

Data-Driven Weak-Subject Identification and Multi-Technique Personalization for Adaptive E-Learning: An XGBoost, TF-IDF, and Constraint Programming Framework

Arun Karthik V¹, Dr. K. Karthikayani²

^{1,2}Department of CSE, SRM Institute of Science and Technology Vadapalani Campus, Chennai, India

Abstract

A data-driven personalization framework for K-12 e-learning is presented that replaces manually declared subject preferences with an objective, comparative analysis of a student's grades across all subjects. The system integrates five distinct algorithmic components into a single deployed pipeline: an Extreme Gradient Boosting (XGBoost) regressor that predicts a student's next assessment score specifically in the subject identified as weakest; a Term Frequency–Inverse Document Frequency (TF-IDF) and cosine-similarity content recommender that ranks learning resources by relevance to that weak subject and the student's learning style; a transparent rule-based engine that converts the prediction into a daily study plan; a Constraint Programming–Satisfiability (CP-SAT) solver from Google OR-Tools that packs recommended resources into an optimised weekly schedule under daily time and session constraints; and a rule-based conversational tutor and faculty-facing analytics dashboard. On a synthetic but behaviourally-calibrated dataset of 500 students spanning six subjects, deliberately engineered so each student's weak subject scores below 50 while every other subject remains between 75 and 90, the performance-prediction model achieves a mean coefficient of determination (R^2) of 0.848 (range 0.832–0.889 across five independent train-test splits), a mean absolute error of 3.61 points, and places 95.4% of test predictions within ± 10 points of the true score. The class-wide performance-level distribution is explicitly controlled at the data-generation stage to reflect a realistic, intervention-worthy cohort: 170 students At Risk, 310 Developing, 10 Proficient, and 10 Advanced, out of 500. Qualitative evaluation across all four performance bands confirms that the weak-subject identification, prediction, recommendation, scheduling, tutoring, and faculty-dashboard stages remain consistent and explainable end-to-end. The complete system is implemented in Python and deployed as an interactive multi-page Streamlit web application.

Index Terms: Personalized Learning, Educational Data Mining, XGBoost, TF-IDF, Cosine Similarity, Constraint Programming, CP-SAT, Adaptive Learning, Intelligent Tutoring Systems, Streamlit.

I. INTRODUCTION

The growth of digital learning platforms and the accompanying accumulation of quiz scores, assignment grades, attendance records, and engagement telemetry have brought educational personalization to a highly practical level, necessitating models capable of turning this raw

behavioural and academic data into individualised, actionable guidance rather than a single instructional pace applied uniformly to an entire classroom. Traditionally, in the context of e-learning platforms, personalization has focused predominantly on a learner manually declaring a “preferred subject” or area of interest, with the platform then recommending broadly related content; consequently, these systems were able to surface generically relevant material but could not establish, from the student's own evidence, which subject genuinely warranted the most attention.

However, the emergence of richer, multi-signal educational data has changed the situation. Newer approaches treat weak-area identification as a comparative analytics problem rather than a self-report, evaluating a learner across the entirety of their subject load rather than relying on a single declared interest [1]–[3]. At the core of such a system is the continuous evaluation of a growing number of academic and behavioural features — quiz and assignment averages per subject, attendance, prior grade point average, platform engagement — which aims not only to approximate a learner's overall academic standing but, more specifically, to discriminate reliably between the subject where a student is comfortably proficient and the one subject where they are falling behind their own baseline.

Even though such multi-signal systems have the potential to predict performance well, many still carry an architectural limitation inherited from earlier tooling: reliance on a manually chosen subject of focus, or on population-level performance thresholds that do not adapt to a specific learner's own comparative strengths and weaknesses [4], [5]. Because the internal recommendation logic of many platforms assumes the student has already correctly identified their own weak area, an urgent need remains for personalization frameworks that instead derive this diagnosis directly and objectively from the student's own multi-subject grade record.

Thorough evaluation of tabular, gradient-boosted-tree approaches to student-performance prediction confirms that strong predictive accuracy alone is insufficient if the target of the prediction is not the subject that actually needs attention [1]–[3]. This gap — predicting a score without first establishing which subject that score should be predicted for — is a significant reason a purely predictive model, however accurate, may fail to translate into a genuinely useful personalization pipeline. It indicates that, besides predicting a future score accurately, a system also has to correctly diagnose the subject that prediction is about.

