Business Failure Prediction Based on a Cost-Sensitive Extreme Gradient Boosting Machine

YAO ZOU\textsuperscript{1}, CHANGCHUN GAO\textsuperscript{1}, AND HAN GAO\textsuperscript{2}

\textsuperscript{1}Glorious Sun School of Business and Management, Donghua University, Shanghai 200051, China
\textsuperscript{2}College of Fashion and Design, Donghua University, Shanghai 200051, China

Corresponding author: Yao Zou (zouyaodhu@mail.dhu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 71874027, in part by the Shanghai Planning Office of Philosophy and Social Science under Grant 2021ECK001, and in part by the Shanghai Sailing Program under Grant 21YF1415900.

ABSTRACT Business failure prediction is very important for the sustainable development of enterprises. Machine learning algorithms, especially ensemble algorithms, have shown great economic benefits in enterprise financial early warning. However, the highly imbalanced class distribution of financial risk data and the unexplainable of most machine learning-based early distress warning models limit their commercial application. To address the above limitations, we enhance the business failure prediction performance by tree-ensemble in a boosting manner. Moreover, to solve the class imbalanced issue in business failure datasets, a weighted objective function, weighted cross-entropy, is embedded into the boosted tree framework, making the weighted XGBoost a cost-sensitive business failure prediction model. Besides, to tackle the second issue, we explore the intrinsic interpretability of the proposed method by visualizing the feature importance and incorporating a partial dependence plot technique to locally interpret the individual business failure event. Experimental results on business failure datasets with different predictive horizons collected from China Security Market Accounting Research (CSMAR) database show the proposed weighted XGBoost is a good solution to reduce the error on recognizing firms in business failure. Furthermore, the visualized feature importance score and partial dependence plot result both demonstrate that the cost-sensitive tree-based ensemble can be a good tool to guide the investors in making rational as well as provide interpretable business prediction results as a reference for the policy-making of the regulators.

INDEX TERMS Business failure prediction, XGBoost, cost-sensitive, interpretability.

I. INTRODUCTION

Early business failure warning is an essential aspect of financial risk preventive management. An efficient financial risk early warning system can provide timely alarms to the managers of a firm, preventing the company from going bankrupt. The implementation of financial risk warnings not only allows for improving asset allocation efficiency but also serves as critical support in ensuring the company’s financial stability. Furthermore, the frequency of business bankruptcies is a significant element that impacts a country’s economy and may even be used to predict if a financial crisis will occur. The strong connections between the phenomenon that many firms are experiencing bankruptcy and the economic growth make risk managers more aware of the importance of bankruptcy risk prevention and control.

As the global economy evolves, bankruptcy forecasting, the goal of which is to analyze a company’s current financial position and growth potential through its financial data, is playing an increasingly important role in the economic development process. Moreover, the update of the Basel Accord further highlights the importance of developing an accurate financial risk warning model. Predictive models for business failure prediction (BFP) can be divided into two types: statistical methods and artificial intelligence algorithms. Traditional statistical models such as multivariate discriminant analysis (MDA) [1], linear regression, logistic regression (LR) were the mainstream solutions in early-stage BFP studies. The arrival of the information age has accelerated the development of financial risk early
warning models, which have progressed from statistical-based methodologies to smarter and more accurate artificial intelligence-based algorithms. Popularized ML-based BFP models includes support vector machine (SVM) [2]–[4], DT [5], [6], KNN [7], [8], naive Bayesian (NB) [9], neural networks (NNs) [10], [11]. Ensemble learning algorithms, which is a collection of solutions that aggregates multiple weak hypotheses into a stronger one, are considered as the mainstream ML-based BFP models [12], [13]. A mountain of evidence has shown the promise of ensemble-based BFP models. Boosting ensemble approaches such as GBDT [14] are considered a popular solution for BFP.

Although it has been proved to be an effective solution to reduce the early business failure warning error through ensemble techniques, most of the early business failure warning models ignore the highly imbalanced distribution of the samples in business failure datasets. For example, in the work of [15], the number of special treatment (ST) samples and the number of non-special treatment (NST) samples are balanced by pairing the ST samples and the NST samples for training and testing. However, there is a extremely imbalance in the real-world bankruptcy dataset, where the ratio of the number of normally operating firms to the number of bankrupt firms is usually close to 90% : 10% or even higher. Several works implemented imbalanced corporate BFP by modifying the distribution of ST and NST samples in business failure datasets, and these methods aim to balance business failure datasets by resampling. These techniques include resampling-based methods such as undersampling [16] and SMOTE [17], [18]. However, the imbalanced BFP based on re-sampling has several inevitable limitations. For instance, undersampling removes irrelevant majority class samples may lead to loss of valuable failure information; oversampling techniques that copy or synthesize minority class samples may lead to the variance of bankruptcy financial ratios.

Based on the above considerations, in this study, we propose a weighted XGBoost (XGBoost-W) to realize cost-sensitive BFP. The proposed method considers XGBoost as the base framework, ensuring the robust optimization of BFP error. Moreover, to tackle the class imbalance issue, a weighted cross-entropy that takes into account the different costs of misclassifying ST firms and NST firms, is introduced into XGBoost to realize cost-sensitive BFP. Besides, to reveal the contribution of financial ratios to the prediction of XGBoost-W and its marginal effect, feature importance scores, and partial dependence plot tools are incorporated to interpret the BFP results.

II. RELATED WORKS
A. ENSEMBLE LEARNING ALGORITHMS FOR BUSINESS FAILURE PREDICTION
This study contributes to the literature of ensemble approaches for BFP. In this context, [19] first developed a Z-score model to predict the bankruptcy of corporates, which was established based on multivariate discriminant analysis by selecting representative financial ratios. Subsequently, [20] introduced logistic regression to estimate the probability of bankruptcy of enterprises. The development of machine learning (ML) technology and artificial intelligence theories have witnessed the boom of BFP models. Many ML-based BFP models have been broadly developed, these studies include SVM [21], [22], DT [5], [14], Naive Bayesian, NN [23], [24], k-nearest neighbors.

In recent years, to pursue accurate BFP, ensemble models, which seek to implement enhance the performance of BFP by combining individual learners to a stronger one, have been warmly discussed [25], [26]. [27] introduced a time weighting mechanism to address the concept drift issue in the BFP task. Besides, AdaBoost and SVM were combined into an ensemble framework to further improve the performance of dynamic BFP. [16] developed a heterogeneous ensemble method and incorporated 5 feature selection approaches to reduce the error of BFP. [28] proposed an adaptive boosting algorithm to predict the contractor’s bankruptcy in Korea. Their experimental results showed the advancement of the boosting ensemble algorithm compared with a NN, decision tree (DT), and SVM. [29] realized the prediction of financial crises of a contractor two and three years before business failure in South Korea from 2007 to 2012 by developing a voting-based ensemble method. In [30]’s work, a soft ensemble model is proposed, which expert system methods and convolutional neural network to recognize the status of an enterprise. [31] considered the random subspace method as an ensemble strategy to extract sentiment and textual information for business failure prediction. Ensemble learning is the act of creating a collection of individual learners and then merging them using an integration technique to minimize individual learners’ mistakes. According to the ensemble strategy, ensemble BFP algorithms can be divided into Bagging ensemble algorithms and boosting ensemble approaches [32]. Boosting ensembles are a collection of algorithms that boosted BFP performance in a sequential ensemble manner. Benefit from the boosting framework, [33] applied the AdaBoost to predict the financial insolvency risk of European companies and showed that the proposed boosting ensemble method can reduce the generalization error by about 30% compared with NN-based BFP models. [34] incorporated unstructured data such as company audit reports and management reports into BFPs, their experimental results demonstrated extreme gradient boosting (XGBoost) is the optimal BFP model among ensemble-based BFP approaches. [35] improved the BFP performance by considering extreme gradient boosting as an ensemble of DTs, and introduced a synthetic feature generation strategy to reflect the high-order statistical characteristics of bankruptcy data. [36] studied the effect of the distributions of positive samples and negative samples on the predictive performance of ensemble models that ensembled MLP, KNN, and C4.5 DT on credit scoring domain and bankruptcy prediction task. [37] introduced a new
conception that named as “financial profile” based on the knowledge of bankruptcy governance. Their results proved the profiles-based bankruptcy prediction models outperform individual ML-based classifiers and ensemble algorithms. The performance improvement has been highlighted by many other works that focus on developing ensemble models, these studies include [16], [27], [29], [38], [39].

