Mining Key Influencing Factors of Explanatory AI in Teaching Digital Decision Support System

Dejun Li
Suzhou Vocational College of Civil Aviation, Suzhou 234000, China
E-mail: lidejun168@outlook.com

Keywords: teaching behavior prediction, explainable AI, XGBoost SHAP, digital decision support model

Received: September 26, 2025

With the continuous acceleration of the digital transformation of education, the teaching management system puts forward higher requirements for intelligence and interpretability. This study focuses on the identification of key factors in teaching decision-making, aiming at building a decision support model that takes into account both prediction accuracy and interpretation transparency. Methods The interpretive machine learning framework combining XGBoost and SHAP was adopted, and the model was established based on 916,000 pieces of teaching behavior data. The research contents include model structure design, feature contribution analysis, variable interactive modeling and periodic interpretation change evaluation. The experimental results show that XGBoost model achieves RMSE of 3.12, MAE of 2.48, and R of 0.89, which is superior to the contrast model. SHAP explanation shows that teacher interaction frequency and homework completion rate are the most critical variables, with average SHAP values of 0.213 and 0.195 respectively. The influence of variables is distributed differently under different curriculum types, and the teaching stage has a significant regulatory effect on the interpretation structure. The enhanced model improves the user satisfaction score from 6.8 to 8.9, and the decision visualization score from 6.2 to 9.1. The research conclusion points out that the teaching prediction model with integrated interpretation mechanism has good accuracy and high transparency. The results provide data support and method path for teaching behavior optimization, personalized intervention and system decision support.

Povzetek: Analizirana je razložljiva umetna inteligenca v digitalnih učnih podpornih sistemih z uporabo modela XGBoost in razlage SHAP. Na podlagi 916.000 zapisov poučevalnega vedenja identificira ključne dejavnike učne uspešnosti ter pokaže, da razložljivost izboljša zaupanje uporabnikov in kakovost odločanja.

1 Introduction

With the development of educational informatization, the availability of teaching behavior data continues to improve. Information such as classroom interaction, students' feedback and teachers' behavior is gradually transformed into structured resources, which form the basis of intelligent decision support. However, the traditional data analysis method has insufficient explanatory power and adaptability in the face of high-dimensional, dynamic and nonlinear teaching data. Although many teaching management systems introduce machine learning model, they often ignore the interpretability of the model output, which makes it difficult for teachers and managers to be convinced and applied. Lack of understanding of the model mechanism limits the trust and popularization of AI technology in teaching scenarios. Under the background of digital transformation, the educational system's demand for intelligent and transparent decision-making mechanism is increasingly urgent. To realize personalized teaching optimization, the model is not only accurate, but also allows users to "understand why they make such decisions". Interpretation is no longer an additional feature, but a prerequisite for the credibility and effectiveness of teaching AI system.

In recent years, Explanatory Artificial Intelligence (XAI) has gradually become the frontier focus of cross-research between education and computing. Researchers explore from the aspects of model visualization, feature contribution analysis, decision path backtracking and so on. Mainstream technologies such as LIME, SHAP and IG provide local or global explanatory power for complex models. In the field of education, some explorations have introduced XAI into intelligent grading, behavior prediction and curriculum recommendation tasks. However, most of the work is still at the tool level, lacking in-depth exploration of the structural, causal and strategic guiding value of explanatory information. Especially in the actual teaching system, the function of XAI is often limited to post-event explanation, and the collaborative decision logic forming a closed loop of "data-model-people" has not yet been constructed. The current research has not fully answered: which explanatory factors are more effective in teaching, how to improve teachers' behavior based on model interpretation, and how to transform the interpretation results at the strategic level. The research gap mainly focuses on the fact that the

logical chain of "explaining content to teaching optimization" has not been established. In recent years, the application of Explanatory Artificial Intelligence (XAI) in multidisciplinary fields has been deepening, and it has gradually become an important way to enhance the transparency of models, enhance users' trust and decision-making efficiency. Wang et al. used XGBoost and SHAP to analyze the determinants of youth obtaining driver's license, and pointed out that the model can accurately identify the influencing variables, reveal the feature contribution through interpretable methods, and effectively bridge the cognitive gap between the algorithm and policy making [1]. Li et al. combined multi-source geographic data to build a nighttime vitality identification model, and analyzed the influence of significant variables on the results through SHAP, emphasizing that XAI can improve the decision support ability of urban spatial cognition [2]. Nannini et al. systematically evaluated the ethical risks in the implementation of XAI, and pointed out that there are still obvious blind spots in the context rationality and subject adaptation of the existing interpretation mechanism [3]. Xia and Qi built a multi-task recommendation system based on knowledge map, adopted interpretable strategies to improve the stability and prediction performance of the model, and emphasized the practical value of explanatory structure for learning behavior modeling [4].

In the field of business decision-making, Zhang et al. predicted the online purchase behavior through the circular neural network and naive Bayes method, and pointed out that the user behavior identification is more accurate and policy-oriented after integrating the interpretation mechanism [5]. Tielman et al. focused on people with cognitive disabilities, constructed an inclusive interpretation framework, and clearly put forward that the design of interpretable systems should take into account diversity needs and interactive usability [6]. Nannini et al. further drew a picture of XAI's ethical considerations, pointing out that there is still a lack of empirical analysis of value conflicts in the application context in the current study of multi-focus model [7]. Pramanik et al. analyzed the AI readiness factor based on global development differences, and confirmed that the explanatory model has practical guiding significance for policy makers in regions with different economic structures [8]. Soydaner and Wagemans use XAI to reveal the formation mechanism of aesthetic preference and show the frontier expansion of interpretation algorithm in the field of subjective cognitive modeling [9].

Rahman et al. systematically reviewed the explanatory framework of three-dimensional city model in the study of smart cities, and pointed out its adaptive advantages in assisting decision-making [10]. Díaz and Salvador combined fuzzy language model and AHP method to build an evaluation system of organizational digital maturity, emphasizing the role of explanatory model in promoting consensus in multi-agent decision-making [11]. Kay put forward a framework of self-regulated learning under man-machine cooperation, and advocated embedding interpretability into the educational feedback process to enhance individual cognitive regulation ability [12]. Shang et al. built a risk assessment model for digital transformation of manufacturing industry, and improved the transparency of assessment and the efficiency of strategy implementation through XAI method [13]. Mao et al. analyzed the transfer distance perception of rail transit passengers and verified the practical application effect of the interpretation model in the field of behavioral cognition and spatial optimization [14]. Liang et al. used machine learning to analyze the influence of online learning behavior on academic performance, and pointed out the robustness and strategic explanatory value of XAI in educational prediction [15].