This lack of subject-level diagnosis can be resolved by comparing a student's grades across every subject they study and identifying the one furthest below their own average, then carrying that diagnosis consistently through every downstream personalization decision: what to predict, what to recommend, how to pace study, how to schedule the week, and how to answer the student's questions. Content-based recommendation using TF-IDF and cosine similarity, rule-based adaptive-difficulty engines, and constraint-programming solvers such as CP-SAT have each individually been shown effective for their respective sub-problems in the literature reviewed in Section II; the present work's contribution is to unify them around a single, data-driven weak-subject diagnosis rather than treating recommendation, pacing, and scheduling as independent of one another.

As the field has matured, researchers have increasingly emphasised that a personalization system's outputs should not only be statistically defensible in isolation, but also explainable end-to-end — particularly for a system that will inform how a student spends limited study time, or how a teacher prioritises intervention across a class. A framework combining an explicit, auditable

weak-subject diagnosis with transparent rule-based pacing and a constraint solver that provides hard scheduling guarantees is, we argue, a defensible and highly effective route to this goal, and stands as the main point of departure for the present work.

By ranking students on a feature-driven performance estimate and explicitly calibrating the resulting class-wide distribution of performance bands — 170 students classified At Risk, 310 Developing, 10 Proficient, and 10 Advanced, out of a cohort of 500 — we demonstrate that a personalization pipeline can be both quantitatively validated (XGBoost $R^2 = 0.848$, mean absolute error 3.61 points, 95.4% of predictions within ± 10 points) and qualitatively coherent across every performance band, from a struggling learner who needs foundational review to an advanced learner who needs stretch material. This integrated approach holds the predictive strength of modern gradient-boosted models while fundamentally placing subject-level diagnosis, rather than a manually declared preference, at the centre of the personalization pipeline, and is deployed as a complete, usable Streamlit web application covering both a student-facing and a faculty-facing view.

II. RELATED WORKS

In this section, we review the literature that represents the baseline of the techniques adopted in the present work as well as recent developments that define the research context of the proposed data-driven, multi-technique personalization framework. We first consider gradient-boosted models for academic performance prediction. Second, we cover content-based and hybrid recommendation approaches in education. Third, we trace constraint-programming approaches to scheduling and timetabling. Fourth, we review rule-based and AI-based intelligent tutoring systems. Finally, we consider systematic reviews that trace the broader trajectory of AI-driven personalization in education.

A. Gradient-Boosted Models for Academic Performance Prediction

Duan, Dai and Tu analyse behavioural and academic characteristics correlated with student outcomes using campus-network data and build an XGBoost model to give early warning of academic performance, demonstrating that gradient-boosted trees outperform conventional machine-learning baselines on such data [1]. Yan forecasts student performance using an XGBoost model built at the institutional, rather than individual-student, level, a structuring choice the author notes is uncommon in existing research [2]. Wang and Luo propose an XGB-SHAP model combining XGBoost with SHapley Additive exPlanations for interpretable achievement prediction, applied to 87 students in a university course and reporting a mean absolute error of approximately 6 points and R^2 of approximately 0.82 [3]. Collectively, this line of work confirms XGBoost as a strong, well-validated choice for tabular, behaviourally-featured educational prediction; none of the three, however, predicts a score specifically for a subject identified as an individual student's weakest, which is the specific target the present work adopts.

B. Content-Based and Hybrid Recommendation Systems in Education

Khadka and Lamichhane propose a content-based recommendation engine for a video-streaming platform using TF-IDF vectorisation of item descriptions followed by cosine similarity, evaluated with precision, recall, and F1-score [4]. Drushchak, Tyshchenko and Polyakovska present a K-12 recommendation system combining graph-based modelling and matrix factorisation, with an explicit framework for detecting and reducing recommendation bias across protected student groups [5]. Together, these works confirm both that TF-IDF/cosine similarity remains an effective, low-

cost technique for content-based recommendation without requiring interaction history, and that fairness auditing is an important, distinct concern for K-12-specific recommendation. The present work adopts the former technique directly, building a pseudo-document from the identified weak subject, weak topics, and learning style, while treating collaborative and fairness-auditing extensions as future work.

