EVA-LA: An Explainable Value-Added Learning Analytics Framework for Higher Education

Fang Liu[1], Yuting Yang[1], Mengxue Hong[1] and Qin Dai[2]

[1] School of Computer Science, Shenyang Aerospace University, Shenyang, China, 110136
liufang@sau.edu.cn
[2] College of Information, Shenyang Institute of Engineering, Shenyang, China, 110136
daiqin@sie.edu.cn

Abstract. Existing learning analytics approaches in higher education provide limited insight into students' incremental academic growth and its underlying mechanisms. To address this gap, we propose an explainable value-added learning analytics framework, which contributes a residual-based value-added modeling paradigm, along with a cross-level interpretability mechanism for explaining student growth. It first constructs a structured learner representation that fuses heterogeneous digital trace data into a unified analytical space spanning behavioral, process, and performance dimensions. To quantify students' academic growth, we build an elastic net-based estimator to model expected outcomes under high-dimensional and collinear predictors, enabling the derivation of individualized value-added scores as adjusted indicators of net learning progress. To enhance interpretability, we further develop Bi-Lens, a bidirectional explanatory framework that couples SHAP-based global attribution with LIME-based local diagnosis, supporting cross-level explanatory coherence and strengthening the identification of factors associated with value-added learning outcomes. Experiments on two real-world courses demonstrate competitive performance (R^2 = 0.541 and 0.368) and stable, well-separated value-added estimates, while yielding interpretable insights for student development.

Keywords: Learning Analytics, Value-Added Evaluation, Explainable Learning, Elastic Net.

1 Introduction

With the widespread adoption of computer-aided online learning platforms, such as Learning Management Systems (LMS), Intelligent Tutoring Systems (ITS), and Massive Open Online Courses (MOOCs), large volumes of heterogeneous digital footprints are continuously generated throughout learners' learning processes [1, 2]. While these data provide rich evidence for understanding learning behaviors and performance, they also make meaningful evaluation increasingly challenging due to their scale and complexity [3, 4]. In higher education, the key issue is not merely to describe learning behaviors or predict outcomes, but to assess students' incremental academic growth under different prior learning performance [5]. In this regard, value-added evaluation is essential because it focuses on learners' progress rather than absolute achievement [6], thus supporting a fairer and more development-oriented assessment. However, although Learning Analytics (LA) has been widely applied in education [1], most existing studies remain centered on descriptive analysis and predictive modeling [7], with insufficient attention to value-added evaluation and the factors underlying learning improvement. Consequently, they are limited in capturing learners' true developmental gains and explaining how such gains are achieved.

In higher education, in particular, LA research has primarily focused on learning outcome visualization, performance comparison, risk detection, and personalized intervention [2, 8, 9]. While these approaches provide valuable insights into students' current performance and learning processes, they pay relatively limited attention to students' developmental progress over time. Value-added evaluation offers a complementary perspective by measuring students' growth after accounting for prior learning performance [10]. This approach has been explored through input-adjusted outcome measures, economic value-added indicators, and multilevel models [6]. Although these studies underscore the importance of development-oriented assessment, they are largely oriented toward institutional performance or traditional statistical evaluation, with limited use of heterogeneous online learning footprints and little attention to interpretability [9, 11]. Therefore, existing research remains insufficient for capturing students' incremental improvement and explaining the factors that drive such improvement within learning analytics settings.

Despite the rapid development of LA in higher education, several challenges remain. Existing research still lacks structured learner modeling for value-added analysis. Educational data are often scattered across heterogeneous platforms such as LMSs, MOOCs, and intelligent tutoring systems, making it difficult to construct a unified representation of learners' behavioral, process, and performance features from heterogeneous digital footprints [11]. Although some studies have explored open learner models [12], they primarily focus on visualization and comparison of knowledge states rather than structured support for fine-grained growth analysis [8]. Meanwhile, current LA research provides limited support for robust value-added evaluation. Most existing studies focus on descriptive statistics, group comparison, or risk prediction, rather than measuring learners' incremental academic growth after accounting for prior learning performance [4]. In addition, the interpretability of value-added results remains inadequate, as many existing methods still face a black-box dilemma and produce abstract evaluation scores without clearly explaining why learners exhibit different levels of growth [13]. These limitations weaken the usefulness of LA for educational diagnosis and intervention, and highlight the need for explainable value-added analysis.

To address the above limitations, this paper proposes EVA-LA, an explainable value-added learning analytics framework for higher education. Centered on value-added evaluation, the framework captures learning gains as deviations from expected performance. Specifically, we construct a structured learner model that integrates heterogeneous digital footprints from multiple educational platforms into a unified representation, providing a process-complete basis for fine-grained growth analysis. On this basis, value-added evaluation is formulated as a regression task, where an elastic net model is employed to estimate expected academic performance from high-dimensional features. Learning gains are then quantified as residual differences between observed and predicted outcomes, enabling a more robust and individualized assessment of academic progress [14]. To enhance interpretability, we further develop Bi-Lens, a dual-perspective explanation framework that combines SHapley Additive Explanations (SHAP) -based global attribution with Local Interpretable Model-agnostic Explanations (LIME) -based local explanations. By linking cohort-level feature importance with instance-level contributions, Bi-Lens provides cross-level consistency in explaining learning gains, facilitating the identification of both common drivers and individual-specific factors underlying student development [15, 16]. Overall, EVA-LA improves both the robustness and interpretability of value-added evaluation, offering a principled and actionable basis for educational diagnosis and targeted intervention.

The main contributions of this paper are as follows:

(1) We define a learner model for value-added learning analytics, which transforms heterogeneous digital footprints from multiple educational platforms into a unified and process-aware representation, thereby supporting fine-grained characterization of students' developmental trajectories.

