College Student Entrepreneurship Risk Assessment Model Based on XGBoost

Ming Guo
School of Physical Education, Wuhan Business University, Wuhan 430056, Hubei, China
Corresponding author: Ming Guo (e-mail: guoming37212003@163.com)

ABSTRACT The core challenge of college student entrepreneurship risk assessment is that traditional models often ignore complex factor interactions, resulting in biased risk predictions. To address this issue, this paper applies an XGBoost model with SHAP-based interpretation. The implementation includes systematic data preprocessing, including cleaning, encoding, and missing-value imputation, followed by a recursive feature elimination strategy combined with XGBoost feature importance to select the optimal subset of predictive variables. Model hyperparameters are optimized through grid search and cross-validation. Finally, the SHAP framework is used to calculate the marginal contribution of each feature and generate global and local attribution graphs, revealing key drivers of risk prediction. Experimental results confirm the effectiveness of the proposed model. In multi-scenario comprehensive evaluation, the average accuracy exceeds 0.85, the misclassification rate is below 0.15, the F1-score exceeds 0.87, Cohen's Kappa exceeds 0.71, and crossdisciplinary generalization is strong, with AUC exceeding 0.91. The study demonstrates that the XGBoost-SHAP framework improves the reliability, transparency, and interpretability of entrepreneurship risk assessment under complex nonlinear feature interactions.

INDEX TERMS XGBoost, shapley additive explanations, feature selection, risk assessment, interpretable machine learning

I. INTRODUCTION

As an important fulcrum of the national innovation-driven development strategy, college student entrepreneurship has become a key path for higher education reform and the transformation of scientific and technological achievements, particularly driving innovation in sectors like the industry through the development of sustainable materials and smart manufacturing technologies [1,2]. The early stability of entrepreneurial projects is directly related to the efficiency of resource input and the vitality of the innovation ecosystem [3]. However, entrepreneurial behavior is influenced by multiple dynamic factors such as individual ability, team collaboration, and market environment, and the causes of risk are highly nonlinear and heterogeneous [4-6]. Building a scientific risk assessment system not only helps to precisely identify individuals with potential risks, but also provides a quantitative basis for university entrepreneurship education intervention and incubation platform resource allocation.

Currently, college student entrepreneurship risk assessment faces multiple technical bottlenecks in modeling practice. At the data level, information sources cover multimodal channels such as academic records, financing behavior, team composition, and project descriptions. There are problems such as structural heterogeneity, inconsistent dimensions, and complex missing patterns. Direct input is prone to cause model bias [7,8]. In terms of feature dimensions, variables include both quantifiable financial indicators and subjective soft factors such as entrepreneurial motivation and leadership perception. This type of mixed input places higher demands on the model's feature expression ability [9,10]. In algorithmic applications, most methods struggle to simultaneously capture nonlinear relationships and identify high-order interactions. Furthermore, feature selection, relying solely on the model's inherent importance ranking, can mistakenly delete low-frequency variables with potential interactive value, resulting in information omission [11,12]. Existing frameworks generally lack refined attribution analysis of individual samples, failing to answer the critical question of "why this entrepreneur is identified as high-risk" [13,14]. Assessment results remain at the probabilistic output level, lacking actionable explanations and providing specific guidance for intervention. These challenges collectively constitute core obstacles in current modeling practices, necessitating the integration of high-performance algorithms and interpretable technologies to build a risk identification system that combines both accuracy and transparency.

This study aims to develop a college student entrepreneurship risk assessment model with both high predictive performance and strong interpretability to address the limitations of existing methods, a requirement increasingly critical for

the industry in managing complex production and supply chain risks where clear, actionable insights are essential. To address issues such as inconsistent formats and missing value interference from multi-source heterogeneous data, a systematic approach of data cleaning, encoding conversion, and missing value processing is implemented to ensure input quality. To address the redundancy and noise applied by high-dimensional features, a screening mechanism combining recursive feature elimination with the built-in importance metric of XGBoost is proposed. This method iteratively optimizes feature subsets, retaining the most discriminative variable combinations and reducing model complexity. Furthermore, a grid search and cross-validation (CV) strategy is used to fine-tune key hyperparameters, improving model classification accuracy and generalization. Finally, the SHAP interpretation framework is applied on the optimal XGBoost model. The TreeExplainer algorithm is used to calculate the Shapley value of each feature, generating a global feature importance ranking and local individual attribution maps, revealing the direction and strength of the marginal contribution of different factors to risk prediction. The innovation of this research lies in the deep integration of XGBoost's powerful learning capabilities with SHAP's theoretical interpretability, forming a complete analytical chain from data preprocessing to risk attribution. Unlike traditional assessment models that only provide output results, this method reveals individual risk attribution through SHAP analysis, providing transparent and traceable decision-making basis for risk assessment.

II. RELATED WORK

In the field of risk modeling, studies have attempted to improve predictive performance by applying statistical learning and machine learning methods. Some scholars have used logistic regression and discriminant analysis to construct scorecard models, achieving risk classification through linear combinations of feature variables [15,16]. However, these methods have limited ability to model nonlinear relationships and are unable to cope with the interaction effects commonly found in entrepreneurial data. Algorithms such as support vector machines and random forests have been used to improve classification boundary demarcation. Among them, random forests, thanks to their ensemble learning mechanism, have demonstrated good robustness in risk modeling and disaster risk assessment [17,18]. Support vector machines are often used in network security intrusion detection to assess risk [19,20]. Some studies have used neural networks to explore deep feature representations, achieving high AUC values on specific sample sets. However, the model suffers from a serious lack of interpretability, limiting its application in decision support scenarios [21,22]. In response to the problems of high feature dimension and information redundancy, Luo N summarized the existing customs risk assessment methods and combined fuzzy logic and neural networks to build an import and export enterprise risk assessment model, which could effectively handle data uncertainty and improve assessment efficiency, and had strong feasibility and reference value [23].

To address the dual challenges of modeling high-dimensional nonlinear data and lack of interpretability, interpretable machine learning frameworks have gradually become a research frontier. XGBoost, a decision tree ensemble algorithm based on gradient boosting, has demonstrated excellent performance in structured data classification tasks. Its regularization mechanism effectively suppresses overfitting, and its split gain calculation supports automatic feature selection. It has been successfully applied in the field of financial risk management [24,25]. These practices demonstrate that the XGBoost architecture combined with SHAP has the dual ability to handle complex interactions and generate readable explanations. However, existing applications are mostly concentrated in mature enterprises or financial scenarios, and specific modeling for university student entrepreneurs remains a gap. Furthermore, most studies fail to systematically optimize the feature selection process and directly use all variables for modeling, which can easily apply noise. While some work incorporates feature selection, it lacks coordinated design with the model's internal mechanisms, failing to fully leverage the iterative advantages of recursive elimination and importance feedback. Therefore, there is an urgent need to develop an integrated evaluation framework for college student entrepreneurship that integrates dynamic feature selection and explainable reasoning. This paper proposes an end-to-end modeling process that integrates data preprocessing, dynamic feature selection, parameter tuning, and attribution analysis to form a closed-loop evaluation system.

III. ENTREPRENEURIAL RISK MODELING BASED ON THE XGBOOST MODEL

Figure 1 illustrates the three core layers of the XGBoost-based college student entrepreneurship risk assessment architecture. Data input integrates heterogeneous data from multiple sources, such as academic records and financial indicators. In the core analysis engine, the XGBoost model performs deep feature learning and risk prediction, while the SHAP framework provides prediction deconstruction and interpretation [26,27], forming a two-way intelligent interaction. The risk intelligence output layer generates four key outputs: global feature importance, individual attribution analysis, interaction effect detection, and a risk decision matrix. The entire system transforms black-box predictions into transparent decisions, providing the university entrepreneurial ecosystem with a risk assessment solution that combines predictive accuracy and decision-making explainability, supporting precise intervention and optimal resource allocation.