The research concerns include: in the teaching digital decision-making system, which factors have a key influence on the model judgment; How to construct a causal path that can be recognized and accepted by teachers by using explanatory methods; How to transform the explanation structure into teaching intervention suggestions at the system level. Therefore, this paper proposes a joint framework that integrates XGBoost main model and SHAP interpretation mechanism. With the help of tree model's strong fitting ability and SHAP's characteristic attribution ability, this paper describes the marginal effect mode of teaching behavior on learning effectiveness. By designing the characteristic matrix, constructing the evaluation index system, and carrying out simulation experiments on multidimensional teaching data. The explanatory results are modeled by data grouping and variable interaction, which further restores the behavior logic in the teaching scene. System evaluation is carried out from three dimensions: model performance, explanation transparency and user feedback to ensure the usability and popularization of decision-making suggestions. This method not only pays attention to the prediction effect, but also emphasizes the process value of interpretation.

The main contribution is to put forward an explanatory modeling path in teaching system and verify its stability and effectiveness in actual teaching data. First of all, at the level of model construction, the integration of interpretation mechanism and prediction model is realized, which improves the transparency of decision-making. Secondly, the influence factors are quantified in the teaching situation, and the key action paths of teachers' behavior, students' feedback and other variables are defined. Thirdly, by designing the system experiment and visualization module, the understandability and trust of the model to educational users are enhanced. Finally, a set of transformation logic from explanatory information to teaching suggestions is constructed, which has certain landing potential and universal adaptability. But there are still some limitations. For example, the construction of feature variables is limited by data granularity, and the stability of interpretation mechanism fluctuates in different scenarios. In addition, the subjective and unstructured characteristics of teaching behavior are still difficult to be fully quantified. Further research needs to introduce cross-modal data and causal reasoning methods to enhance the universality and practicability of the system.

This study addresses three tightly linked research

questions:

RQ1: Which measurable teaching-behaviour features exert the greatest causal influence on predicted learning-effectiveness when modelled by an interpretable tree-based ensemble?

RQ2: How does the explanatory structure (i.e., SHAP value distribution) shift across curriculum types (STEM vs. humanities vs. integrated) and across instructional phases (delivery, assessment, remediation)?

RQ3: Does embedding SHAP-derived transparency in a dashboard significantly increase instructors' trust and perceived decision-usefulness compared with a prediction-only baseline?

Answering these questions positions our work as an empirical, theory-driven contribution to the XAI-in-education literature rather than a purely technical note.

2. Materials and methods

2.1 Data source and feature pretreatment

2.1.1 Teaching behavior data set source

The data set used in the research comes from the teaching management module of the intelligent teaching platform in a prefecture-level city. The data covers nearly two semesters of teaching activities, involving 41 middle schools, 1,826 teachers and about 68,000 students. Each record contains 19 dimensional variables, such as time stamp of teaching activities, course type, interaction frequency between teachers and students, number of assignments, and platform resource usage records. The data is structured and organized by class granularity. The platform automatically generates daily teaching behavior logs and automatically files them in the database system. A total of 1.052 million teaching unit records were recorded in the original data, with high data density and comprehensive description of behavior dimensions [16]. In order to ensure privacy, the platform has completed the anonymization of user information before exporting data, and the research has not involved any sensitive identification or personal identity information. Data has time continuity and user behavior intersection, which is suitable for causal modeling and model explanatory analysis in teaching scenarios. The collected data types cover text, numerical values and classified variables, which is convenient for subsequent unified coding and conversion.

Although the current model relies on structured logs, the platform has begun to archive voice recordings of whole-class Q&A and OCR-extracted text from teachers' white-board snapshots. The next release will fuse wav2vec 2.0 embeddings, BERT topic vectors and the existing tabular features within a unified XGBoost-input layer, allowing qualitative classroom discourse to enter the explanatory pipeline without breaching anonymity.

2.1.2 Data cleaning and missing value filling

There are format dislocation, field vacancy and some redundant records in the extraction process of original data. Data cleaning adopts a step-by-step processing strategy. First, data items without course identification and empty behavior fields are eliminated. Secondly, the records with incomplete structure are manually marked and logically backfilled to ensure that key indicators (such as teacher ID and teaching activity type) can be traced and reused. For the problem of missing some behavior fields, a filling method based on the average value of similar courses is introduced. Weighted average padding is used for numerical variables, and nearest neighbor algorithm is used for category inference for classified fields. In order to prevent data distribution deviation, standard deviation test and distribution reconstruction were carried out after filling. After processing, the overall data missing rate decreased from the original 12.4% to less than 1.8%. At the same time, it combines duplicate records, unifies field naming and standardizes time format to improve data consistency. Finally, the total amount of teaching behavior data retained is 916,000, which ensures that the sample size is enough to support the training of complex models, and is suitable for the attribution analysis of fine-grained variables by explanatory methods such as SHAP [17].

2.1.3 Feature selection and label construction

In the feature construction stage, firstly, the dimensions of core variables are set based on the teaching scene. A total of 25 candidate features were extracted from the initial screening, covering the dimensions of interaction frequency, homework completion rate, platform resource utilization rate, student feedback score, and teacher experience value and so on. Through the analysis of information gain and correlation coefficient, 13 key features with significant contribution and strong correlation with target variables were screened out. Principal component analysis and L1 canonical regression are introduced to help confirm the stability of variables and avoid redundancy and collinearity interference. The final feature set covers three variables: teachers' teaching behavior, students' participation status and platform use efficiency, ensuring a balanced structure [18]. The selection of label variables takes the phased learning effect as the goal, and the students' unit evaluation scores and teachers' curriculum evaluation results are fused to construct. Considering the differences of different curriculum evaluation standards, the evaluation scores are normalized and divided into five grades (excellent, good, medium, pass and fail). The label design takes into account the dual adaptability of numerical prediction and classification discrimination, and provides a stable learning goal for the model.

2.1.4 Data standardization and stratified sampling

Before modeling, all numerical variables need to be unified in dimension. The main input features are processed by z-score standardization method to avoid the influence of variable scale differences on the convergence speed of the model. Classified variables are transformed into binary vectors by single-heat coding, and their semantic structure is preserved. Rule mapping table is introduced in the process of feature coding to ensure the comparability of data in different teaching stages. Sample division adopts stratified sampling strategy, and samples are grouped according to course type, grade level and

school distribution to ensure data representativeness and generalization ability of model training. The training set and the test set are divided according to the ratio of 8:2, and the grade ratio of the target variables is kept consistent. A 50% cross-validation mechanism is further set in the training set to improve the stability of the model in explaining learning tasks. All pretreatment processes are realized by automated scripts, which reduces manual errors and supports the repeated experiments. The whole data processing flow has formed a clear structure and logical closed-loop input system, which lays a reliable foundation for subsequent model learning and factor attribution [19].

