



From Signals to Action: Explainable AI for Engagement-Responsive Instructional Support in Digital Higher Education

Andino Maselena^{1,*}, Meinhaj Hussain², Aygul Z. Ibatova³

¹Institut Bakti Nusantara, Lampung, Indonesia

²Rennier University, Ireland

³Tyumen Industrial University, Russia

Emails: andino.maselena@ibnus.ac.id; meinhaj@rennier.online; aigoul@rambler.ru

Abstract

Artificial intelligence is increasingly used to monitor learning processes in higher education; however, many analytics pipelines still terminate at prediction and provide limited support for instructional action. The research establishes an explainable artificial intelligence framework which utilizes digital learning environment behavioral data and contextual information to create customized instructional support solutions. The analysis uses xAPI-Edu-Data dataset which contains 480 records to build engagement index and create support profiles and predict multiclass performance through rule based action allocation. The study tests three classification models using stratified cross validation. The study selects Random Forest as the most effective system because it delivers superior results across all tests. The selected model demonstrates 0.8021 accuracy and 0.8204 macro precision and 0.8010 macro recall and 0.8084 macro F1 score and 0.9140 macro area under the curve on the hold-out sample. The analysis shows that student absence and composite engagement index and gender and student-guardian relationship and support profile and digital resource access are the most important factors that determine student performance. The final decision layer manages student assignment to instructional support plans which contain attendance-first intervention and adaptive engagement support and family-engagement reinforcement and structured progression coaching and challenge-and-extend pathways. The study develops an analytical framework which connects explainable artificial intelligence to digital higher education instructional decision support systems.

Keywords: Artificial intelligence in education; Learning analytics; Higher education; explainable AI; Educational technology; Student support

1. Introduction

Learning analytics and artificial intelligence (AI) now underpin many platforms in higher education, where they are used for monitoring, feedback, personalisation and performance analysis. Recent research suggests that AI is transforming online higher education through predictive modelling, chatbots, personalised support, and decision-focused dashboards (Chiu et al., 2023; Lee et al., 2024; Schmidt et al., 2025). However, at the same time as the technical complexity of these systems is growing, pathways for instruction are not always as well developed. Too often, analytics processing remains limited to visualisations, performance scores, or risk flags, with little to guide the implementation of support decisions.

This is important because the pedagogic potential of AI relates to the extent to which its outputs can inform intervention. Literature reviews on learning analytics and student engagement demonstrate ongoing interest in prediction of student performance and monitoring of student engagement, but also uneven implementation of analytics outputs in pedagogical practice (Bergdahl et al., 2024; Palanci et al., 2024). Research in human-centred AI education also stresses that AI tools should be transparent, explainable and accessible to educators, rather than being merely technical layers that are removed from teaching practice (Alfredo et al., 2024).

The next wave of AI implementation in higher education amplifies this need. Research on generative AI, governance and student support demonstrates that universities are simultaneously innovating and facing issues of ethics, institutional agency, assessment integrity, and inclusivity (Mumtaz et al., 2025; Pierrès et al., 2025; Kleimola et al., 2025). These debates reveal a key requirement: AI in education must be pragmatically meaningful, as well as statistically significant.

In this context, this study proposes an explainable AI approach for *engagement-responsive instructional support*. With the goal of student-performance prediction, the framework integrates four processes: the development of an engagement index, profiling of support, classification of student-performance outcomes (using a multiclass approach), and allocation of actions. We apply the analysis to a public data set to ensure transparency and replicability. So the study responds to a question that is relevant for digital higher education: how can traces from a learning environment be translated into actionable and interpretable support plans for educators and student-success teams?

2. Literature Review

2.1. AI, learning analytics, and student engagement

The use of AI in education has evolved from simple automation to more comprehensive feedback, personalisation and educational support systems. Chiu et al. (2023) present a systematic review of AI in education, revealing significant opportunities for adaptive learning and performance support, as well as methodological and ethical issues. Similar studies in higher education have shown how generative AI is increasingly used for teaching and learning, alongside issues relating to assessment, quality, and governance (Lee et al., 2024; Mumtaz et al., 2025; Schmidt et al., 2025).