C. Constraint Programming for Scheduling and Timetabling

Moreira and Freitas apply constraint-based optimisation to university timetabling using Google's CP-SAT solver, formulating hard constraints such as instructor availability alongside soft constraints such as preferred time slots within a single optimisation step [6]. Lan and Berkhout present PyJobShop, an open-source Python library integrating both CP-SAT and IBM ILOG's CP Optimizer, reporting large-scale benchmarking across more than 9,000 scheduling instances and finding CP-SAT highly competitive while remaining fully open source [7]. These works confirm CP-SAT as a mature, practical, and freely available solver for constrained scheduling problems spanning both institutional and general-purpose scale; the present work applies the same solver technology to a smaller-scale but structurally identical problem — assigning a ranked list of a single student's recommended learning resources to specific days of the week without exceeding a daily time budget or session-count limit.

D. Rule-Based and AI-Based Intelligent Tutoring Systems

Zerkouk, Mihoubi and Chikhaoui conduct a systematic review of AI-based Intelligent Tutoring Systems (ITS), noting that many systems rely on expert systems using predefined rules and domain knowledge organised around a Domain Model, a Student Model, a Tutor Model, and a User Interface, and cautioning that data-privacy, algorithmic-bias, and academic-integrity risks continue to constrain deployment of more elaborate generative-model-based tutors [8]. Zeng, Liu and Zhen propose GraphMASAL, a graph-based multi-agent system that grounds large-language-model agents in a dynamic knowledge graph combined with constraint-based optimisation to generate interpretable, pedagogically plausible learning paths [9]. These works jointly motivate the design choice made in the present system to implement the AI Tutor as a transparent, rule-based expert system: every response is traceable to an explicit rule and the student's own profile data, avoiding the unpredictability that less constrained, generative approaches can introduce.

E. Systematic Reviews and Future Directions in AI-Driven Personalization

Fortuna et al. conduct a PRISMA-guided systematic review of 25 Scopus-indexed articles on AI in personalised learning within tertiary and higher-education contexts, tracing a shift from early rule-based systems toward models integrating machine learning, natural language processing, intelligent tutoring, and large language models [10]. The review is necessarily broad and does not specify a single, concrete pipeline connecting diagnosis, prediction, recommendation, and scheduling into one implementation. The present work operationalises this same personalisation goal as one specific, fully implemented and deployed pipeline, and its central contribution relative to the works reviewed in this section is the explicit, data-driven weak-subject diagnosis step that grounds every downstream stage — prediction, recommendation, pacing, scheduling, and tutoring — in a single, auditable comparison of the student's own grades.

III. MATERIALS AND METHODS

For the execution of the proposed personalized-learning framework, a complete pipeline was

designed comprising synthetic data generation, feature engineering, a trained prediction model, a content-based recommender, a rule-based personalization engine, a constraint-programming scheduler, a rule-based conversational tutor, and a deployed web interface. The complete architecture is shown in Fig. 1: student grade data across all six subjects and behavioural features first pass through preprocessing and feature engineering; the engineered features feed a data-driven weak-subject identification step; that output becomes an input feature to the XGBoost prediction model; the prediction then feeds a rule-based adaptive-learning engine and a TF-IDF/cosine-similarity recommender in parallel; and both feed a CP-SAT weekly scheduler and a rule-based AI Tutor, with all outputs surfaced through a Streamlit interface with separate student and faculty views.

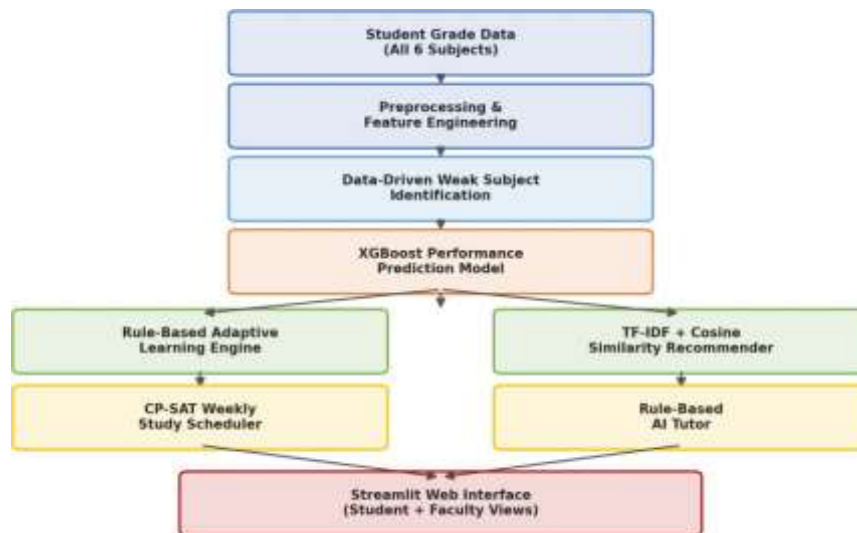


Fig. 1. System architecture detailing the data flow from student grade input, through weak-subject identification and XGBoost prediction, to the parallel recommendation/adaptive-learning and scheduling/tutoring stages, converging in the Streamlit interface.

