



AN AI-BASED PERSONALIZED LEARNING SYSTEM

TUSHAR MHARGUDE¹

¹ Department of Engineering

Corresponding author: tushar@gmail.com

ARTICLE ID IJREK-2026-00008	ARTICLE TYPE Research Article	RESEARCH AREA Information Technology	PUBLICATION Volume 1, Issue Issue 1 (2026)
RECEIVED 02 October 2026	REVISED 05 October 2026	ACCEPTED 02 October 2026	PUBLISHED 14 October 2026

Abstract

Personalized learning requires more than predicting whether a learner will pass or fail. A useful learning system must connect performance evidence to topic-level diagnosis, an appropriate level of challenge, and recommendations that a learner can understand and act upon. This paper proposes an AI-based personalized learning framework that combines student-performance prediction, topic-wise weakness detection, recommendation rules, and an explainability layer. The framework is designed to integrate previous examination marks, quiz scores, assignment performance, attendance, number of attempts, learning-activity completion, and time-on-task when these variables are available. For empirical benchmarking, the verified UCI Student Performance dataset is selected as a reproducible public baseline. The UCI repository reports 649 instances and 33 attributes; the dataset supports classification and regression and contains final grade G3 together with prior-period grades and school, demographic, social, and study-related variables. Because the public dataset does not contain genuine LMS event logs or topic-tagged assessment records, it is used only for the performance-prediction component. Topic-level recommendation is specified as a deployable framework requiring richer course/LMS data. Five supervised models—Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and Support Vector Machine—are proposed for controlled comparison using stratified cross-validation and Accuracy, Precision, Recall, F1-score, ROC-AUC, and confusion matrices. Recommendation quality is to be evaluated separately with Precision@K, Recall@K, and NDCG@K after implementation. No experimental accuracy or recommendation score is reported in this manuscript because the dataset could not be downloaded into the execution environment and no completed experiment is being represented as performed. The main contribution is therefore a reproducible research design that explicitly links prediction to actionable, explainable learning recommendations while identifying the data requirements and evaluation boundaries of the public benchmark.

Keywords: personalized learning; educational data mining; student performance prediction; machine learning; learning analytics; recommendation systems; explainable AI; student risk detection

1.2 PROBLEM STATEMENT

Conventional student-performance systems commonly aggregate marks into a single score or display descriptive dashboards. Such approaches can overlook localized weaknesses and may provide limited guidance for what a student should do next. The research problem addressed here is: How can heterogeneous student-performance indicators be transformed into a predictive and explainable personalized learning pathway that identifies performance risk, detects weak topics, and recommends an appropriate learning action without overstating what the available data can support?

1.3 Motivation Three practical considerations motivate the framework. First, early identification of students who may need support can make intervention more timely. Second, topic-level diagnosis is more actionable than a single aggregate grade because learning resources are organized around concepts, units, or skills. Third, explainability matters in education because students and educators need to understand the evidence behind a recommendation. Research on explainable learning analytics has emphasized the need to move from prediction toward understandable, actionable feedback [5]–[7].

1.4 RESEARCH GAP

Prior research demonstrates extensive use of classification and regression for student-performance prediction, with Decision Trees, Support Vector Machines, Random Forests, and other ensemble approaches appearing frequently [1]–[4]. Other work has advanced explainable or prescriptive learning analytics, including explanations of predictive decisions and automated action recommendations [5], [6]. Nevertheless, three gaps remain relevant to the present study. First, predictive benchmarking and personalized recommendation are often evaluated as separate

problems. Second, many public benchmarks lack the topic-tagged assessment and fine-grained activity data required to validate genuine topic recommendations. Third, student-facing dashboards can emphasize analytics without sufficiently connecting predictions to pedagogically meaningful next actions [8], [9]. The proposed framework addresses these gaps by defining an explicit interface between prediction, topic diagnosis, recommendation, and explanation. It also makes the data limitation explicit: a benchmark dataset can validate the predictive module, but it cannot be used to claim validated topic-level personalization unless topic-level ground truth is available.