A. SCREENING AND ENCODING OF MULTI-SOURCE FEATURE VARIABLES

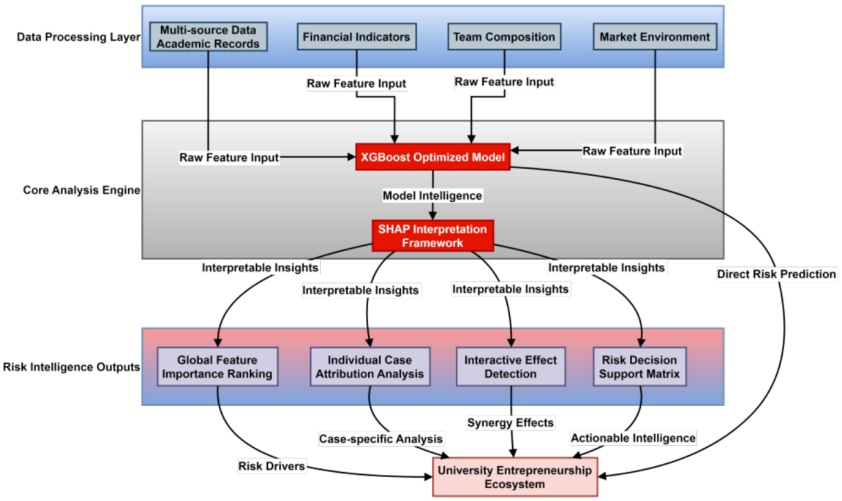

FIGURE 1. College student entrepreneurship risk assessment based on XGBoost

1) Cleaning and Missing Value Handling of Multi-source Heterogeneous Data

This study is based on a sample of N = 1,842 university student entrepreneurs tracked by a university's entrepreneurship incubation center from 2018 to 2023. High-risk labels were jointly determined by the incubation platform's expert group based on objective events such as project termination, no progress for six consecutive months, or failed financing, within a 12-month window after project initiation. The final positive sample (high-risk) proportion was 18.7% (n = 345), and the negative sample proportion was 81.3% (n = 1,497). The data was divided into training, validation, and test sets in a 7:1.5:1.5 ratio, using stratified sampling to ensure consistent risk proportions across sets. All features were limited to observable variables with a time frame of $t \leq 3$ months to strictly avoid future information leakage. The model prediction threshold is determined by maximizing the Youden exponent, rather than the default 0.5, in order to balance sensitivity and specificity The data covers three dimensions: individual background, project progress, and external environment, with a total of 15 raw variables collected. Due to the diversity of data collection channels, including the academic affairs system, entrepreneurial incubation platform logs, financing records, and questionnaire filling, the field formats are not unified; the record granularity varies significantly; there are varying degrees of omissions and anomalies. To ensure the reliability of the model input, variables with a missing field rate exceeding 30% are first removed to avoid systematic bias applied by interpolation. For the remaining variables, different strategies are implemented based on data type and distribution characteristics: for continuous indicators, a median-based imputation method is used to reduce the interference of outliers on the mean estimate. For categorical variables, an "unknown" category is applied as an independent value to preserve the potential information conveyed by the missing values and prevent misclassification as an existing category. At the same time, outliers that significantly deviate from the acceptable range are identified and corrected, and traceability correction is performed using timestamp logs.

After completing basic cleanup, variables are further restructured. The original data contains a large number of non-numeric fields. Although these variables contain important discriminative information, they cannot be directly input into the gradient boosting model. To this end, all categorical variables are one-hot encoded, converting them into binary indicator variable vectors. This operation maps each categorical value into an independent 0-1 dimension, enabling the model to discriminate equally between unordered categories and avoiding the interference of false order relationships caused by artificial ordering. This operation not only meets XGBoost's requirement for numeric input features, but also enhances the model's sensitivity to categorical differences. Considering that some high-cardinality categorical variables may cause dimensionality inflation, one-hot encoding is only performed on variables with clear semantic classification and low cardinality. For high-cardinality categorical variables (such as 'project domain' or 'school major'), target encoding is used for numerical transformation, that is, the average risk label of positive samples in that category is used as the encoded value, and after fitting on the training set, it is applied to the validation and test sets to avoid data leakage. All variables are ultimately input into the model in numerical form to ensure XGBoost compatibility. Through this encoding mechanism, unstructured text descriptions are converted into a computable vector space representation, establishing a unified modeling foundation for heterogeneous information.

2) Numerical Feature Standardization and Input Structure Unification

After completing variable encoding, all features are converted to numerical form, but there are still significant differences in dimension and order of magnitude between

different indicators. Some variables show a high amplitude range, while others are concentrated in a smaller numerical range. Although tree-based models are theoretically scale-invariant, to ensure numerical stability among heterogeneous features and to maintain consistency with subsequent SHAP value comparisons and cross-feature interaction visualizations, Z-score standardization was still applied to all continuous variables, transforming them into a scale-uniform form with a mean of 0 and a standard deviation of 1, but without changing their original distribution. The specific transformation formula is as follows:

$$x' = \frac{x - \mu}{\sigma} \tag{1}$$

x represents the original feature value; μ and σ are the sample mean and standard deviation of the feature in the training set, respectively; x' is the output value after standardization. This transformation preserves the original data distribution while keeping the features of each dimension in a comparable range, improving the stability of the model in gradient calculation and splitting gain evaluation. The standardization parameters (that is, μ and σ) are derived only based on the statistics of the training set and applied to the validation set and test set to simulate the data inflow process in the real prediction scenario and prevent information leakage.

Ultimately, the cleaned, encoded, and standardized dataset forms a uniformly structured feature matrix, with each row corresponding to a university student entrepreneur and each column representing a processed numerical feature variable. This matrix not only meets the basic input format requirements of the XGBoost algorithm but also lays a solid foundation for subsequent feature screening and model training in terms of data quality. The entire preprocessing process strictly adheres to the logical chain of "cleaning first, then encoding, and then standardization", ensuring the traceability and reproducibility of the conversion from raw heterogeneous data to modeling input, effectively supporting the accuracy and robustness of risk prediction in high-dimensional and complex environments.

B. INPUT DIMENSION OPTIMIZATION BASED ON RECURSIVE FEATURE ELIMINATION

1) Feature Importance Quantification Mechanism Based on Model Feedback

To deal with the interference of redundant variables in high-dimensional feature space on classification performance, an evaluation system with XGBoost endogenous feature importance index as the core is constructed as the decision basis for the recursive feature elimination process. In the initial training stage, the XGBoost classifier is trained with the complete feature set after preprocessing, and the evaluation criterion based on split gain is used to quantify the discriminant contribution of each variable in the tree structure construction process. This index reflects the structural role of each feature in the model decision path by counting the total amount of loss function reduction brought about by each feature when it appears as a split node in all decision trees. Its calculation method can be expressed as:

$$I_f = \sum_{t=1}^{T} \sum_{s \in S_t(f)} \Delta L_s \tag{2}$$

I_f represents the importance score of feature f; T is the total number of decision trees in the ensemble; $S_t(f)$ represents the set of all nodes in the t-th tree that are split by feature f; ΔL_s is the reduction in the loss function before and after the node s is split. This indicator is directly related to the model optimization objective and can be used as a rough screening criterion in recursive feature elimination; however, it is easily affected by feature cardinality and collinearity, and cannot accurately reflect the true attribution. Reliable feature contribution evaluation and interaction analysis are completed by the subsequent SHAP framework. After obtaining the importance ranking of all features, a recursive elimination process is initiated. In each iteration, the single feature with the lowest current importance is identified and removed, ensuring that each variable reduction is based on dynamic model feedback rather than static thresholds. This strategy avoids information gaps that can be caused by bulk elimination and ensures the stability of the feature subset evolution process. After the elimination operation is complete, the XGBoost model is immediately retrained on the updated feature set, and the importance weights of the remaining features are simultaneously updated. This closed-loop mechanism ensures that the importance ranking always reflects the relative contribution of the current feature combination, overcoming the lag bias caused by fixed ranking. Through continuous iteration, the input dimension is gradually compressed, suppressing the distraction of model attention by noise variables, and improving the focus of feature representation and discriminative efficiency.