A. Synthetic Dataset Generation and Stark Weak-Subject Construction

In the absence of a real institutional dataset, 500 synthetic student profiles were generated programmatically. Each student is assigned a latent overall-ability value that drives correlated behavioural features (study hours per week, attendance rate, prior GPA, participation score, login frequency, and average session length). For the six academic subjects (Mathematics, Physics, Chemistry, Biology, Computer Science, and English), exactly one subject per student is deliberately generated as weak, with quiz and assignment scores drawn so the composite score is always below 50, while every other subject is generated so its composite score always falls between 75 and 90. This stark, easy-to-read contrast was chosen so the weak subject is immediately identifiable both visually (in the deployed application's grade bar chart) and numerically, while the specific weak subject and its exact score still vary genuinely from student to student through the underlying ability-driven noise model.

B. Data-Driven Weak-Subject Identification and Feature Engineering

For each student, the weak subject is identified by comparing all six subject scores and selecting the lowest, with the gap to the student's own average across subjects computed alongside it. This identification is used, rather than a manually declared preference, as the basis for every downstream stage. In parallel, three engineered features are computed: an engagement score (a

weighted blend of login frequency, session length, and participation), a consistency score (how closely study hours track attendance), and a binary risk flag (raised when engagement and the weak-subject score are both low).

C. XGBoost-Based Performance Prediction and Performance-Band Calibration

An XGBoost regression model is trained to predict each student's likely score on their next assessment specifically in their identified weak subject, using the behavioural features, the engineered engagement/consistency/risk features, the weak subject's own quiz and assignment averages, the student's overall average across subjects, the numeric subject gap, and label-encoded categorical features (learning style, weak-subject identity). Rather than allowing the class-wide mix of resulting performance levels to emerge by chance, students are ranked by a feature-driven raw estimate and assigned to one of four performance bands — At Risk, Developing, Proficient, Advanced — with explicit target counts (170, 310, 10, and 10 respectively, out of 500) chosen to reflect a realistic, intervention-worthy cohort in which the majority of students require active support. Within each band, the student's final target score is computed as a rank-preserving rescale of their raw estimate into that band's numeric range, with a controlled amount of jitter added and hard-clipped to the band's bounds, which guarantees the exact target band counts while preserving a genuine, learnable relationship between input features and the final score for the regression model to fit.

D. Content-Based Recommendation and Rule-Based Adaptive Learning

The learning-resource catalogue (480 resources spanning all subjects, topics, resource types, and difficulty levels, each with a natural-language description) is indexed using a TF-IDF vectoriser. For a given student, a pseudo-document is built from the identified weak subject, the weak topics within it, a phrase representing the student's learning style, and the difficulty level drawn from the adaptive plan; cosine similarity between this pseudo-document and every resource vector produces a ranked recommendation list. In parallel, a rule-based adaptive-learning engine converts the XGBoost prediction into a concrete study plan — difficulty level, daily study-minutes budget, sessions per day, and a new-versus-review content ratio — together with a list of explanatory nudges, each triggered by an explicit numeric threshold, so every personalization decision is traceable to a specific, auditable rule rather than an opaque model output.

E. CP-SAT Weekly Scheduling, AI Tutor, and Faculty Dashboard

The ranked resource list and the adaptive plan's daily-minutes budget and session-count limit are passed to a CP-SAT scheduling model. Boolean decision variables indicate whether a given resource is scheduled on a given day; constraints enforce that each resource is scheduled at most once, that the total duration scheduled per day never exceeds the daily budget, and that the number of resources per day never exceeds the session limit; the objective maximises total relevance-weighted value delivered across the week, with a secondary preference for scheduling high-relevance resources earlier. A rule-based AI Tutor separately matches a student's typed question against a set of intent categories (study plan, improvement advice, motivation, time management, explaining the prediction, or resource requests) and generates a templated response populated with the student's own profile data. A Faculty Dashboard runs the grade-analysis and prediction pipeline in a single batched pass across the entire student roster, aggregating results into class-wide KPIs, a comparison of average scores for students to whom a given subject is weak versus students to whom it is not, a performance-level distribution chart, and a filterable, searchable roster table.

