminative performance rather than excellent multiclass discrimination.

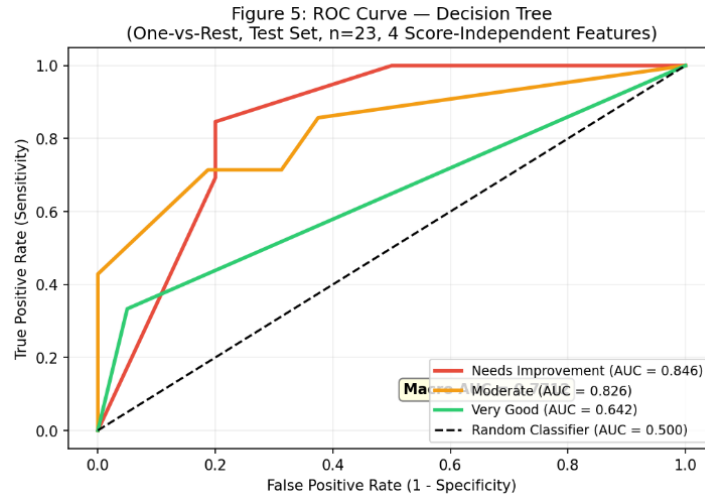


Figure 5. ROC Curves for Multiclass Student Performance Prediction (Decision Tree, One-vs-Rest, n=23)

Explainable AI (XAI) Analysis

SHAP (SHapley Additive exPlanations) was integrated to quantify the contribution of each score-independent behavioral feature to the Decision Tree model's predictions and to support interpretation of the model for educational decision-making. Figure 6 presents the global SHAP feature importance obtained from the Decision Tree model trained on four score-independent behavioral features, displayed as mean absolute SHAP values aggregated across all 111 students. Learning Duration contributed the highest mean absolute SHAP value (0.3183), substantially dominating the other features. Quiz Frequency (0.0427) and Activity Level (0.0361) showed comparatively smaller contributions, while Study-Time Pattern showed the lowest contribution (0.0109). The dominance of Learning Duration suggests that the amount of time students invest per quiz session is the most influential behavioral indicator distinguishing performance categories among score-independent features. It is acknowledged that the overall predictive contribution of these four features is limited, as reflected in the macro ROC-AUC of 0.771, which indicates limited-to-moderate discriminative performance. This finding is consistent with the theoretical expectation that score-independent behavioral features provide a partial, rather than complete, signal for predicting academic performance categories defined by quiz scores.

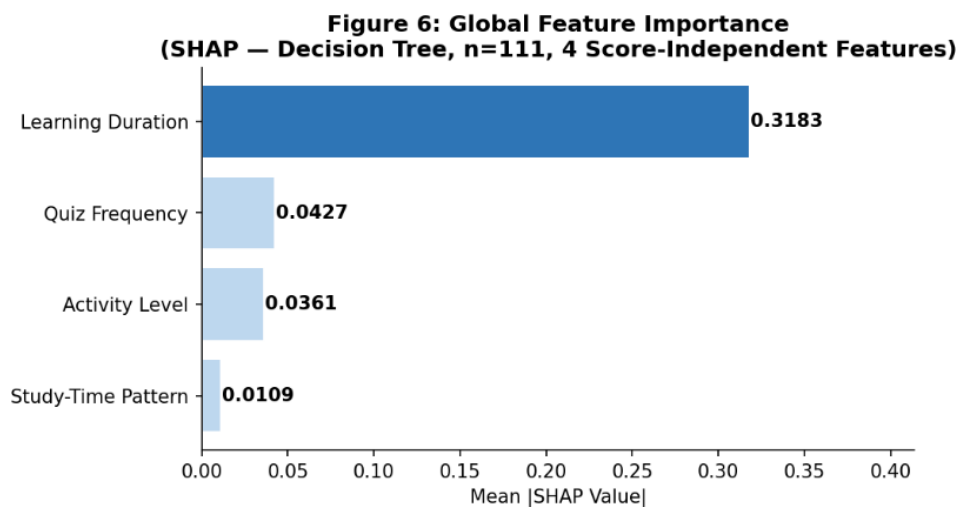


Figure 6. Global Feature Importance — SHAP (Decision Tree, n=111 students, 4 Score-Independent Features)

The global SHAP analysis provides interpretable information regarding the relative importance of the behavioral factors associated with the Decision Tree predictions. These findings may assist teachers in interpreting students' observed learning behaviors, but they do not by themselves demonstrate the effectiveness of any particular intervention.

Early Warning System Implementation

To operationalize the analytical findings, the Decision Tree predictive model and SHAP explanations were embedded into a web-based dashboard prototype (Figure 7). The dashboard implements a retrospective risk-prioritization mechanism: a student is identified as potentially at-risk when the predicted probability for the Needs Improvement category exceeds 50%, based exclusively on the four score-independent behavioral features (Quiz Frequency, Learning Duration, Activity Level, and Dominant Study-Time Pattern). The 50% threshold was manually specified and was not empirically optimized in this study. Answer Accuracy, Learning Consistency, and Learning Efficiency are displayed only as descriptive indicators and are not used as predictive inputs. Because the prototype operates on retrospective aggregated data and has not been validated using a prospective temporal design, it should be regarded as a retrospective behavioral monitoring and risk-prioritization prototype rather than a validated prospective early warning system. Future work should implement temporal validation, threshold optimization using sensitivity/specificity analysis, and prospective evaluation.