2) Feature Subset Termination Criteria

To prevent the loss of key information due to excessive elimination, a cross-validation framework is applied as a performance monitoring mechanism for the feature screening process to ensure that the generalization ability of the model after each round of variable deletion is effectively evaluated. After each feature is removed, and the model is retrained, the average classification performance index under cross-validation is calculated, focusing on the changing trends of the AUC value and F1-score. AUC is the core indicator for measuring the overall discrimination ability of the classifier. Its calculation is based on the ranking consistency between the predicted probability and the true label, and is defined as:

$$\mathrm{AUC} = \frac{1}{N_+ N_-} \sum_{i \in P} \sum_{j \in N} I\left(p_i > p_j\right) \tag{3}$$

N_+ and N_- are the number of positive and negative samples, respectively; P and N are the corresponding sample

index sets; p_i and p_j are the predicted probabilities output by the model; I is the indicator function. This indicator is highly robust to class imbalance and is suitable for the sparse distribution of highrisk samples in college student entrepreneurship data.

The screening process continues until the AUC value decreases for three consecutive rounds or the F1-score decreases by more than the preset tolerance threshold, and the feature combination corresponding to the previous round is retained as the optimal subset. This termination strategy balances model simplicity and predictive effectiveness, achieving a dynamic balance between dimensionality reduction and performance preservation. The resulting retained feature set not only significantly reduces input space complexity but also enhances the model's sensitivity to key risk drivers. The entire recursive elimination process, under the dual constraints of model performance feedback and statistical validation, ensures interpretable and reproducible feature screening results, providing a high signal-to-noise ratio input foundation for subsequent parameter optimization and attribution analysis.

Table 1 lists the termination criteria and their specific configurations used in the recursive feature elimination process, including a three-round decrease in AUC or a decrease in F1-score exceeding 0.005 as the stopping criteria, ensuring that model performance does not significantly degrade due to excessive feature removal. Cross-validation uses 5-fold stratified sampling to ensure the consistency of sample distribution across all categories. The feature elimination strategy removes the least important feature in each round, and the model is retrained immediately after each deletion to dynamically update the feature contribution ranking. All thresholds and process settings are based on the model's generalization performance on the validation set, ensuring a stable and reproducible feature selection process. The resulting retained feature subset maximizes discriminative information while reducing dimensionality.

TABLE 1. Termination criteria

Stopping Criterion	Threshold Setting	Monitoring Metric Source
Consecutive AUC decrease	$\geq$3 rounds	Mean of 5-fold CV
F1-score drop tolerance	>0.005	Macro-averaged validation set
Number of cross-validation folds	5	Stratified K-Fold splitting
Feature removal strategy	Remove 1 lowest per round	Split gain-based importance
Model retraining frequency	After each removal	Immediately upon feature update

C. GRID SEARCH AND CROSS-VALIDATION TUNING OF MODEL HYPERPARAMETERS

1) Systematic Traversal of the Multidimensional Hyperparameter Space

To fully unleash the learning potential of the XGBoost model in the entrepreneurial risk classification task, a structured search space is constructed for its key hyperparameters, and a grid search strategy is used for systematic traversal. Three parameters with significant impact on model performance are optimized: the learning rate, which controls the iteration step size; the maximum depth, which determines the complexity of individual trees; the coefficients for adjusting the L1 (Regalpha) and L2 (Reglambda) regularization penalties. The candidate set for the learning rate covers multiple discrete values, ranging from low to medium, to balance convergence stability and training efficiency. The maximum depth is set to multiple integer values within a reasonable range to avoid overfitting caused by excessively deep trees while retaining the ability to capture nonlinear patterns. The regularization parameter is logarithmically scaled to enhance the suppression of sparse features and noisy variables. All parameter combinations form a high-dimensional discrete space, and their Cartesian product forms the complete set of configurations to be evaluated. For each set of hyperparameter configurations, the model training process is initiated, and a five-fold crossvalidation mechanism is simultaneously implemented to evaluate generalization performance. Specifically, the training data is evenly divided into five mutually exclusive subsets, four of which are selected for model fitting and the remaining one for performance verification. After five iterations, the average of each evaluation metric is taken as the final performance metric for that configuration. This validation method effectively mitigates the accidental bias applied by single-class partitioning and improves the robustness of parameter selection. The evaluation process focuses on the classifier's comprehensive discriminative ability under imbalanced data, using a multi-dimensional evaluation based on AUC and F1-score. By mapping crossvalidation results to hyperparameter combinations, a performance response surface is constructed, identifying parameter regions with an overall advantage and ultimately selecting the optimal configuration. This process ensures that the model not only has a good fit on the current sample but also exhibits stable predictive consistency on unseen data.

Figure 2 (AUC Heatmap): the horizontal axis represents the learning rate, with values labeled as 0.01, 0.05, 0.10, 0.20, and 0.30; the vertical axis represents the maximum depth, with values labeled as 3, 5, 7, 9, and

11. The center of each cell indicates the AUC value for the corresponding combination, with two decimal places. When the learning rate increases from 0.01 to 0.10, and the tree depth increases to 7, the AUC reaches its highest value (0.89). However, when the learning rate is too high (0.30), or the tree depth is too shallow or too deep (3 or 11), the AUC shows a significant downward trend. This pattern indicates that, in this simulation, a moderate learning rate combined with an appropriate tree depth achieves the optimal balance

between fitting ability and generalization stability.

Figure 3 shows the response of F1-score to the regularization parameter of the XGBoost model. In the preliminary experiment with fixed learning rate and tree depth, and only tuning of regularization parameters, the F1 score reached a peak of 0.700 near log10(RegAlpha) ≈ -2 (RegAlpha = 0.01) and log10(RegLambda) ≈ 1 (RegLambda = 10). Analysis reveals that moderate L1 regularization can suppress redundant noise in the high-dimensional feature space, sparsifying the model weights and reducing the risk of overfitting. Moderate L2 regularization, on the other hand, constrains the weight amplitude, preventing the model from being overly sensitive to local feature fluctuations. If the L1 or L2 regularization is too strong, the model becomes overly smooth or sparse, weakening the influence of key features and resulting in a lower F1-score. If the regularization is too weak, the model is prone to overfitting the training data, failing to generalize to highprecision examples, and compromising both recall and precision. The three-dimensional surface clearly reveals the interaction mechanism of the regularization parameters, suggesting that in the scenario of college student entrepreneurship risk assessment, the parameter combination of medium L1 (0.01) combined with relatively strong L2 (10) should be preferred to take into account the stability and discrimination ability of the model, achieve a balance between precision and recall, and thus obtain the optimal F1-score.

2) Optimal Parameter Locking Strategy Based on Loss Function Optimization

After completing the full grid traversal, the final optimal parameter set needs to be locked from a large number of candidate configurations. The selection criteria not only rely on the absolute value of the performance index, but also need to be comprehensively judged in combination with the model complexity and training dynamics. To this end, an early stopping mechanism is applied to assist each round of training process, set a fixed upper limit for the number of iterations, and monitor the trend of binary logarithmic loss on the validation set:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{p}_i\right) + \left(1 - y_i\right)\log\left(1 - \hat{p}_i\right)\right] \quad (4)$$

N is the total number of samples; y_i is the true label of the i-th sample (0 or 1); $\hat{p}_i$ is the corresponding predicted probability output by the model. This loss function is directly related to the probability calibration quality of the classifier, and its decline process reflects the degree of approximation of the model to the target distribution. In each cross-validation fold, if the validation loss does not show a significant decrease for multiple consecutive rounds, the iteration is terminated early to prevent invalid training from wasting resources and reduce the risk of overfitting.