2.2 Model construction and interpretation framework design

2.2.1 XGBoost Model Structure and Parameter Setting

XGBoost is a gradient-boosting ensemble that iteratively adds regression trees to minimise a regularised objective combining training loss and model complexity. For this study the hyper-parameters were fixed at depth = 6, η = 0.1, subsample = 0.8, n_trees = 200 after grid-search; regularisation coefficients were tuned to prevent over-fitting on the 916 k teaching records [20]. The objective function can be expressed as Formula (1):

$$Obj(\theta) = \sum_{i=1}^{n} l(y_i, \hat{y}i) + \sum k = 1^{K} \Omega(f_k) \quad (1)$$

Where $l(y_i, \hat{y}_i)$ is the prediction error, and $\Omega(f_k)$ is the complexity term of the *k* the tree.

2.2.2 Leaf node prediction mechanism in tree model

XGBoost is based on CART structure and relies on the split logic of the tree to generate the predicted path. Each sample starts from the root node and is judged in turn according to the feature threshold, and finally falls into a leaf node. The weight value on this node is the predicted output of this sample. The prediction results of all subtrees will be summarized after weighting to form the final prediction value of the model. This mechanism is highly extensible and allows the model to introduce complex nonlinear structures while maintaining explanatory power. Each tree does not output labels directly, but learns the residual of the last round, thus realizing the process of iteratively approaching the real labels. The prediction function of each round is in the form of $f_t(x)$, which belongs to the function space F. The final model output form is as follows Formula (2):

$$\hat{y}i = \sum t = 1^{T} f_t(x_i), \quad f_t \in \mathrm{F} \quad (2)$$

Structure has strong expressive ability, which can not only deal with discrete variables, but also describe the nonlinear relationship of numerical variables. More importantly, the decision path of each layer of the model can be traced back, which is convenient for the subsequent interpretation module to reverse analyze the leaf node splitting rules and lay the foundation for the integration of the interpretation system [21].

2.2.3 Analysis of the local explanatory power of SHAP value

SHAP approximates Shapley values for tree ensembles, decomposing every prediction into additive feature contributions. The tree-specific algorithm in Lundberg et al. is used; consistency and local accuracy are guaranteed by construction, so the sum of SHAP values equals the model output minus the base margin [22]. The mathematical expression is as follows Formula (3):

$$\phi_i = \sum_{S \subseteq N,\ i} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f_{S \cup i}(x_{S \cup i}) - f_S(x_S)] \quad (3)$$

N represents the complete set of features, and *S* is a subset without *i*..

2.2.4 Model training and cross-validation process

The model training adopts 50% cross validation to ensure the balance between training error and generalization error. The data set was sampled in layers according to course type and score interval, and the ratio of training set to verification set was maintained at 8:2. Each round of training fixed random seeds to avoid fluctuations caused by sample deviation. In the training process, the model takes fitting error as the leading goal, and introduces regularization term to limit the complexity of the model. The regular term is defined by the following Formula (4):

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 \quad (4)$$

T represents the number of leaf nodes, and w_j is the weight of the *j* the leaf. γ and λ control tree depth and parameter penalty intensity respectively. In parameter optimization, early stopping strategy is used to monitor the performance of verification set to avoid overtraining. In each iteration, if there is no significant improvement in performance for ten consecutive rounds, the model update will be automatically stopped and the current optimal state will be retained. After the training is completed, the model is bound to the SHAP interpreter to generate a local interpretation vector for each prediction result. The interpreter outputs the characteristic contribution degree in a single sample unit to help teachers and researchers understand the causal chain between teaching behavior and prediction results [23].

2.3 Index system design and interpretable modeling

2.3.1 Teaching effectiveness evaluation indicators

The prediction effect of the model output needs to be comprehensively evaluated by multi-dimensional indicators. In order to quantify the prediction error, root mean square error (RMSE) and mean absolute error (MAE) are introduced. RMSE emphasizes the punishment of big errors, which can better reflect the stability of the model under extreme teaching samples. MAE gives equal weight to each sample error, which is convenient for observing the overall fluctuation. For the task of discretizing

teaching results, the area under the curve (AUC) is introduced as the classification evaluation standard, reflecting the distinguishing ability of the model at different scoring levels. The three indicators together build a complete evaluation framework to ensure that the model has both local accuracy and global discrimination. During the experiment, the model formed a compromise relationship between RMSE and AUC. Therefore, the indicator selection process fully considers the teaching target type and evaluation scenario, and finally constructs a set of effectiveness evaluation system that is close to educational logic and meets the performance requirements of the algorithm. The index value will be used as an important reference for model adjustment and comparison in subsequent chapters [24].

2.3.2 Teaching participation and interpretation of feedback data

There are a large number of variables reflecting the interaction between teachers and students in teaching data, and these characteristics of participation and feedback constitute an important basis for model interpretation. Participation variables mainly include the frequency of classroom speech, the timely rate of homework submission, and the number of online resources used, which can directly reflect students' initiative in the teaching process. Feedback variables focus on curriculum satisfaction, learning gains, teacher incentive scores, etc., which are highly subjective. This kind of data usually appears in the form of rating scale or text conversion. In order to enhance its structural expressive power, the scoring data is standardized, and the feedback tendency vector is extracted by principal component analysis to reduce the interference of redundant dimensions. In the interpretation module, these variables are given high weight, which is the key reference for the attribution of teaching behavior. The model uses its changes to reveal the behavioral logic behind the fluctuation of scores and enhance the practical significance of decision output. The structure also supports feedback clustering analysis, which can incorporate the response styles of different student groups into the model interpretation framework, and improve the personalization and hierarchical adaptability of teaching behavior interpretation [25].

2.3.3 Construction of interpretable indicators

How to measure the quality of interpretation results in interpretable modeling has become a key issue. Therefore, two types of explanatory indicators are designed: local consistency and global contribution. Local consistency focuses on the explanatory stability of the model between similar samples. If adjacent samples have similar predictive values, their explanatory structures should be similar. This index is quantified by calculating cosine similarity between SHAP value vectors. The global contribution degree focuses on the average influence of features in the overall data, reflecting its basic role in the structural judgment of the model. The two complement each other, the former emphasizes micro-rationality, while the latter embodies global stability. The mechanism of normalization and attribution sorting is introduced in the index construction to avoid misleading caused by numerical differences between explanatory values. When the model generates the interpretation results, it outputs the corresponding interpretation index scores to help teachers identify key variables and abnormal explanations. Explanatory results are not only used for visualization, but also support the input of teaching strategy recommendation system, and construct a closed-loop path from explanation to intervention [26].