In learning analytics, the construct of student engagement is prominent. Bergdahl et al. (2024) demonstrate that in higher-education learning analytics, engagement is multi-faceted and often

not well-defined. Palanci et al. (2024) observe similar phenomena in distance education, where engagement and performance prediction are central analytical targets, but the intervention actions following those analyses are often not well enough defined. This problem is significant for higher education, where we can easily gather behavioral trace data, but support strategies are difficult to standardise.

2.2. Explainability and intervention-oriented analytics

The second group of studies focuses on explainability, transparency and human-centered AI. Alfredo et al. (2024) propose that learning analytics and AI should support human agency and make the learning analytics process comprehensible to teachers. Ghimire et al. (2024) show that explainable machine-learning approaches can be used for prediction in education while increasing explainability. Research on self- and socially shared regulation, as well, suggests that AI and analytics are most effective when they enhance instruction and support rather than simply classifying performance (Kleimola et al., 2025; Kim et al., 2025).

There are also recent developments on intervention-aware analytics. Alalawi et al. (2025a) assess a learning-analytics intervention framework that connects student prediction to teaching actions, while Alalawi et al. (2025b) build on this work to develop a larger architecture that integrates machine learning and teaching actions. These works demonstrate that the educational effectiveness of prediction is enhanced when analytical models are integrated with an action system.

2.3. Research gap

The literature provides substantial evidence on AI adoption, learning analytics, and explainability in higher education; however, fewer studies integrate behavioral signal engineering, interpretable segmentation, multiclass prediction, and practical action allocation within a single reproducible workflow. In response to this gap, the present study proposes an explainable AI architecture that uses digital trace data to produce differentiated instructional support plans rather than risk labels alone.

Table 1: Selected post-2022 studies informing the analytical framing

Study	Year	Focus	Design	Main contribution
Chiu et al.	2023	AI in education	Systematic review	Synthesizes opportunities, challenges, and future research directions for AI in education.
Bergdahl et al.	2024	Student engagement in learning analytics	Systematic review	Clarifies the multidimensional nature of engagement and highlights actionability challenges.
Palanci et al.	2024	Learning analytics in distance education	Systematic review	Documents the increasing use of analytics for engagement, prediction, and instructional redesign.
Alfredo et al.	2024	Human-centred learning analytics	Systematic review	Emphasizes interpretability, human oversight, and responsible human–AI collaboration.
Ghimire et al.	2024	Explainable AI for student-score estimation	Empirical modeling	Demonstrates the value of explainable machine-learning models in educational prediction.

Study	Year	Focus	Design	Main contribution
AlGhamdi	2024	AI-generated feedback	Blinded empirical study	Shows that AI feedback can improve technical-writing performance in computing education.
Lee et al.	2024	Generative AI in higher education	Educator survey	Examines opportunities and risks of generative AI from the perspective of teaching staff.
Alalawi et al.	2025a	Prediction-action intervention framework	Intervention study	Evaluates an action-oriented learning-analytics framework in practice.
Alalawi et al.	2025b	AI and pedagogical intervention	Framework development	Integrates predictive analytics with pedagogical response for student-support design.
Kleimola et al.	2025	Learning analytics and self-regulation	Qualitative study	Explores how analytics can support self-regulated learning in higher education.
Kim et al.	2025	AI and socially shared regulation	Conceptual-empirical study	Discusses how AI may support collaborative regulation of learning.
Mumtaz et al.	2025	Ethical AI use in higher education	Survey study	Examines ethical readiness for AI adoption among future business leaders.
Pierrès et al.	2025	Inclusive AI in higher education	Interview study	Highlights inclusive-design requirements for students with disabilities.
Schmidt et al.	2025	Institutional AI integration	Empirical study	Identifies challenges and strategic responses in higher-education AI implementation.
Deng et al.	2025	ChatGPT learning outcomes	Systematic review and meta-analysis	Provides quantitative evidence on the educational effects of ChatGPT-based interventions.