### B. IMBALANCED BUSINESS FAILURE PREDICTION

In financial risk prediction challenges, such as credit scoring, when failing clients make up a tiny fraction of all customer groups. In the task of BFP, failure events are scarce cases compared with normal operating firms. The asymmetric distribution of positive samples and negative samples leads to class imbalance is a prevalent issue in the financial risk prediction domain. The class imbalance issue can be briefly described as the samples of different categories being unequally distributed. In a business failure case, the class of NST firms is the dominant class whose number is significantly larger than that class of ST firms. However, in a BFP task, the minority class is more concerned since misclassifying a NST firm takes a higher cost than misclassifying a ST firm. This makes imbalanced BFP a challenge in this study.

Imbalanced BFP can be achieved in two directions: resampling strategies and cost-sensitive solutions. Resampling strategies address the class imbalance issue of the business failure dataset by modifying the raw distribution of firms that are in business failure and enterprises that are in normal operation. [18] performed a synthetic minority over-sampling technique (SMOTE) to balance the distribution of business failure dataset, and combined it with an ensemble framework to realize imbalanced BFP. [40] proposed a KNN-based oversampling strategy to generate minority class samples for Chinese tourism business failure warning. [41] combined clustering-based under-sampling strategy and a boosting ensemble framework to BFP, and provided reference value for fraud detection, credit scoring and other fields/domains. [42] incorporated SMOTE algorithm into a Geometric mean-aware boosting framework to alleviate the class imbalance problem in the bankruptcy prediction task. In their study, different class imbalance rates of bankruptcy datasets are investigated and the results showed the effectiveness of the hybrid imbalanced learning approach. The above works focus on the early warning of imbalanced business failure based on resampling strategy. However, as stated by [42], oversampling operations, which balance the skew distribution by copying or synthesizing minority class samples may result in the over-fitting issue while increasing the learning from the over-sampled training set. Under-sampling strategies, which implement imbalanced BFP by removing majority class samples may lead to the loss of valuable information of majority class samples, further trigger the under-fitting on majority class samples.

Based on the above considerations, some efforts have been contributed to the imbalance of BFP tasks by taking the different costs of misclassifying a NST firm and a ST firm into account, these methods are cost-sensitive approaches. [43] designed an improved AdaBoost, named AsymBoost that introduced cost-sensitive learning conception into a boosting framework to transform the optimization objective from maximizing predictive accuracy to minimizing the total misclassification cost. A cost-sensitive SVM is proposed in [44]’s work, which assigned a higher weight to the misclassified majority class samples, thus modifying the decision boundary and the optimal classification hyperplane. To overcome the problem that the misclassification cost matrix is known and fixed for model training and evaluation in the previous literature, a heterogeneously integrated selection and multi-objective optimization framework are used for cost-sensitive BFP to adapt to the uncertainty of misclassification cost. To overcome the issue that traditional cost-sensitive BFP implemented based on the assumption that misclassification cost matrix is known and fixed, [45] proposed a heterogeneous ensemble selection framework and combined it with a multi-objective optimization algorithm for cost-sensitive BFP, making it adapt to the circumstances when the misclassification cost is uncertain. [46] established an example-dependent cost-sensitive logistic regression algorithm to reduce the misclassification error on minority class samples, thus further saving 10% misclassification cost on a vehicle dataset from a European manufacturer. [47] assigned different costs to each example and proposed an example-dependent cost-sensitive DT to finance-related fields such as credit scoring, fraud detection, and direct marketing, which maximized the cost savings.

In this study, we proposed a weighted XGBoost (XGBoost-W) to realize cost-sensitive BFP. XGBoost-W considers XGBoost as a robust training framework, which minimizes the prediction error between the estimated probability of bankruptcy of firms and the real operating status of listed firms. To alleviate the class imbalance issue, a weighted objective loss function is introduced into the boosting framework, enabling the cost-insensitive XGBoost into a cost-sensitive version. Moreover, two interpretation mechanisms are introduced into this work to explore which factor is the main driving force to the prediction of a cost-sensitive BFP model, and to investigate the marginal effect of financial ratios on the decision results. Different from [48]’s work that realized imbalanced P2P loan evaluation by manipulating predictive threshold in a cost-sensitive manner, we modify the training pattern of XGBoost in the objective, driving the BFP based on XGBoost-W towards skew-insensitive fashion. Moreover, we associate the misjudgment costs of risky and risk-free firms with the training stage into the optimization objective, making XGBoost-W distinct from the similar work of [49] that implements cost-sensitive URL detection by minimizing the overall misclassification error.

The above studies motivate our contributions are: (1) XGBoost, which is an ensemble of DTs in a boosting manner, is considered as the base framework for BFP, providing the possibility to explore the interpretability of a BFP model;
(2) a weighted cross-entropy is embedded into a boosting framework, enables the proposed boosted tree a cost-aware solution for imbalance BFP; (3) The tree-based cost-sensitive framework allow us further interpret the predictions of XGBoost-W, providing an explicit answer for decision-makers what elements should be taken into account to measure the business failure in the decision-making process, thus guiding investors making rational investments [50].

III. METHODOLOGY

A. EXTREME GRADIENT BOOSTING

XGBoost is a new gradient boosting algorithm that was proposed by [51], which has been popularized by ML researchers and players of ML competitions because of its efficient parallel training and significant improvement of ML-based applications. XGBoost is a variant of GDBT while GBDT is an ensemble approach that combined gradient boosting optimization strategy and DT classifiers. That is, multiple DTs are combined into a gradient boosting framework to iteratively optimize the training target. Given a training business failure dataset \( \{(x_i, y_i)\}_{i=1}^{N} \), where \( N \) is the number of training samples, \( x_i \) denotes the features of \( i \)-th financial sample, \( y_i \) represents the label of \( i \)-th samples. GBDT ensembles multiple DTs in an additive manner, which is:

\[
F_{M}(x) = \sum_{m=1}^{M} \alpha_{m} f_{m}(x),
\]

(1)

\( F_{M}(x) \) denotes GBDT; \( M \) is the number of DTs, \( f_{m}(x) \) represents \( m \)-th base learners, \( \alpha_{m} \) control the weight for \( m \)-th base learner. To reduce the complexity of optimizing a boosted tree, GBDT employs a forward stagewise algorithm to get the solution of its ensembled structure. Therefore, the additive ensemble manner of GBDT can be re-expressed as:

\[
\hat{y}_{i}^{(m)} = \sum_{m=1}^{M} f_{m}(x_i) = \hat{y}_{i}^{(m-1)} + f_{m}(x_i), f_{m} \in F, \quad i \in N,
\]

(2)

where \( \hat{y}_{i}^{(m)} \) is the prediction that is generated by GBDT. ML algorithms realize BFP by ensemble minimizing the empirical risk, which in XGBoost, can be defined as:

\[
L^{(m)} = \sum_{i=1}^{N} l(y_i, \hat{y}_{i}^{(m)}) + \Omega(f_m),
\]

(3)

\( l \) measures the loss between the real label \( y_i \) and its prediction \( \hat{y}_{i}^{(m)} \). To alleviate the over-fitting problem, XGBoost adds a regularization term \( \sum_{k=1}^{K} \Omega_{k}(f_k) \) for loss function to control the complexity of each classification and regression tree (CART), of which \( \Omega(f_m) \) can be specifically denoted as \( \Omega(f_m) = \gamma T + \frac{1}{2} \lambda \| w \|^{2} \). \( T \) is the number of leaf nodes, \( w \) represents the score at the leaf node, \( \gamma, \lambda \) are the regularization coefficients.