(2) It formulates value-added evaluation as a regression-based paradigm that estimates expected academic performance and defines learning gains as residual deviations from this baseline, advancing value-added analysis from descriptive comparison toward more rigorous growth measurement.

(3) We propose Bi-Lens, a dual-perspective explainer that integrates SHAP-based global attribution with LIME-based local explanations. By establishing a consistent linkage between cohort-level patterns and instance-level evidence, we enhance the interpretability of factors associated with learning gains.

The paper is structured as follows. Section 2 presents the related work of this study. Section 3 presents the proposed explainable value-added learning analytics framework. Section 4 presents the experimental and analysis results. Finally, the conclusion is in Section 5.

2 Related Work

2.1 Learning Analysis

Learning analytics (LA) has emerged as a rapidly growing field that applies data analysis techniques to understand and optimize learning processes. Li et al. conducted a comprehensive survey on predicting at-risk students through LA, highlighting the field's evolution from descriptive analysis toward more sophisticated predictive applications [7]. Almalawi et al. reviewed educational predictive models, revealing increasing use of machine learning to predict at-risk students and academic outcomes for early intervention [9]. Pan et al. reviewed LA-integrated instructional interventions on learning management systems, finding descriptive analysis dominant and failure prediction also widely studied [17]. These capabilities position LA as a promising foundation for advancing more sophisticated assessment approaches that capture both learning processes and outcomes.

2.2 Value-Added Evaluation in Education

Value-added evaluation, as an assessment approach emphasizing learning progress, has established a relatively mature theoretical and practical foundation internationally. Eryilmaz et al. conducted empirical comparisons of various statistical models for value-added evaluation, providing references for methodological selection in value-added assessment [18]. Basileo et al. validated the effectiveness of teacher evaluation models in predicting value-added effects through state-level analysis [19]. However, Amrein-Beardsley et al. identified methodological concerns regarding the validity, reliability, and bias of existing value-added evaluation systems [20]. Recently, researchers have begun to explore predictive machine learning models for value-added evaluation. Li et al. indicated that methods such as tree-based models and neural networks outperform traditional statistical models in capturing nonlinear relationships in the learning process [21]. Almalawi et al. demonstrated that regularization-based regression models can effectively reduce estimation errors and enhance the stability and reliability of value-added scores when processing high-dimensional longitudinal educational data [9]. These studies provide methodological support for constructing value-added evaluation models within a regression-based framework.

2.3 Explainable Artificial Intelligence in Education

The application of explainable AI (XAI) in education has grown rapidly, focusing on opening the "black box" of predictive models to better understand learner behavior and cognition. Wong et al. employed SHAP value analysis to examine token contributions in collaborative problem-solving diagnosis, highlighting the risk of models relying on spurious correlations for decision-making in educational applications [22]. Larasati et al. employed SHAP to compare predictive drivers of student success between disabled and non-disabled learners, revealing that key influencing factors differ significantly across groups, underscoring the importance of fairness in learning analytics [23]. Ouared et al. proposed the XAI-Profile framework, which integrates explainable machine learning with interactive visualization to transform raw learning data into interpretable and actionable learner profiles, enhancing educators' trust and usability [24]. Similarly, Maathuis et al. evaluated multiple XAI methods in higher education prediction, demonstrating that glass-box models achieve comparable performance to black-box models while providing greater interpretability [12]. These studies collectively highlight the growing role of XAI in supporting transparent and equitable educational decision-making.

3 Explainable Value-Added Learning Analytics Methodology

This section presents the proposed EVA-LA framework, which integrates a structured learner model, elastic net-based value-added evaluation, and a Bi-Lens explainer to support interpretable academic analysis.

3.1 The Framework of EVA-LA

As shown in Fig.1, this paper proposes the explainable value-added learning analytics framework. Based on the learner model, we introduce elastic net regression to achieve precise value-added calculation and construct a Bi-Lens explainer to enable interpretation of value-added sources, thereby providing a comprehensive solution that combines robust growth measurement with transparent diagnostic insights.

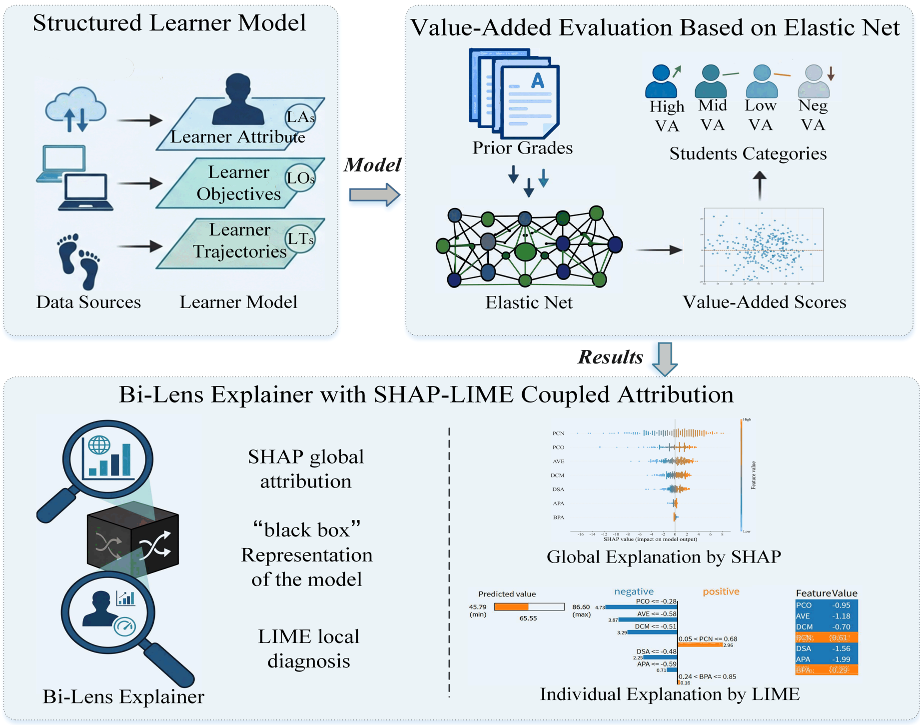


Fig. 1. The Framework of Explainable Value-Added Learning Analytics

3.2 The Structured Definition of the Learner Model

The Learner Model is used to structurally describe learners’ static attributes, learning objectives, and learning process trajectories [25]. Its formal definition is: $LM = \{LAs, LOs, LTs\}$, with each attribute set described as follows:

LAs (*Learner Attribute Set*): A set of learner attributes.
LOs (*Learning Objective Set*): A set of learning objectives for knowledge components.
LTs (*Learning Trace Set*): A set of learning trajectories.

Learner attribute: The learner attribute includes information such as school registration, classes, and majors. As education is targeted at all learners, the universal learner model should not only include individual learners but also cohorts. Therefore, we define three granularities of learners: individual learner, class learner, and major learner. The attributes of different granularity learners are:

Individual learner attributes' LA = {name, student ID, gender, class, birthplace},

Class learner attributes' LA = {class ID, student set, major},

Major learner attributes' LA = {major ID, class set}.

Learning objectives: Learning objectives are the learning tasks and guiding principles for learners, providing a crucial benchmark for assessing learning outcomes. These objectives set the knowledge components that learners must master and the required mastery degree. This article sets learning objectives on a per-course basis.

Learning Objectives $LO = \{CourseId, KCs, Level\}$.

Where *CourseID* is the ID of the course, *KCs* is a collection of knowledge components within the course, and *Level* describes the cognitive level that learners need to achieve for each knowledge component.

$KCs= \{kci，i=1,2...,n\}$,

$Level = \{lei \mid lei \in$(Memory, Understanding, Mastery, Analysis, Creation…), $i = 1, 2, ..., n\}$.

Where *kci* and *lei* refer to the ith knowledge component of a course and its cognitive level that the teaching objective requires learners to achieve. n is the total number of knowledge components in the course.

Learning Trajectory: The learning trajectory documents the entire learning process of a learner over time, providing the most direct description of the learning behavior. On a per-course basis, the learning trajectory encompasses learning objects, learning actions, learning outputs, and time.

Learning trajectory $LT =\{CourseID, LTas, LActs, LAccs, LTs\}$.

Where *LTas* is learning tasks, *LActs* is learning actions, *LAccs* is the completion status of learning tasks, and *LTs* is the start and end times of tasks.

$LTas = \{ tai \mid tai\in$ {teaching videos, assignment, experiments, tests, …}, $0 < i \leq m\}$;

$LActs = \{ acti \mid acti\in$ (watching, browsing, querying, submitting), $0 < i \leq m\}$;

$LAccs = \{(scorei$ or $accomi, duti) \mid scorei \in [0,100]$ or $accomi \in \{0,1\}, duti \in$ R, $0 < i \leq m\}$;

$LTs = \{(tib, tie) \mid tib<tie, 0 < i \leq m\}$.

Where m is the total number of learning tasks for a course. By nature, task i has two possible completion statuses: its score (*scorei*) or whether it is completed (*accomi*). *duti* represents the duration of task i's completion, *tib* represents the start time of task i, and *tie* represents the end time of task i.

3.3 Value-Added Evaluation based on Elastic Network

Value-added evaluation aims to measure students' academic growth by isolating the influence of prior learning performance. In this study, we formulate value-added evaluation as a regression task, where a student's expected performance is estimated based on their prior learning features. The value-added score is defined as the residual between the actual performance and the predicted expectation. Based on both teacher experience and the correlation coefficient matrix calculated from the courses, we observe that certain course grades exhibit high correlations. This collinearity issue is further exacerbated when integrating data from multiple learning platforms, which significantly increases the feature dimension. To address this issue, we employ Elastic Net regression, which introduces both L1 and L2 regularization [26]. This enables it to identify key prerequisite courses by performing feature selection. At the same time, it smoothly shrinks the coefficients of correlated feature groups, effectively mitigating multicollinearity. These properties align well with the modeling demands of value-added evaluation scenarios. Based on this formulation, we construct the Elastic Net-based value-added evaluation model as follows.

We define the input dataset as $D = \{(x_i, y_j)\}$, $i \in 1,2,...,N$, where x_i is the feature of learner i, $y_i \in [0,100]$ is his score, and N is the number of samples. The features consist of prerequisite course grades and some learning data from the current course.

For the given features, the Elastic Net regression model estimates coefficients β through a combination of L1 and L2 regularization, with the objective function shown in Equation (1):

$$L(\beta) = \frac{1}{2N}\sum_{i=1}^{N}(y_i - x_i^T\beta)^2 + \lambda\left[\alpha \parallel \beta \parallel_1 + \parallel \beta \parallel_2^2\right] \quad (1)$$

where $\lambda > 0$ controls the regularization strength, $\alpha \in [0,1]$ balances the ratio between L1 and L2 and T denotes the transpose operation. We employ coordinate descent to solve this optimization problem. For the j-th coefficient β_j, with other coefficients fixed, the partial derivative is given by Equation (2):

$$\frac{\partial L}{\partial \beta_j} = -\frac{1}{N}\sum_{i=1}^{N} x_{ij}(y_i - \hat{y}_i^{(j)}) + \lambda(1-\alpha)\beta_j + \lambda\alpha \cdot sign(\beta_j) \quad (2)$$

where x_{ij} denotes the j-th feature value of learner i and $\hat{y}_i^{(j)} = \sum_{k \neq j} x_{ik}\beta_k$ is the prediction excluding feature j. Setting the partial derivative to zero yields the soft-thresholding update formula presented in Equation (3):