3. Conceptual Foundation and Analytical Framework

3.1. Analytical premise

The proposed framework is built on the premise that observable digital behavior contains signals that can be aggregated, interpreted, and translated into differentiated instructional action. In digital higher education, engagement is not directly observable as a latent construct; rather, it is approximated through trace indicators such as interaction frequency, resource access, and participation behavior. Let $\mathbf{x}_i \in \mathbb{R}^p$ denote the predictor vector for student i , and let the outcome variable $y_i \in \{H, M, L\}$ represent the observed performance class, where H , M , and L denote high, medium, and low performance, respectively.

To provide a concise behavioral representation, the study constructs a composite engagement index from four numerical behavioral indicators: raised hands, visited resources, viewed announcements, and discussion activity. For student i , the index is defined as

$$EI_i = \frac{1}{4} \sum_{k=1}^4 z_{ik}, \quad (1)$$

where z_{ik} is the standardized value of behavioral variable k for student i . This index does not replace the original variables in all analyses; rather, it provides an interpretable summary that

supports both segmentation and model interpretation.

The framework then segments students into support profiles by combining EI_i with attendance and contextual indicators. In operational terms, the segmentation distinguishes digitally active students, low-touch absentees, moderate-progress students, and support-sensitive students. This intermediate layer serves two functions. First, it reduces the distance between raw behavioral signals and educational interpretation. Second, it creates an additional explanatory feature that can be used in the predictive model.

3.2. Mathematical specification of the predictive layer

The classification problem is addressed as a multiclass learning task. For multinomial logistic regression, the conditional class probability for student i and class c is given by

$$P(y_i = c \mid \mathbf{x}_i) = \frac{\exp(\beta_c^\top \mathbf{x}_i)}{\sum_{j \in \{H, M, L\}} \exp(\beta_j^\top \mathbf{x}_i)}. \quad (2)$$

The predicted class is obtained by

$$\hat{y}_i = \arg \max_{c \in \{H, M, L\}} P(y_i = c \mid \mathbf{x}_i). \quad (3)$$

For ensemble tree models, class assignment follows majority voting across T trees:

$$\hat{y}_i = \arg \max_{c \in \{H, M, L\}} \frac{1}{T} \sum_{t=1}^T \mathbb{I}(h_t(\mathbf{x}_i) = c), \quad (4)$$

where $h_t(\mathbf{x}_i)$ is the prediction of tree t and $\mathbb{I}(\cdot)$ is the indicator function. Logistic regression, Random Forest, and Extra Trees are estimated using the same train-test split and preprocessing logic so that model comparison remains internally consistent.

Model performance is assessed using accuracy, macro-precision, macro-recall, macro- F_1 , and macro-area under the receiver operating characteristic curve. For class c ,

$$Precision_c = \frac{TP_c}{TP_c + FP_c}, \quad Recall_c = \frac{TP_c}{TP_c + FN_c}, \quad (5)$$

and the macro- F_1 score is computed as

$$F_1^{macro} = \frac{1}{C} \sum_{c=1}^C \frac{2 \cdot Precision_c \cdot Recall_c}{Precision_c + Recall_c}, \quad (6)$$

where $C = 3$.

3.3. Action translation and support-plan allocation

To move from prediction to action, the framework applies a priority score that ranks students according to intervention intensity. For student i , the score is defined as

$$PS_i = 0.45 \hat{p}_{iL} + 0.20 A_i + 0.20(1 - \widetilde{VR}_i) + 0.15(1 - \widetilde{RH}_i), \quad (7)$$

where \hat{p}_{iL} is the predicted probability of class L , A_i is the absence indicator, \widetilde{VR}_i is normalized resource visitation, and \widetilde{RH}_i is normalized hand-raising behavior. The score is not intended to replace the class prediction; instead, it ranks students within and across predicted classes for support prioritization.

The decision layer then allocates cases to a small portfolio of interpretable instructional plans. Students predicted as low performing with high absence receive an *attendance-first intervention*; students predicted as low performing without high absence are assigned to *structured progression coaching*; students predicted as medium performing with weak parental-engagement signals receive *family-engagement reinforcement*; high-performing students are assigned to a *challenge-and-extend pathway*; and remaining cases are routed to *adaptive engagement support*. The framework therefore preserves predictive rigor while ensuring that the final analytical output is an actionable support plan.