Combined with equation (2) and equation (3), the loss function of XGBoost can be re-expressed as:

\[
L^{(m)} = \sum_{i=1}^{n} l(y_i, \hat{y}_{i}^{(m-1)} + f_{m}(x_i)) + \Omega(f_m).
\]

(4)

To solve equation (4), we approximate the loss function in by performing a Taylor expansion in the function space, and get:

\[
L^{(m)} = \sum_{i=1}^{n} \left[ l(y_i, \hat{y}_{i}^{(m-1)}) + g_i f_{m}(x_i) + \frac{1}{2} h_i f_{m}^{2}(x_i) \right] + \Omega(f_m)
\]

\[
= \sum_{i=1}^{n} \left[ g_i f_{m}(x_i) + \frac{1}{2} h_i f_{m}^{2}(x_i) \right] + \gamma T + \frac{1}{2} \lambda w^{2} + C,
\]

(5)

where \( g_i = \frac{\partial l(y_i, \hat{y}_{i}^{(m-1)})}{\partial \hat{y}_{i}^{(m-1)}}, h_i \) represents the second derivative of the loss function, \( T \) is the number of lead nodes. XGBoost real-izes performance-boosting by iteratively adding new DTs to the previous ensemble model, and each DT accomplishes the fitting of the residual generated by the previous ensemble model. Let \( I_{j} = \{ i | q_{j}(x_i) = j \} \) define the set of the samples at leaf nodes, we further derive equation (5) to get the optimal score \( w_{i}^{*} \) at leaf node \( i \), where \( w_{i}^{*} = -\frac{g_{i}}{h_{i}+\lambda} \) and \( H_{i} = \sum_{j \in I_{j}} h_{i} \). Taking the optimal leaf score into equation (5), the minimized loss function can be obtained:

\[
L^{(m)} = -\frac{1}{2} \sum_{j=1}^{T} H_{j} \frac{G_{j}^{2}}{H_{j} + \lambda} + \gamma T + C.
\]

(6)

Compared with classical GBDT algorithm that adopts mean square error as the splitting to determine the structure of each DT in GBDT algorithm, XGBoost employs a greedy algorithm to grow a DT. Let \( L_{L} \) and \( L_{R} \) be the sample set at the left node and the right node after splitting, \( I = I_{L} \cup I_{R} \), the information gain after node splitting can be computed:

\[
Gain = \frac{1}{2} \left[ \frac{G_{L}^{2}}{H_{L} + \lambda} + \frac{G_{R}^{2}}{H_{R} + \lambda} + \frac{(G_{L} - G_{R})^{2}}{H_{L} + H_{R} + \lambda} \right] - \gamma,
\]

(7)

where \( G_{L} = \sum_{i \in I_{L}} g_{i}, H_{L} = \sum_{i \in I_{L}} h_{i} \) sums the first derivative and the second derivative for the left subtree, \( G_{R} = \sum_{i \in I_{R}} g_{i}, H_{R} = \sum_{i \in I_{R}} h_{i} \) sums the first and the second derivative for the right subtree.

B. WEIGHTED CROSS-ENTROPY FOR XGBoost

A classical XGBoost implements BFP by minimizing the expected risk:

\[
R(i|x) = \sum_{j=1}^{C_{i,j}} C_{i,j} p(j|x),
\]

(8)

where \( p(j|x) \) is a posterior probability estimated from empirical training set \( X = \{(x_{n}, y_{n})\}_{n=1}^{N} \); \( C_{i,j} \) represents the
cost of predicting \( k \) into class \( j \). According to equation (8), the prediction can be determined by minimizing the expected loss function, which is:

\[
\hat{y}^* = \arg \min_{\hat{y}} \mathbb{E}[R(\hat{y}|x)]= \arg \min_{\hat{y}} \mathbb{E}[E_{X,Y}[c_{i,j}]].
\] (9)

\( X \) and \( Y \) denote the input space and output space of the XGBoost-based BFP model. Due to \( p_{ij}(x) \) is a posterior probability that can be estimated from the empirical training samples, we define the empirical risk as:

\[
R_{ij} = E_{X,Y}[l] = \frac{1}{N} \sum_{n=1}^{N} l(p_{n}, y_n, c).\] (10)

where \( p_n \) is the predicted probability of classifying \( n \)-th business failure sample as a ST firm. \( l() \) is a loss function that optimizes the misclassification error of a BFP task. In the original implementation of XGBoost, a cost-insensitive loss function, cross-entropy is introduced, which makes equation (9) as:

\[
l_c(p, y) = -y \log p.\] (11)

Combining equation (9) and equation (10), it can be seen a cost-insensitive XGBoost implements BFP based on \( c_{1,0} = 0, c_{0,1} = 1 \), where \( c_{1,0} \) represents the cost of misclassifying a ST firm into NST firm, \( c_{1,0} \) is the cost of misclassifying a NST firm into a ST firm. In a practical BFP task, it is widely acknowledged that misclassifying a ST firm takes a higher cost than misclassifying a NST firm, which is \( c_{1,0} > c_{0,1} \). The implementation \( c_{1,0} > c_{0,1} \) is equal to transforming the cost-insensitive XGBoost into a cost-sensitive version:

\[
l_{wc}(p, y, c) = \log p + w_{0,1}(1 - y) \log (1 - p),\] (12)

where \( w_{1,0} \) represents the weights of misclassified ST samples, \( w_{0,1} \) represent the weights of misclassified NST samples.

**C. INTERPRETATION OF XGBoost-W**

1) **FEATURE IMPORTANCE**

Feature Importance (FI) measures the effect of a given feature on the predicted response. It specifies the extent to which each feature contributes to the predicted outcome of the entire dataset. The higher a feature importance score indicates the greater contribution of the feature. Business failure forecasting can be modeled as a binary classification problem. In the implemented weighted XGBoost, XGBoost-W integrates multiple DTs in a boosting way, and each DT is grown by recursively node splitting where feature selection based on impurity reduction criterion is performed. XGBoost-W is a cost-sensitive version of XGBoost that retains the gradient boost optimization pattern of XGBoost. This provides the possibility to further visualize the feature importance score to explore the global interpretability of XGBoost-W. Therefore, in a boosted tree that composed of \( M \) DTs, the importance score of the \( j \)-th feature can be computed by averaging the feature importance over \( M \) trees:

\[
I_{j}^2 = \frac{1}{M} \sum_{m=1}^{M} I_{m}^j(T_m).\] (13)

where \( I_{j}^2 \) denotes the importance score of \( j \)-th feature based on XGBoost-W. \( I_{j}^2(T_m) \) computes the importance score of \( j \)-th feature on \( m \)-th DT \( T_m \), which can be specifically expressed as:

\[
I_{j}^2(T_m) = \sum_{k=1}^{K} \Delta_k^j l(v(k) = j),\] (14)

where \( K \) is the number of leaf nodes on \( m \)-th DT, \( l(v(k) = j) \) judges whether the selected splitting feature on \( k \)-th split node is \( j \)-th feature. \( \Delta_k^j \) represents the impurity reduction after node splitting with \( k \)-th splitting feature. In a classification task, CART considers the Gini index as the impurity criterion while in a regression task, a square error is considered as the impurity criterion.