The final optimal parameter combination is ranked by the cross-validation average AUC value, and the consistency is verified by combining F1-score and logloss. When multiple configurations have similar performance, combinations with high regularization strength and moderate tree depth are prioritized to enhance model robustness and interpretability. Once selected, this parameter set is used to retrain the final model on the entire training data, providing a high-precision, low-variance prediction foundation for subsequent SHAP attribution analysis. The entire tuning process, through a combination of intensive sampling and rigorous validation, achieves a transition from coarse-grained empirical setting to refined parameter configuration, significantly improving the model's discriminative accuracy and external applicability in complex entrepreneurial risk scenarios.

D. RISK ATTRIBUTION CONSTRUCTION WITH SHAP MECHANISM

1) Calculation of Individual-level Feature Contributions Based on TreeExplainer

To overcome the limitations of the "black-box" output of traditional risk scoring models and enhance the traceability and decision-support value of prediction results, the TreeExplainer algorithm within the SHAP (SHapley Additive exPlanations) framework is used to perform attribution analysis on trained XGBoost models. Based on cooperative game theory, this method considers model predictions as the result of the "cooperation" of input features. By calculating the mean marginal contribution of each feature across all possible feature subset combinations and assigning it a corresponding Shapley value, it achieves fair distribution of prediction results. Specifically, for each sample's prediction output, TreeExplainer leverages the tree structure of the XGBoost model to efficiently traverse all decision paths, identify the participation patterns of each feature in node splitting, and combine path probabilities with leaf node weights to precisely estimate the direction and strength of its contribution to the final prediction's deviation from the baseline value.

The core of this process is to construct the marginal influence function of each feature, which is mathematically expressed as follows:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!}\left[f(S \cup \{i\}) - f(S)\right] \quad (5)$$

ϕ_i represents the Shapley value of the feature; F is the entire feature set; S is any subset that does not contain feature i; $f(S \cup \{i\})$ and $f(S)$ represent the model prediction expectations before and after adding feature i to the feature set S. Formula (5) ensures that the attribution results meet axiomatic properties such as efficiency, symmetry, and consistency, and avoids misallocation of contributions due to feature correlation. TreeExplainer significantly reduces the combinatorial explosion complexity of the original Shapley calculation through dynamic pruning and path compression technology, making it suitable for practical application sce-

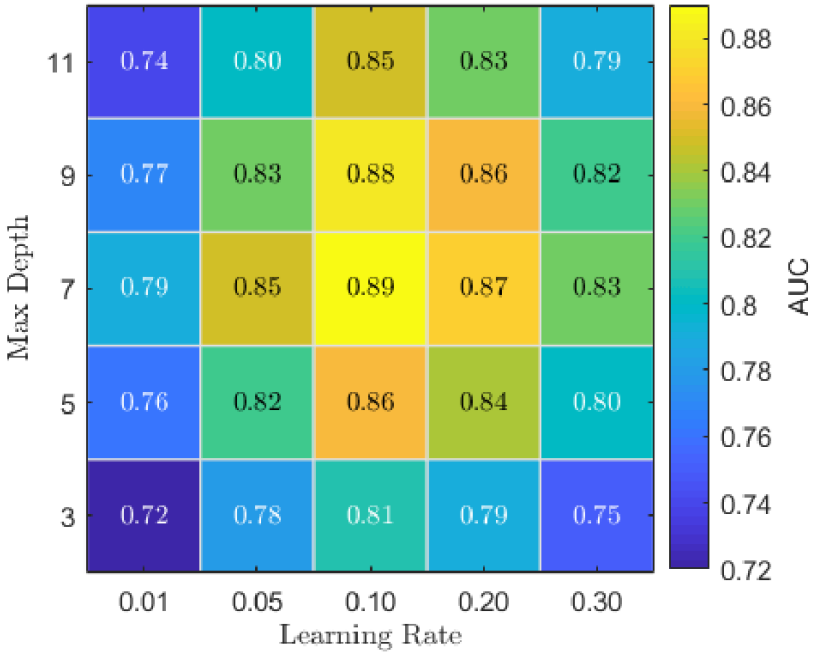

FIGURE 2. AUC across learning rates and maximum depths

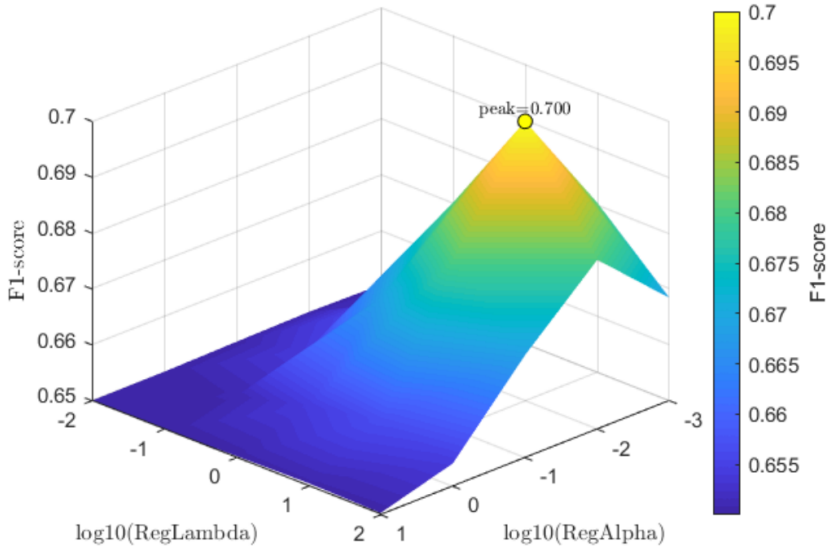

FIGURE 3. Response of F1 score to regularization (log scale)

narios of large-scale tree ensemble models. Finally, each dimension of each sample input is assigned a real-valued SHAP contribution score. A positive value indicates that the prediction is pushed towards a highrisk direction, while a negative value indicates an inhibitory effect, This establishes a fine-grained contribution attribution mapping from model output to input variables.

2) Generation and Semantic Analysis of Global and Local Attribution Maps

After calculating the SHAP values for individual samples, a multi-level attribution visualization system is constructed to reveal systematic patterns and case-specific mechanisms in risk identification. First, a global feature importance ranking is generated based on the mean SHAP absolute value of all samples. This ranking differs from the model's built-in importance metric in that it is based on the actual impact of the feature on the prediction outcome, rather than the frequency of splits or cumulative gain. Therefore, it better

reflects the true weight of the variable in the decision logic. A horizontal bar chart displays the average contribution of each feature, and a color gradient is used to indicate the direction of its influence (red and blue correspond to positive drivers of high and low values, respectively). This creates an intuitive global attribution map, helping to identify dominant risk factors across populations.

Furthermore, a local attribution explanation map is constructed for specific individual samples, demonstrating how the prediction outcome is driven by the specific values of each feature. This map starts with the baseline prediction value (that is, the average output of all features) and overlays the SHAP values of each feature, ultimately leading to the actual prediction outcome and forming a traceable "explanatory path". On this basis, a dependency analysis mechanism is applied to examine the interaction effects between high-contribution features and other variables. By plotting the scatter distribution of SHAP values and feature values, nonlinear response patterns, such as threshold effects or saturation intervals, are identified. To assist in identifying potential conditional dependency patterns, this paper introduces an exploratory measure:

$$\text{Dependence}(i,j) = \text{Corr}\left(\phi_i, x_i \mid x_j\right) \tag{6}$$

Where x_i is the original value of the target feature, and x_j is the potential interaction feature. This formula calculates the local partial correlation coefficient between SHAP value ϕ_i and x_i within a fixed x_j interval, used to visualize the direction and intensity of the interaction effect. It should be noted that this is a heuristic indicator, which does not have causal or statistical inference power,and only serves to qualitatively enhance the dependency graph. This analysis helps to reveal the synergistic or offsetting mechanisms hidden behind the main effects and enhance the depth of explanation. Ultimately, the attribution map not only serves model auditing and credible verification, but also provides actionable intervention clues for university entrepreneurship guidance personnel, achieving closed-loop support from "risk warning" to "precision intervention".