1.5 Objectives Design a student-performance analysis framework that accepts academic and learning-activity indicators. Compare Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, and Support Vector Machine models for performance-category prediction. Define a topic-level weakness-detection mechanism using assessment evidence mapped to learning topics. Develop a personalized recommendation mechanism that selects topics, resources, and difficulty levels from the learner state. Provide interpretable reasons for recommendations using feature contributions and rule-based diagnostic evidence. Specify reproducible evaluation protocols for prediction and recommendation quality without fabricating experimental outcomes.

1.6 RESEARCH CONTRIBUTIONS

Integrated framework: a modular pipeline linking student data, prediction, topic diagnosis, recommendation, and dashboard delivery. Prediction benchmark design: a controlled comparison of five widely used supervised learning algorithms. Topic-level diagnostic layer: a formal mechanism for converting item/topic scores into weakness and mastery indicators. Explainable recommendation: each recommendation is paired with evidence such as low topic mastery, repeated errors, or an observed performance gap. Data-boundary specification: an explicit distinction

between what the verified public benchmark can test and what requires richer LMS/course data.

2.1 EXISTING PERSONALIZED LEARNING SYSTEMS

Personalized and adaptive learning systems seek to adjust content, feedback, sequencing, or difficulty to learner characteristics and observed performance. Earlier intelligent tutoring research established the value of adapting instructional interaction to individual learners, while contemporary systems increasingly use data-driven learning analytics. The current research direction extends personalization from static learner profiles toward continuously updated evidence derived from assessments and learning activities. A critical distinction is between personalization as content selection and personalization as decision support. A content recommender may select an item similar to a learner's past interactions; a pedagogical system should additionally consider mastery, prerequisites, difficulty, recency, and the intended learning objective. This motivates the proposed architecture, in which the recommendation engine receives a structured learner state rather than only a model probability.

2.2 EDUCATIONAL DATA MINING

EDM applies data-mining and machine-learning techniques to educational data for tasks such as prediction, clustering, knowledge estimation, and pattern discovery. Reviews have identified previous grades, class performance, e-learning activity, demographic information, and social variables among recurring feature groups [2], [3]. The choice of variables is context dependent: a variable that predicts performance in one institution may not generalize to another because assessment design, curriculum, learner population, and platform behavior differ.

2.3 MACHINE LEARNING FOR STUDENT PERFORMANCE PREDICTION

Student-performance prediction has been investigated using linear and logistic models, decision trees, support vector machines, nearest-neighbor methods, neural networks, and ensemble learning. Zhang et al. [1] reviewed the student-performance-prediction pipeline from data collection through practical application and emphasized the importance of fair comparisons and feature selection. Roslan and Chen [2] similarly reported the frequent use of classification methods and Decision Trees in the literature. Alwarthan [4] found that academic variables such as assessment grades and CGPA are commonly used, while ensemble approaches are frequently reported in comparative studies. The UCI Student Performance dataset has also been used as a public benchmark. Cortez and Silva originally studied secondary-school performance using demographic, social, school, and grade-related variables [10]. The UCI repository notes that predicting final grade G3 without earlier grades G1 and G2 is more difficult, but can be more useful, highlighting an important methodological distinction between retrospective prediction and early-warning prediction [11].

2.4 RECOMMENDATION SYSTEMS IN EDUCATION

Educational recommendation systems can recommend learning resources, activities, exercises, or sequences. Unlike generic product recommendation, educational recommendation should preserve pedagogical constraints. A useful recommendation should be related to a learning objective, appropriate to the learner's current state, and compatible with prerequisite relationships. In the proposed framework, recommendations are generated from topic mastery and predicted performance rather than from popularity alone. Explainable learning analytics provides an additional bridge between prediction and recommendation. Li and Weegar [6] describe an approach in which explainable machine learning is used to provide data-driven feedback and action recommendations. Khosravi et al. [5] argue that educational XAI has distinctive requirements concerning stakeholders, presentation of explanations, human-centred design, and potential harms. These findings support the use of short, evidence-based explanations rather than opaque recommendation labels.

2.5 RESEARCH GAP

The literature supports the individual components of the proposed system, but the research challenge is their disciplined integration. Prediction, diagnosis,

recommendation, and explanation have different targets and should therefore have separate validation criteria. Moreover, dashboard research has identified weaknesses in theoretical grounding, actionable feedback, and evaluation [8]. The present framework responds by defining explicit interfaces and metrics for each stage, while avoiding the unsupported claim that a benchmark with final grades alone can validate topic-level personalization.

