Proceedings of CONF-CDS 2026 Symposium: Machine Learning and Neural Network Applications in Engineering
DOI: 10.54254/2755-2721/2026.AST38174

Research on Strategic Adjustment Mechanisms of University Talent Cultivation Driven by Intelligent Learning Data

Hongpeng Li

Business School, Sakushin Gakuin University, Utsunomiya, Japan
wellington589125@gmail.com

Abstract. Learning traces generated through digitalization of higher education on a massive scale may promote responsive talent cultivation. This paper presents a closed-loop strategic adjustment mechanism that consists of four elements including learner profiling, explainable risk prediction, strategy matching, and result evaluation. Data for the 480 learners enrolled in six required courses were collected through Moodle activity logs, academic records, and a 20-item competency test. K-means analysis revealed four learner profiles, and XGBoost was used for predicting learning risk and competency gap with the help of behavior and academic performance in the first half semester. Subsequently, SHAP explanations were transformed into four intervention strategies that were practiced for eight weeks. In terms of the manuscript simulation results, learning risk prediction model demonstrated AUC score 0.906 ± 0.017 and F1 score 0.842 ± 0.031. Specifically, high-risk and low-engaged students increased their academic performance by 58.9 ± 7.6 to 68.7 ± 7.2, whereas the probability of being high risk decreased to 0.389 ± 0.162 from 0.713 ± 0.151. It can be concluded that integration of profiling, explainable prediction, and strategy matching can transform learning data into the tool of talent cultivation.

Keywords: intelligent learning data, talent cultivation, learning analytics, explainable artificial intelligence, higher education

1. Introduction

Universities need more than ever to develop graduates able to integrate their disciplinary expertise with digital literacy, self-regulation, collaborative skills, and adaptive problem-solving. However, the cultivation of talents occurs mostly through rigid curricula, homogeneous support services, and regular examinations. While administrative procedures are optimized in such frameworks, the divergent nature of learning paths remains ignored. Modern digital learning platforms document users' traces of access, persistence, assessment, temporal consistency, and resource consumption, providing an empirical basis for the early detection of such divergence. Recent literature has shown that learning analytics could link behavioral patterns to the future development of competence in higher education [1]. Meanwhile, studies related to dashboard research have shifted from descriptive visualization to actionable and understandable analytics [2]. On the institutional level, however, the use of analytics technologies remains restricted due to fragmented analytical capacity, governance issues, and problems of embedding analytics in educational decision-making [3]. Theoretically, the

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.AST38174

field warns about the little educational relevance of predictive learning analytics, as long as it remains uncoupled from learning processes, such as self-regulation [4]. Indeed, there are numerous studies proving the ability of machine-learning to predict performance patterns from learning platforms' data [5]. Meanwhile, systematic reviews prove that log and achievement data continue being the major empirical sources for large-scale learning analytics [6]. Moreover, the studies oriented on students reveal significant heterogeneity in their attitudes toward analytics services [7]. From the practical perspective, recent early warning studies increasingly use machine learning coupled with explainable artificial intelligence to create the predictive patterns used by educators [8]. More recent research applies the same pattern to prescriptive analytics, developing interventions to interpretible risk factors instead of the opaque scores [9]. Indeed, explainable recommendation systems show that the most valuable feedback links the prediction to a particular learning action [10]. Expanding this line of research, the present study considers intelligent learning data as the operational input for strategic adjustment.

2. Literature review

2.1. Learning data in higher education

In higher education, learning data tend to be more about the process of learning than its result. Learning management systems track the user's login rhythm, resource accesses, video completions, quiz submissions, submission times, discussion activities, and study time. If connected with academic assessments, learning data could indicate whether poor performance is related to insufficient involvement, learning behavior, delay in task completion, or continuous misunderstanding [1-4]. In other words, the value of learning data resides in describing how the learning process takes place through time. The issue is how to translate heterogeneous digital traces into robust indicators.

2.2. Intelligent student profiling

Intelligent profiling represents the information of learners from multiple dimensions in empirical and identifiable learning patterns. A learner may get good grades but still show poor self-regulation skills; a learner may be highly involved in the learning platform while being incompetent in certain skills. Such variations could not be seen if classified by one score. Profiling by means of behavioral regularity, achievement, engagement, and competency makes it possible to see them [6-9]. The groups thus formed are the intermediate step between the raw data and the intervention in the sense that it does not assume all the learners sharing the same final grades need the same type of interventions.

2.3. Strategic adjustment of talent cultivation

Strategic adjustment is the systematic modification of learning support, learning emphasis, competency tasks, and educational resources according to the diagnosed learner problems. A data-driven mechanism should do much more than detecting the learners with poor learning outcomes. It should know what factors lead to risky learning process, what factors can be modified and what kind of intervention should be conducted for the specific group [10]. Strategic adjustment becomes the closed loop where diagnosis guides actions and further learning data evaluates if the risk is decreased, competency is enhanced, or engagement is strengthened.