IV. VERIFICATION OF MODEL RISK PREDICTION PERFORMANCE

A. FEATURE SCREENING

During the evaluation process, a systematic missing value analysis is performed on 15 initial variables based on a real-world project dataset. These variables include GPA (Grade Point Average), Internship, Funding, TeamSize, ParentEnt (Parental Entrepreneurial Background), Motivation, Leadership, ProjInnov (Project Innovation), MarketSize, CompIntensity (Competitive Intensity), WebTraffic, SocialMed (Social Media Influence), BizPlanLen (Business Plan Length), AdvisorExp (Advisor Experience), and PitchScore. Variable screening is performed by calculating the missing percentage for each variable and strictly applying a 30% missingness threshold. Median imputation is used for retained continuous variables, while "Unknown" is applied as a new feature category for categorical variables. This ensures that all processing operations are based on the actual data distribution characteristics. Ultimately, an optimized feature set is obtained for subsequent modeling.

Figure 4 illustrates the missing value analysis and processing strategy during the data cleaning phase. To protect data privacy and facilitate visualization, the variable names and missing rate values in Figure 4(a) are displayed schematically after anonymization, but their missing patterns are consistent with the real data. The 15 variables and their missing rates used in the actual modeling were all derived from real records of the university entrepreneurship support system, using a horizontal bar chart. Simulated data shows that variables such as "ParentEnt", "PitchScore", and "Funding" have the highest missingness rates (60%, 50%, and 40%, respectively), while variables such as "TeamSize" and "ProjInnov" have lower missingness rates. The black dashed line indicates the 30% missingness rate threshold; variables above this threshold are directly removed to prevent systematic bias applied by highly missing variables. Figure 4 (b) presents the distribution of missing values using differentiated missing value handling strategies using a pie chart. The three sections of the pie chart represent the number of continuous variables imputed with the median (seven: GPA, TeamSize, WebTraffic, CompIntensity, MarketSize, BizPlanLen, and AdvisorExp); the number of categorical variables treated by applying an "unknown" category (five: Motivation, Leadership, Internship, ProjInnov, and SocialMed); the number of variables removed (three, serving as a reference). This data demonstrates how missing values in mixed-type data are handled, ensuring data integrity for subsequent analyses.

Figure 5 (a) shows the changing importance of 12 entrepreneurial risk features during the recursive feature elimination process. The horizontal axis represents the number of iterations, and the vertical axis represents the specific variables. The color represents the relative importance score. As the iterations progress, lowcontribution variables such as BizPlanLen, AdvisorExp, and SocialMed are gradually eliminated, while core variables such as GPA, TeamSize, Leadership, and MarketSize maintain their high importance after multiple rounds of iteration. This indicates that in risk assessment, factors such as individual academic performance, team size, and leadership have sustained discriminative value, while external and weakly correlated factors contribute only a limited amount. The seven feature subsets ultimately retained have clear theoretical basis in existing research. Academic performance (GPA) is often regarded as a proxy indicator of an individual's learning ability and self-discipline, and is related to an entrepreneur's adaptability and resilience; team size directly affects resource mobilization capabilities and decision-making efficiency, and is a key antecedent of organizational stability; perceived leadership reflects the influence and coordination ability of core members, and has been proven to significantly affect

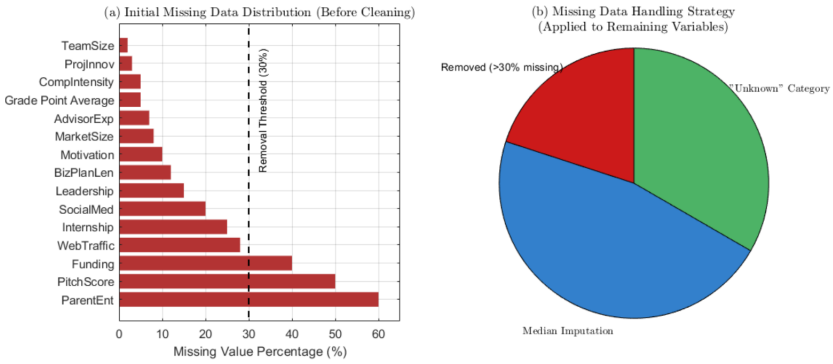

FIGURE 4. Missing value analysis and preprocessing strategy

entrepreneurial intentions and project survival; market size represents the attractiveness of the external environment and is one of the final retained basic dimensions of risk assessment; project innovation reflect endogenous drivers and is core element for predicting the potential of startups. In addition, internship experience reflects practical skills and the depth of industry exposure, while entrepreneurial motivation characterizes the strength of intrinsic drive. Together, these two factors constitute the psychological and behavioral basis of an individual's risk resilience and are also identified by the model as stable discriminant factors. Therefore, this feature combination not only optimizes processes through data-driven approaches but also aligns with the mainstream theoretical framework for entrepreneurial risk among college students. Under the premise of using the determined optimal hyperparameter configuration (learning rate = 0.10, maximum depth = 7), Figure 5 (b) shows the trends of AUC and F1 scores as the number of retained features decreases. When the number of features decreases from 12 to 7, AUC and F1 scores reach their peak values (AUC $\approx$ 0.88 and F1 $\approx$ 0.79, respectively), indicating that the model's recognition ability and classification accuracy are improved after removing redundant features. However, as the number of features continues to decrease, both metrics decline rapidly, indicating that excessive compression leads to information loss and weakens the model's discriminative power. Ultimately, the top seven features are retained. The overall trend suggests that moderate feature selection can enhance model performance, but excessive elimination destroys the multidimensional interactive information of risk factors.

B. COMPARISON OF ACCURACY AND FALSE POSITIVE COST

A description should be inserted before line 499, or the wording in lines 499–501 should be revised as follows: To assess the robustness of the model under different resource constraints, the sample is divided into three subgroups based on the actual amount of funds received by the project (not the "funding" category): less than 50,000 yuan, 50,000 to 100,000 yuan, and more than 100,000 yuan. This amount information is fully recorded by the incubation platform's financial system, with a missing rate of less than 3%, and can be used for stratified analysis of the entire sample.Within each subgroup, the classification accuracy and misclassification rate (that is, the total proportion of samples that are misclassified as the opposite category) of each model are independently calculated. Five-fold cross-validation is repeated three times to obtain the mean and standard deviation of performance. Comparison models include the proposed XGBoost+SHAP ensemble framework, as well as popular traditional XGBoost, random forest (RF), support vector machine (SVM), and LightGBM (Light Gradient Boosting Machine). All indicators are compared fairly while maintaining consistent data distribution, ensuring that the evaluation results reflect the essential differences in the models rather than sampling bias.