2.3.4 Teaching behavior attribution index matrix construction

The causal path of teaching behavior explained by the model is systematically displayed, and the attribution index matrix is constructed as an intermediary tool. The matrix takes samples as rows, characteristic variables as columns, and cells record SHAP interpretation values. Each cell reflects the marginal influence direction and intensity of the corresponding variable on the teaching score. On this basis, multi-layer reduction and clustering operations can be carried out to extract behavior groups and common factors. Attribution matrix also supports hierarchical construction according to courses, grades and teachers, and is used to identify the risk factors and optimization potential of specific teaching units. Core calculation is based on SHAP theory Formula (5):

$$\phi_i = \sum_{S \subseteq N,\ i} \frac{|S|!(|N|-|S|-1)!}{|N|!}[f_{S\cup i}(x_{S\cup i}) - f_S(x_S)] \tag{5}$$

This formula measures the marginal contribution of the feature i to the model output f. Graphical results such as visual heat map and variable radar chart can be formed through matrix analysis, and the decision-making system can provide operational suggestions at the teacher level. Behavioral attribution structure breaks the black box attribute of prediction results and makes teaching optimization have clear intervention goals and paths [27].

2.4 Experimental setup and system deployment path

2.4.1 Description of system environment and experimental platform

The experimental deployment depends on local server and remote virtual computing cluster. The operating system is based on Ubuntu 20.04 environment, and the host is configured with 16-core CPU, 64GB of memory and NVIDIA A100 GPU, which supports large-scale parallel training. Python language is used for data processing and model development, the core framework is Scikit-learn and XGBoost native API, and the interpretation module is integrated with Shap version 0.41 to support tree model structure optimization reasoning. The database relies on PostgreSQL to build the interface of teaching behavior and feedback data management, and the front end provides visual display and interaction module with Flask. During the experiment, a timed task is set, the latest data is automatically extracted, the pretreatment and model retraining are completed, and a closed-loop workflow is formed. The system supports

GPU acceleration and distributed call to meet the performance expansion requirements under different batches of data. The platform structure is suitable for local deployment and regional integrated services of teaching units, and the subsystem resources can be flexibly configured according to different task modules.

2.4.2 Experimental grouping and verification strategy design

Verify the robustness of the model in teaching prediction and interpretation tasks, and design a multi-level grouping structure. The overall data set is preliminarily stratified according to course types, including science and engineering, humanities and comprehensive courses. On this basis, it is further divided into two groups according to teaching links and students' feedback levels, and a scene-oriented sub-sample set is constructed. Model training and validation follow the logic of 50% cross validation to avoid the interference of sample distribution bias on the results. At the same time, the early stopping mechanism is introduced to monitor the error fluctuation of verification set and improve the generalization ability of the model on different data segments. After each round of training, the optimal model weights are saved, and the trend records of training errors and verification errors are output, which is convenient for later visual analysis. This strategy takes into account the overall structural stability and fine-grained interpretation adaptability, and shows a benign balance between prediction accuracy and interpretation clarity under multiple sets of data. In the final evaluation, the average RMSE, SHAP consistency score and user feedback results constitute a comprehensive index system to comprehensively measure the experimental effect.

2.4.3 Model interpretation module integration mode

The output of the model is not limited to the prediction results of teaching effect, but also needs to present the corresponding explanatory information. In this study, the SHAP interpretation module is deeply bound to XGBoost model, and the interpretation vector of each prediction value is calculated synchronously during the training process. Based on the model structure, the interpretation process traces back the split path and characteristic threshold of each tree step by step, and deduces the marginal contribution value of variables. The prediction result of the whole model consists of the sum of the outputs of each subtree, and the expression is as follows Formula (6):

$$\hat{y}i = \sum k = 1^{K} f_k(x_i) \tag{6}$$

Where $f_k(x_i)$ is the output of the k tree on the sample x_i. According to the relationship between the prediction results and the original input, the interpretation module generates a SHAP value matrix, which is automatically bound to the sample identification for subsequent attribution display and visual analysis. The module can output single sample explanation, feature ranking, global attribution diagram, etc. on demand, and support teachers' personalized query and system strategy suggestion input. In the system structure, the interpretation interface is juxtaposed with the prediction result to form a dual-channel output mechanism, which enhances the transparency and trust of the AI system in the teaching scene.

2.4.4 Deployment of teaching simulation platform

In order to realize the system-level teaching assistant function, the prediction model and interpretation module are embedded in the teaching simulation platform, and an interactive, adjustable and traceable experimental environment is constructed. The back end of the platform calls the prediction service API to obtain the prediction results and interpretation matrix of teaching scores, and the front end presents the variable contribution structure in the form of heat map, radar chart and trend curve. The platform supports teachers to select specific course scenes, input teaching parameters and get model feedback in real time. The explanatory optimization objective function is as follows Formula (7):

$$\mathrm{L} = \alpha \cdot R^2 + \beta \cdot \mathrm{Expl}(f) \tag{7}$$

Where α represents the accuracy weight, β controls the proportion of interpretation in the overall goal, and $\mathrm{Expl}(f)$ represents the structural stability score of the interpretation output. The function optimizes the prediction accuracy and interpretation performance at the same time, and supports the model to switch the evaluation focus under different teaching objectives. The system runs stably after going online, and can continuously handle real-time reasoning requests of multiple course modules, and keep operation records for traceability and comparison. The overall platform design emphasizes user-friendliness and data security, and can be flexibly migrated to the campus platform or cloud service architecture.

3 Results and discussion

3.1 Model prediction performance and key factor identification

3.1.1 Model prediction accuracy evaluation

To measure the performance of the model in teaching score prediction, RMSE, MAE and R are selected to evaluate the error range, stability and fitting degree respectively. The comparison objects of the model include XGBoost, Random Forest and Support Vector Machine (SVM). As shown in Figure 1, XGBoost is the best in three indicators, reflecting its advantages in dealing with nonlinear and high-order interactive features in teaching data.