$$\beta_j = \frac{S(\frac{1}{N}\sum_{i=1}^{N} x_{ij}(y_i - \hat{y}_i^{(j)}), \lambda\alpha)}{\frac{1}{N}\sum_{i=1}^{N} x_{ij}^2 + \lambda(1-\alpha)} \quad (3)$$

where $S(\frac{1}{N}\sum_{i=1}^{N} x_{ij}(y_i - \hat{y}_i^{(j)}), \lambda\alpha)$ is the soft-thresholding operator. We iteratively update all coefficients until convergence to obtain optimal estimates β^*. Based on the estimated Elastic Net model, the theoretical score of learner i is predicted as shown in Equation (4):

$$\hat{y}_i = \beta_0^* + x_i^T\beta^* \quad (4)$$

where β_0 is the estimated intercept term. The value-added score VE_i is defined as the residual between the actual score and the predicted score, as presented in Equation (5):

$$VE_i = y_i - \hat{y}_i \quad (5)$$

3.4 The Bi-Lens Explainer

Although elastic net regression can accurately quantify value-added scores, its "black box" nature makes it difficult to interpret the results. To this end, EVA-LA proposes a

Bi-Lens explainer, which integrates SHAP for global feature contribution analysis and LIME for local instance-level explanation. This design enhances the transparency and actionability of evaluation results, providing both macro-level insights into key growth drivers and fine-grained diagnostic evidence for individual students.

SHAP is an explainability method derived from Shapley values in game theory [27]. Its core logic involves calculating the marginal contribution of each feature to the model's prediction, enabling globally consistent and axiomatically verifiable explanations. Its advantage lies in the independent calculation of each feature's contribution, unaffected by other features, thus providing more precise explanations. For the value-added prediction model $f(x)$, SHAP constructs an explanatory model $g(x)$ that satisfies the additivity axiom. The model prediction value for sample x, as explained by SHAP, is given by Equation (6). In the value-added evaluation context of this study, SHAP can compute the average feature contribution across all samples to obtain global importance rankings, identifying the key factors that most significantly impact learners' overall value-added.

$$g(x) = \phi_0 + \sum_{i=1}^{p} \phi_i \tag{6}$$

where ϕ_0 is the average of the prediction values for all training samples, ϕ_i is the SHAP contribution value of the i-th feature to the prediction value of the current sample, and p is the total number of input features.

LIME is a model-agnostic local explainability method [28]. Its core approach involves approximating the prediction results of the original complex model within the local neighborhood of the sample to be explained, employing a simple, interpretable linear model to generate specific explanations for individual samples. For the target sample, the optimal local explanation model is obtained through the following equation (7). In the value-added evaluation context of this study, LIME can provide targeted explanations for typical individual learners, such as those with high or low value-added, uncovering personalized factors affecting these learners'value-added. This compensates for the limitation of SHAP's global explanations in capturing individual-specific features, thereby providing a more precise basis for personalized teaching interventions.

$$\xi(x) = arg \min_{g \in G} L(f, g, \pi_x) + \Omega(g) \tag{7}$$

Where G represents the set of interpretable models, f is the original complex model, $\pi_x$defines the neighborhood around sample$x$, $L(f, g, \pi_x)$measures how well g approximates f in that neighborhood, and $\Omega(g)$ penalizes the complexity of g to ensure interpretability.

4 Experimental Results and Discussion

4.1 Datasets

The dataset is a real learning performance dataset from the School of Computer Science at Shenyang Aerospace University. It comprises comprehensive undergraduate course records for all students enrolled from the 2019 cohort to the 2023 cohort. Each student has multiple pieces of information, including the student number, class, name, course names, and grades. In total, the dataset contains 106,087 pieces of data for 359 subjects. We encrypt the student IDs to protect privacy.

4.2 Value-Added Evaluation Model Settings

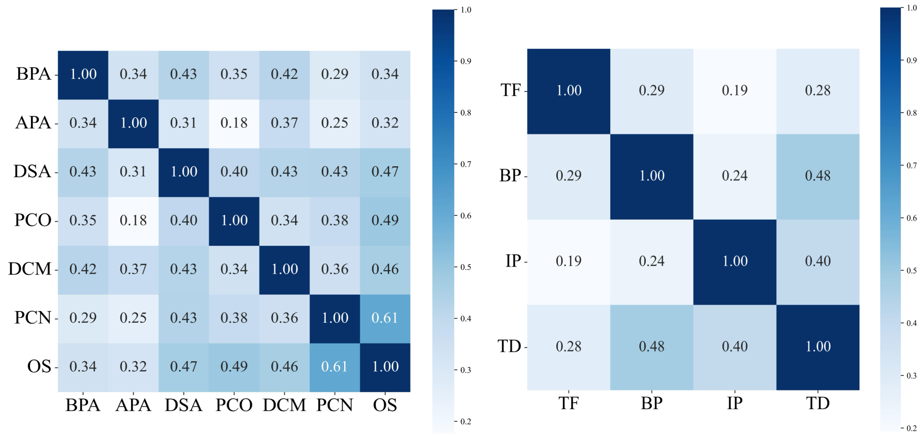
To comprehensively evaluate the performance of the proposed model, the dataset is split 8:2 for training and testing, with 10-fold cross-validation employed. This section describes baseline methods and evaluation metrics used in the comparative experiments.

Baseline Models. Five regression methods are selected for performance comparison. (1) Ridge Regression employs L2 regularization to mitigate multicollinearity. (2) Lasso Regression utilizes L1 regularization to perform feature selection. (3) Elastic Net is the method proposed in this study, which combines L1 and L2 regularization. (4) XGBoost is a gradient boosting ensemble that captures nonlinear relationships. (5) Random Forest is a bagging ensemble that enhances prediction stability.