4. Materials and Methods

4.1. Dataset and variables

The analysis uses the public xAPI-Edu-Data dataset, which contains 480 student observations and records of demographic, contextual, and behavioral characteristics derived from a digital learning environment. The original target variable is a three-class performance label (H, M, L). Numerical behavioral indicators include raised hands, visited resources, announcement views, and discussion participation. Contextual indicators include educational stage, grade level, topic, section, semester, student-guardian relation, parental-survey participation, parental-school satisfaction, and student absence days.

4.2. Preprocessing and experimental design

The analysis proceeds in four stages. First, the dataset is cleaned and recoded, and the numerical behavioral variables are standardized to construct the engagement index. Second, support profiles are derived using engagement, attendance, and support-related variables. Third, multiclass prediction models are estimated. Categorical variables are one-hot encoded, numerical variables are standardized where needed, and the sample is divided into training and hold-out subsets using stratified sampling. Fourth, the selected model is interpreted using permutation-importance analysis and translated into support plans through the rule-based action layer.

Model comparison is based on stratified cross-validation. The final model is chosen according to macro- F_1 because the three-class outcome requires balanced treatment of all classes rather than accuracy alone. All analyses are reproducible and are included in the project package.

5. Results

5.1. Descriptive characteristics of the sample

The sample comprises 142 students in class H , 211 in class M , and 127 in class L . The support-profile segmentation yields four analytically distinct groups: 189 digitally active students, 114 low-touch absentees, 100 moderate-progress students, and 77 support-sensitive students. The descriptive evidence shows marked behavioral separation across classes. Students in class H

display higher levels of hand-raising, resource visits, announcement views, and discussion activity than students in class *L*, while class *M* occupies an intermediate position. Absence patterns follow the opposite direction, with the low-performance group containing a larger proportion of students with more than seven absence days.