3.1 OVERALL SYSTEM ARCHITECTURE

The proposed system is organized as a pipeline with feedback. Student records are collected and normalized; features are extracted into academic, engagement, and topic-level representations; a predictive model estimates a performance category; topic-level analysis identifies learning gaps; the recommendation engine maps gaps to learning actions; and the dashboard displays the learner state, recommendations, and explanations. New assessment evidence can then update the learner state. Student Data Data Collection Data Preprocessing Feature Extraction ML Performance Prediction Topic-Wise Performance Analysis Weak Topic Detection Recommendation Engine Personalized Learning Path Student Dashboard Optional LMS inputs attendance · activity · time Explainability layer feature reasons · rules

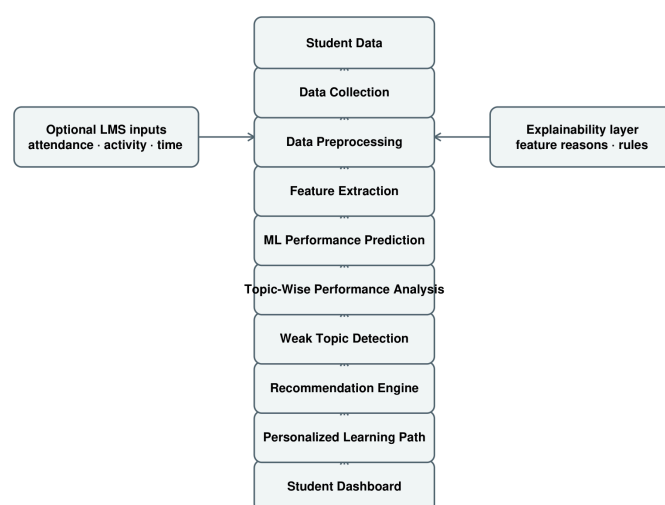


Fig. 1. Conceptual architecture of the proposed AI-based personalized learning system.

3.3 DATA PREPROCESSING

Schema validation: verify column types, value ranges, duplicate records, and missing-value status against the UCI documentation. Target construction: define performance categories before model fitting. A reproducible three-level design may use Low ($G3 < 10$), Moderate ($10 \leq G3 < 15$), and High ($G3 \geq 15$) on the 0–20 scale; thresholds must be declared before evaluation and should be sensitivity-tested. Leakage control: remove G3 from all predictors and, in the early-warning condition, remove G1 and G2 as specified. Categorical encoding: use one-hot encoding for nominal variables and retain ordinal variables only when their documented ordering is meaningful. Scaling: standardize continuous variables for Logistic Regression and SVM; tree-based models do not require scaling, although a common preprocessing pipeline can still be used when correctly configured. Validation: use stratified k-fold cross-validation inside the training set and keep the test set untouched until final evaluation.

3.4 FEATURE SELECTION

Table 2. Mapping between available benchmark evidence and the proposed learner model.

Model	Reason for inclusion
Logistic Regression	Interpretable linear baseline; provides class probabilities and coefficient-based evidence.
Decision Tree	Readable decision rules; useful as an interpretable nonlinear baseline.
Random Forest	Bagged trees reduce variance and capture nonlinear interactions in tabular data.
Gradient Boosting	Sequential tree ensemble that can model nonlinear boundaries and interactions.
Support Vector Machine	Strong margin-based classifier; useful for moderate-sized, high-dimensional encoded data.

3.5 MACHINE LEARNING MODELS

Model Reason for inclusion Interpretable linear baseline; provides class probabilities and coefficient-based evidence. Readable decision rules; useful as an interpretable nonlinear baseline. Bagged trees reduce variance and capture nonlinear interactions in tabular data. Sequential tree ensemble that can model nonlinear boundaries and interactions. Strong margin-based classifier; useful for moderate-sized, high-dimensional encoded data. Table 3. Candidate models for the performance-prediction benchmark. Hyperparameters must be tuned using only the training folds. A reproducible starting grid can include regularization strength for Logistic Regression and SVM, maximum depth and minimum leaf size for Decision Tree, number of trees and maximum depth for Random Forest, and learning rate, number of estimators, and tree depth for Gradient Boosting. The final search space should be reported with software versions and random seeds.