Figure 6 compares the classification accuracy and misclassification rate of five models at different levels of entrepreneurship funding. Figure 6(a) shows that as entrepreneurship funding increases, the accuracy of each model generally declines slowly, reflecting the more subtle risk patterns and greater identification complexity of high-funded projects. The proposed XGBoost+SHAP model maintains optimal or near-optimal accuracy across all funding groups, with an average value exceeding 0.85. This demonstrates that its integrated advantages in feature interpretation and decision path optimization effectively enhance generalization. The misclassification rate results in Figure 6(b) further confirm this trend. The low-funded group has generally lower misclassification rate, indicating that this group's risk characteristics are more pronounced and easier to identify. Notably, the traditional SVM's misclassification rate increases significantly in the high-funded group, exposing its limitations in modeling nonlinear structures and sparse features. The width of the error bars across the three funding groups varies slightly, indicating that the models perform stably under multiple cross-validation cycles. The XGBoost+SHAP model maintains consistently narrow error bands, with an average misclassification rate below 0.15, demonstrating strong predictive consistency. Overall, funding size not only affects the difficulty of risk

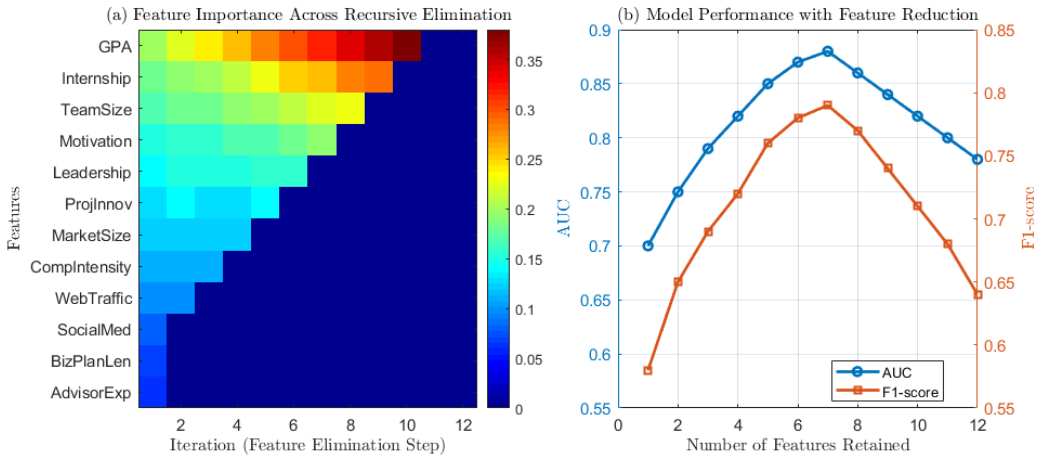

FIGURE 5. Recursive feature elimination and model performance changes

identification but also reveals the differences in the adaptability of different algorithms in heterogeneous populations. This model, through its interpretability enhancement mechanism, achieves robust identification under diverse funding scenarios.

C. STABILITY AND ROBUSTNESS TESTING UNDER VARYING TEAM SIZE

To test the model's adaptability under different organizational structure complexities, the samples are divided into three categories based on the number of entrepreneurial teams: 5 or fewer, 6-15, and more than 15. Within each subgroup, the F1 score and Cohen's Kappa coefficient are independently calculated for each model. The F1 score comprehensively reflects the balance between precision and recall in classification, while the Cohen's Kappa coefficient measures the degree of consistency between predicted results and true labels, and is particularly useful in scenarios with uneven class distribution. All metrics are evaluated based on the mean and standard deviation of three 5-fold cross-validation runs to ensure statistical reliability. By analyzing the performance fluctuations of each model under three types of team sizes, its sensitivity to changes in organizational size is identified, and the robustness and generalization ability of the model in the diversity of real entrepreneurial ecosystems are evaluated. Figure 7 shows the performance differences of various models in terms of F1 scores and Cohen's Kappa coefficients for different team sizes, revealing the nonlinear impact of organizational complexity on the performance of risk prediction models. An inverted U-shaped trend is observed overall: in small teams ($\leq$5 people), where decision chains are short and behavioral patterns are concentrated, all models demonstrate high discriminative ability. However, when teams expand to 6–15 people, communication costs increase and role divisions become more complex, leading to increased information noise and diluted risk signals. The performance of all models declines significantly, reflecting the most significant modeling challenges in medium-sized teams. In large teams with more than 15 people, some projects demonstrate institutionalized management and functional specialization, leading to more structured risk identification paths and improved model performance, with the XGBoost+SHAP framework showing a significant improvement. The average F1 score of the proposed models is above 0.87, and the average Cohen's Kappa coefficient is above 0.71. This phenomenon suggests that interpretability mechanisms can help identify the contributions of key decisionmakers or core functional modules within large teams, mitigating the discriminative attenuation caused by "group ambiguity". In contrast, traditional models such as SVM and random forests struggle to recover in large groups, exposing their limitations in modeling high-dimensional collaborative structures. The error bar distribution shows that XGBoost+SHAP achieves the minimum variance across all groups, demonstrating its stability and robustness across organizational structures and validating the adaptability of fusion attribution analysis in complex entrepreneurial contexts.

D. MEASURING GENERALIZATION ABILITY IN INTERDISCIPLINARY CONTEXTS

To examine the model's adaptability and generalization boundaries across entrepreneurial groups from different disciplines, the sample is divided into three groups based on the entrepreneurs' disciplines: science and engineering, economics and management, and humanities and arts. The stability of each model is systematically evaluated under heterogeneous knowledge structures and entrepreneurial logics. Within each subgroup, the AUC is independently calculated to measure the model's overall discriminative ability. The Brier score is also applied to assess the calibration quality of the predicted probabilities (lower is better), balancing classification efficiency and confidence reliability. Means and standard deviations for all metrics are obtained based on three 5-fold cross-validation runs. By analyzing the performance fluctuation patterns of each model in three types of subject groups, its sensitivity to the expression of subject-specific characteristics is identified, thereby revealing the

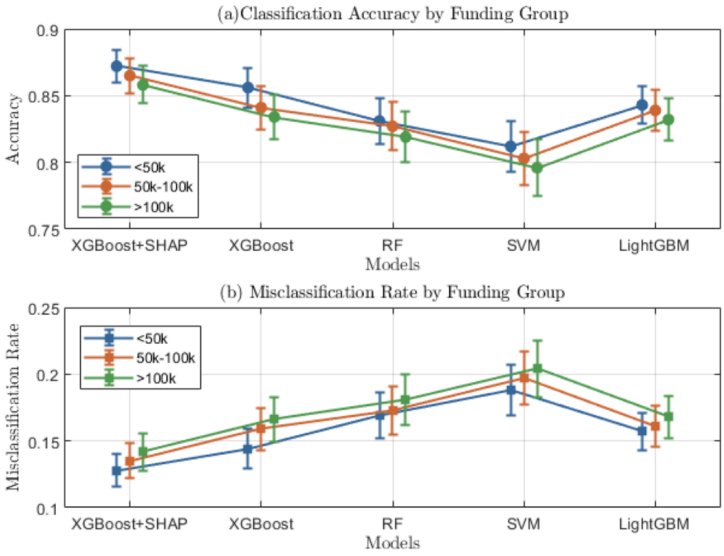

FIGURE 6. Accuracy and misclassification rate

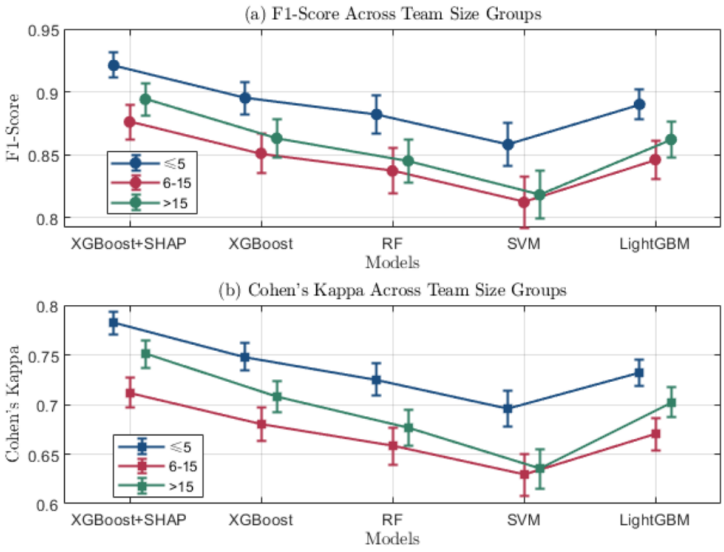

FIGURE 7. F1 score and Cohen's Kappa coefficient

applicability boundaries and transferability of the algorithm in the context of diversified higher education.