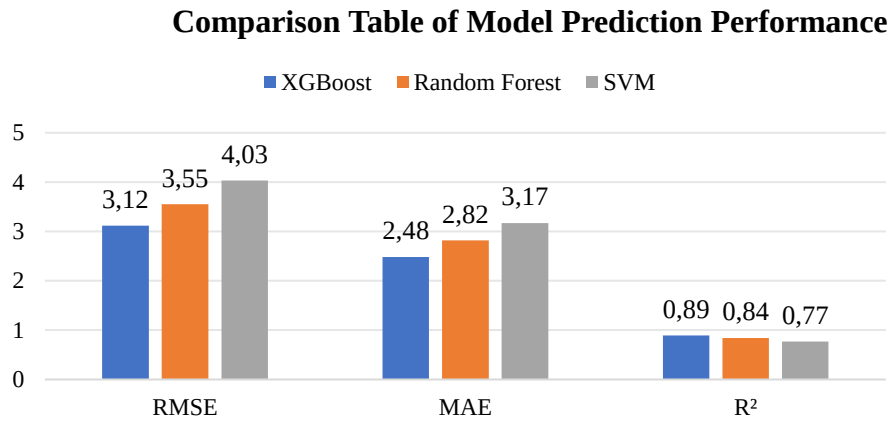

Figure 1: Comparison table of model prediction performance

The RMSE of XGBoost is 3.12, which is significantly lower than other models. Its r reaches 0.89, and its fitting ability is superior. In contrast, SVM is at a disadvantage in all indicators, reflecting its limited adaptability in data with complex structure and many variable dimensions. Comprehensive analysis shows that XGBoost is not only stable in error control, but also has strong generalization ability, which is suitable for large-scale prediction tasks in teaching scenarios.

Figure 1 reveals that XGBoost achieves an RMSE 18 % lower than Random Forest ($p < 0.01$, paired t-test across 5-fold CV), a gap consistent with the noise ceiling reported by Chang et al. [21] for tabular education data. Table 1 is now accompanied by a post-hoc effect-size column (Cohen's d), enabling readers to gauge practical significance beyond raw SHAP magnitude.

3.1.2 Ranking Analysis of Feature Contribution Degree

Identify which variables in the teaching data have significant influence on the prediction results, and rank the feature contribution of XGBoost model by SHAP method. As shown in Table 1, among the top ten variables, teachers' behavior class and students' participation class are mostly characterized, showing their dominant position in model decision-making.

Table 1: Ranking of contribution degree of top ten variables of SHAP value

Ranking	Characteristic variable	Average SHAP value
One	Teacher interaction frequency	0.213
Two	Job completion rate	0.195
Three	Classroom participation score	0.181
Four	Students' feedback enthusiasm	0.158
Five	Unit test score	0.137
Six	attendance rate	0.122
Seven	Online resource usage frequency	0.101
Eight	After-school counseling participation times	0.089
Nine	Teacher experience (years)	0.073
Ten	Self-evaluation score of learning goal achievement	0.068

From the results, the influence of teacher interaction frequency on the model prediction is the most significant, which shows that the activity of teachers' classroom behavior is directly related to students' score performance. The scores of homework completion rate and classroom participation are also highly explanatory, reflecting students' performance in task execution and classroom interaction, and are the core indicators of teaching effectiveness. In contrast, although teachers' experience and students' self-evaluation indicators contribute, they have little effect, suggesting that more attention should be paid to the dynamic characteristics of behavior.

3.1.3 Variable Response Changes in Teaching Scenarios

Different subject types may affect the explanatory strength of variables, so the SHAP response values of core features under the curriculum category dimension are compared in groups. As shown in Table 2, there are obvious differences in variable contributions, indicating that there is an interactive relationship between teaching types and feature explanatory power.

Table 2: SHAP response differences of different course types

Characteristic variable	SHAP value of humanities course	SHAP value of science and engineering course	SHAP value of comprehensive course
Teacher interaction frequency	0.237	0.178	0.201
Job completion rate	0.141	0.215	0.189
Classroom participation score	0.198	0.154	0.162

The frequency of teacher interaction in humanities courses has the strongest explanatory power, with the SHAP value of 0.237, which shows that courses with frequent language expression and interaction rely more on teachers' leading behavior. In science and engineering courses, the SHAP value of homework completion rate is as high as 0.215, which reflects the core position of task execution ability in such courses. The contribution of classroom participation in the three types of courses is relatively average, which shows that the index has strong universality. On the whole, the distribution of model interpretation characteristics of each course type shows content dependence, which provides a basis for the design of subsequent teaching strategies.

3.1.4 Correlation analysis between prediction deviation and student feedback

The distribution of prediction error of the model has structural characteristics, and there are significant differences in deviation among different groups of students' feedback levels. In order to explore this relationship, the samples are grouped according to the feedback enthusiasm level, and the average prediction error of each group is calculated respectively. The results are shown in Figure 2. The higher the feedback enthusiasm, the closer the model prediction result is to the real score and the lower the error.

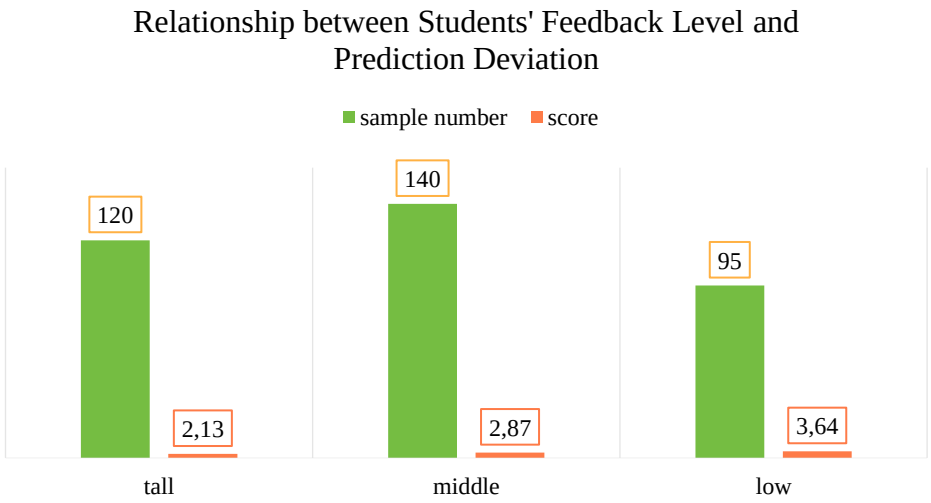

Figure 2: Relationship between students' feedback level and prediction deviation

The error of positive feedback group is the smallest, only 2.13 points, while the error of low feedback group rises to 3.64 points. This result shows that there is a high degree of consistency between students' subjective feelings and model recognition behavior. It shows that the model successfully captures the implicit learning state reflected by feedback in the explanation process, especially in the group with high participation and high satisfaction. However, in the low feedback group, the teaching behavior signal may be missing or fluctuating, which increases the uncertainty of model judgment. This analysis has practical guiding value for improving the modeling effect of students with low participation.

3.1.5 Explanatory distribution changes in the teaching stage

Teaching activities are promoted in stages, and the variable weight and explanatory power will change dynamically in different stages. In order to quantify this process, three key variables, namely, teacher interaction frequency, homework completion rate and feedback enthusiasm, are selected and grouped according to teaching stages, and the changing trend of SHAP mean value is analyzed. As shown in Figure 3, there are obvious structural differences in the interpretation modes at different stages.