**D. PARTIAL DEPENDENCE PLOT**

A partial dependence plot (PDP) shows the marginal impact of a given attribute on the prediction results of a machine learning model. PDP is a tool that visualizes how a given variable affects the prediction (linear, monotonic, or more complex), while feature importance score only determines which the contributions of business failure feature to the prediction [52]. Although the degree of feature contribution intuitively shows the impact of each feature on the prediction results, this interpretable mechanism cannot well provide the marginal contribution of each feature to the prediction results, nor can it well answer how each feature affects the input response. PDP provides the possibility to solve the above limitations. In the BFP task, PDP visualizes the marginal impact of each financial variable on the BFP results, which answers how the given variable (linear and nonlinear) affects the prediction of the model by averaging the predicted value of the model under the given attribute.

In this study, we employ a weighted XGBoost to realize cost-sensitive BFP. Given a feature \( x_i \), the local dependency function of a business failure predictor \( F \) under a given feature can be expressed as:

\[
F_j(x_j) = \mathbb{E}_{S \setminus j} [F(x_j, x_{S\setminus j})] = \int F(x_j, x_{S\setminus j})dP(x_{S\setminus j}).\] (15)

where \( S \) represents the complete feature set, \( S \setminus j \) denotes the feature subset that excludes \( j \)-th feature. \( P(x_{S\setminus j}) \) expressed as the marginal probability density of \( x_{S\setminus j} \).

Specifically, the above formula can be further estimated from the empirical training data:

\[
\hat{F}_j(x_j) = \frac{1}{N} \sum_{i=1}^{N} F(x_j, x_i, S\setminus j),\] (16)

\( N \) is the number of samples in a training set, \( x_i, S\setminus j \) is a feature vector of \( i \)-th sample that composed by feature subset \( S \setminus j \).
IV. EXPERIMENTAL SETTINGS

A. BUSINESS FAILURE PREDICTION DATASETS

The business failure of a company usually refers to the situation that the operating cash flow of a company cannot supersede the negative net assets of the firm [5]. In empirical studies, most scholars equate bankrupt enterprises with financially troubled enterprises. However, Altman distinguished bankruptcy from concepts such as business failure, insolvency, liquidation, and default, arguing that bankruptcy favors the “legal meaning” of business failure but weakens the “economic meaning” of business failure [53]. In reality, bankruptcy is a lengthy legal process, there can be a significant time lag between official insolvency and financial failure. A company in business failure will usually file for bankruptcy only when the business failure is severe enough, and the relevant authorities will review the bankruptcy petition to decide whether to approve it. Based on the analysis mentioned above and the imperfection of China’s bankruptcy law, we use companies that have been listed as ST by the CSMAR for two consecutive years of losses as a sample of business failure companies. According to the stock listing rules issued by the Shanghai Stock Exchange and the Shenzhen Stock Exchange, the stock transactions of listed companies with abnormal financial or other conditions are subject to “special treatment”. These stocks labeled as ST indicate a serious deterioration in the company’s financial situation. Moreover, given such a mechanism, it is useless for one to predict a company’s business failure with data collected from the current year when it receives the ST label. According to the study of [54], the financial ratios are well-differentiated on 1-year and 3-year horizon, making the model insensitive to short-term changes caused by the economic environment, thus, in this study, we implemented the models based on financial data obtained 1, 2, 3, and 4 years before the companies received the ST label. A 4-year predictive horizon means that we used the financial data of the year T-4 to predict whether the firm was labeled as ST in year T. Based on the previous studies, this paper collects a total of 47 financial indicators to predict business failure. 2826 listed companies are selected as the experimental instances from the Shenzhen Stock Exchange and the Shanghai Stock Exchange of China, the distributions of 4 predictive horizons are summarized in Tab. 1.

To collect financial features, 47 accounting ratios are selected from the CSMAR Database. Tab. 2 shows the concrete definitions of 47 accounting ratios in each aspect of solvency, operational capabilities, profitability, business development capacity, index per share, as well as relative value index.

B. BUSINESS FAILURE PREDICTION MODELS

In this study, to verify the effectiveness of BFP, we select multiple classical BFP models that include statistical-based BFP approaches LDA and LR, individual ML-based classifier DT, and its ensembles such as Bagging-based ensemble method RF and boosting ensemble algorithms such as AdaBoost, GBDT, LightGBM [55], XGBoost [56], for comparison. Moreover, to further validate the advantages of cost-sensitive BFP algorithms, we further conduct an empirical comparison among imbalanced BFP models. These models include resampling-based strategies such as random oversampling (ROS), random undersampling (RUS), SMOTE, cost-sensitive approaches including AdaCost [57], MetaCost [58], cost-sensitive boosted trees, hybrid methods such as RUSBoost [59] and SMOTEBest [60].

C. EVALUATION METRICS

To evaluate the performance of imbalanced BFP, we select several broadly utilized BFP metrics, which include AUC, Type-I error, Type-II error, and Gmean. AUC is a metric that measures the area under the receiver operation curve (ROC), while ROC is a graphical metric whose x-axis represents false-positive rate (FPR) and its y-axis is the true-positive rate (TPR). A ML-based BFP models predict the ST and NST samples into 4 groups according to the confusion matrix that is shown in Tab. 4: TP, FP, TN, FN. TP is the number of ST samples that are accurately predicted; FP denotes the number of NST samples that are wrongly predicted; TN counts the number of NST samples that are correctly classified; FN is the number of ST samples that are incorrectly classified. Therefore, FPR, also described as Type-I accuracy denotes the proportion of accurately predicted NST samples, TPR, also known as Type-II accuracy, calculates the proportion of accurately predicted ST samples.

Type-I error measures the proportion of the misclassified NST samples; which is defined as:

\[ e_1 = \frac{TN}{TN + FP}. \]  

Type-II error computes the proportion of the misclassified ST samples, which can be calculated by:

\[ e_2 = \frac{TN}{TP + FN}. \]  

Gmean is a comprehensive metric that combines both Type-I accuracy and Type-II accuracy:

\[ Gmean = \sqrt{TPR \times FPR}. \]

D. SIGNIFICANCE TEST

Different from error-minimization-based optimization BFP models, in this study, a cost-minimization-based approach, weighted XGBoost is proposed for cost-sensitive corporates BFP. To get the optimal BFP framework, we first finetune

| Period | sample | Number of firms | Ratio |
|--------|--------|----------------|-------|
|        |        | Non-fail | Fail |       |
| T-1    | 36181  | 34511    | 1670 | 95.38%: 4.62% |
| T-2    | 30027  | 28545    | 1482 | 95.06%: 4.94% |
| T-3    | 29007  | 27413    | 1594 | 94.55%: 5.45% |
| T-4    | 26043  | 24537    | 1506 | 94.21%: 5.78% |
By finetuning the hyperparameter of XGBoost, we determine the optimal XGBoost that minimizes the error between predictions of XGBoost and labels. Based on the finetuned robust XGBoost, we finetune to transform the XGBoost into a weighted version, giving XGBoost the cost-sensitive learning power. In this study, $\alpha$ is an important parameter that transforms the XGBoost into a cost-sensitive fashion while the costs of misclassifying a ST firm and a NST firm are not observed due to the privacy of financial information of a listed company. We determine the cost ratio that represents the cost of misclassifying a ST firm to the cost of misclassifying a NST firm by finetuning $\alpha$ while $r = \frac{\alpha}{1 - \alpha}$. We first finetune $\alpha$ from the interval [0.7, 1] with searching stride 0.01, next we perform a further search from the searching space [0.99, 0.999] with searching stride 0.001. By finetuning the hyperparameter to determine the optimal cost ratio $r$.

