Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 9(1): 974–997.
DOI: 10.3934/math.2024049
Received: 21 October 2023
Revised: 15 November 2023
Accepted: 20 November 2023
Published: 05 December 2023

Research article

A novel profit-based validity index approach for feature selection in credit risk prediction

Meng Pang[1,2] and Zhe Li[1,*]

[1] School of Business, Liaocheng University, Liaocheng 252059, China
[2] Graduate School, Lyceum of the Philippines University, Batangas City 4200, Philippines

* **Correspondence:** Email: li_zhe@yeah.net, Tel: +8615762476319.

Abstract: Establishing a reasonable and effective feature system is the basis of credit risk early warning. Whether the system design is appropriate directly determines the accuracy of the credit risk evaluation results. In this paper, we proposed a feature system through a validity index with maximum discrimination and commercial banks' loan profit maximization. First, the first objective function is the minimum validity index constructed by the intra-class, between-class, and partition coefficients. The maximum difference between the right income and wrong cost is taken as the second objective function to obtain the optimal feature combination. Second, the feature weights are obtained by calculating the change in profit after deleting each feature with replacement to the sum of all change values. An empirical analysis of 3,425 listed companies from *t*-1 to *t*-5 time windows reveals that five groups of feature systems selected from 614 features can distinguish between defaults and non-defaults. Compared with 14 other models, it is found that the feature systems can provide at least five years' prediction and enable financial institutions to obtain the maximum profit.

Keywords: credit risk; validity index; feature combination; feature weights; loan profit
Mathematics Subject Classification: 68M25

1. Introduction

Failure to accurately identify customers' credit risk before granting credit is the main reason for non-performing loans in commercial banks, and whether the construction of the feature system is reasonable directly affects the evaluation results' effectiveness. The construction of a credit risk feature system involves at least the following two aspects:

The first problem is feature combination selection. The discriminative ability of a single feature is strong, but a feature system consisting of individual features is not necessarily optimal. Furthermore, 2^m - 1 feature combinations exist for m features, and different feature combinations will yield different evaluation results and loan profits. Therefore, we aim to address an empirical issue: constructing the feature selection model and selecting the optimal feature combination with respect to discriminative ability and loan profit.

The second problem is the determination of feature weights. Different weights given to the same group of features will lead to different evaluation results, and reasonable feature weights can also provide an essential reference for bank credit approval. Therefore, we aim to answer another empirical question: determining optimal feature weights and ensuring the evaluation results are reasonable and effective.

Motivated by these reasons, a new feature system based on the validity index and maximum profit is proposed. The validity index determines which features enter the prediction model, and the loan profit is the criterion for evaluating the quality of selected features.

This paper is most related to those of Liu et al. [1] and Kozodoi et al. [2] regarding feature selection. In [1], the validity index was constructed based on the between-class coefficient. In [2], the loss of the customer predicted to be non-default is 0. The difference in this study is the construction of the objective functions. We take the minimum validity index constructed by the intra-class, between-class, and partition coefficients as the first objective function and the maximum difference between right income and wrong cost as the second objective. Regarding feature weights, the authors in [3–5] proposed to take maximum accuracy or minimum error as the objective function. In this paper, the feature weights are obtained by the influence of the features on loan profit.

The primary contributions of this work are as follows:

(1) A new feature combination selection method based on the validity index and loan profit to ensure that the selected feature combination considers both the discriminative ability and loan profit, which addresses the deficiency that existing research does not account for the actual business objectives of banks.

(2) The feature weights are determined by considering the impact of the feature on loan profit. Every change in profit to the sum of all change values is taken as the feature weight, which addresses the shortcoming of existing research that only considers classification accuracy.

(3) A novel evaluation function is proposed based on the loan profit. The profit function consists of the right income and wrong. Because of taking the real loan profit as an estimation method, the feature combination selected is more likely to obtain more profits than the feature combination obtained by the maximum accuracy or minimum error.

The features selected by the model proposed in this paper have better discriminative ability and can obtain the maximum profit than the other 14 feature selection methods. The feature selection method based on the validity index proposed in this paper is effective.