Evaluation Metrics. Model performance is evaluated from two perspectives. (1) Prediction Accuracy: Root Mean Square Error (RMSE) measures deviation between predictions and actual values; Mean Absolute Error (MAE) measures the average absolute error; and the Coefficient of Determination (R^2) quantifies the explained variance. (2) Residual Analysis: Residual Standard Deviation (RSD) evaluates the prediction residual stability.

4.3 Value-Added Evaluation Model Performance Validation

This section aims to validate the performance of the proposed Elastic Net-based value-added evaluation model. Two courses with distinct disciplinary attributes are selected for experimentation: Course_one is Operating System (OS), and Course_two is Theoretical Development (TD). Based on correlation analysis, we identify the prerequisite courses for OS as Principles of Computer Organization (PCO), Data Structures and Algorithms (DSA), Principles of Computer Networks (PCN), Discrete Mathematics (DCM), Basic Programming A (BPA), and Advanced Programming A (APA). For TD, the prerequisite courses include Theoretical Frontiers (TF), Basic Principles (BP), and Introduction to Principles (IP). As illustrated in Figure 2, the correlation coefficient between PCN and OS reaches 0.61, and that between BP and TD is 0.48, indicating strong feature correlations among the prerequisite courses for both target courses.



(a) Course_one (b) Course_two

Fig. 2. Correlation Coefficient Heatmap of Prerequisite Course Grades on two courses

The performance evaluation consists of two parts: prediction accuracy and residual analysis. The prediction accuracy results are shown in Table 1, and the residual analysis results are shown in Figure 3.

(1) Prediction Accuracy Analysis. As shown in Table 1, Elastic Net achieves the best comprehensive performance on both courses. On Course_one, Elastic Net attains the highest test R^2 (0.541) and the lowest RMSE (9.12), while its MAE (7.19) is comparable to that of the best-performing model. On Course_two, Elastic Net outperforms all other baseline models across all three metrics, with a test R^2 of 0.368, an RMSE of 5.98, and an MAE of 4.69. In comparison, Ridge Regression and Lasso Regression show slightly inferior performance on certain metrics, while XGBoost exhibits significant overfitting on both courses, and Random Forest demonstrates the weakest overall prediction accuracy. These results indicate that Elastic Net maintains stable prediction accuracy advantages across courses of different disciplinary types.

Table 1. Model Performance on Student Value-Added Evaluation on two courses

Course	Model	Train R^2	Test R^2	RMSE	MAE
Course_one	Ridge	0.501	0.541	9.13	7.18
	Lasso	0.494	0.536	9.17	7.21
	Elastic Net	0.502	0.541	9.12	7.19
	XGBoost	0.575	0.522	9.31	7.37
	Random Forest	0.463	0.476	9.75	7.53
Course_two	Ridge	0.322	0.365	5.99	4.72
	Lasso	0.311	0.349	6.07	4.86
	Elastic Net	0.323	0.368	5.98	4.69
	XGBoost	0.479	0.312	6.24	4.70
	Random Forest	0.370	0.346	6.08	4.93

(2) Residual Analysis of value-added scores. As shown in Figure 3, Elastic Net consistently achieves the lowest residual standard deviation across both courses. On Course_one, the residual standard deviation of Elastic Net is 9.095, lower than that of Ridge Regression (9.101) and Lasso Regression (9.143), and significantly better than that of Random Forest (9.269) and XGBoost (9.709). On Course_two, Elastic Net again achieves the lowest residual standard deviation (5.952), outperforming all other baseline models. This benefit stems from Elastic Net's effective control of model complexity through collaborative regularization constraints, thereby achieving more robust error performance while maintaining prediction accuracy.

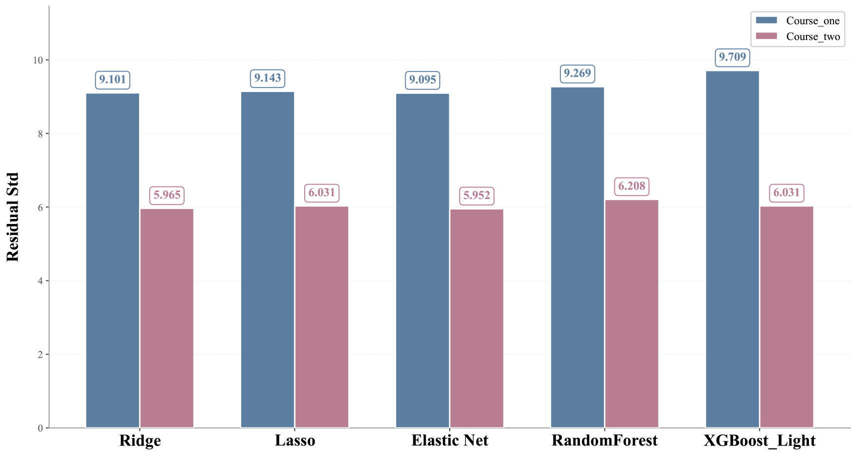


Fig. 3. Comparison of Residual Standard Deviation Across Different Models on two courses

In summary, Elastic Net demonstrates superior performance across prediction accuracy and residual analysis, validating its effectiveness and cross-disciplinary applicability as the core algorithm for value-added evaluation.

4.4 Interpretable Analysis of Evaluation Results

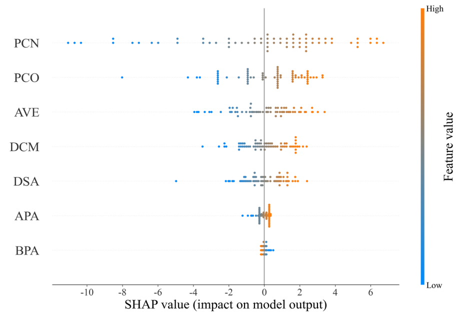
This section conducts interpretable analyses of the value-added evaluation results from both global and local perspectives.