V. EXPERIMENTAL RESULTS

A. MISCLASSIFICATION ERROR MINIMIZATION VS XGBoost-W

To test the effectiveness of embedding cost-sensitive objective function into the XGBoost framework for imbalanced BFP, a comprehensive comparison among classical BFP approaches and XGBoost-W are investigated. The classical BFP approaches cover statistical-based methods, ML-based individual predictors, and ensemble algorithms. Tab. 5 shows the performance comparison of XGBoost-W and classical error-minimization-based BFP models. As can be viewed from Tab. 5, XGBoost-W gets the optimal score on Type-II error and Gmean, compared with error-minimization-based BFP models, the Type-II error achieved by XGBoost-W has been significantly reduced. Gmean is a comprehensive metric that is computed based on Type-I error and Type-II error. XGBoost-W and LightGBM-W improve the imbalance learning ability to get a better Gmean score based on reducing the misclassification error of ST samples, which indicates the effectiveness of XGBoost-W and LightGBM-W.
TABLE 5. Performances of XGBoost-W and benchmark BFP models.

| Algorithm | T-1 | T-2 |
|-----------|-----|-----|
| AUC       | e-I | e-II| Gmean | AUC | e-I | e-II | Gmean |
| LDA       | .8868 | .0091 | .7964 | .4491 | .9093 | .0091 | .7240 | .5229 |
| LR        | .9187 | .0081 | .8024 | .4427 | .9215 | .0100 | .5497 |
| DT        | .9105 | .0123 | .7246 | .5216 | .9280 | .0093 | .6883 | .5557 |
| RF        | .9462 | .0083 | .7335 | .5141 | .9542 | .0118 | .5584 | .6606 |
| AdaBoost  | .9407 | .0119 | .6766 | .5653 | .9490 | .0127 | .5260 | .6841 |
| GBDT      | .9473 | .0115 | .6138 | .6179 | .9527 | .0149 | .5325 | .6786 |
| LightGBM  | .9495 | .0110 | .6497 | .5886 | .9485 | .0120 | .5487 | .6677 |
| XGBoost   | .9510 | .0112 | .6347 | .6010 | .9538 | .0127 | .5130 | .6934 |
| LightGBM-W | .9495 | .1115 | .1108 | .8888 | .9405 | .1261 | .0942 | .8897 |
| XGBoost-W | .9550 | .1402 | .0689 | .8947 | .9351 | .1043 | .0909 | .9024 |

Fig. 1 shows the performance comparison of classical BFP models and cost-sensitive boosted trees for BFP datasets with different predictive horizons. From Fig. 1, it can be seen that if we take the area covered by the 4 metrics of BFP models on the radar maps as a more comprehensive metric, the larger coverage area of LightGBM-W and XGBoost-W compared to other classical BFP models further demonstrates the advantage of cost-sensitive cross-entropy for the boosted framework. Moreover, as can be seen from Fig. 1(a)–Fig. 1(d), XGBoost-W and LightGBM-W get balanced predictive performance on while other BFP models get type-II accuracy score and a high type-II accuracy score, implying XGBoost-W and LightGBM-W both good choices for reducing the misclassification error on business failure firms. In contrast, traditional BFP models, which implement predicting the probability of business failure by minimizing the misclassification error, cannot effectively deal with an imbalanced BFP, leading to the balanced BFP models focusing more on the optimization of Type-I accuracy. Type-I accuracy, also known as true negative rate, measures the proportion of the accurately predicted NST samples. Such results further confirm that error minimization-based BFP models are not suitable for imbalanced BFP. Compared with error minimization-based methods, cost-minimization-based BFP models improve the performance by significantly enhancing the Type-II accuracy, making the prediction results of LightGBM-W and XGBoost-W more balanced in all 4 metrics, which effectively solves the problem of skew-sensitivity in imbalanced BFP.

B. COMPARISON OF IMBALANCED BUSINESS FAILURE PREDICTION APPROACHES

Tab. 6 presents the performance comparison of the proposed BFP models and baseline imbalanced BFP algorithms. As can be seen from Tab. 6, XGBoost-W performs best on AUC and Gmean metrics, indicating that XGBoost-W is not only capable of predicting the probability of ST but also provides an accurate BFP result on an imbalanced BFP task. On the 4 predictive horizons, RUS and ROS are considered as suboptimal solutions for imbalanced BFP since their performance significantly outperforms other resampling-related imbalanced strategies such as SMOTE, RUSBoost, and SMOTEBoost in terms of Gmean and AUC. Compared with resampling-based imbalanced BFP models, MetaCost gets higher AUC scores and lower Gmean scores on the 4 BFP datasets, indicating MetaCost can be a good option to predict the probability of ST while it is not a good solution for ST/NST discrimination. AdaCost gets the optimal Type-II error on T-1, T-2, and T-3 predictive horizons, indicating AdaCost can be a good option to optimize the accuracy of predicting ST firms. However, AdaCost gets the highest Type-I error score, indicating AdaCost cannot well balance the tradeoff of misclassifying ST firms and incorrectly predicted NST samples. In contrast, MetaCost gets the opposite prediction results, which gets higher Type-II error and lower Type-I error, implying that MetaCost is a cost-sensitive solution that prefers to predict ST samples into NST, which may lead to the financial losses of investors. In the comparison of cost-sensitive BFP algorithms, XGBoost-W
FIGURE 1. Radar maps of BFP models for different predictive horizons.

TABLE 6. Performance comparison between XGBoost-W and imbalanced BFP models.

| Algorithm       | T-1         | T-2         | T-4         |
|-----------------|-------------|-------------|-------------|
|                 | AUC  | A-I  | A-II | Gmean | AUC  | A-I  | A-II | Gmean |
| ROS             | .9326 | .8980 | .8653 | .8815 | .9462 | .9134 | .8669 | .8898 |
| RUS             | .9414 | .8893 | .8743 | .8818 | .9602 | .9063 | .8799 | .8930 |
| SMOTE           | .9339 | .8967 | .8843 | .8870 | .9437 | .9143 | .8506 | .8819 |
| RUSBoost        | .9408 | .8998 | .8939 | .8878 | .9488 | .9097 | .8766 | .8930 |
| SMOTEBoost      | .9211 | .9140 | .7994 | .8548 | .9350 | .9124 | .8409 | .8759 |
| AdaCost         | .9214 | .2719 | .9910 | .5191 | .9323 | .5011 | .9805 | .7009 |
| MetaCost        | .9441 | .9939 | .2246 | .4724 | .9526 | .9909 | .3669 | .6029 |
| LightGBM-W      | .9495 | .8885 | .8892 | .8888 | .9469 | .8739 | .9058 | .8897 |
| XGBoost-W       | .9510 | .8598 | .9311 | .8947 | .9531 | .8957 | .9091 | .9024 |

VOLUME 10, 2022
and LightGBM-W are regarded as good methods that well balance the Type-I error and Type-II error, getting higher Gmean scores.

Moreover, as can be seen in the performance comparison between different predictive horizons, XGBoost-W and LightGBM-W get better predictive performance on T-1 and T-2 horizons compared with T-3 and T-4 horizons, demonstrating that both T-1 and T-2 financial ratios can be selected as financial features to realize accurate BFP. An appropriate business failure warning window is T-2 and T-1, untimely early warning of business failure may lead to the failure of BFP, theeffectiveness of financial data will decrease over time.