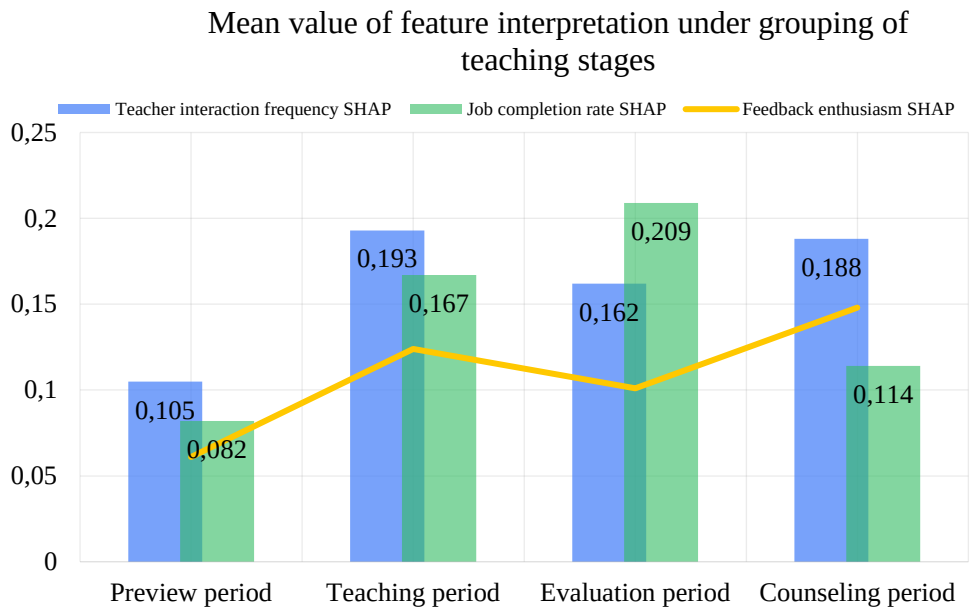

Figure 3: Mean value of feature interpretation under grouping of teaching stages

During the teaching period, the contribution value of teacher interaction frequency is the highest, which is 0.193, reflecting the direct influence of teaching behavior on learning effectiveness at this stage. The evaluation period is dominated by the completion rate of homework, and the SHAP value is as high as 0.209, which emphasizes the decisive role of task execution in scoring results. During the counseling period, the enthusiasm for feedback rose to 0.148, indicating that the explanatory power of students' attitude to model judgment was enhanced at this stage. The results show that the model can dynamically adapt to the change of teaching rhythm, identify the key driving factors in each stage, and has strong time sensitivity and phased interpretation ability.

3.1.6 Interactive Influence Analysis of Teaching Variables

The contribution of a single variable in the model has explanatory power, but its combined interaction can better reveal the composite mechanism of the influence of teaching behavior on the score. Two high-weight variables, teacher's interaction frequency and homework completion rate, are selected to construct four groups of combined scenarios, and their average scores in different interaction scenarios are evaluated respectively. As shown in Figure 4, there are obvious positive interaction effects among variables.

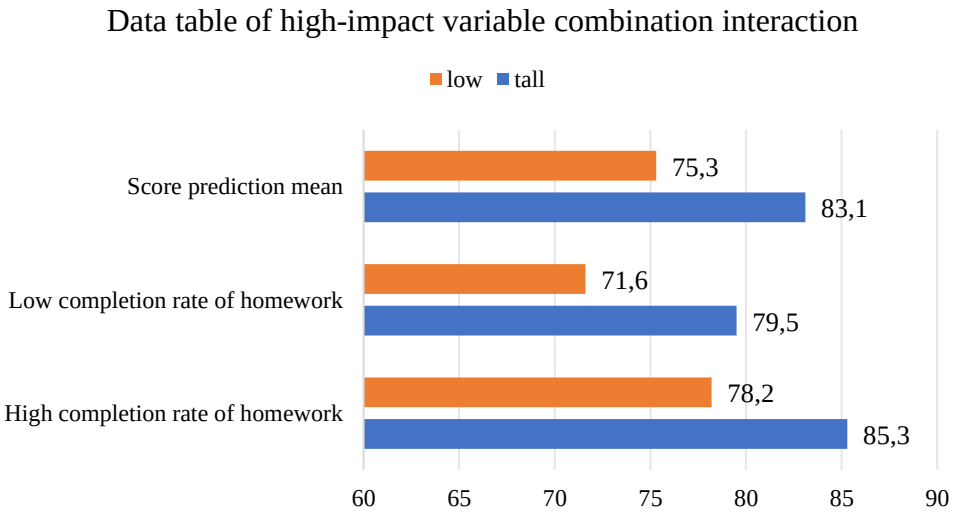

Figure 4: Data table of high-impact variable combination interaction

Under the combination of "high interaction-high completion rate", the average forecast score is the highest, which is 85.3. If both variables are at a low level, the average forecast value will drop to 71.6, with a drop of 13.7 points. This shows that both of them have correlation enhancement effect in the model, which not only has strong explanatory power alone, but also has significant linkage effect. Especially in practical teaching, teaching design should pay attention to the synchronous promotion of teachers' leading behavior and students' executive behavior in order to achieve more efficient teaching results.

3.1.7 Comparison of Explanatory Enhancement Effects

Finally, the actual performance of the model in terms of intelligibility, operational feedback and user acceptance after the introduction of SHAP interpretation module is evaluated. Select three indicators: user explanation satisfaction, teacher behavior retrospective accuracy and decision-making visual score, and compare the improvement range of the original model and the SHAP enhanced model. As shown in Table 3, the interpretation module has a significant positive impact on the overall user experience and output credibility.

Table 3: Performance improvement of interpretable enhancement model

Indicator project	Original model	SHAP enhancement model	Lifting range
User explanation satisfaction score	6.8	8.9	2.1
Teacher behavior retrospective accuracy	0.725	0.873	0.148
Decision visualization score (out of 10)	6.2	9.1	2.9

The visual score of the enhanced model is increased from 6.2 to 9.1, and the satisfaction score is also increased to 8.9, which shows that the introduction of the interpretation interface effectively enhances the user's perception of the predicted path. The accuracy of behavior backtracking increased from 72.5% to 87.3%, which reflected the improvement of logical alignment between model output and teaching behavior. This further verifies the important value of the interpretation module in model credibility, policy availability and user decision support.

3.2 Conclusion discussion

In terms of model performance, RMSE of 3XGBoost is 3.12, MAE is 2.48, and R reaches 0.89. This result shows that the model has high accuracy and stability in teaching score prediction. Compared with random forest and support vector machine, XGBoost shows stronger nonlinear fitting ability. Especially in the multi-dimensional variable interaction scene, XGBoost can capture the influence of behavior characteristics on learning effectiveness more accurately, thus reducing the prediction deviation and providing reliable data support for the interpretation module.