To facilitate subsequent differentiated analysis and interpretability modeling for different types of learners, this study classifies learners into levels based on their value-added scores within each course separately. According to quartiles, learners are divided into four levels:

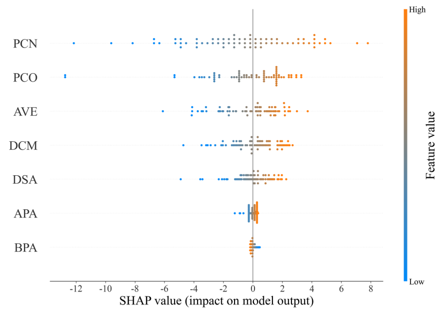
Category A: Value-added scores in the top 25%;
Category B: Value-added scores between 25% and 50%;
Category C: Value-added scores between 50% and 75%;
Category D: Value-added scores in the bottom 25%.

Cohort-Level interpretable analysis. Based on the classification results, the SHAP global feature importance analysis for the four types of learners is shown in Figure 4

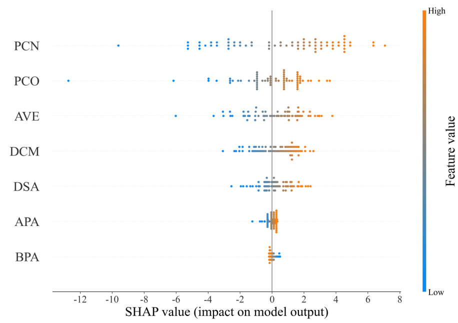
(Course_one) and Figure 5 (Course_two). Features are ranked by importance, with each point's position representing its impact on the classification result (i.e., SHAP value), and color depth reflecting the magnitude of the feature value.



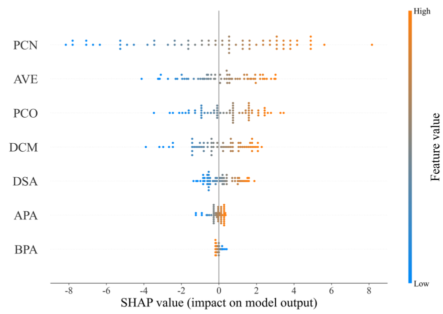
(a) Course_one_Category A



(b) Course_one_Category B



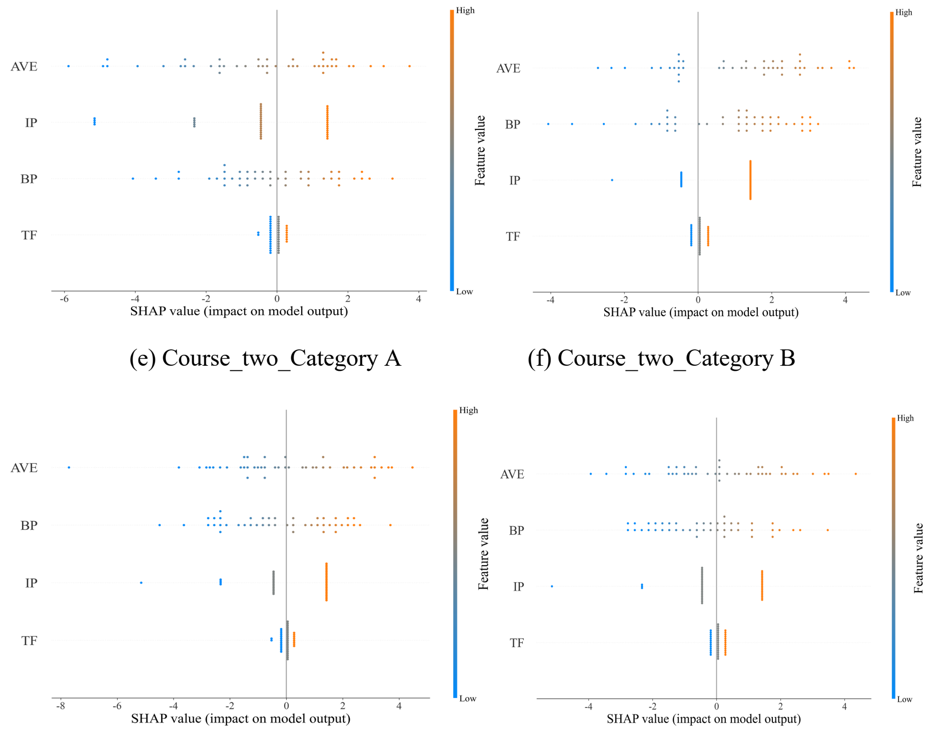
(c) Course_one_Category C



(d) Course_one_Category D

Fig. 4. The summarized impact of all features on students in each category on Course_one

Category A students show the greatest contribution from PCN and PCO to their value-added, with strong core foundations driving their academic growth beyond expectations. Advanced resources, such as operating system kernel development, should be provided to further amplify their value-added advantages. Category B students show prominent contributions from PCN, but DCM and DSA cluster near zero, indicating that algorithmic foundations fail to provide effective support for their value-added. Efforts should focus on strengthening these foundational areas to enhance their value-added potential in subsequent courses. Category C students exhibit negative SHAP values for PCO, indicating that weak foundations in this prerequisite course drag down their value-added. Prerequisite diagnostics and reduced project difficulty are needed to mitigate the negative factors constraining their value-added. Category D students show all features left-skewed with negative contributions, reflecting that systemic knowledge deficiencies lead to the lowest value-added outcomes. Remedial instruction and tiered academic support are required to fundamentally improve their learning foundations.