The subsequent chapters of this paper are organized as follows. Section 2 provides a literature review. Section 3 introduces the new feature selection method. Section 4 presents the results of comparing the dataset of Chinese listed companies, and the robustness of the proposed model is verified. Finally, Section 5 presents the conclusions and future work.

2. Related work

Scholars have performed numerous studies on the construction of feature systems for credit risk evaluation, primarily focusing on feature selection and the determination of feature weights.

(1) Research on feature selection for credit risk evaluation.

The single feature selection method defines the importance degree of each feature and selects the features with high importance to compose the feature set. According to the different evaluation functions, it can be divided into four categories. First, relief and its extended feature selection algorithm [6] based on distance measurement. Second, mutual information, information gain, etc. [7] based on information measurement. Third, the Pearson correlation coefficient [8], Fisher score [9] and other methods are used based on the importance of features. Fourth, focus, LVF and other algorithms [10] based on consistency measures are used to find the smallest subset that has the same classification and discrimination ability as the whole set on the premise of ensuring classification ability.

The feature combination selection method has 2^m - 1 combinations for m features, considering the joint effect of features on default status. Xia et al. [11] proposed a feature selection model based on a gradient boosting decision tree, and a Bayesian network was used to optimize the parameters of the decision tree to select the optimal feature combination. Jadhav et al. [12] graded the features based on information gain, obtained the optimal feature combination through the embedded method, and obtained the credit rating through K-nearest neighbors, naive Bayes, etc. Arora and Kaur [13] proposed a feature combination selection method based on the Bolasso model and combined it with four models, such as random forest and decision tree, to prove its robustness and effectiveness. Liu et al. [1] proposed a new validity index and combined it with the support vector machine. The empirical results show that the feature combination selected by the validity index has higher accuracy. Kozodoi et al. [2] proposed a profit-based credit feature selection method and demonstrated that it can obtain the maximum profit in the case of a small number of features.

The existing studies focus more on the single feature selection, less on feature combination selection, and even less on bank profit. We construct a profit-maximization feature combination selection method based on the validity index proposed by Liu et al. [1]. Profit is defined as the difference between the right income and the wrong cost. The heuristic algorithm is used to obtain the optimal feature combination from 2^m - 1 feature combinations.

(2) Research on feature weighting for credit risk evaluation.

The weight is determined by the feature's degree of variation. On the one hand, the weight comes from the subjective weighting of experts, such as the analytic hierarchy process (AHP) and Delphi method. For example, Gu et al. [14] proposed a credit evaluation method based on the analytic hierarchy process (AHP) and data envelopment analysis (DEA). On the other hand, the feature weight is determined by the connection degree of each feature or the amount of information provided. For example, Li et al. [15] proposed a combination of the entropy weight method and fuzzy comprehensive method to construct a credit evaluation model and divide the credit ratings of microenterprises. Karaaslan and Özden [16] established a credit rating prediction model based on the grey model and grey correlation, which shows that the feature weights obtained by grey correlation can provide a reference for the variable selection of the grey model.

The weight is determined by the relationship between the feature and default status. Zhu et al. [17] take the minimum sum of the average distance between good and best applicants and the average distance between bad and worst applicants as the objective function to calculate the optimal feature weights. Chen et al. [18] applied the neighborhood rough set to an unbalanced data set and determined the feature weights according to the upper and lower boundaries. Panday et al. [19] proposed an imwk-means clustering algorithm to carry out unsupervised clustering, took the minimum weighted Euclidean distance within the class as the objective function, and adopted the Lagrange multiplier to solve for the optimal feature weights and the number of clusters.

The weight is given through the relationship between the feature and accuracy. Serrano-Silva et al. [20] proposed a combination of a heuristic algorithm with the NAC model to obtain feature weights. Through the comparison of 10 credit data sets, they found that the classification model after weighting has better classification ability. Mercadier and Lardy [21] took the reduced precision of random forest after random arrangement as the feature weight, and the mean square error proved the effectiveness of this method. Mateos-García et al. [5] proposed the SWAN model based on the K-nearest neighbour approach to improve the feature weights. The empirical results show that the setting of weights can significantly improve the model's accuracy.

According to the above studies, the existing feature weighting mainly focuses on three aspects, with little consideration of profit. Based on the optimal feature combination obtained with respect to profit maximization, the feature weights are determined following the approach of Mercadier and Lardy [21]. We calculate the feature weights by the change in profit after deleting each feature with replacement to the sum of all change values.