Figure 7. Early Warning System Dashboard and Risk Map

By visualizing these behavioral patterns through the dashboard, the prototype provides educators with a structured overview of students' historical learning trajectories, supporting more informed pedagogical decision-making. The proposed dashboard may assist teachers in monitoring students' accumulated behavioral learning patterns and identifying learners who may benefit from additional academic support, such as personalized remedial materials or additional guidance sessions. It is important to emphasize that these capabilities operate on retrospective behavioral data and should not be interpreted as real-time monitoring or prospective prediction. Future work should evaluate whether early-semester behavioral data can prospectively predict end-of-semester outcomes.

Several methodological limitations must be acknowledged. First, the dataset originated from a single senior high school, which may constrain the generalizability of the findings to broader educational contexts [28]. Second, the study relied exclusively on LMS quiz interaction data, excluding external variables such as socioeconomic background or psychological factors. Third, while XAI significantly improved model transparency, the recommendation engine remained rule-based rather than dynamically adaptive. Fourth, and critically, this study employed a random stratified train-test split rather than a temporal validation strategy. In an operational early warning system, behavioral data from an earlier observation window (e.g., July–August 2025) would be used to predict student outcomes in a subsequent independent period (e.g., September–October 2025). The absence of temporal validation means that the reported predictive performance metrics may overestimate real-world prospective utility, since the model was trained and tested on data from the same overall observation period. Future research should implement temporally structured validation to establish the genuine prospective validity of the proposed framework. Fifth, answer accuracy was excluded

from the feature set to avoid potential target leakage, which may have reduced predictive performance compared to prior studies that included this feature. Despite these constraints, the study demonstrates the practical viability of integrating Behavioral Patterns Learning Analytics and XAI to support transparent, data-driven educational decision-making in secondary education [28].

CONCLUSION

This study developed a Behavioral Patterns Learning Analytics framework augmented by Explainable Artificial Intelligence (XAI) to analyze students' digital learning behaviors within a senior high school LMS. To address target leakage, three score-derived features (Answer Accuracy, Learning Consistency, and Learning Efficiency) were excluded from the final predictive feature set. The models were trained on four score-independent behavioral features: Quiz Frequency, Learning Duration, Activity Level, and Dominant Study-Time Pattern. Among the evaluated algorithms, the Decision Tree classifier achieved the highest accuracy (73.9%), macro F1-score (0.638), and macro ROC-AUC (0.771) on the test set ($n = 23$), compared with Random Forest (69.6%, 0.631, 0.669) and Logistic Regression (56.5%, 0.455, 0.656). The limited-to-moderate discriminative performance indicates that score-independent behavioral features provide only a partial signal for predicting performance categories defined by quiz scores. SHAP analysis identified Learning Duration as the most influential feature by mean absolute SHAP value (0.3183). The study remains limited by its single-institution dataset, rule-based recommendation mechanism, and random rather than temporal train-test split. Future research should validate the framework using temporally structured train-test separation, larger multi-institutional datasets, and prospective outcomes that are temporally independent of the behavioral observation window.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to SMK Negeri 4 Padalarang for facilitating the implementation of this research. The authors also thank the participating teachers and students for their cooperation and valuable contributions throughout the study. Special appreciation is extended to the supervisors for their guidance and constructive feedback during the research process.

AUTHOR CONTRIBUTION STATEMENT

NO: Conceptualization, methodology, software development, investigation, data collection, formal analysis, visualization, and writing-original draft preparation. YP: Investigation, data collection, software development support, and writing-review and editing. MV: Methodology support, formal analysis, visualization, and writing-review and editing. LR: Supervision, validation, methodology review, and writing-review and editing. All authors have read and approved the final version of the manuscript.

AI DISCLOSURE STATEMENT

The author used ChatGPT (OpenAI) during the preparation of this work for language editing and translation purposes. After using the tool, the author thoroughly reviewed and edited the content as needed and takes full responsibility for the content of the publication.

REFERENCES

- [1] Mukul, E., & Büyüközkan, G. (2023). Digital transformation in education: A systematic review of education 4.0. *Technological forecasting and social change*, 194, 122664.
- [2] Bradley, V. M. (2021). Learning Management System (LMS) use with online instruction. *International Journal of Technology in Education*, 4(1), 68-92.
- [3] Nirsal, N. (2025). Digital Transformation Through Learning Management System Implementation In Secondary Schools: Analysis Of Impact, Challenges, And Adaptation Strategies. *MSJ: Majority Science Journal*, 3(3), 268-275.