3.6 STUDENT PERFORMANCE PREDICTION

Let x_i denote the feature vector of student i and y_i the performance category derived from G3. The classifier estimates $p(y_i=c | x_i)$ for each category c . The predicted category is $\text{argmax}_c p(y_i=c | x_i)$. For early-risk support, the system should retain the probability distribution rather than only the class label. A student with a high probability of the low-performance class can be flagged for additional diagnostic analysis, but the probability is not interpreted as a causal statement about the student's ability.

3.7 WEAK TOPIC IDENTIFICATION

Topic-level diagnosis is implemented only when assessment items are mapped to topics. For student i and topic t , define a mastery score M_{it} as the weighted mean of normalized assessment outcomes associated with t . Let r_{it} represent recency and let a_{it} represent attempt evidence. A simple diagnostic score can be defined as: $M_{it} = \sum_j w_j s_{ijt} / \sum_j w_j$, where s_{ijt} is the normalized score for assessment item j mapped to topic t and w_j combines assessment importance and recency. A topic is considered a candidate weakness when M_{it} is below a predefined mastery threshold or when its score is substantially below the student's overall topic average. Thresholds should be selected using development data and validated against expert-labelled or future-performance outcomes. This formulation deliberately avoids calling a topic 'weak' solely because a predictive model assigns a low probability to a final grade. Topic weakness should be supported by topic-specific evidence. If topic tags are absent, the system must return 'insufficient topic evidence' rather than infer a topic weakness from unrelated demographic or aggregate variables.

3.8 PERSONALIZED RECOMMENDATION ALGORITHM

The recommendation engine maps the learner state to candidate learning resources. Each resource r is associated with one or more topics, prerequisites, a difficulty level d_r , and a resource type. The engine computes a suitability score: $S(i,r) = \alpha(1 - M_{i,t(r)}) + \beta R_{i,t(r)} + \gamma D_{i,r} + \delta P_{i,r} - \lambda C_{i,r}$, where M is topic mastery, R is predicted risk or urgency, D is difficulty compatibility, P is prerequisite readiness, and C is a constraint penalty such as excessive repetition or an unmet prerequisite. The weights α , β , γ , δ , and λ must be fixed using a development set or domain-informed calibration and then held constant during test evaluation. Recommendations are generated by ranking eligible resources after applying

prerequisite and safety constraints. Difficulty is selected using a conservative progression rule: learners below mastery receive foundational material; learners near mastery receive guided practice; learners demonstrating mastery receive more challenging or transfer-oriented tasks. This is a pedagogical policy, not a claim that the machine-learning model has discovered the optimal teaching strategy. 3.9 Explainability The system should expose two complementary explanations. First, model-level explanations summarize which features influenced the predicted performance category, using coefficients for Logistic Regression or model-agnostic/permutation or SHAP-style explanations where appropriate. Second, recommendation-level explanations state the direct learner evidence behind an action. For example: 'Practice Topic A because recent topic-tagged quiz performance is below the mastery threshold; begin with the foundational exercise because prerequisite Topic B is not yet mastered.' Such explanations should be concise, evidence-based, and distinguish association from causation. Educational XAI literature warns that explanations are not automatically trustworthy merely because they are mathematically faithful to a model [5], [7]. Therefore, future validation should include educator review of explanation relevance and stability, not only predictive performance.

4. SYSTEM ARCHITECTURE

Table 4. Functional decomposition of the proposed architecture.