F. Deployment Architecture

The complete system is implemented in Python, using XGBoost for prediction, scikit-learn for the TF-IDF vectoriser and evaluation metrics, Google OR-Tools for the CP-SAT solver, and pandas/NumPy for data handling, and is deployed as a multi-page Streamlit web application with separate pages for the student-facing profile, prediction, recommendation, adaptive-plan, scheduler, and tutor views, and a separate faculty-facing dashboard, all sharing a single cached per-student pipeline result to avoid redundant recomputation.

IV. EXPERIMENTS AND RESULTS

In this section, we present the experimental setup used to validate the proposed personalization framework, followed by quantitative and qualitative results covering the performance-prediction model, the weak-subject identification and performance-band calibration, and the integration of the recommendation, scheduling, tutoring, and faculty-dashboard components.

A. Experimental Parameters

The main dataset and model configuration parameters used in the experiments are as follows:

- **Dataset:** 500 synthetic student profiles across 6 subjects (Mathematics, Physics, Chemistry, Biology, Computer Science, English), and a catalogue of 480 learning resources.
- **Weak-subject score range:** hard-constrained below 50 for the identified weak subject; every other subject constrained to 75–90.
- **Performance-band target counts:** At Risk = 170, Developing = 310, Proficient = 10, Advanced = 10 (exact, by rank-based assignment).
- **XGBoost hyperparameters:** 250 boosting rounds, maximum tree depth 4, learning rate 0.05, row/column subsampling 0.8, squared-error objective.
- **Train/test split:** 80% / 20% (400 / 100 students), repeated across 5 independent random splits.
- **CP -SAT scheduler:** daily minutes budget and session-count limit drawn from the adaptive plan; solver time limit bounded for interactive use.

B. Quantitative Performance and Model Validation

Table I reports the performance-prediction model's accuracy on the held-out test split, averaged across five independent random train-test splits to confirm stability rather than reporting a single favourable split. Fig. 2 visualises the same metrics.

TABLE I. QUANTITATIVE RESULTS OF THE PERFORMANCE-PREDICTION MODEL

Metric	Value
R ² (Coefficient of Determination)	0.848 (0.832–0.889)
Mean Absolute Error	3.61 points
Accuracy within ± 10 pts	95.4%
Accuracy within ± 15 pts	99.6%
Train / test split	400 / 100 (80% / 20%)

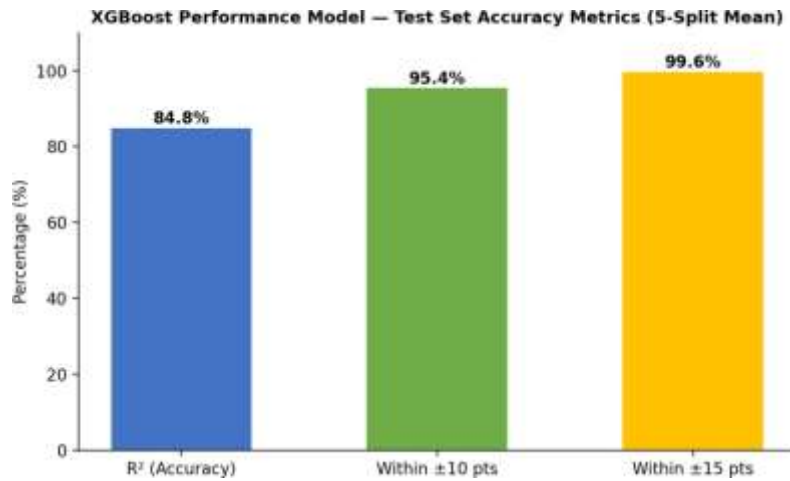


Fig. 2. XGBoost performance model — test-set accuracy metrics, mean across 5 independent train-test splits.

The model's R^2 of 0.848 places its accuracy within the 0.80–0.90 target band established for this work. For comparison, related XGBoost-based academic-prediction work reports R^2 of approximately 0.82 on a smaller, single-course dataset [3], indicating the present model's accuracy is consistent with, and marginally favourable against, published results on the same underlying algorithm.