(3) Research on credit risk prediction methods.

In existing work for credit risk prediction, most searchers adopted popular two-way classification methods directly, such as mathematical statistics methods, artificial intelligence methods, and ensemble learning methods.

The most common mathematical statistics methods summarize a classification rule through many existing samples, including multivariate discriminant analysis [22], naive Bayes [23], linear discriminant analysis [24], and fuzzy logistic regression [25]. Multivariate discriminant analysis is a supervised learning method that aims to find a discriminant function that maximizes the separation between the groups [26]. Naive Bayes is a simple and efficient algorithm based on Bayes' theorem, assuming independence between the features [27]. The key idea behind LDA is to transform the original high-dimensional feature space into a lower-dimensional space while preserving the discriminatory information between the classes [28]. Fuzzy logistic regression is an extension of traditional logistic regression that incorporates fuzzy logic principles to handle uncertainty and ambiguity in data [29].

Credit risk prediction methods based on artificial intelligence show greater advantages. The traditional models include artificial neural networks [30], support vector machines [31], and extreme learning machines [32]. Deep learning has gained significant attention and popularity in recent years due to its remarkable performance in various applications. Zhang et al. [33] developed a fraud detection system using a deep learning architecture and an advanced feature engineering process based on homogeneity-oriented behavior analysis. Ala'raj et al. [34] proposed a model based on the bidirectional Long-Short Term Memory (LSTM) model to give each customer the probability of a missed payment during the next month. Zhao et al. [35] proposed a novel imputation method named Multiple Generative Adversarial Imputation Networks for accurate missing data prediction. The experimental results on real-world data demonstrate that the proposed method is superior to other popular imputation methods.

Ensemble learning methods refer to training multiple classification models and combining multiple results for better results. Currently, the main ensemble learning methods are bagging [36] and boosting [37], and the learning strategies are average, majority voting, and stacking [38]. Shen et al. [39] developed a new deep-learning ensemble credit risk evaluation model to deal with imbalanced credit data combined with the long-short-term-memory (LSTM) network and the adaptive boosting (AdaBoost) algorithm. Forough and Momtazi [40] proposed an ensemble model based on the sequential modeling of data using deep recurrent neural networks and a novel voting mechanism based on an artificial neural network to detect fraudulent actions. Belhadi et al. [41] proposed a hybrid ensemble machine learning approach to forecast the credit risk associated with SMEs' agriculture 4.0 investments. Two core approaches were used, i.e., the Rotation Forest algorithm and the Logit Boosting algorithm.

The existing credit risk prediction models, especially the ensemble models, have been proven to have higher classification accuracy. We construct an ensemble model, which is composed of six models: logistic regression [42], linear discriminant analysis [43], decision tree [44], support vector machine [45], backpropagation neural network [46] and KNN [47].

3. Feature system construction

3.1. Basic parameters of the validity index

The validity index constructed by the intra-class, between-class, and partition coefficients is used to test the feature combination's discriminative ability. The basic parameters of the validity index are calculated as follows.

(1) Calculation of the distance from each sample to the class centroid.

Let d_{ic} be the distance between the ith sample and the cth class centroid, h be the number of features, x_{ij}^{c} be the jth feature data of the ith sample in the cth class, and p_{cj} be the jth feature data in the cth class centroid. The d_{ic} is defined as:

$$d_{ic}^{2}=\sum_{j=1}^{h}(x_{ij}^{c}-p_{cj})^{2} \tag{1}$$

where the cth class centroid is a vector composed of the mean of all samples in class c.

(2) Calculation of the membership degree.

The membership degree μ_{ic} of the ith sample belonging to class c is defined as:

For any i and c, if d_{ic}>0, then,

$$\mu_{ic}=\frac{1}{\sum_{k=1}^{r}(\frac{d_{ic}}{d_{ik}})^{2}} \tag{2}$$

In Eq (2), if there are i and c such that d_{ic}=0, then μ_{ic}=1; if there is k such that $k\neq c$, then μ_{ic}=0.

(3) Calculation of the partition coefficient.