Component	Input	Output	Purpose
Data collection	Academic records; LMS events where available	Validated learner records	Acquire evidence
Preprocessing	Raw records	Clean feature matrix	Remove errors and leakage
Prediction	Feature vector	Class probabilities	Estimate performance/risk
Topic analysis	Topic-tagged assessments	Mastery vector	Locate learning gaps
Recommendation	Mastery + risk + resource metadata	Ranked actions	Select next learning steps
Explanation	Prediction/recommendation evidence	Reason statements	Make decisions interpretable
Dashboard	Learner state + recommendations	Student-facing view	Support reflection and action

5. ALGORITHM / PSEUDOCODE

Algorithm 1: Personalized Learning Recommendation Input: Student record X Optional topic-tagged assessment set A Resource catalogue R with topic, prerequisite and difficulty metadata Trained performance classifier f 1. Validate X and remove unavailable/leakage-prone fields. 2. Transform X using the preprocessing pipeline fitted on training data. 3. Obtain performance probabilities $p = f(X)$. 4. If topic-tagged assessments A exist: a. Normalize assessment outcomes. b. Compute topic mastery M for each topic. c. Identify candidate weak topics where $M[t] < \text{mastery_threshold}$. Else: return "Topic-level recommendation unavailable" with a data-coverage notice. 5. For each eligible resource r in R : a. Check prerequisite readiness. b. Compute topic gap, risk, difficulty compatibility and repetition penalty. c. Compute suitability $S(\text{student}, r)$. 6. Rank eligible resources by S . 7. Select top- K resources subject to diversity and prerequisite constraints. 8. Generate a short explanation for each selected resource using direct evidence. 9. Log model version, feature version, recommendation score and timestamp. 10. Return performance status, weak-topic evidence, recommended learning path and explanations. Output: Personalized learning path L and explanation set E .

6. EXPERIMENTAL SETUP

Table 5. Proposed experimental protocol.

Stage	Protocol
Data split	Stratified train/test split; hold out test set until final evaluation
Cross-validation	Stratified 5-fold CV on training data
Preprocessing	Fold-fitted encoding/scaling; no information from validation/test folds
Model selection	Nested or training-fold hyperparameter search using identical candidate grids
Prediction tasks	Three-class G3 category; optionally binary risk as a secondary analysis
Ablation	Full-information condition versus early-warning condition without G1/G2
Reproducibility	Report Python/scikit-learn versions, random seed, feature list, preprocessing and hyperparameters
Recommendation experiment	Requires topic-tagged assessment/resource logs and held-out future interactions or expert relevance labels

7. EVALUATION METRICS

Table 6. Evaluation measures for prediction and recommendation.

Results to be obtained after implementation. No experimental performance values are reported here. The execution environment used to prepare this manuscript did not have network access for downloading the UCI data, and reporting unexecuted model outputs would violate the requirement not to fabricate results. The following table specifies exactly what should be filled after the experiment.					
Model	Accuracy	Precision (macro)	Recall (macro)	F1 (macro)	ROC-AUC
Logistic Regression	TBD	TBD	TBD	TBD	TBD
Decision Tree	TBD	TBD	TBD	TBD	TBD
Random Forest	TBD	TBD	TBD	TBD	TBD
Gradient Boosting	TBD	TBD	TBD	TBD	TBD
Support Vector Machine	TBD	TBD	TBD	TBD	TBD

8. RESULTS AND ANALYSIS

Results to be obtained after implementation. No experimental performance values are reported here. The execution environment used to prepare this manuscript did not have network access for downloading the UCI data, and reporting unexecuted model outputs would violate the requirement not to fabricate results. The following table specifies exactly what should be filled after the experiment. Model Accuracy Precision (macro) Recall (macro) F1 (macro) ROC-AUC TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD TBD Table 7. Prediction results to be obtained after implementation; TBD values are intentionally not fabricated. The analysis should not simply select the model with the highest accuracy. If class distributions are uneven, macro-F1 and per-class recall may provide more useful evidence. The confusion matrix should be inspected for systematic confusion between adjacent performance categories. The full-information and early-warning conditions should be reported separately because they answer different research questions. For the recommendation layer, a separate experiment should use a dataset containing item-to-topic mappings, learner attempts, timestamps, resource metadata, and either observed future resource interactions or expert relevance labels. Precision@K, Recall@K, and NDCG@K should be computed on held-out interactions. If no such ground truth exists, recommendation quality should be assessed through expert review or a controlled learner study rather than inferred from prediction accuracy. 9. Discussion The proposed framework treats personalization as a sequence of evidence transformations rather than as a single machine-learning prediction. This distinction is important for methodological validity. A classifier estimates a future outcome from observed variables; it does not, by itself, establish which topic caused the outcome or which intervention will improve it. Topic diagnosis therefore requires topic-specific assessment evidence, and recommendation requires a resource catalogue with pedagogical metadata. The five selected algorithms provide complementary inductive biases. Logistic Regression supplies an interpretable linear baseline.