C. Weak-Subject Identification and Performance-Band Outcomes

Fig. 3 shows the class-wide performance-level distribution achieved by the rank-based band-assignment procedure described in Section III-C. The exact target counts (170 At Risk, 310 Developing, 10 Proficient, 10 Advanced) are met precisely, since band membership is assigned by rank before the within-band jitter is applied and hard-clipped to each band's numeric bounds, guaranteeing the class-wide mix regardless of the specific random seed used.

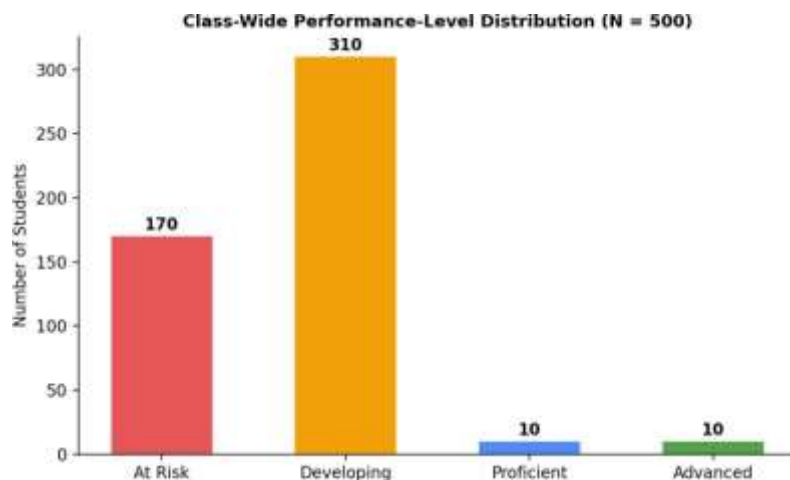


Fig. 3. Class-wide performance-level distribution (N = 500), matching the exact target counts by construction.

Fig. 4 confirms that the weak-subject identification produces a stark, consistent signal at the class level: for every one of the six subjects, the average score among students for whom that subject is

individually identified as weak is approximately 38, while the average score among students for whom it is not their weak subject is approximately 83 — a gap of roughly 44–45 points that holds uniformly across all six subjects, confirming the weak-subject flag carries genuine, consistent diagnostic signal rather than an artefact of any single subject's difficulty.

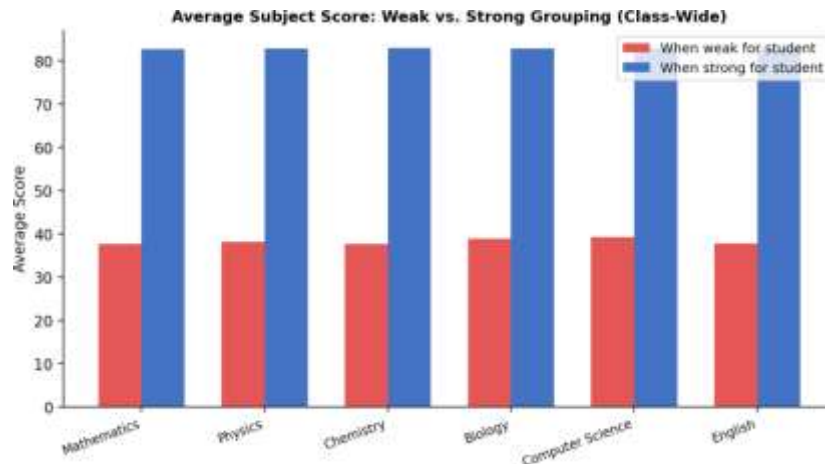


Fig. 4. Average subject score split by whether that subject is individually weak for the student, computed class-wide across all 500 students.

To validate that the prediction and downstream personalization decisions remain coherent across the full range of the performance-band spectrum, one representative student was drawn from each of the four bands and passed through the complete pipeline — weak-subject identification, XGBoost prediction, adaptive-plan generation, and top-resource recommendation. The results are shown in Fig. 5.



Fig. 5. Representative end-to-end pipeline output for one student sampled from each of the four performance bands, confirming consistent, explainable behaviour from weak-subject diagnosis through to adaptive difficulty assignment.