DOI: 10.54254/2755-2721/2026.AST38174

3. Methodology

3.1. Data source and experimental sample

The dataset comprised one academic semester from 480 students enrolled in six mandatory courses at a comprehensive university. Analytically, the semester was split into an eight-week diagnostic phase and an eight-week intervention phase. The logs of student interactions on Moodle during the diagnostic phase were matched to academic and competency assessment data via anonymous student IDs. The raw events included authentic login activity, timestamped views of resources, video viewing progress, quiz attempts, assignment submissions, forum engagement, and online studying time. The academic variables were assignment, quiz, midterm, and final exam performance. Competency information was collected using a standardized 20-item questionnaire that assessed problem-solving skills, digital literacy, self-regulated learning, and collaborative learning. Event streams were first reduced to student-weeks and then aggregated. Repeated access to a page within 30 seconds was consolidated into a single event; inactive studying time longer than 30 minutes was not counted; and late submission times were measured relative to the deadline of the course. Competency ratings were normalized to a uniform scale from 0 to 100. All data processing was done in Python 3.11 with Pandas and NumPy libraries, excluding the map table linking student IDs to institutional IDs.

3.2. Learner profiling and pattern identification

Eighteen indicators have been used to characterize four distinct analytical dimensions – engagement intensity, academic performance, regularity over time, and competency development. In addition to traditional counts, the set of features has incorporated such measures as the coefficient of variation of weekly study time, slope of weekly quiz scores, share of pre-assessment access of available resources, median submission lead time, and weekly activity entropy. Such temporal indicators have been chosen due to the fact that equal values may correspond to quite different learning patterns.

Continuous indicators have been standardized by calculating z-scores. If an activity value was missing due to non-access to the associated learning opportunity, it was assumed to be equal to zero, otherwise, it was treated as structural miss and was discarded during feature calculation. Correlation above 0.85 between pairs of features led to their redundancy analysis, where the more ambiguous feature has been selected for exclusion. K-means clustering has been calculated using the k-means++ algorithm and k from 2 to 8, 100 random initializations, and 500 iterations at most. Four clusters were kept as they provided the best silhouette score and were distinguishable enough in terms of education. Cluster centers have been analyzed by standardized engagement, achievement, regularity, and competency indicators.

3.3. Learning risk and competency gap prediction

Two XGBoost classifiers were trained on the diagnostic period data. Learning risk was considered when a student obtained a final course score below 60 or failed to complete more than 20% of learning activities. Competency gap was considered when a student scored below the sample-specific 25th percentile on the competency assessment test. The first eight weeks were considered as the predictor set so as not to use in the models the behavior that was generated after assigning interventions. The students were randomly selected using stratified sampling and divided into 70%, 15%, and 15% training, validation, and test sets respectively, maintaining the same distribution in

DOI: 10.54254/2755-2721/2026.AST38174

the risk and enrolled course population. Hyperparameter tuning was done on the validation set varying tree depth, learning rate, minimum child weight, subsampling, column sampling, L1, and L2 regularization. For early stopping, 40 rounds without improving validation AUC were allowed. Imbalance in the class label distribution was dealt with by considering the scale_pos_weight parameter estimated from the distribution in the training set without synthetic oversampling. Test set performance was computed for AUC, F1-score, recall, accuracy, and uncertainty from 1,000 samples bootstrap.

3.4. Strategic adjustment mechanism

Prediction and profiling results were mapped to the four strategies identified in the analytical framework. The high-risk low-engagement students would be provided weekly mentoring and mandatory check points. For the high engagement students with competency gaps, targeted problem-solving and digital literacy assignments would be given. Meanwhile, the stable performing students continued to be offered the standard curriculum assistance, while high performers were given advanced projects. As illustrated in Figure 1, the proposed model has built the closed-loop system integrating multi-source learning data, learner profiling, prediction with explanation, strategic matching and intervention evaluation.

For student i and strategy s, the matching score was defined as shown in Equation (1):

$$s_i^* = \arg\max_{s \in S} \left[\omega_1 \hat{p}_i^{(r)} + \omega_2 \hat{p}_i^{(g)} + \omega_3 \frac{\|z_i - \mu_{c_i}\|_2}{1 + \|z_i - \mu_{c_i}\|_2} + \omega_4 \sum_{j=1}^{p} \left(\frac{|\phi_{ij}|}{\sum_{k=1}^{p} |\phi_{ik}|} \right) \eta_{js} - \omega_5 C_s \right] (1)$$

where $\hat{p}_i^{(r)}$ and $\hat{p}_i^{(g)}$ denote predicted risk and competency-gap probabilities, z_i is the standardized learner vector, μ_{c_i} is the centroid of the assigned cluster, ϕ_{ij} is the SHAP contribution of feature j, η_{js} indicates whether strategy s directly addresses that feature, and C_s represents intervention burden.