C. STATISTICAL SIGNIFICANCE TEST

To give a view of the positions of imbalanced approaches on BFP and detect whether the proposed weighted XGBoost is statistically superior to other imbalanced BFP models, we further perform the Friedman-Nemenyi test. Friedman test is a nonparametric statistical test to tell whether there is a significant test among imbalanced BFP models. Friedman statistic is defined as:

\[
\chi^2 = \frac{12D}{N_c(N_c + 1)} \left[ \sum_{n=1}^{N_c} \text{AvR}_{nc}^2 - \frac{N_c(N_c + 1)}{4D} \right],
\]

(20)

where \(D\) is the number of experimental datasets, \(N_c\) represents the number of classifiers for BFP. \(\text{AvR}_{nc}\) denotes the average rank computed over all involved datasets of \(n\)-th classifier, which can be specifically defined as \(\text{AvR}_{nc} = \frac{1}{N} \sum_{i=1}^{N} r_i^{(n)c}\), where \(r_i^{(n)c}\) is the average metrics score that average by \(r_i^{(n)c} = \frac{1}{4} \left( r_{i,\text{AUC}}^{(n)c} + r_{i,\text{A}\text{--}\text{I}}^{(n)c} + r_{i,\text{A}\text{--}\text{II}}^{(n)c} + r_{i,\text{Gmean}}^{(n)c} \right)\), where \(r_{i,\text{AUC}}^{(n)c}\) denotes the AUC rank of \(n\)-th classifier on the \(i\)-th dataset, \(r_{i,\text{A}\text{--}\text{I}}^{(n)c}\) calculates the Type-I accuracy rank of \(n\)-th classifier on the \(i\)-th dataset, \(r_{i,\text{A}\text{--}\text{II}}^{(n)c}\) computes the Type-II rank of \(n\)-th classifier on the \(i\)-th dataset, \(r_{i,\text{Gmean}}^{(n)c}\) reports the Gmean rank of \(n\)-th classifier on the \(i\)-th dataset. The null hypothesis (there is no significant difference among the proposed weighted GBDTs and baseline BFP classifiers) is rejected when the computed statistic score \(\chi^2\) is greater than a critical value at a significance level. Once the null hypothesis is rejected, a post-hoc procedure, the Nemenyi test is followed to detect the significant difference among pair-wise comparisons. To report the differences among pair-wise comparisons, we computed the critical difference values at different significance levels by following Demarsa’s [61] work:

\[
CD_{\alpha} = q_{\alpha,\infty,N_c} \sqrt{\frac{N_c(N_c + 1)}{12D}}.
\]

(21)

where \(q_{\alpha,\infty,N_c}\) is the critical value at the significance level \(\alpha\), which is sampled from a studentized range distribution.

According to equation 20, we first compute \(\chi_{\text{AUC}}^2 = 12.497\), \(\chi_{\text{A}\text{--}\text{I}} = 12.112\), \(\chi_{\text{A}\text{--}\text{II}} = 12.638\), \(\chi_{\text{Gmean}} = 12.657\) rejecting the null hypothesis at the significance level 99%. Next, we perform the Nemenyi test to find the pair-wise differences among proposed imbalanced methods and other BFP baseline models. We compute \(q_{0.01} = 3.5265\), \(q_{0.05} = 3.0309\), \(q_{0.1} = 2.7799\) to get CD values at different significance levels. According to Eq. 21, we calculate \(CD_{0.01} = 4.3191\), \(CD_{0.05} = 3.7121\), \(CD_{0.1} = 3.4047\).

Fig. 2 shows the ranking score of imbalanced approaches on different metrics and their CDs + lowest ranking. Fig. 2(a) shows the ranking score of imbalanced approaches on AUC and their CDs + lowest rankings; Fig. 2(b) is the ranking score of imbalanced approaches on Type-I error and their CDs + lowest rankings; Fig. 2(c) illustrates the ranking score of imbalanced approaches on Type-II error and their CDs + lowest rankings; Fig. 2(d) provides the ranking score of imbalanced approaches on AUC and their CDs + lowest rankings.

As can be seen from Fig. 2, XGBoost gets the optimal ranking score on AUC Gmean, indicating the effectiveness of the cost-sensitive boosted tree. As can be observed from Fig. 2(a), XGBoost-W gets the optimal AUC ranking and Gmean rankings over the 4 BFP predictive horizons. On the Type-II error rankings, XGBoost-W gets the second-best ranking among imbalanced BFP models, indicating the superiority of XGBoost-W. As can be seen from Fig. 2(a), if we considered XGBoost-W as a comparison baseline, XGBoost-W outperforms SMOTE, AdaCost, and SMOTE at the significance level \(\alpha = 0.01\), and is superior to RUSBoost at the significance level \(\alpha = 0.05\). Moreover, as can be seen from Fig. 2(b) and Fig. 2(d), AdaCost gets the worst Type-I error and best Type-II error ranking. Gmean is a comprehensive metric that is computed based on Type-I error and Type-II error, the low Type-I error ranking, and Type-II error ranking lead to the poor Gmean ranking of AdaCost, demonstrating AdaCost is not a good solution for imbalanced BFP. From the ranking score in Fig. 2(b), it can be seen that if we regard MetaCost as the benchmark, LightGBM-W, XGBoost-W, and AdaCost are significantly worse than MetaCost with 99% confidence. From the comparison of Fig. 2(c), it can be viewed that AdaCost outperforms SMOTE, SMOTE-Boost, and MetaCost on optimizing the recognition of ST firms at the significance level \(\alpha = 0.01\). By reducing the Type-II error, XGBoost-W gets a top 2 ranking score on the Type-II error rankings, improving the Gmean ranking to top 1 among imbalanced BFP models. If we consider XGBoost as the comparable BFP model, XGBoost statistically outperforms SMOTE, SMOTEBoost, AdaCost, and MetaCost at the significance level \(\alpha = 0.01\), indicating that XGBoost-W is statistically optimal among imbalanced BFP models.

D. INTERPRETING THE PREDICTIONS OF XGBoost-W

To investigate the global contributions of financial features on the predictions of XGBoost-W, we visualize the importance scores of financial features for the T-2 predictive horizon as a template, which is shown in Fig. 3. As can be seen from Fig. 3, \(X_{40}\), whose feature importance ranking ranks
FIGURE 2. Statistical significance test of imbalanced BFP models and their scores of CDs + lowest ranking.

Y. Zou et al.: Business Failure Prediction Based on Cost-Sensitive Extreme Gradient Boosting Machine

FIGURE 3. A template of feature importance of XGBoost-W.

FIGURE 4 shows the feature importance ranking of different predictive horizons. Fig. 4(a) is the feature importance ranking of the T-1 predictive horizon; Fig. 4(b) shows the