Table 2 shows the AUC values and Brier scores of various models across different disciplinary backgrounds, revealing differences in their generalization abilities across heterogeneous entrepreneurial populations. Models perform best in the science and engineering sample, reflecting the relatively clear risk profiles of technology-driven projects, making them easier to model and identify. Business and management performs second, as their business logic is more structured but still subject to market dynamics. Humanities and arts entrepreneurial projects, lacking standardized financial pathways and quantifiable outcomes, result in sparse risk signals and significant nonlinearity, leading to a decrease in the discriminative performance of all models. Notably, the proposed XGBoost+SHAP framework achieves the highest AUC values and lowest Brier scores across all three disciplinary categories, with AUC values exceeding 0.91 and Brier scores below 0.17, demonstrating its greater adaptability in capturing entrepreneurial risk patterns. The data validates the SHAP attribution mechanism's effectiveness in enhancing decision transparency and stability in weakly structured feature spaces. Random forests and support vector machines exhibit significant fluctuations in crossdisciplinary transfer, particularly in probability calibration, exposing their limitations in modeling sparse, highorder interactive features. Overall, XGBoost+SHAP not only excels in discriminative ability but also demonstrates

exceptional performance in predictive confidence reliability, demonstrating its potential for robust and reliable risk assessment within a multidisciplinary ecosystem.

TABLE 2. Generalization ability measures

Model	Discipline	AUC (Mean ± Std)	Brier Score (Mean ± Std)
XGBoost+SHAP	Science & Engineering	0.932 ± 0.011	0.142 ± 0.008
	Economics & Management	0.925 ± 0.013	0.151 ± 0.009
	Humanities & Arts	0.918 ± 0.015	0.163 ± 0.011
XGBoost	Science & Engineering	0.906 ± 0.014	0.168 ± 0.010
	Economics & Management	0.899 ± 0.016	0.177 ± 0.012
	Humanities & Arts	0.882 ± 0.018	0.194 ± 0.014
RF	Science & Engineering	0.887 ± 0.017	0.185 ± 0.013
	Economics & Management	0.880 ± 0.019	0.192 ± 0.015
	Humanities & Arts	0.863 ± 0.021	0.210 ± 0.017
SVM	Science & Engineering	0.864 ± 0.020	0.208 ± 0.016
	Economics & Management	0.858 ± 0.022	0.215 ± 0.018
	Humanities & Arts	0.839 ± 0.024	0.236 ± 0.020
LightGBM	Science & Engineering	0.901 ± 0.015	0.172 ± 0.011
	Economics & Management	0.894 ± 0.017	0.181 ± 0.013
	Humanities & Arts	0.876 ± 0.019	0.199 ± 0.015

Note: std stands for standard deviation.

E. DYNAMIC ADAPTABILITY IN TIME-EVOLVING CONTEXTS

To evaluate the model's dynamic adaptability over the lifecycle of entrepreneurial projects, a longitudinal tracking dataset is constructed, selecting three key time points: 3 months, 6 months, and 12 months after project inception. At each time point, risk predictions are made using the cumulatively updated feature information up to that point. The recall rate of each model is calculated to measure its ability to identify real high-risk cases. Furthermore, a Predictive Consistency Index (PCI) is defined, that is, the ratio of the same project being assigned the same risk level at two consecutive time points, to assess the temporal stability of the model output. This design effectively tests the model's discriminative persistence and logical coherence under dynamic information evolution, revealing its applicability and credibility in long-term monitoring scenarios.

Table 3 shows the recall and prediction consistency index of various models at different stages of project development, revealing the dynamic evolution of their performance over time. As the observation period extends from 3 to 12 months, the recall of all models exhibits varying degrees of change. However, the proposed XGBoost+SHAP framework maintains a high recall level (average above 0.89) and continues to improve in the later stages, demonstrating its ability to effectively capture the gradual nature of risk accumulation. More importantly, the PCI metric demonstrates that the model's logical stability in continuous predictions significantly outperforms the comparison methods. Although the consistency of all models declines over time, XGBoost+SHAP maintains a high consistency of 91.6% at 12 months, exceeding both traditional XGBoost and support vector machines. This performance reflects its stronger inherent consistency in updating feature weights and adjusting decision paths, preventing drastic fluctuations in judgments due to dynamic data changes. In contrast, the performance of random forests and LightGBM exhibits significant attenuation in long-term tracking, exposing the ensemble's sensitivity to temporal perturbations. Support vector machines, lacking a probabilistic output mechanism, exhibit the weakest dynamic adaptability. XGBoost+SHAP, leveraging an interpretability-guided feature attribution mechanism, achieves a synergistic combination of high sensitivity and stability, demonstrating its dual advantages of discriminative power and reliability in long-term entrepreneurial risk monitoring.

TABLE 3. Dynamic adaptability in time-evolving contexts

Model	Prediction Time Point	Recall (Mean ± Std)	PCI (%)
XGBoost+SHAP	3 months	0.892 ± 0.013	94.7
	6 months	0.905 ± 0.011	92.3
	12 months	0.918 ± 0.009	91.6
XGBoost	3 months	0.864 ± 0.015	88.2
	6 months	0.851 ± 0.017	83.5
	12 months	0.837 ± 0.019	76.8
RF	3 months	0.843 ± 0.018	85.6
	6 months	0.829 ± 0.020	80.1
	12 months	0.812 ± 0.022	73.4
SVM	3 months	0.816 ± 0.021	82.3
	6 months	0.794 ± 0.023	75.6
	12 months	0.768 ± 0.025	68.2
LightGBM	3 months	0.858 ± 0.016	87.1
	6 months	0.842 ± 0.018	82.7
	12 months	0.825 ± 0.020	75.3

V. CONCLUSIONS

This paper constructs an XGBoost model for college student entrepreneurship risk assessment that incorporates the SHAP interpretation mechanism, providing an approach also vital for the industry to analyze and explain the high-dimensional factors driving complex operational risks. Through systematic data preprocessing, recursive feature elimination, and hyperparameter fine-tuning, it significantly improves classification performance and generalization capabilities. Empirical results demonstrate that the model exhibits superior stability and discriminative accuracy across diverse scenarios of entrepreneurship funding, team size, disciplinary background, and time evolution, demonstrating particularly strong adaptability for non-technical projects and large-scale teams. By applying TreeExplainer for attribution analysis, the model transitions from "black-box prediction" to "explainable decision-making", revealing the direction and marginal contribution of key risk factors. This provides visual support for university entrepreneurship education interventions and the precise allocation of incubation resources. Cross-model comparisons show that the proposed method outperforms traditional methods such as XGBoost, random forest, support vector machine, and LightGBM in multiple metrics, including AUC, F1-score, Brier score, and prediction consistency. It also maintains high recall and logical coherence during long-term dynamic monitoring. This research not only improves the technical effectiveness of risk identification but also, through its interpretability enhancement mechanism, promotes the transformation of entrepreneurial support systems from experience-driven to data-cognition-driven, a shift that holds significant practical value for modernizing decision-making and operational risk management within the industry.