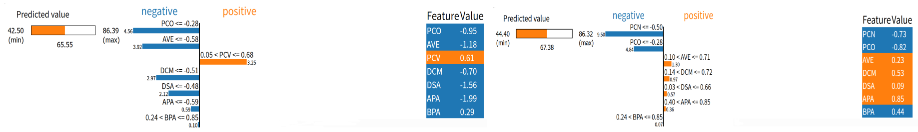
(e) Course_two_Category A (f) Course_two_Category B

(g) Course_two_Category C (h) Course_two_Category D

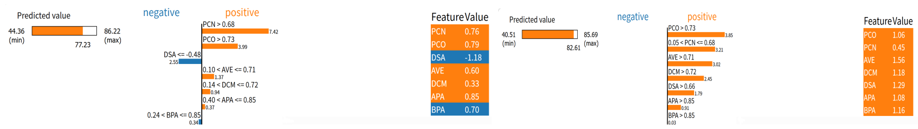
Fig. 5. The summarized impact of all features on students in each category on Course_two

Students in Category A show high positive SHAP values for IP and BP, indicating that strong foundational knowledge drives their high value-added growth. They should be provided with advanced in-depth modules to further tap their academic potential. Category B students gain moderate support from Average Prerequisite, but other courses contribute little to their value-added. Consolidated foundational training is needed to enhance their value-added potential. Students in Category C are negatively impacted by weak prerequisite knowledge across all courses, which drags down their value-added. Prerequisite diagnostics and reduced learning difficulty are required. Students in Category D receive no positive contributions from any courses, reflecting systemic deficiencies that result in the lowest value-added outcomes. Stratified track-based teaching is needed to rebuild foundational competence.

Individual student's interpretable analysis. This study adopts a paired case analysis method for both courses. On Course_one, controlling for actual grades, typical cases with identical grades but significant value-added differences are selected from each learner category. LIME local interpretation is used to compare prediction logic and diagnose value-added sources, as shown in Figure 6. The same analysis is applied to Course_two, as presented in Figure 7.



(a) Course_one_Category A (b) Course_one_Category B

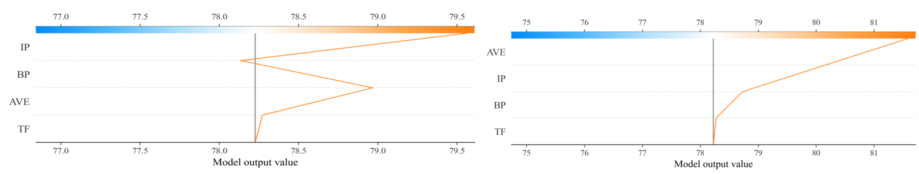


(c) Course_one_Category C (d) Course_one_Category D

Fig. 6. LIME Local Interpretation Comparison for Four Learner Categories on Course_one

The student in Category A shows positive contributions only from PCN. The model gives a low prediction based on weak foundations, and the student achieves a high value-added. The student in Category B shows negative contributions from PCN and PCO balanced by positive compensation from DCM and DSA, resulting in modest value-added. The student in Category C shows high impacts from PCN and PCO, pushing the prediction upward, but actual performance is lower, leading to negative value-added. This student may have relied on rote memorization for exams. The student in Category D shows positive contributions from all prerequisite courses, yet actual performance is low, resulting in deeply negative value-added. This student may reflect test-taking skills rather than genuine competence.

(e) Course_two_Category A (f) Course_two_Category B



(g) Course_two_Category C (h) Course_two_Category D

Fig. 7. LIME Local Interpretation Comparison for Four Learner Categories on Course_two

The student in Category A has only TF as a slight positive contributor, while the remaining features contribute negatively, lowering the model's prediction and resulting in high added value. This suggests that the student possesses a certain ability to transfer knowledge. For the student in Category B, IP has the most significant negative impact, outweighing the positive effect of BP and leading to high added value. It is recommended that this student strengthen their foundational knowledge in prerequisite courses to improve. The student in Category C shows that IP is the strongest positive contributor, with only BP exerting a minor negative pull. It is suggested that this student target and address specific knowledge gaps. For the student in Category D, all features contribute positively, consistently driving up the model's prediction. However, their actual academic performance is low, resulting in negative added value. This student may be facing issues related to test-taking strategies or exam anxiety.

Across both courses, students with traditionally high performance yet negative value-added are identified, alongside students with weak prior foundations who achieved unexpectedly high value-added gains. This demonstrates that value-added evaluation focuses on learners' progress rather than absolute achievement, thereby supporting fairer and more development-oriented assessment. Furthermore, SHAP global analysis and LIME individual analysis form a cross-level mutual validation mechanism, where cohort-level patterns are consistently corroborated at the individual level. This bidirectional validation constructs a value-added evaluation interpretation system that possesses both internal consistency and external transferability.

5 Conclusion

This study proposes the EVA-LA framework to achieve explainable value-added learning analytics, addressing issues such as imprecise value-added quantification and difficulty in interpreting evaluation. We construct a structured learner model, develop the Elastic Net-based value-added evaluation to separate prior foundations from net learning gains, and build a Bi-Lens explainer to achieve multi-level transparency of evaluation results. Experimental results on two real-world courses validate the framework's effectiveness in both robust growth measurement and interpretability, demonstrating its ability to identify key growth drivers across learner categories and to provide individualized diagnostic insights. This method realizes scientific assessment and personalized explanation of students' academic growth, effectively supporting the intelligence and fairness of educational decision-making. Future work will explore the integrated analysis of multimodal educational data to promote the development of educational evaluation systems toward greater intelligence and transparency.

Acknowledgments. This paper is partly supported by the Project of Liaoning Province Educational Science under the "14th FiveYear Plan" (JG25DB369, JG25DB364), the Reformation Research Project of Graduate Education and Teaching in Liaoning Province (LNYJG2023348).