be calculated by business profit/number of ordinary shares at the end of the year, the top-ranking of operating profit per share represents the level of a company’s operating income for each share held by the company’s common shareholders. The higher its value means that the company has better product sales, technical and management capabilities, strong profitability. Therefore, the higher the operating profit per share, the smaller the chance of business failure. Other financial ratios including earnings per share (X37), net profit rate of current assets (X21), net assets per share (X41), retained earnings per share (X15), net profit rate of total assets (X20), tangible asset-liability ratio (X6), undistributed profits per share (X44), accounts receivable turnover rate (X12), operating profit rate before interest and tax (X30), which covers the financial ratio groups of per share index, profitability, solvency, and operating ability. Moreover, the importance score of financial features may be considered as a feature selection approach to reduce the dimension of financial ratios for empirical BFP studies, providing theoretical support for banks and other lending structures to understand the factors that affect the occurrence of debt default.
feature importance scores of the T-2; Fig. 4(c) represents the feature importance scores of T-3 predictive horizons; Fig. 4(d) illustrates the feature importance scores of T-4 predictive horizons. As can be seen from Fig. 4, on the T-1 predictive horizon, retained earnings per share ($X_{45}$), asset impairment loss/ business income ($X_{29}$), growth rate of owner’s equity ($X_{10}$), operating profit rate before interest and tax ($X_{30}$), cash ratio ($X_3$) rank top 5, demonstrating the importance of these 5 financial ratios for T-1 predictive horizon. On the T-2 horizon, the most important features are operating profit per share ($X_{40}$), earnings per share ($X_{37}$), net profit rate of current assets ($X_{21}$), net assets per share ($X_{44}$), retained earnings per share ($X_{45}$) for XGBoost-W to predict the business failure of listed firms. On the T-3 horizon, the top 5 financial ratios are operating profit per share ($X_{40}$), earnings per share ($X_{37}$), undistributed profits per share ($X_{44}$), retained earnings per share ($X_{45}$), accounts receivable turnover days ($X_{13}$). On the T-4 horizon, operating profit per share ($X_{40}$), earnings per share ($X_{37}$), net profit rate of total assets ($X_{20}$), undistributed profits per share ($X_{44}$), retained earnings per share ($X_{45}$) occupy the top 5 of feature importance ranking. By comparing the ranking of feature importance on the T-4 BFP datasets, it can be seen that the feature importance rankings computed based on the T-1 dataset are quite different from those on the other three datasets. The more important feature is retained earnings per share ($X_{45}$), while on the other three data sets, the first two important features are operating profit per share ($X_{40}$) and earnings per share ($X_{37}$), and the importance of these two features in the model is significantly higher than other features. Moreover, it can be concluded from Fig. 4(a) to Fig. 4(d) that $X_{45}$, which ranks top 5 for different predictive horizons, is an important financial ratio that participates in the prediction of XGBoost-W. On the T-2, T-3, and T-4 predictive horizons, operating profit per share ($X_{40}$), earnings per share ($X_{37}$), and undistributed profits per share ($X_{44}$) all ranks into top 5, demonstrating the importance of financial ratios of index per share for predicting the business failure of Chinese listed corporates.

To explore the local interpretation results of XGBoost-W, we further select the top 8 features to investigate the marginal effect of the financial ratios on the predictions of
FIGURE 5. PDPs of the top 8 financial ratios on the T-2 predictive horizon.

XGBoost-W. Fig. 5 shows the 1D PDPs of the top 8 financial ratios on the T-2 predictive horizon computed by XGBoost-W. As can be observed from Fig. 5(a), the financial ratios operating profit per share ($X_{40}$) and earnings per share ($X_{37}$) are inversely proportional to the predicted probability of enterprise business failure. As can be seen from Fig. 5(a) and Fig. 5(b), with the increase of operating profit per share and earnings per share, the possibility of business failure of the enterprise gradually decreases. This is because operating profit per share and earnings per share reflect the operating status of the enterprise and measure the profit level of common shares. The larger the value, the better the operation of the enterprise, the stronger the profitability of the enterprise, and the lower the possibility of business failure. Fig. 5(c) presents the marginal effect of net profit rate of current assets ($X_{21}$) to the predictions of XGBoost-W. As can be seen from Fig. 5(c), the financial ratio, net profit rate of current assets ($X_{21}$) plays an inversely proportional role in the probabilistic prediction of business failure. This can be explained as: with the increase of the net profit rate of current assets, the possibility of a business failure occurring is low. This is because the net profit rate of current assets reflects...
the efficiency of current assets. The larger the net profit rate of current assets, the higher the management efficiency of current assets, and the lower the financial risk faced by the enterprise. As can be seen from Fig. 5(d), Fig. 5(e), and Fig. 5(f), net assets per share (X_{41}), undistributed profits per share (X_{44}), retained earnings per share (X_{45}) are inversely proportional to the probability of business failure. Among them, the net asset per share reflects the actual value of an asset contained in each share of a listed company. The higher the net asset per share, the lower the financial risk of the company and the lower the possibility of business failure; the undistributed profit per share is the undistributed profit owned by one share of the enterprise. An enterprise gets great autonomy over the undistributed profit per share, avoiding the further deterioration of the enterprise’s financial dilemma; retained earnings per share is the sum of undistributed profits per share and surplus reserves. Greater retained earnings per share implies that more funds can be utilized by the enterprise to make up for losses, thus reducing the possibility of business failure. Moreover, as can be seen from Fig. 5(g), a lower net profit rate of total assets will result in Business failure, when its value is smaller than 0 (indicates a low input-output level), the probability of business failure of a corporate approaches to 1. The net profit rate of total assets reflects the comprehensive utilization of enterprise assets and also measures the profitability of the enterprise by using the total creditors and owners’ equity. The higher the total asset net profit of the enterprise, the better the profitability of the enterprise, thus the lower the possibility of business failure. As can be seen from Fig. 5(h), tangible asset-liability ratio (X_{6}) is positively correlated with the probability of business failure. This financial ratio reflects the ability of enterprises to repay debts. The higher the tangible asset-liability ratio, the lower the debt repayment ability of enterprises, and the higher the possibility of business failure.

To investigate the feature interaction effect of financial ratios, we visualize the 2D PDPs of top financial ratios on the T-2 predictice horizon. Fig. 6 shows the second-order interaction of financial ratios visualized by the 2D PDP tool. Since operating profit per share (X_{40}) reflect the profitability per share, tangible asset-liability ratio (X_{6}) that measures the solvency of an enterprise, and net profit rate

FIGURE 6. Interaction effects of financial ratios visualized by 2D PDP tool.
of total assets (X_{21}) is one of the important factors that assess the profitability of a firm, play the most important contributions on the predictions of XGBoost-W, we visualize the 2D interaction effect among these financial ratios, which is shown in Fig. 6. Fig. 6(a) shows the interaction results of financial ratios Operating profit per share (X_{40}) and Tangible asset-liability ratio (X_6); Fig. 6(b) provides the 2D interaction results of financial ratios net profit rate of total assets (X_{21}) and tangible asset-liability ratio (X_6); Fig. 6(c) illustrates the 2D interaction results of financial ratios operating profit per share (X_{40}) and net profit rate of total assets (X_6). As can be seen from Fig. 6(a), The greater the operating profit per share (X_{40}), the lower the tangible asset-liability ratio (X_6), and the less likely business failure would occur. Specifically, when the operating profit per share (X_{40}) is greater than 1.53 and the tangible asset-liability ratio (X_6) is less than 0.01, the probability of business failure is relatively small. As can be seen from Fig. 6(b), a higher net profit rate of current asset (X_{21}) and a lower tangible asset-liability ratio (X_6) imply a lower probability that business failure a firm would suffer. Furthermore, as can be seen from Fig. 6(c), a lower operating profit per share and a lower net profit ratio of current assets may lead to a business failure event. Concretely, when the operating profit per share (X_{40}) of the enterprise is greater than 1.53 and the net profit rate of current assets (X_{21}) of the enterprise is greater than 3.78, XGBoost-W tends to classify such samples into “ST” firms.

The above interpretation evidence of XGBoost-W suggests that in the process of enterprise risk management, enterprise managers should focus on three indicators: operating profit per share (X_{40}), tangible asset-liability ratio (X_6) that is one of indicator measure the solvency, and net profit ratio of current assets (X_{21}) that indicates the profitability of a firm. The tangible asset-liability ratio is an extension of the asset-liability ratio and an objective indicator to evaluate the solvency of enterprises. The intangible assets of an enterprise, such as trademarks, patent rights, non-patented technologies, goodwill, etc., may not be used to repay debts. They shall be regarded as insolvent assets and deducted from the total assets. Compared with the intangible asset-liability ratio, the tangible asset-liability ratio is a financial ratio that establishes the analysis of enterprise debt-paying capacity based on a more practical and reliable guarantee, can be a more objectively financial ratio to evaluate the debt-paying ability of an enterprise. In this study, as can be seen from the interpretation in Fig. 6, enterprise managers should be cautious about the financial ratio Tangible asset-liability ratio and keep its value below 0.01, thus alleviating the risk of business failure. The net profit ratio of current assets reflects the management of enterprise current assets. The greater the profit ratio of current assets, the better the management efficiency of current assets. At the same time, this financial ratio is also affected by the net sales profit and the turnover ratio of current assets, which can be regarded as a comprehensive ratio that reflects the financial situation of a corporate. Besides, an interesting finding is that when the net profit of current assets is greater than 3.78, the risk of business failure of enterprises is reduced. Therefore, enterprise management must expand sales, save costs, improve the turnover ratio of current assets, enhance the management efficiency of current assets, thus avoiding the occurrence of business failure.