- [4] Jo, I. H., Park, Y., Yoon, M., & Sung, H. (2016). Evaluation of online log variables that estimate learners' time management in a Korean online learning context. *International Review of Research in Open and Distributed Learning*, 17(1), 195-213.
- [5] Matcha, W., Uzir, N. A. A., Gašević, D., & Pardo, A. (2019). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE transactions on learning technologies*, 13(2), 226-245.
- [6] Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and higher education*, 28, 68-84.
- [7] Tsai, Y. S., Moreno-Marcos, P. M., Tammets, K., Kollom, K., & Gašević, D. (2018, March). SHEILA policy framework: informing institutional strategies and policy processes of learning analytics. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 320-329).
- [8] Palanci, A., Yilmaz, R. M., & Turan, Z. (2024). Learning analytics in distance education: A systematic review study. *Education and Information Technologies*, 29(17), 22629-22650.
- [9] Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational technology research and development*, 68(4), 1961-1990.
- [10] Yau, J. Y. K., & Ifenthaler, D. (2024). Human-centred learning analytics and AI in education: A systematic review. *Computers and Education: Artificial Intelligence*, 6, 100203.
- [11] Gunasekara, S., & Saarela, M. (2024). Explainability in educational data mining and learning analytics: An umbrella review. In *Proceedings of the 17th International Conference on Educational Data Mining* (pp. 887-892).
- [12] Türkmen, G. (2025). The review of studies on explainable artificial intelligence in educational research. *Journal of Educational Computing Research*, 63(2), 277-310.
- [13] Holstein, K., Aleven, V., & Rummel, N. (2020, June). A conceptual framework for human-AI hybrid adaptivity in education. In *International conference on artificial intelligence in education* (pp. 240-254). Cham: Springer International Publishing.
- [14] Gomez, M. J., Armada Sanchez, A., Albaladejo-González, M., Garcia Clemente, F. J., & Ruipérez-Valiente, J. A. (2025). Utilising explainable AI to enhance real-time student performance prediction in educational serious games. *Expert Systems*, 42(3), e70008.
- [15] Janahi, Y., & Obeidat, M. (2025). Predictive Analytics and AI in Education A Systematic Literature Review on Identifying and Supporting At-Risk Students.
- [16] Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-designing a real-time classroom orchestration tool to support teacher-AI complementarity. *Grantee Submission*.
- [17] Khor, E. T., & K, M. (2023). A systematic review of the role of learning analytics in supporting personalized learning. *Education sciences*, 14(1), 51.
- [18] Hur, P., Lee, H., Bhat, S., & Bosch, N. (2022). Using machine learning explainability methods to personalize interventions for students. In *Proceedings of the 15th International Conference on Educational Data Mining* (p. 438)..
- [19] Paolucci, C., Vancini, S., Bex II, R. T., Cavanaugh, C., Salama, C., & de Araujo, Z. (2024). A review of learning analytics opportunities and challenges for K-12 education. *Heliyon*, 10(4).
- [20] Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., & De Laet, T. (2020, March). Learning analytics dashboards: The past, the present and the future. In *Proceedings of the tenth international conference on learning analytics & knowledge* (pp. 35-40).

- [21] Altukhi, Z. M., & Pradhan, S. (2025). Systematic literature review: Explainable AI definitions and challenges in education. *arXiv preprint arXiv:2504.02910*.
- [22] Romero, C., & Ventura, S. (2024). *Educational data mining and learning analytics: An updated survey*. *arXiv*.
- [23] Tiukhova, E., Vemuri, P., López Flores, N., Islind, A. S., Óskarsdóttir, M., Poelmans, S., Baesens, B., & Snoeck, M. (2024). *Explainable Learning Analytics: Assessing the Stability of Student Success Prediction Models by Means of Explainable AI*. *Decision Support Systems*, 182, 114229.
- [24] Alfredo, R., Echeverria, V., Jin, Y., Yan, L., Swiecki, Z., Gašević, D., & Martinez-Maldonado, R. (2024). Human-centred learning analytics and AI in education: A systematic literature review. *Computers and Education: Artificial Intelligence*, 6, 100215.
- [25] Cukurova, M., & Miao, F. (2024). *AI competency framework for teachers*. UNESCO Publishing.
- [26] McKenney, S., & Reeves, T. C. (2021). Educational design research: Portraying, conducting, and enhancing productive scholarship. *Medical education*, 55(1), 82-92.
- [27] Drugova, E., Zhuravleva, I., Zakharova, U., & Latipov, A. (2024). Learning analytics driven improvements in learning design in higher education: A systematic literature review. *Journal of Computer Assisted Learning*, 40(2), 510-524.
- [28] Bergdahl, N., Bond, M., Sjöberg, J., Dougherty, M., & Oxley, E. (2024). Unpacking student engagement in higher education learning analytics: a systematic review. *International Journal of Educational Technology in Higher Education*, 21(1), 63.
- [30] Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, 2(3), 160.
- [29] Stinar, F., Xiong, Z., & Bosch, N. (2024). An Approach to Improve "k"-Anonymization Practices in Educational Data Mining. *Journal of Educational Data Mining*, 16(1), 61-83.