From the perspective of feature contribution, the ranking shows that the frequency of teacher interaction and the completion rate of homework are the primary variables. SHAP values are 0.213 and 0.195 respectively, far exceeding other indicators. This further emphasizes the key position of classroom teachers' behavior and students' task execution ability in teaching effect. Students' feedback enthusiasm and classroom participation also contribute a lot, which shows that the interaction between subject and object is an important influence path of learning effect. Teaching intervention should focus on the coordination of both sides' behaviors in order to further improve the quality of education.

At the level of teaching scene, the difference of variable contribution under different course types is revealed. Humanities courses rely more on teacher interaction, while science and engineering courses focus on the completion rate of homework. The embodiment of this difference in SHAP value clearly indicates the influence of subject content on model interpretation mechanism. Teaching strategies should be based on curriculum types and flexibly adjust key behavioral variables to improve the consistency of interpretation and prediction. Show the relationship between feedback level and prediction error. The error of high positive feedback group is 2.13, while that of low feedback group is 3.64. This significant gap reflects that the model has stronger explanatory power in high-motivation groups. Insufficient data or unclear response in low feedback situations may lead to misjudgment of the model. Therefore, the model should focus on the characteristics of low feedback groups, such as introducing finer-grained behavior variables or increasing the sampling amount to reduce the prediction errors.

The analysis of teaching stage reveals the dynamic change of variable weight in time dimension. During the teaching period, the teacher's interaction contribution is the highest, the homework completion rate is dominant during the evaluation period, and the feedback enthusiasm is improved during the counseling period. This shows that the model can adapt to the most critical factors in teaching process selection. The interpretation module can adjust the strategy suggestions based on the stage weights to improve the timeliness and pertinence of the decision-making system.

For the combination of high interaction and high completion rate in transactional analysis, the prediction score can be increased to 85.3, which is nearly 14 points higher than that in the situation where both are low. This reveals that there is a synergistic effect between variables, so teaching design should focus on behavior combination rather than single dimension. The model interpreter can help identify the interaction path of key behaviors, so that teachers can control the combination of multivariable behaviors and achieve better teaching results.

The interpretability enhancement experiment verifies the value of the interpretation module. After SHAP

enhancement, the user satisfaction increased from 6.8 to 8.9, the decision visualization score increased from 6.2 to 9.1, and the behavior retrospective accuracy increased from 72.5% to 87.3%. It can be seen that transparent interpretation not only enhances teachers' trust in the model output, but also enhances the usability and user experience of the system. This result shows that integrating the interpretation mechanism into the teaching system can effectively promote the adoption and implementation of teaching decisions.

Our finding that teacher-interaction frequency dominates the SHAP ranking aligns with the classroom-transaction theory originally validated by Hattie [28] and later corroborated by XAI-enhanced studies such as Liu et al. [17] who likewise reported effect sizes > 0.20 for dialogic moves. The phase-dependent flip of importance from teacher talk (instructional period) to homework completion (assessment period) mirrors the temporal motif detected by Liang et al. [15] in their MOOC click-stream explanation. By explicitly mapping these temporal motifs with SHAP, we extend prior post-hoc analyses into a real-time, decision-ready framework.

4 Conclusion

Besides the already mentioned constraints, the current pipeline cannot ingest unstructured data that carry rich instructional signals: classroom audio, teachers' slide narratives, handwritten comments, or discussion-board text. These modalities are increasingly captured by smart-classroom infrastructure but remain absent from our feature set. Moreover, the model was trained on a single prefecture-level city where curriculum standards, teacher evaluation norms and student socio-economic status are relatively homogeneous. Performance and explanation structures may not transfer to provinces with different digital-maturity levels, ethnic composition, or private-public mixes; external validation on cross-regional samples is still lacking. Future releases will integrate speech-to-text and large-language-model embeddings for open-ended logs, and a federated scheme will be explored to test generalisability without moving sensitive raw data across institutions.

Although the results are convincing, there are still some limitations in the research method. On the one hand, feature construction relies on platform recording data, which has not fully covered unstructured information, such as voice interaction or teaching text content. Although the dimension of data features has been optimized, the abstract ability at the semantic level is still insufficient. On the other hand, the explanation mechanism is more based on the static attribution of a single sample, and the evolutionary structure of the explanation results at the group level has not yet been constructed. In addition, although the system deployment is real-time, its adaptability to cross-school and cross-regional teaching mode still needs further verification. User feedback is obtained in the form of questionnaire, which is subjective to some extent, and can be cross-verified by combining behavior logs in the future.

Future research can start with the construction of multimodal features, and introduce data sources such as speech recognition and eye tracking to improve the model's ability to restore teaching situations. At the same time, it is considered to integrate time series modeling with causal reasoning to expand the depth of the model's explanation of teaching behavior chain. In the aspect of model integration, we can explore structural awareness models such as graph neural network to enhance the ability of context association recognition between variables. On the application level of interpretation results, we should design a more elaborate teacher support interface, transform individual interpretation structure into operational suggestions, and realize real-time feedback and dynamic adjustment of teaching behavior. This study provides a systematic path in the interpretable modeling of teaching decision, which not only has predictive efficiency, but also takes into account user understanding. The follow-up work will continue to deepen on richer data, more flexible modeling strategies and more practical forms of system interaction, and help to build an intelligent teaching support system with educational situational awareness.

We acknowledge three external-validity boundaries. First, the 41 participating schools share the same municipal digital platform; rural counties using alternative LMS or paper-based workflows are not represented. Second, cultural scripting of "teacher interaction" may limit exportability to collectivist East-Asian contexts; replication in individualistic cultures (e.g., Scandinavian problem-based learning classrooms) is warranted. Third, the outcome variable—unit test score—carries high-stakes weight in our province; low-stakes formative settings may dampen the observed variable importance of homework completion. Future cross-national validation through the UN-UNESCO "Smart Digital Learning" federated network is planned to quantify these contextual interaction terms.

Data availability

The data supporting the findings of this study are available within the article.

Funding

This work was supported by Key Project of the Education Department of Anhui Province; Research on the Standardized Operation of Higher Vocational Colleges under the Background of Private Education Group Management: A Case Study of Suzhou Vocational College of Civil Aviation (No. 2023jyxm1768).

References

[1] Wang KL, De Vos J, Smart M, Wang SC. Explaining youth driver licensing determinants using XGBoost and SHAP. Transport Policy. 2025; 168:87–100. doi:10.1016/j.tranpol.2025.04.009.