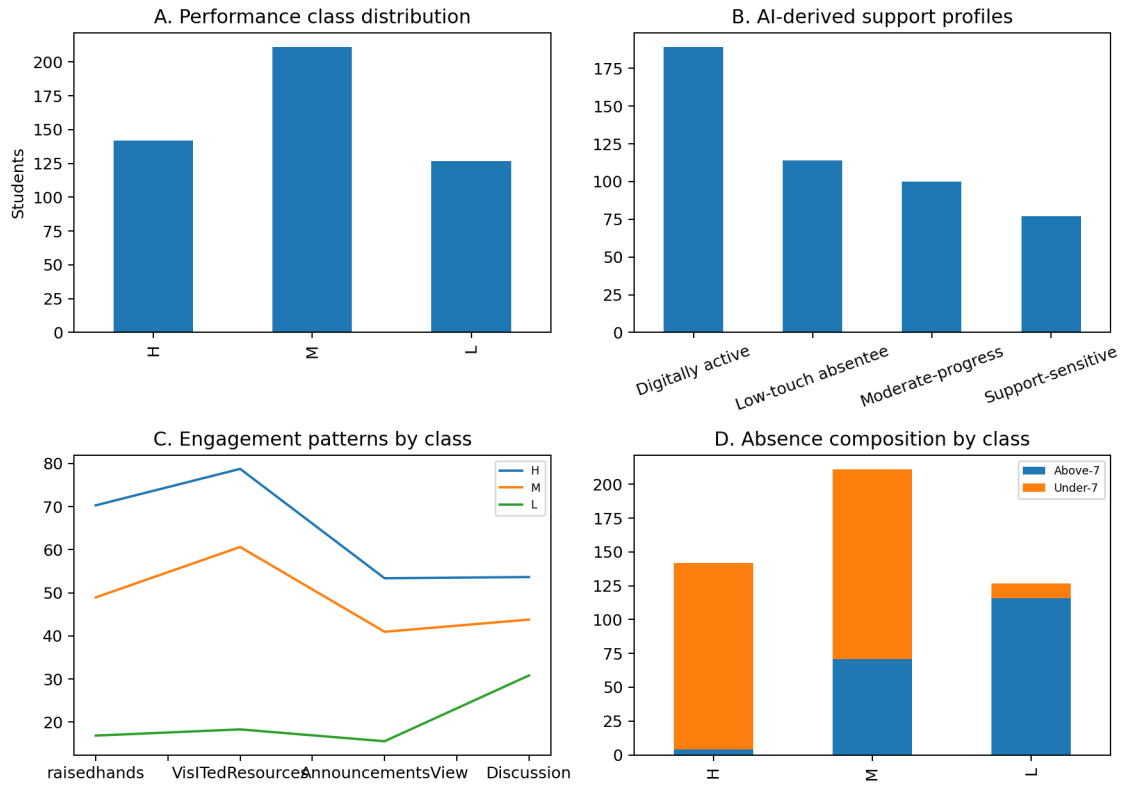


Figure 1. Descriptive diagnostics of performance, support profiles, engagement, and absence

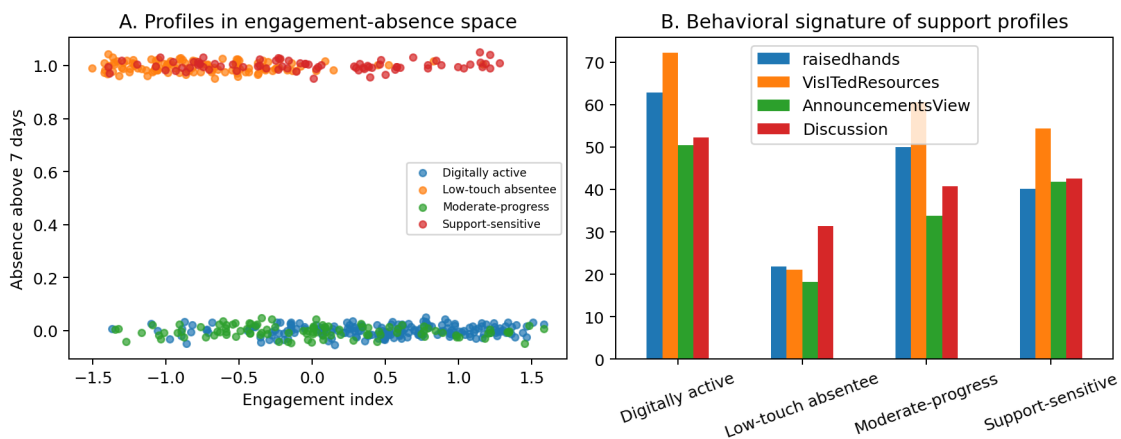


Figure 2. Behavioral signatures of the support profiles

5.2. Comparative model performance

Table 2 reports the cross-validated performance of the candidate models. Random Forest provides the strongest overall macro- F_1 score (0.8048), followed by Extra Trees (0.8033) and logistic regression (0.7757). Although the two ensemble models perform similarly, Random Forest also

delivers the highest macro-precision and is therefore selected for final hold-out evaluation.

Table 2: Cross-validated model performance

Model	Accuracy	Macro-Precision	Macro-Recall	Macro- F_1
Random Forest	0.8000	0.8161	0.8004	0.8048
Extra Trees	0.7979	0.8039	0.8064	0.8033
Logistic Regression	0.7688	0.7835	0.7726	0.7757