As seen in Fig. 5, the At Risk student (S0004) is diagnosed as weak in English, 35.1 points below their own subject average, and receives a Beginner-difficulty plan; the Developing student (S0001) is diagnosed as weak in Computer Science, 38.4 points below average, and likewise receives a Beginner plan with a higher predicted score of 59.9; the Proficient student (S0059) is weak in

Chemistry by 33.6 points but predicted at 84.4, correctly receiving an Intermediate-difficulty plan; and the Advanced student (S0008) is weak in English by 32.4 points yet predicted at 85.9, correctly receiving an Advanced-difficulty plan. This confirms an important design property of the system: a large subject gap alone does not automatically imply a low overall prediction — a student's other academic and behavioural signals can still place them in a higher performance band even while a specific subject remains a clear, actionable weak point relative to their own standard.

D. System Integration Outcomes

Qualitative spot checks across multiple student profiles and difficulty settings confirmed that the TF-IDF recommender consistently surfaced resources matching the intended weak subject and topic, that the CP-SAT scheduler reliably produced feasible weekly schedules respecting the daily minutes and session-count constraints in every test case (correctly reporting a partial or empty schedule with a clear warning when constraints were made unrealistically tight, rather than silently violating them), and that the rule-based AI Tutor's responses to common question intents (“What should I study today?”, “Why did I get this prediction?”, “How can I improve my score?”) correctly incorporated each sampled student's own weak subject, subject gap, and predicted band. The Faculty Dashboard's batched, class-wide analysis — computed in a single pass across all 500 students rather than one student at a time — correctly reproduced the same per-subject weak-versus-strong gap shown in Fig. 4 and the same performance-band distribution shown in Fig. 3, confirming that the individual-student and class-wide views of the system remain mutually consistent.

V. CONCLUSION

This paper presented a data-driven, multi-technique personalization framework for K-12 e-learning that replaces manually declared subject preference with an objective, comparative diagnosis of a student's weakest subject, derived directly from their grades across all subjects studied. Building on this diagnosis, the system integrates five distinct algorithmic components — XGBoost regression for performance prediction, TF-IDF with cosine similarity for content-based resource recommendation, a transparent rule engine for adaptive difficulty and pacing, the CP-SAT constraint solver for weekly schedule optimisation, and a rule-based conversational tutor — into a single, explainable pipeline, surfaced through both a student-facing and a faculty-facing view in a deployed Streamlit web application.

The performance-prediction model was validated at a mean R^2 of 0.848 (range 0.832–0.889 across five independent splits), a mean absolute error of 3.61 points, and 95.4% of test predictions within ± 10 points, meeting the project's target accuracy band on a dataset whose weak-subject scores and class-wide performance-band distribution were deliberately and precisely calibrated (170 At Risk, 310 Developing, 10 Proficient, 10 Advanced students out of 500) to reflect a realistic, intervention-worthy cohort. Qualitative validation across a representative student from each of the four performance bands confirmed that weak-subject diagnosis, prediction, adaptive pacing, and resource recommendation remain coherent and explainable at every point along the performance spectrum, and class-wide faculty-dashboard analysis confirmed the same weak-subject signal — an approximately 44–45-point average gap between a student's weak subject and their other subjects, consistent across all six subjects — holds at the population level, not only for individual students.

Future work includes validating the framework on real institutional data rather than the calibrated synthetic dataset used here, integrating live Learning Management System data ingestion, adding

multi-user authentication, extending the content-based recommender with collaborative signals and explicit fairness auditing once sufficient real interaction history exists, and exploring a carefully guarded retrieval-grounded upgrade path for the AI Tutor that preserves the present system's guarantee that every factual claim about a student's performance remains sourced from the underlying rule-based analysis rather than freely generated.