AUTHOR CONTRIBUTIONS

Ming Guo designed, collected and analyzed the data, and drafted the manuscript. Ming Guo conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Ming Guo participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by Innovation and Practice of Entrepreneurship and Employment Guidance for Art Major Students under the background of Coordination between Schools and Enterprises.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] G. Zichella and T. Reichstein, "Students of entrepreneurship: Sorting, risk behaviour and implications for entrepreneurship programmes," Management Learning, vol. 54, no. 5, pp. 727-752, 2023, doi: 10.1177/13505076221101516.

[2] Q. Zhou, "Research on the problems and countermeasures of the cultivation of adult college students' innovation and entrepreneurship ability in the internet era," Open Access Library Journal, vol. 8, no. 7, pp. 1-12, 2021, doi: 10.4236/oalib.1107718.

[3] F. Bian, C. H. Wu, L. Meng, and S. B. Tsai, "A study on the relationship between entrepreneurship education and entrepreneurial intention," International Journal of Technology, Policy and Management, vol. 21, no. 1, pp. 1-19, 2021, doi: 10.1504/IJTPM.2021.114306.

[4] M. Tas, H. B. Saydaliev, and S. Kadyrov, "Impact of collaborative, entrepreneurship education on the financial risk taken by university students," Industry and Higher Education, vol. 36, no. 5, pp. 595-603, 2022, doi: 10.1177/09504222211068294.

[5] E. Gurel, M. Madanoglu, and L. Altinay, "Gender, risk-taking and entrepreneurial intentions: assessing the impact of higher education longitudinally," Education + Training, vol. 63, no. 5, pp. 777-792, 2021, doi: 10.1108/ET-08-2019-0190.

[6] C. Bandera, S. C. Santos, and E. W. Liguori, "The dark side of entrepreneurship education: A Delphi study on dangers and unintended consequences," Entrepreneurship Education and Pedagogy, vol. 4, no. 4, pp. 609-636, 2021, doi: 10.1177/2515127420944592.

[7] Y. Su and Z. Liu, "A study of ChatGPT empowering college students' innovation and entrepreneurship education in the context of China," International Journal of New Developments in Education, vol. 5, no. 13, pp. 1-7, 2023, doi: 10.25236/IJNDE.2023.051301.

[8] J. J. Cater, M. Young, M. Al-Shammari, and K. James, "Re-exploring entrepreneurial intentions and personality attributes during a pandemic," Journal of International Education in Business, vol. 15, no. 2, pp. 311-330, 2022, doi: 10.1108/JIEB-04-2021-0050.

[9] B. Utami, M. M. Ibrahim, V. Violin, D. E. Subroto, D. Destari, and E. D. Astuti, "The role of entrepreneurship education, attitude toward risk and digital literacy on entrepreneurial intention among gen Z," Journal of Social Science and Business Studies, vol. 3, no. 1, pp. 407-411, 2025, doi: 10.61487/jssbs.v3i1.136.

[10] M. Fanea-Ivanovici and H. Baber, "The role of entrepreneurial intentions, perceived risk and perceived trust in crowdfunding intentions," Engineering Economics, vol. 32, no. 5, pp. 433-445, 2021, doi: 10.5755/j01.ee.32.5.29300.

[11] K. B. Kahn, "Innovation is not entrepreneurship, nor vice versa," Journal of Product Innovation Management, vol. 39, no. 4, pp. 467-473, 2022, doi: 10.1111/jpim.12628.

[12] K. Krichen and H. Chaabouni, "Entrepreneurial intention of academic students in the time of COVID-19 pandemic," Journal of Small Business and Enterprise Development, vol. 29, no. 1, pp. 106-126, 2022, doi: 10.1108/JSBED-03-2021-0110.

[13] J. R. Anderson, "The role of subjective norms in developing entrepreneurial intentions in university students," Journal of Strategy and Management, vol. 16, no. 4, pp. 643-653, 2023, doi: 10.1108/JSMA-10-2022-0190.

[14] T. T. A. Ngo, P. T. Truong, Y. L. Tran, V. Tran, N. T. Tran, A. M. Huynh, et al., "Factors affecting entrepreneurial intention of college students: An empirical study from Vietnam," The Journal of Asian Finance, Economics and Business, vol. 9, no. 5, pp. 147-156, 2022, doi: 10.13106/jafeb.2022.vol9.no5.0147.

[15] A. Levy and R. Baha, "Credit risk assessment: a comparison of the performances of the linear discriminant analysis and the logistic regression," International Journal of Entrepreneurship and Small Business, vol. 42, no. 1-2, pp. 169-186, 2021, doi: 10.1504/IJESB.2021.112265.

[16] W. Wang and M. J. Guedes, "Firm failure prediction for small and medium-sized enterprises and new ventures," Review of Managerial Science, vol. 19, no. 7, pp. 1949-1982, 2025, doi: 10.1007/s11846-024-00742-4.

[17] M. S. Uddin, G. Chi, M. A. M. Al Janabi, and T. Habib, "Leveraging random forest in micro-enterprises credit risk modelling for accuracy and interpretability," International Journal of Finance & Economics, vol. 27, no. 3, pp. 3713-3729, 2022, doi: 10.1002/ijfe.2346.

[18] Z. Zhu and Y. Zhang, "Flood disaster risk assessment based on random forest algorithm," Neural Computing and Applications, vol. 34, no. 5, pp. 3443-3455, 2022, doi: 10.1007/s00521-021-05757-6.

[19] Y. K. Saheed, M. O. Arowolo, and A. U. Tosho, "An efficient hybridization of k-means and genetic algorithm based on support vector machine for cyber intrusion detection system," International Journal on Electrical Engineering and Informatics, vol. 14, no. 2, pp. 426-442, 2022, doi: 10.15676/ijeei.2022.14.2.11.

[20] M. Arunkumar and K. A. Kumar, "GOSVM: Gannet optimization based support vector machine for malicious attack detection in cloud environment," International Journal of Information Technology, vol. 15, no. 3, pp. 1653-1660, 2023, doi: 10.1007/s41870-023-01192-z.

[21] X. Li, J. Wang, and C. Yang, "Risk prediction in financial management of listed companies based on optimized BP neural network under digital economy," Neural Computing and Applications, vol. 35, no. 3, pp. 2045-2058, 2023, doi: 10.1007/s00521-022-07377-0.

[22] M. Roba and O. K. Moulay, "Risk management in using artificial neural networks," SocioEconomic Challenges, vol. 8, no. 2, pp. 302-313, 2024, doi: 10.61093/sec.8(2).302-313.2024.

[23] N. Luo, H. Yu, Z. You, Y. Li, T. Zhou, Y. Jiao, et al., "Fuzzy logic and neural network-based risk assessment model for import and export enterprises: A review," Journal of Data Science and Intelligent Systems, vol. 1, no. 1, pp. 2-11, 2023, doi: 10.47852/bonviewJDSIS32021078.

[24] S. Clark, X. Zhu, Z. Wang, R. Mehta, J. Blake, and X. Wei, "Construction of a Supply Chain Credit Risk Evaluation Model for Manufacturing Enterprises Using XGBoost," International Journal of Management Science Research, vol. 8, no. 5, pp. 1-8, 2025, doi: 10.53469/ijomsr.2025.08(05).01.

[25] M. Hernes, J. Adaszyński, and P. Tutak, "Credit risk modeling using interpreted XGBoost," European Management Studies, vol. 21, no. 3, pp. 46-70, 2023.

[26] S. Ergün, "Explaining XGBoost predictions with SHAP value: A comprehensive guide to interpreting decision tree-based models," New Trends in Computer Sciences, vol. 1, no. 1, pp. 19–31-19–31, 2023, doi: 10.3846/ntcs.2023.17901.

[27] M. Rzychoń, A. Żogała, and L. Rog, "SHAP-based interpretation of an XGBoost model in the prediction of grindability of coals and their blends," International Journal of Coal Preparation and Utilization, vol. 42, no. 11, pp. 3348-3368, 2022, doi: 10.1080/19392699.2021.1959324.