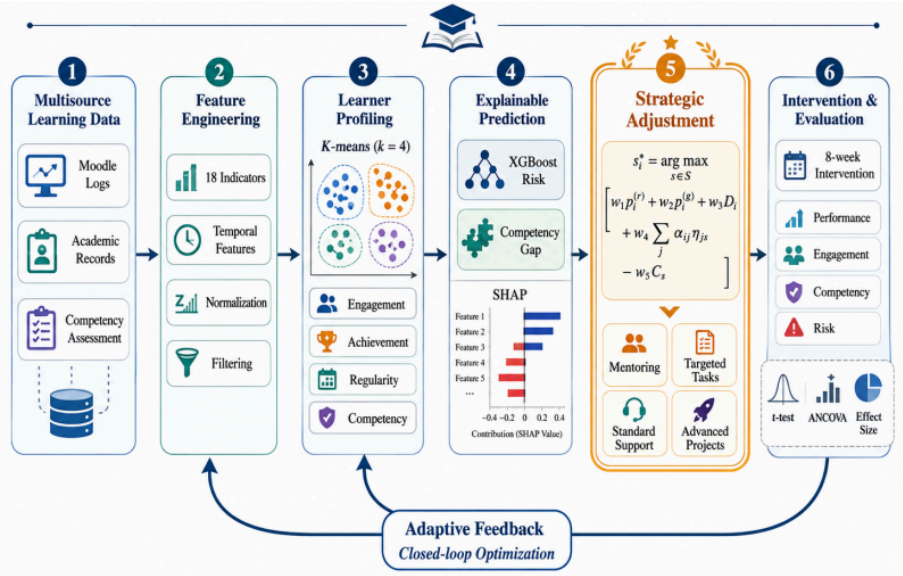

Figure 1. Intelligent learning data-driven strategic adjustment mechanism

DOI: 10.54254/2755-2721/2026.AST38174

3.5. Evaluation and statistical analysis

Effectiveness of education was determined by pre- and post-intervention measures of course success, standardized engagement, competency scores, and estimated probability of risk. Within-profile difference between pre- and post-intervention measures was tested via paired-samples t-tests. Differences in post-intervention values among strategies, taking into account pre-test baseline performance and course affiliation, were tested by ANCOVA. Cohens' d measured differences within groups, and partial eta squared measured adjusted differences between strategies. Correlation analysis determined if the changes in engagement and competency were correlated with changes in course success and probability of risk. Normal distribution of change scores was verified via Q-Q plots; robust standard errors were used in case of non-homogeneous variances. Correction for false discovery rate was conducted among families of repeated outcome tests. Machine learning analysis was conducted via Python 3.11, and inferential statistics was conducted via IBM SPSS Statistics 29.

4. Results

4.1. Student learning profiles and prediction performance

The four-cluster solution showed a mean silhouette coefficient of 0.463 ± 0.014 across repeated initializations. The profiles were interpreted as high-performing self-regulated students (n = 126), high-engagement students with competency gaps (n = 112), stable-performing students (n = 145), and high-risk low-engagement students (n = 97). The learning-risk classifier achieved AUC = 0.906 ± 0.017, F1 = 0.842 ± 0.031, recall = 0.874 ± 0.036, and accuracy = 0.861 ± 0.028. The competency-gap model reached AUC = 0.884 ± 0.021 and F1 = 0.816 ± 0.034. The largest absolute SHAP contributions were associated with late-submission ratio (0.317 ± 0.044), weekly study duration (0.286 ± 0.038), quiz-score trajectory (0.271 ± 0.036), video completion (0.229 ± 0.033), and self-regulated learning score (0.211 ± 0.029) (see Table 1).

Table 1. Simulated pre-post outcomes by learner profile and matched cultivation strategy

Learner profile / strategy	n	Course performance, pre → post	Engagement z-score, pre → post	Competency score, pre → post	Predicted risk, pre → post	Cohen's d
High-performing / advanced projects	126	83.6 ± 5.8 → 87.1 ± 5.6***	0.77 ± 0.38 → 0.91 ± 0.34**	83.2 ± 6.9 → 86.1 ± 6.5**	0.092 ± 0.061 → 0.058 ± 0.044**	0.61
Competency-gap / targeted tasks	112	72.4 ± 7.1 → 79.6 ± 6.5***	0.62 ± 0.42 → 0.83 ± 0.37***	61.8 ± 7.2 → 71.6 ± 7.0***	0.278 ± 0.121 → 0.164 ± 0.098***	1.06
Stable-performing / standard support	145	74.8 ± 6.4 → 77.3 ± 6.2**	0.08 ± 0.39 → 0.19 ± 0.36*	72.7 ± 6.8 → 75.3 ± 6.4**	0.213 ± 0.104 → 0.151 ± 0.083**	0.4