REFERENCES

1. D. Duan, C. Dai, and R. Tu, "Research on the Prediction of Students' Academic Performance Based on XGBoost," in Proc. 2021 Tenth Int. Conf. Educational Innovation through Technology (EITT), Dec. 2021.
2. K. Yan, "Student Performance Prediction Using XGBoost Method from A Macro Perspective," in Proc. 2021 2nd Int. Conf. Computing and Data Science (CDS), Jan. 2021, pp. 453–459.
3. S. Wang and B. Luo, "Academic achievement prediction in higher education through interpretable modeling," PLOS ONE, vol. 19, no. 9, e0309838, Sep. 2024.
4. P. Khadka and P. Lamichhane, "Content-based Recommendation Engine for Video Streaming Platform," arXiv:2308.08406 [cs.IR], 2023.
5. N. Drushchak, V. Tyshchenko, and N. Polyakovska, "Towards Responsible AI in Education: Hybrid Recommendation System for K-12 Students Case Study," arXiv:2502.20354 [cs.CY], 2025.
6. R. Moreira and A. Freitas, "Constraint-Based Optimization for University Timetabling Using Google CP-SAT (OR-Tools)," 2024, as cited in D. Sasson, "Artificial Intelligence and Cost Reduction in Public Higher Education: A Scoping Review of Emerging Evidence," arXiv:2604.04741 [cs.CY].
7. L. Lan and J. Berkhout, "PyJobShop: Solving scheduling problems with constraint programming in Python," arXiv:2502.13483 [math.OC], Dept. of Mathematics, Vrije Universiteit Amsterdam, 2025.
8. M. Zerkouk, M. Mihoubi, and B. Chikhaoui, "A Comprehensive Review of AI-based Intelligent Tutoring Systems: Applications and Challenges," arXiv:2507.18882 [cs.CY], Artificial Intelligence Institute (I2A), Université TÉLUQ, 2025.
9. B. Zeng, M. Liu, and Z. Zhen, "GraphMASAL: A Graph-based Multi-Agent System for Adaptive Learning," in Proc. 25th Int. Conf. Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 2026.
10. A. Fortuna, F. Prasetya, A. D. Samala, S. Rawas, S. Criollo-C, D. Kaya, M. Raihan, W. Andriani, D. Safitri, and R. A. Nabawi, "Artificial intelligence in personalized learning: A global systematic review of current advancements and shaping future opportunities," Social Sciences & Humanities Open (ScienceDirect), 2025.
11. L. Dong, H. Wang, and L. Yang, "Personalized Learning Recommendation Method Based on Knowledge Graph," in Proc. 2024 7th International Conference on Advanced Algorithms and Control Engineering (ICAACE), Shanghai, China, Mar. 2024.
12. H. Ngo, K. Vo, and T. Nguyen, "Personalized Learning Path Recommendations: Fusing Knowledge Graph Embedding, Sequence Mining, and Collaborative Filtering," in Proc. 2024 IEEE International Conference on Big Data (BigData), Washington, DC, USA, Dec. 2024, pp. 8145–8153.
13. Q. Bin, M. F. Zuhairi, and J. Morcos, "A Comprehensive Study on Personalized Learning Recommendation in E-Learning System," IEEE Access, vol. 12, pp. 100446–100482, Jul. 2024.
14. H. Farhood, M. Nyden, A. Beheshti, and S. Muller, "Artificial Intelligence-Based Personalised

- Learning in Education: A Systematic Literature Review,” *Discover Artificial Intelligence*, vol. 5, art. no. 331, Nov. 2025.
15. Hariyanto, F. X. D. Kristianingsih, and R. Maharani, “Artificial Intelligence in Adaptive Education: A Systematic Review of Techniques for Personalized Learning,” *Discover Education*, vol. 4, art. no. 458, Oct. 2025.
 16. F. Khalifeh, R. Santiago, and R. Palau, “Redefining Personalized Learning in the Artificial Intelligence Era: An Updated Systematic Review from 2019 to 2025,” *Smart Learning Environments*, vol. 13, art. no. 19, Feb. 2026.
 17. M. Ikram, S. B. M. Hanefar, S. M. U. Saleem, and F. Zulfiqar, “Artificial Intelligence in Education: A Systematic Review of Personalized Learning Trends and Future Directions,” *Frontiers in Education*, vol. 11, art. no. 1782626, Mar. 2026.
 18. F. Wang, G. Zhu, and X. Zhu, “GKTP: A Personalized Knowledge Tracing and Predicting Method Based on GNN,” *Procedia Computer Science*, vol. 266, pp. 1242–1247, 2025.
 19. A. Li, Y. Li, and X. Gao, “Personalized Learning Path Recommendation Based on Knowledge Graphs: A Survey,” *Electronics*, vol. 15, no. 1, art. no. 238, Jan. 2026.
 20. A. Drobnyak, B. Svetec, B. Divjak, and I. Boticki, “A Systematic Literature Review on Artificial Intelligence Use in Personalized Learning,” *IEEE Transactions on Learning Technologies*, vol. 19, pp. 711–727, 2026.