VI. CONCLUSION AND FUTURE WORK

Establishing a precise BFP system is an important topic for the financial management of listed companies. Taking into account the influence variety of financial ratios and the enrichment of financial datasets, many BFP methodologies have been developed to update the BFP models driven by the accumulation of financial data. However, they virtually completely disregard the issue of class imbalance throughout the BFP modeling procedure. Moreover, to pursue precise BFP, most works neglect the importance of the interpretability of BFP systems. In this study, we overcome these issues based on the advanced ensemble framework XGBoost. Moreover, we tackled the class imbalance problem by modifying the XGBoost into a weighted version XGBoost-W, transforming the error-minimization-based pattern of XGBoost into a cost-sensitive one. Besides, to interpret the prediction results of XGBoost-W, we incorporated feature importance calculation mechanism and partial dependence plot to answer which financial ratio is the main driven force for the prediction of XGBoost-W and depicted the complex relationship among financial ratios and the response of the XGBoost-W. Experimental results on financial datasets that collected from Chinese companies listed on Shanghai and Shenzhen stock exchanges with different predictive horizons showed that cost-sensitive BFP models significantly reduced the error of misclassifying ST samples compared with classical balanced BFP approaches such as statistical-based methods LDA and LR, advanced ensemble algorithms such as GBDT, LightGBM, and XGBoost. Furthermore, the empirical comparison among imbalanced approaches further demonstrated XGBoost-W is the best cost-sensitive solution for imbalanced BFP, which outperformed resampling-based methods such as ROS, RUS, and SMOTE as well as cost-sensitive algorithms including AdaCost and MetaCost. Based on that, the feature importance score of XGBoost-W further suggested that index per share-related financial ratios have significant contributions to the precise prediction of business failure. This may inspire risk managers and policy-makers to pay more attention to the control of the index per share-related financial ratios of listed companies. Last but not the least, the partial dependence plots of financial ratios reflect the positive and negative effects of financial variables on the prediction results. This can be a motivation for BFP model developers to flexibly optimize the BFP model according to the consistency between the interpretation results and the practical business failure situation.

Though XGBoost-W achieved a promising BFP result compared with classical imbalanced BFP models. Its predictive performance can be further enhanced by introducing an
advanced finetuning technique that alternative grid search. This motivates our first future work focus on investigating the performance of other hyperparameter optimization methods such as random search and Bayesian optimization [62].

Second, in this study, to balance the performance and interpretability of a BFP model, two post-interpretation algorithms are introduced to interpret the prediction results of XGBoost-W. Model-agnostic interpretation algorithm, PDP is introduced into this study to reveal the relationship between business failure features and predictive responses. However, a more precise interpretable BFP model should be derived from its intrinsic interpretability. Therefore, in our future work, some tree-based pruning mechanisms such as [63], [64] will be introduced to simplify the decision pattern of tree-based ensemble methods while getting an accurate BFP result.

Third, in this study, 47 financial ratios are collected from 2001 to 2019 for BFP. The accumulation of financial data allows us to expand the future BFP research to large-scale BFP and investigate the impact of non-financial ratios on BFP performance, thus guiding policy-makers to make more scientific decisions.

REFERENCES
[1] E. U. Hassan, Z. Zainuddin, and S. Nordin, “A review of financial distress prediction models: Logistic regression and multivariate discriminant analysis,” Indian-Pacific J. Accounting Finance, vol. 1, no. 3, pp. 13–23, Jul. 2017.
[2] J. Sun, H. Li, and H. Adeli, “Concept drift-oriented adaptive and dynamic support vector machine ensemble with time window in corporate financial risk prediction,” IEEE Trans. Syst., Man, Cybem., Syst., vol. 43, no. 4, pp. 801–813, Jul. 2013.
[3] H. Li, C.-J. Li, X.-J. Wu, and J. Sun, “Statistics-based wrapper for feature selection: An implementation on financial distress identification with support vector machine,” Appl. Soft Comput., vol. 19, pp. 57–67, Jun. 2014.
[4] M.-Y. Chen, “Predicting corporate financial distress based on integration of decision tree classification and logistic regression,” Expert Syst. Appl., vol. 38, no. 9, pp. 11261–11272, Sep. 2011.
[5] R. Geng, I. Bose, and X. Chen, “Prediction of financial distress: An empirical study of listed Chinese companies using data mining,” Eur. J. Oper. Res., vol. 241, no. 1, pp. 236–247, Feb. 2015.
[6] J. Sun, H. Li, Q.-H. Huang, and K.-Y. He, “Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches,” Knowl.-Based Syst., vol. 57, pp. 41–56, Feb. 2014.
[7] C.-H. Cheng, C.-P. Chan, and Y.-J. Sheu, “A novel purity-based k nearest neighbors imputation method and its application in financial distress prediction,” Eng. Appl. Artif. Intell., vol. 81, pp. 283–299, May 2019.
[8] H. Li, J. Sun, and B.-L. Sun, “Financial distress prediction based on OR-CBR in the principle of k-nearest neighbors,” Expert Syst. Appl., vol. 36, no. 1, pp. 643–659, Jan. 2009.
[9] S. Y. Kim and A. Upneja, “Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models,” Expert Syst. Appl., vol. 36, pp. 354–362, Jun. 2014.
[10] W.-S. Chen and Y.-K. Du, “Using neural networks and data mining techniques for the financial distress prediction model,” Expert Syst. Appl., vol. 36, no. 2, pp. 4075–4086, Mar. 2009.
[11] P. Ravishanker and V. Ravi, “Financial distress prediction in banks using group method of data handling neural network, counter propagation neural network and fuzzy ARTMAP,” Knowl.-Based Syst., vol. 23, no. 8, pp. 823–831, Dec. 2010.
[12] M. Pavlicko, M. Durica, and J. Mazanec, “Ensemble model of the financial distress prediction in visegrad group countries,” Mathematica, vol. 9, no. 16, p. 1886, Aug. 2021.
YAO ZOU received the B.S. degree in business administration from the School of Management, Shanghai University of Engineering Science, Shanghai, China, in 2013. He is currently pursuing the Ph.D. degree with the Glorious Sun School of Business and Management, Donghua University. His research interests include decision-making, enterprise financing, and credit risk management.

CHANGCHUN GAO was born in Huadian, Jilin, China, in 1964. He received the M.S. degree in economics from the Harbin Institute of Technology, in 1996, and the Ph.D. degree in economics from Jilin University, in 1996. From 1999 to 2004, he was a Professor, the Deputy Dean, and the Chairperson of the Academic Committee, School of Humanities, Harbin Institute of Technology. Since 2004, he has been a Professor with the Department of Management Engineering and Science and the Department of Business Administration, Donghua University. He is the author of ten books and more than 100 articles. His current research interests include decision analysis and innovation of cultural and creative industries.

HAN GAO received the Ph.D. degree in management science and engineering from Donghua University, China, in 2017. He is currently a Lecturer at the College of Fashion and Design, Donghua University. He has expertise in the domains of business analytics and industrial economics.