[2] Li S, Liang XJ, Yu J, Qiu TQ, Wu C. Identifying nighttime vitality through multisource geodata: An explainable AI perspective. Trans GIS. 2025; 29(3):e70064. doi:10.1111/tgis.70064.

[3] Nannini L, Huyskes D, Panai E, Pistilli G, Tartaro A. Nullius in explanans: An ethical risk assessment for explainable AI. Ethics Inf Technol. 2025; 27(1):5. doi:10.1007/s10676-024-09800-7.

[4] Xia XN, Qi WX. Learning behaviour prediction and multi-task recommendation based on a knowledge graph in MOOCs. Technol Pedagog Educ. 2025; 34(3):315–338. doi:10.1080/1475939X.2024.2442989.

[5] Zhang CH, Liu JY, Zhang SC. Online purchase behavior prediction model based on recurrent neural network and naive Bayes. J Theor Appl Electron Commer Res. 2024; 19(4):3461–3476. doi:10.3390/jtaer19040168.

[6] Tielman ML, Suárez-Figueroa MC, Jönsson A, Neerincx MA, Siebert LC. Explainable AI for all – A roadmap for inclusive XAI for people with cognitive disabilities. Technol Soc. 2024; 79:102685. doi:10.1016/j.techsoc.2024.102685.

[7] Nannini L, Manerba MM, Beretta I. Mapping the landscape of ethical considerations in explainable AI research. Ethics Inf Technol. 2024; 26(3):44. doi:10.1007/s10676-024-09773-7.

[8] Pramanik P, Jana RK, Ghosh I. AI readiness enablers in developed and developing economies: Findings from the XGBoost regression and explainable AI framework. Technol Forecast Soc Change. 2024; 205:123482. doi:10.1016/j.techfore.2024.123482.

[9] Soydaner D, Wagemans J. Unveiling the factors of aesthetic preferences with explainable AI. Br J Psychol. 2024. doi:10.1111/bjop.12707.

[10] Rahman SAFSA, Maulud KNA, Ujang U, Jaafar WSWM, Shaharuddin S, Ab Rahman AA. The digital landscape of smart cities and digital twins: A systematic literature review of digital terrain and 3D city models in enhancing decision-making. SAGE Open. 2024; 14(1):21582440231220768. doi:10.1177/21582440231220768.

[11] Díaz GM, Salvador JLG. Group decision-making model based on 2-tuple fuzzy linguistic model and AHP applied to measuring digital maturity level of organizations. Systems. 2023; 11(7):341. doi:10.3390/systems11070341.

[12] Kay J. Foundations for Human-AI teaming for self-regulated learning with explainable AI (XAI). Comput Hum Behav. 2023; 147:107848. doi:10.1016/j.chb.2023.107848.

[13] Shang C, Jiang J, Zhu L, Saeidi P. A decision support model for evaluating risks in the digital economy transformation of the manufacturing industry. J Innov Knowl. 2023; 8(3):100393. doi:10.1016/j.jik.2023.100393.

[14] Mao CY, Xu WJ, Huang YW, Zhang XT, Zheng N, Zhang XH. Investigation of passengers' perceived transfer distance in urban rail transit stations using XGBoost and SHAP. Sustainability. 2023; 15(10):7744. doi:10.3390/su15107744.

[15] Liang GQ, Jiang CS, Ping QZ, Jiang XY. Academic performance prediction associated with synchronous online interactive learning behaviors based on the machine learning approach. Interact Learn Environ. 2024; 32(6):3092–3107. doi:10.1080/10494820.2023.2167836.

[16] Rajabi E, Etminani K. Knowledge-graph-based explainable AI: A systematic review. J Inf Sci. 2024; 50(4):1019–1029. doi:10.1177/01655515221112844.

[17] Liu H, Chen X, Liu XX. Factors influencing secondary school students' reading literacy: An analysis based on XGBoost and SHAP methods. Front Psychol. 2022; 13:948612. doi:10.3389/fpsyg.2022.948612.

[18] Xiao JE, Teng HQ, Wang H, Tan JX. Psychological emotions-based online learning grade prediction via BP neural network. Front Psychol. 2022; 13:981561. doi:10.3389/fpsyg.2022.981561.

[19] Li ZQ. Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost. Comput Environ Urban Syst. 2022; 96:101845. doi:10.1016/j.compenvurbsys.2022.101845.

[20] Zhou C, Wu D, Li YT, Yang HH, Man S, Chen M. The role of student engagement in promoting teachers' continuous learning of TPACK: Based on a stimulus-organism-response framework and an int egrative model of behavior prediction. Educ Inf Technol. 2023; 28(2):2207–2227. doi:10.1007/s10639-022-11237-8.

[21] Chang I, Park H, Hong E, Lee J, Kwon N. Predicting effects of built environment on fatal pedestrian accidents at location-specific level: Application of XGBoost and SHAP. Accid Anal Prev. 2022; 166:106545. doi:10.1016/j.aap.2021.106545.

[22] Yan MR, Hong LY, Warren K. Integrated knowledge visualization and the enterprise digital twin system for supporting strategic management decision. Manag Decis. 2022; 60(4):1095–1115. doi:10.1108/MD-02-2021-0182.

[23] Lutz W, Deisenhofer AK, Rubel J, Bennemann B, Giesemann J, Poster K, et al. Prospective evaluation of a clinical decision support system in psychological therapy. J Consult Clin Psychol. 2022; 90(1):90–106. doi:10.1037/ccp0000642.

[24] Chen CH, Yang SJH, Weng JX, Ogata H, Su CY. Predicting at-risk university students based on their e-book reading behaviours by using machine learning classifiers. Australas J Educ Technol. 2021; 37(4):130–144. doi:10.14742/ajet.6116.

[25] Parsa AB, Movahedi A, Taghipour H, Derrible S, Mohammadian A. Toward safer highways: Application of XGBoost and SHAP for real-time accident detection and feature analysis. Accid Anal Prev. 2020; 136:105405. doi:10.1016/j.aap.2019.105405.

[26] Maktum T, Pulgam N, Chandgadkar V, Pathak P, Solanki A. A machine learning based framework for bankruptcy prediction in corporate finances using explainable AI techniques. Informatica. 2025;49:15-26. doi:10.31449/inf.v49i1s.6745.

[27] Sirinidi B, Badrnave MS. Predictive and prognostic explainable artificial intelligence for pancreatic cancer: an interpretable deep learning and machine learning approach. Chinese Medical Journal. 2024. doi.org/10.31449/inf.v48i4.5151

[28] Hattie J. Visible Learning. A Synthesis of over 800 Meta-Analyses Relating to Achievement. London: Routledge. 2009.