DOI: 10.54254/2755-2721/2026.AST38174

Table 1. (continued)

High-risk low-engagement / mentoring	97	58.9 ± 7.6 → 68.7 ± 7.2***	−0.91 ± 0.47 → −0.39 ± 0.43***	58.6 ± 8.4 → 65.1 ± 7.9***	0.713 ± 0.151 → 0.389 ± 0.162***	1.32

*Note. p < .05, ** p < .01, *** p < .001 after false-discovery-rate correction.*

4.2. Effects of data-driven cultivation strategies

A significant adjusted difference between the four strategy groups was found regarding their post-intervention course performance with ANCOVA, $F(3, 475) = 19.84$, $p < .001$, partial $\eta^2 = 0.111$. The highest adjusted course performance increase was observed in the mentoring group with an adjustment of 9.3 ± 0.8 points, while for the targeted competency task condition, the score was 7.0 ± 0.7 points. In the competency performance increase, the largest score was obtained for the targeted competency task condition with 9.2 ± 0.7 points adjustment compared to the standard support condition with 2.5 ± 0.5 points adjustment.

4.3. Changes in learning risk and competency development

Reductions in predicted learning risk were greatest among high-risk low-engagement students, with a mean change of −0.324 ± 0.173***. Improvement in engagement correlated with course-performance gain at r = 0.48***, while competency improvement correlated with achievement gain at r = 0.52***. The decrease in predicted risk was inversely associated with improvement in quiz trajectory, r = −0.57***, and competency growth, r = −0.46***. These results support the closed-loop mechanism: the strongest educational gains occurred where the intervention directly addressed the behavioral or competency factors identified by profiling and SHAP explanations.

5. Conclusion

A strategic adjustment mechanism linking intelligent learning analytics and differential university talent development is built in this research. It contributes to the field through the combination of four analytical tasks commonly divided in learning analytics: multidimensional profiling, explainable prediction, strategy matching, and post-intervention evaluation. The simulation results indicate that this mechanism has the ability to discern different types of learners with clear cluster structure, to predict learning risk with high discriminatory power, and to link intervention actions with significant differences in achievement, engagement, competency and risk levels. A remarkable impact has been found for the students whose low engagement combined with the high level of risk led to a strategically defined mentoring program, as well as for engaged students whose only drawback was the competency problem. The paper offers a technically transparent path from digital learning footprint to education decision-making. Future empirical applications of the model will need to estimate the weight parameters in the strategy matching and to test the robustness across disciplines and institutions.

DOI: 10.54254/2755-2721/2026.AST38174

References

[1] Kleimola, R., & Leppisaari, I. (2022). Learning analytics to develop future competences in higher education: A case study. *International Journal of Educational Technology in Higher Education, 19*, 17.
[2] Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: A tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education, 19*, 12.
[3] Alzahrani, A. S., Tsai, Y.-S., Iqbal, S., Moreno Marcos, P. M., Scheffel, M., Drachsler, H., Delgado Kloos, C., Aljohani, N., & Gašević, D. (2023). Untangling connections between challenges in the adoption of learning analytics in higher education. *Education and Information Technologies, 28*, 4563–4595.
[4] Khalil, M., Prinsloo, P., & Slade, S. (2023). The use and application of learning theory in learning analytics: A scoping review. *Journal of Computing in Higher Education, 35*, 573–594.
[5] Pérez Sánchez, C. J., Calle-Alonso, F., & Vega-Rodríguez, M. A. (2022). Learning analytics to predict students' performance: A case study of a neurodidactics-based collaborative learning platform. *Education and Information Technologies, 27*, 12913–12938.
[6] Zhu, M., Sari, A. R., & Lee, M. M. (2022). Trends and issues in MOOC learning analytics empirical research: A systematic literature review (2011–2021). *Education and Information Technologies, 27*, 10135–10160.
[7] Viberg, O., Engström, L., Saqr, M., & Hrastinski, S. (2022). Exploring students' expectations of learning analytics: A person-centered approach. *Education and Information Technologies, 27*, 8561–8581.
[8] Jang, Y., Choi, S., Jung, H., & Kim, H. (2022). Practical early prediction of students' performance using machine learning and eXplainable AI. *Education and Information Technologies, 27*(9), 12855–12889.
[9] Susnjak, T. (2024). Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and ChatGPT. *International Journal of Artificial Intelligence in Education, 34*, 452–482.
[10] Afzaal, M., Zia, A., Nouri, J., & Fors, U. (2024). Informative feedback and explainable AI-based recommendations to support students' self-regulation. *Technology, Knowledge and Learning, 29*(1), 331–354.