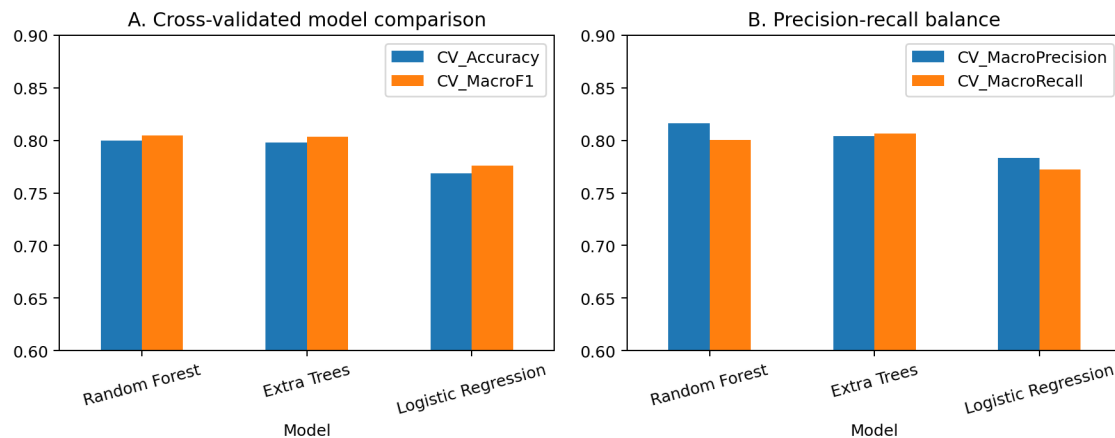


Figure 3. Cross-validated benchmarking of candidate models

5.3. Hold-out performance of the selected model

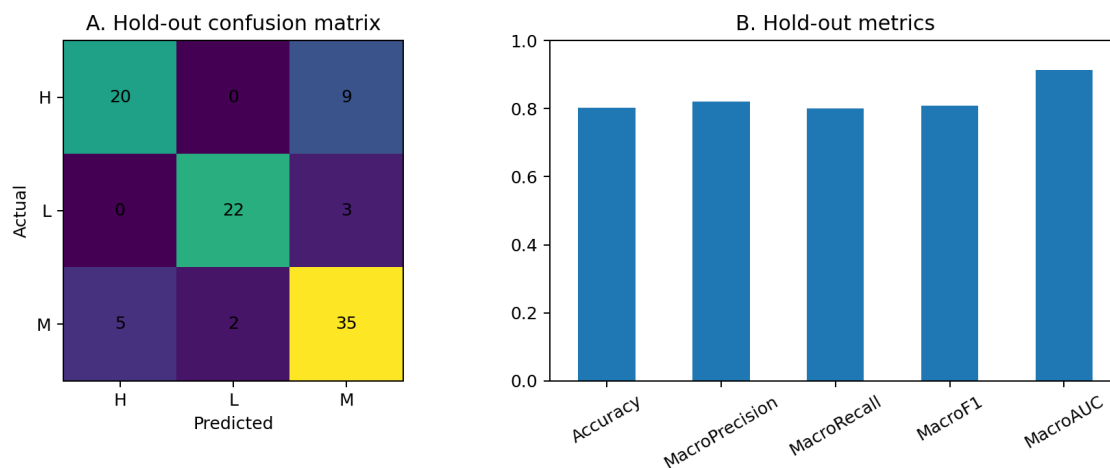
On the hold-out sample, the selected Random Forest model achieves an accuracy of 0.8021, a macro-precision of 0.8204, a macro-recall of 0.8010, a macro- F_1 score of 0.8084, and a macro-AUC of 0.9140. The confusion matrix indicates the strongest separation for class L , while residual confusion remains between classes H and M , which is plausible given that medium- and high-performing students share some behavioral characteristics.

Table 3: Hold-out performance of the selected Random Forest model

Metric	Value
Accuracy	0.8021
Macro-Precision	0.8204
Macro-Recall	0.8010
Macro- F_1	0.8084
Macro-AUC	0.9140

Table 4: Class-specific hold-out performance

Class	Precision	Recall	F_1
H	0.7442	0.7879	0.7654
L	0.9375	0.8824	0.9091
M	0.7797	0.7321	0.7552

**Figure 4.** Hold-out performance of the selected model

5.4. Predictor importance and support-plan portfolio

Permutation-importance analysis identifies `StudentAbsenceDays` as the most influential predictor, followed by the composite `EngagementIndex`, `gender`, `Relation`, `SupportProfile`, `VisITedResources`, and `raisedhands`. The ranking suggests that attendance and sustained participation provide the strongest signals for performance differentiation in this dataset.

The action-allocation layer translates the predictive results into a small portfolio of support plans. Within the hold-out sample, the largest assigned category is *adaptive engagement support* (31 students), followed by *challenge-and-extend pathway* (25), *attendance-first intervention* (23), *family-engagement reinforcement* (16), and *structured progression coaching* (1). This distribution indicates that actionable analytics in higher education should not be limited to identifying severe risk; it should also distinguish cases that require moderate support, attendance stabilization, enrichment, or family-linked engagement strategies.