References

1. Paulsen, L., Lindsay, E.: Learning analytics dashboards are increasingly becoming about learning and not just analytics—A systematic review. Education and Information Technologies 29(1), 1–32 (2024)
2. Matcha, W., Gašević, D., Pardo, A.: A systematic review of empirical studies on learning analytics dashboards: Trends and challenges. Journal of Learning Analytics 10(1), 1–28 (2023)
3. Ifenthaler, D., Yau, J.Y.K.: Utilising learning analytics to support study success in higher education: A systematic review. Educational Technology Research and Development 68(4), 1961–1990 (2020)
4. Umer, R., Susnjak, T., Mathrani, A., Suriadi, S.: Current stance on predictive analytics in higher education: Opportunities, challenges and future directions. Interactive Learning Environments 31(6), 3503–3528 (2023)
5. Steedle, J.T., Bradley, R., Zimmerman, S.: Literature review on the value-added measurement in higher education. OECD Publishing, Paris (2013)
6. Raudenbush, S.W., Bryk, A.S.: Hierarchical linear models: Applications and data analysis methods. 2nd edn. Sage Publications, Thousand Oaks (2002)
7. Li, K.C., Wong, B.T.M., Liu, M.: A survey on predicting at-risk students through learning analytics. International Journal of Innovation and Learning 36(5), 1–15 (2024)
8. Barria-Pineda, J., Guerra-Hollstein, J., Brusilovsky, P.: A fine-grained open learner model for an introductory programming course. In: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization (UMAP 2018), pp. 53–61. ACM, New York (2018)
9. Almalawi, A., Soh, B., Li, A., Samra, H.: Predictive models for educational purposes: A systematic literature review. Big Data and Cognitive Computing 8(12), 187 (2024)
10. Cunha, J.M., Miller, B.A., Murray, D., White, K.: Can "value added" methods improve the measurement of college performance? In: Context for Success, pp. 1–15. HCM Strategists LLC, Washington (2013)
11. Mangaroska, K., Vesin, B., Kostakos, V., Brusilovsky, P., Giannakos, M.N.: Architecting analytics across multiple e-learning systems to enhance learning design. IEEE Transactions on Learning Technologies 14(2), 173–188 (2021)
12. Bull, S., Kay, J.: SMILI: A framework for interfaces to learning data in open learner models, learning analytics and related fields. International Journal of Artificial Intelligence in Education 26(1), 293–331 (2016)
13. Tao, L., Cukurova, M., Song, Y.: Learning analytics in immersive virtual learning environments: A systematic literature review. Smart Learning Environments 12(1), 43 (2025)
14. Bertolini, R., Finch, S.J., Nehm, R.H.: Enhancing data pipelines for forecasting student performance: Integrating feature selection with cross-validation. International Journal of Educational Technology in Higher Education 18(1), 1–20 (2021)
15. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems 30 (NIPS 2017), vol. 30, pp. 4765–4774. Curran Associates, Red Hook (2017)
16. Nazat, S., Abdallah, M.: XAI-based feature ensemble for enhanced anomaly detection in autonomous driving systems. arXiv preprint arXiv:2410.15405 (2024)
17. Pan, Z., Biegley, L., Taylor, A., Zheng, H.: A systematic review of learning analytics–incorporated instructional interventions on learning management systems. Journal of Learning Analytics 11(2), 1–21 (2024)

18. Öz, S., Eryılmaz, M.: Estimating teachers' value-added effects using a generalized persistence model. International Journal of Assessment Tools in Education 11(3), 415–433 (2024)
19. Basileo, L.D., Toth, M.D.: Evidence for the Marzano Focused Model: Reforming teacher evaluation. Marzano Evaluation Center, Palm Beach Gardens (2020)
20. Amrein-Beardsley, A., Geiger, T.: Methodological concerns about the Education Value-Added Assessment System (EVAAS): Validity, reliability, and bias. SAGE Open 10(2), 1–15 (2020)
21. Syed Mustapha, S.M.F.D.: Machine learning models for academic performance prediction: Interpretability and application in educational decision-making. Frontiers in Education 10, 1632315 (2025)
22. Larasati, R.: Explainable AI approach to compare predictive drivers of student success for disabled and non-disabled learners. In: Proceedings of the International Conference on Artificial Intelligence and Teacher Education (ICAITE 2025), pp. 1–8. IEEE, Kyoto (2025)
23. Ouared, A., May, M., Piau-Toffolon, C., Dugué, N.: XAI-Profile: Explainable AI for transparent learner profiling in digital education. HAL Preprint, hal-05490103, pp. 1–15. HAL Open Archive, Lyon (2026)
24. Maathuis, H., Glazenborg, J., Grol, M., Quille, K.: Educators' evaluation of explanation types in XAI for higher education dropout. In: Proceedings of the 2026 ACM Conference on Explainable Artificial Intelligence in Education (XAIEdu 2026), pp. 112–126. ACM, New York (2026)
25. Hassan, A., et al.: A perspective on explainable artificial intelligence methods: SHAP and LIME. Advanced Intelligent Systems 6(12), 2300304 (2024)
26. Ozen, Z., Pereira, N., Bright, S.: Exploring critical predictors of math and science achievement for high achieving students in TIMSS data: Application of elastic-net logistic regression. Journal of Advanced Academics 36(4), 695–713 (2025)
27. Molnar, C.: Interpreting Machine Learning Models With SHAP. 1st edn. (2023)
28. Khatun, M.R., Mim, M.A., Tasin, M.M., Hossain, M.M.: A hybrid framework of statistical, machine learning, and explainable AI methods for school dropout prediction. PLoS One 20(9), e0331917 (2025)