Decision Trees provide explicit nonlinear rules. Random Forest and Gradient Boosting represent tree ensembles that can model interactions. SVM supplies a margin-based alternative that can perform well on structured, moderate-sized feature spaces. A fair comparison should use identical splits, folds, and preprocessing logic, and should report computational cost as well as predictive metrics. The framework also supports a distinction between risk detection and intervention. A risk flag should trigger further diagnostic analysis, not an automatic high-stakes decision. Recommendations should be framed as learning support and should remain revisable as new evidence arrives. This is consistent with research emphasizing explainability, learner agency, and human-centred educational AI [5], [6]. Another important consideration is fairness and feature governance. Demographic or family variables may improve prediction while being unsuitable as direct determinants of what a student should study. A production system should therefore document feature purpose, test subgroup performance, monitor drift, and avoid converting correlated background variables into deterministic judgments about learner ability. 10. Limitations The verified UCI benchmark is relatively small and originates from two Portuguese secondary schools; its results should not be generalized to university populations or other curricula without validation. The public benchmark does not contain true topic-tagged assessment evidence, LMS event streams, completion-rate histories, or time-on-task suitable for validating the complete personalized recommendation pipeline. G3 prediction can be strongly influenced by G1 and G2; therefore, full-information prediction is not equivalent to early-warning prediction. Performance categories require researcher-defined thresholds; different thresholds can change class balance and model metrics. A predictive association does not establish that changing a feature will cause improved academic performance. Recommendation evaluation requires ground truth about relevance or subsequent learning benefit. Ranking metrics alone do not prove that recommendations improve learning outcomes. Educational data can contain privacy, consent, fairness, and governance concerns. A real deployment requires institutional approval, data minimization, access control, and transparent communication to learners.

11. FUTURE SCOPE

Collect a longitudinal LMS dataset containing assessment items mapped to learning outcomes/topics, attempts, timestamps, resource interactions, completion, attendance, and time-on-task. Extend the learner model with temporal methods that can represent changing mastery rather than treating each student as a static feature vector. Compare rule-based, content-based, collaborative, and hybrid educational recommenders using held-out future interactions. Study calibration and uncertainty so that low-confidence predictions trigger human review rather than overconfident automated recommendations. Evaluate fairness across relevant learner groups and monitor whether model performance or recommendation quality drifts across cohorts and academic terms. Conduct a controlled learner study to determine whether personalized recommendations improve learning outcomes, engagement, or self-regulation relative to non-personalized support. Investigate counterfactual or causal methods for distinguishing variables associated with performance from interventions that plausibly change performance. Integrate educator feedback into the recommendation loop so that teachers can override, annotate, or refine recommendations and their reasons. 12. Conclusion This paper presented a research-oriented framework for an AI-based personalized learning system that connects student-performance prediction with topic diagnosis, recommendation, and explanation. The framework explicitly separates what can be validated with a public benchmark from what requires richer course and LMS data. The UCI Student Performance dataset was verified as a reproducible benchmark with 649 instances and 33 attributes, and five supervised learning models were selected for a controlled prediction comparison. No experimental accuracy, F1, ROC-AUC, or recommendation values have been fabricated. Instead, the manuscript provides a complete experimental protocol, evaluation metrics, model-selection procedure, recommendation formulation, and a results template to be populated after execution. The central research proposition is that useful personalization requires an explicit bridge from predictive evidence to topic-level learning needs and then to constrained, explainable actions. Future empirical work should validate that complete loop using longitudinal, topic-tagged educational data and should measure not only predictive performance but also recommendation relevance, fairness, learner outcomes, and the stability of explanations.

Table 1. Verified characteristics of the public benchmark used to design the predictive experiment.