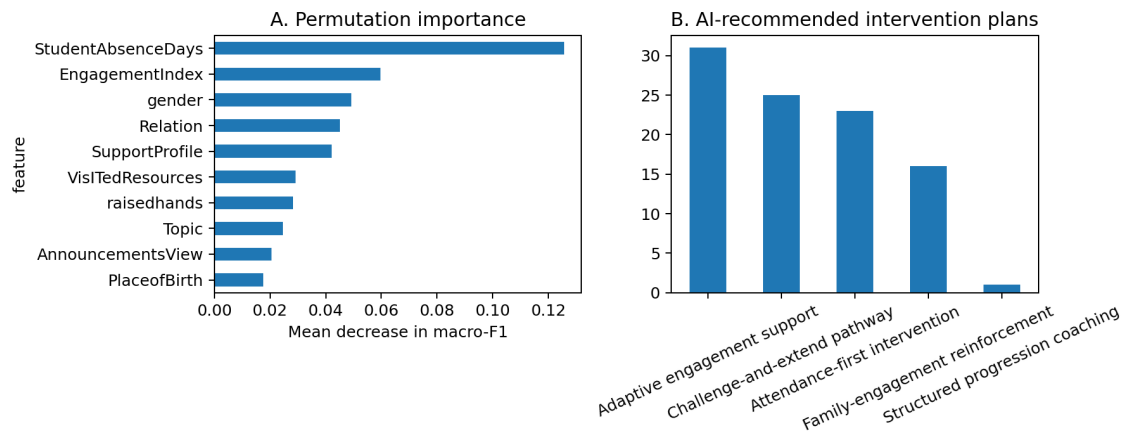


Figure 5. Predictor importance and distribution of AI-recommended support plans

6. Discussion

The empirical results support three main observations. First, simple digital-trace variables contain substantial information for the analysis of student performance. Attendance, resource access, and participation behavior provide the strongest predictive signals, which is consistent with recent work identifying engagement behavior as a central component of learning analytics in higher education (Bergdahl et al., 2024; Palanci et al., 2024).

Second, the explanatory value of AI improves when prediction is embedded within intermediate analytical layers. The engagement index and the support-profile segmentation provide more interpretable representations of student behavior than a single predictive probability alone. This result is consistent with human-centred learning-analytics research, which argues that AI systems should make their logic intelligible to educators and preserve space for pedagogical judgment (Alfredo et al., 2024). It also accords with intervention-oriented frameworks showing that the utility of student prediction increases when it is directly coupled with pedagogical response (Alalawi et al., 2025a, 2025b).

Third, the action-allocation layer illustrates how AI can support differentiated instructional planning. The resulting support portfolio is heterogeneous: some students are best addressed through attendance stabilization, some through adaptive engagement support, some through family-linked reinforcement, and some through extension-oriented pathways. This matters because educational response is rarely one-dimensional. In practical settings, undifferentiated alerts may overstate some cases and under-specify others. A portfolio-based approach provides a more usable basis for instructional coordination.

Several limitations should be acknowledged. The dataset is moderate in size and represents one specific educational context. The support plans are analytically inferred rather than validated through field implementation. In addition, the action-allocation rules are deliberately interpretable and therefore simpler than optimization-based assignment procedures. Future research may extend the framework to larger contemporary datasets, compare alternative intervention-allocation mechanisms, and examine how generative-AI tools may be integrated into the execution of support plans.

7. Practical Implications

The results have direct implications for AI-enabled education technology management which higher education institutions use to manage their educational technology systems. First, institutions should distinguish clearly between predictive modeling and support design. A model that performs well statistically but does not lead to a credible response mechanism has limited practical value. Second, analytical systems should combine attendance, behavioral, and support-related signals rather than depending on a single engagement indicator. Third, instructional support should be managed as a portfolio. The advising units together with instructors and student-success teams need multiple response modes which include stabilization and coaching and reinforcement and enrichment pathways.

These implications become especially important for educational institutions which want to operate AI-enabled student-support systems through transparent and ethical governance. From this perspective, explainability functions as a technical attribute which educators need to make informed decisions based on educational data.