Property	Verified value / interpretation
Official dataset name	Student Performance (UCI Machine Learning Repository)
Official source	UCI Machine Learning Repository, dataset ID 320
DOI	10.24432/C5TG7T
Instances	649
Attributes reported by UCI	33
Task types	Classification and regression
Missing values	None reported by UCI
Primary target for this study	G3, final grade on the dataset's 0–20 scale
Relevant predictors	G1/G2 prior grades, study time, failures, absences, school/social variables and other recorded attributes
Critical limitation	No genuine topic-tagged LMS events, quiz-attempt histories, completion-rate time series, or time-on-task variable suitable for validating the full recommendation layer

REFERENCES

- [1] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, "Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis," *Frontiers in Psychology*, vol. 12, 698490, 2021. DOI: 10.3389/fpsyg.2021.698490.
- [2] M. H. B. Roslan and C. J. Chen, "Educational Data Mining for Student Performance Prediction: A Systematic Literature Review (2015–2021)," *International Journal of Emerging Technologies in Learning*, vol. 17, no. 5, pp. 147–179, 2022. DOI: 10.3991/ijet.v17i05.27685.
- [3] C. Chaka, "Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: an overview of reviews," *Journal of e-Learning and Knowledge Society*, vol. 18, no. 2, pp. 58–69, 2022. DOI: 10.20368/1971-8829/1135578.
- [4] S. A. Alwarthan, "Predicting Student Academic Performance at Higher Education Using Data Mining: A Systematic Review," *Applied Computational Intelligence and Soft Computing*, vol. 2022, Article 8924028, 2022. DOI: 10.1155/2022/8924028.
- [5] H. Khosravi, S. Buckingham Shum, G. Chen, C. Conati, Y.-S. Tsai, J. Kay, S. Knight, R. Martinez-Maldonado, S. Sadiq, and D. Gašević, "Explainable Artificial Intelligence in education," *Computers and Education: Artificial Intelligence*, vol. 3, 100074, 2022. DOI: 10.1016/j.caeai.2022.100074.
- [6] X. Li and R. Weegar, "Explainable AI for Data-Driven Feedback and Intelligent Action Recommendations to Support Students Self-Regulation," *Frontiers in Artificial Intelligence*, vol. 4, 723447, 2021. DOI: 10.3389/frai.2021.723447.
- [7] W. Matcha, N. A. A. Uzir, D. Gašević, and A. Pardo, "A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective," *IEEE Transactions on Learning Technologies*, vol. 13, no. 2, pp. 226–245, 2020. DOI: 10.1109/TLT.2019.2916802.
- [8] L. Paulsen and E. Lindsay, "Learning analytics dashboards are increasingly becoming about learning and not just analytics - A systematic review," *Education and Information Technologies*, vol. 29, pp. 14279–14308, 2024. DOI: 10.1007/s10639-023-12401-4.
- [9] M. Qiang, Z. Liu, and R. Zhang, "Explainable AI in education: integrating educational domain knowledge into the deep learning model for improved student performance prediction," *Scientific Reports*, vol. 16, Article 9515, 2026. DOI: 10.1038/s41598-026-40538-y.
- [10] P. Cortez and A. M. G. Silva, "Using data mining to predict secondary school student performance," in *Proceedings of the 5th Annual Future Business Technology Conference*, 2008.
- [11] P. Cortez, "Student Performance," UCI Machine Learning Repository, dataset 320, 2008. DOI: 10.24432/C5TG7T. Official repository: <https://archive.ics.uci.edu/dataset/320/student+performance>
- [12] A. A. Abu Saa, M. Al-Emran, and K. Shaalan, "Factors Affecting Students' Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques," *Technology, Knowledge and Learning*, vol. 24, pp. 567–598, 2019. DOI: 10.1007/s10758-019-09408-7.
- [13] S. Alnasyan, M. Basher, and M. Allassafi, "The power of deep learning techniques for predicting student performance in virtual learning environments: A systematic literature review," *Computers and Education: Artificial Intelligence*, vol. 6, 100231, 2024. DOI: 10.1016/j.caeai.2024.100231. Source-verification note: Dataset metadata and DOI were checked against the UCI Machine Learning Repository. The literature references were cross-checked against publisher, repository, or scholarly-index records available during preparation. The manuscript intentionally reports no unexecuted experimental results.

Publication archival note: Official scholarly contribution published under CC BY 4.0 International License. Permanent record on <https://www.ijrek.org> | Article ID: IJREK-2026-00008