8. Conclusion

This research designed and evaluated an explainable AI approach to engagement-responsive support in digital higher education. The study integrated the creation of engagement indices, the segmentation of support profiles, the multiclass prediction of support profiles and the allocation of support actions in a replicable workflow using the public xAPI-Edu-Data dataset. The Random Forest model achieved good hold-out results, while the intermediate profile and action level ensured interpretability and relevance for action.

The results demonstrate that AI in education technology may be structured around decision support rather than prediction. The association of digital traces with actionable support profiles provides an explainable foundation for higher education intervention.

References

- 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, Article 100215. <https://doi.org/10.1016/j.caeai.2024.100215>
- Alalawi, K., Athauda, R., Chiong, R., & Renner, I. (2025a). Evaluating the student performance prediction and action framework through a learning analytics intervention study. *Education and Information Technologies*, 30, 2887–2916. <https://doi.org/10.1007/s10639-024-12923-5>
- Alalawi, K., Athauda, R., & Chiong, R. (2025b). An extended learning analytics framework integrating machine learning and pedagogical approaches for student performance prediction and intervention. *International Journal of Artificial Intelligence in Education*, 35(3), 1239–1287. <https://doi.org/10.1007/s40593-024-00429-7>
- AlGhamdi, R. (2024). Exploring the impact of ChatGPT-generated feedback on technical writing skills of computing students: A blinded study. *Education and Information Technologies*, 29, 18901–18926. <https://doi.org/10.1007/s10639-024-12594-2>
- 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, Article 63. <https://doi.org/10.1186/s41239-024-00493-y>

- Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 4, Article 100118. <https://doi.org/10.1016/j.caeai.2022.100118>
- Deng, R., Jiang, M., Yu, X., Lu, Y., & Liu, S. (2025). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 227, Article 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
- Ghimire, S., Abdulla, S., Joseph, L. P., Prasad, S., Murphy, A., Devi, A., Barua, P. D., Deo, R. C., Acharya, R., & Yaseen, Z. M. (2024). Explainable artificial intelligence-machine learning models to estimate overall scores in tertiary preparatory general science course. *Computers and Education: Artificial Intelligence*, 7, Article 100331. <https://doi.org/10.1016/j.caeai.2024.100331>
- Kim, J., Detrick, R., Yu, S., Song, Y., Bol, L., & Li, N. (2025). Socially shared regulation of learning and artificial intelligence: Opportunities to support socially shared regulation. *Education and Information Technologies*, 30, 11483–11521. <https://doi.org/10.1007/s10639-024-13187-9>
- Kleimola, R., Hirsto, L., & Ruokamo, H. (2025). Promoting higher education students' self-regulated learning through learning analytics: A qualitative study. *Education and Information Technologies*, 30(4), 4959–4986. <https://doi.org/10.1007/s10639-024-12978-4>
- Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, 6, Article 100221. <https://doi.org/10.1016/j.caeai.2024.100221>
- Mumtaz, S., Carmichael, J., Weiss, M., & Nimon-Peters, A. (2025). Ethical use of artificial intelligence based tools in higher education: Are future business leaders ready? *Education and Information Technologies*, 30, 7293–7319. <https://doi.org/10.1007/s10639-024-13099-8>
- Palanci, A., Yilmaz, R. M., & Turan, Z. (2024). Learning analytics in distance education: A systematic review study. *Education and Information Technologies*, 29, 22629–22650. <https://doi.org/10.1007/s10639-024-12737-5>
- Pierrès, O., Darvishy, A., & Christen, M. (2025). Exploring the role of generative AI in higher education: Semi-structured interviews with students with disabilities. *Education and Information Technologies*, 30, 8923–8952. <https://doi.org/10.1007/s10639-024-13134-8>
- Schmidt, D. A., Alboloushi, B., Thomas, A., & Magalhaes, R. (2025). Integrating artificial intelligence in higher education: Perceptions, challenges, and strategies for academic innovation. *Computers and Education Open*, 9, Article 100274. <https://doi.org/10.1016/j.caeo.2025.100274>