The partition coefficient (PC) [48] with h features is calculated as:

$$PC(h)=\frac{1}{r}\sum_{c=1}^{r}\sum_{i=1}^{n}\mu_{ic}^{2} \tag{3}$$

where r represents the number of classes, n represents the number of samples, and μ_{ic} represents the membership degree of the ith sample belonging to class c.

Each sample has two membership degrees of default and non-default. Considering the effect of the higher membership degree in each class in this paper on the distinction degree, we obtain a new partition coefficient $PC'(h)$.

$$PC'(h)=\frac{1}{n}\sum_{i=1}^{n}\max_{1\le c\le r}\mu_{ic} \tag{4}$$

(4) Calculation of the intra-class coefficient and between-class coefficient.

The partition coefficient determines the features' distinction degree from the membership degree perspective, but insufficient attention is given to intra-class and between-class data. We construct intra-class and between-class coefficients from the perspective of data structure.

Let r be the number of classes, $\overline{\mathbf{p}}_c$ be a vector composed of the cth class centroid, $\overline{\mathbf{p}}_k$ be a vector composed of the cth class centroid, n be the number of all samples, and m_c be the number of the cth class. The between-class coefficient $BC(c)$ with h features is calculated as:

$$D(h)=\frac{2}{r(r-1)}\sum_{\substack{1\le c,k\le r\\ c\ne k}}\left\|\overline{\mathbf{p}}_c-\overline{\mathbf{p}}_k\right\|_2^2 \tag{5}$$

$$\beta(h)=\frac{1}{n}\sum_{c=1}^{r}m_c\left\|\overline{\mathbf{p}}_c-\overline{\mathbf{p}}_0\right\|_2^2 \tag{6}$$

$$BC(h)=D(h)/\beta(h) \tag{7}$$

Let r be the number of classes, m_c be the number of the cth class, m_c be the number of the cth class, μ_{ic} be the membership degree of the ith sample belonging to class c, $\mathbf{x}_i^c$ be the ith sample in the *cth* class, and $\overline{\mathbf{p}}_c$ be a vector composed of the cth class centroid. The intra-class coefficient $IC(c)$ with h features is calculated as follows:

$$IC(h)=\frac{1}{r}\sum_{c=1}^{r}\sum_{i=1}^{m_c}\mu_{ic}^2\left\|\mathbf{x}_i^c-\overline{\mathbf{p}}_c\right\|_2^2 \tag{8}$$

3.2. Optimal feature combination selection

3.2.1. First feature selection based on the validity index

The partition coefficient $PC(h)'$, intra-class coefficient $IC(h)$, and between-class coefficient $BC(h)$ are integrated to construct the following validity index $VI(h)$.

$$VI(h)=e^{-BC(h)/IC(h)}+(1-PC'(h)) \tag{9}$$

The $BC(h)/IC(h)$ in the first term represents the ratio of the between-class and intra-class coefficients. The larger the $BC(h)/IC(h)$ value is, the better the discriminative ability of h features. For $0<e^{-BC(h)/IC(h)}\le 1$, the smaller its value is, the better the discriminative ability of h features. According to Eq (4), the smaller the 1-PC'(h) in the second term is, the better the discriminative ability of h features.

Therefore, the smaller the validity index is, the better the feature combination containing h features can distinguish between defaults and non-defaults.

Compared with existing work [1], the validity index constructed in this paper has the following two advantages:

First, the validity index proposed in this paper fully considers the separation degree within classes and the separation degree between classes. It can select the optimal feature combination by traversing all feature combinations when the number of features is determined.

Second, the validity index proposed in this paper adds a partition coefficient to measure the distinction degree of all samples, which avoids the deficiency that existing studies only focus on the separation degree of classes.

The amount of calculation is extremely large for dozens or even hundreds of features [49]. To reduce the computational burden, we select the optimal feature combination through the backward selection sequence [50] of the heuristic search.

Step 1: The validity index value $VI(h)$ is computed for all h features.

Step 2: Each feature is deleted one at a time, h validity indices are computed for all subsets with h-1 features, and the feature combination with the minimum validity index is the optimal feature combination.

Step 3: Each feature among the remaining h-1 features is deleted one at a time, and the optimal feature combination is obtained to form a subset with h-2 features.

Step 4: Steps 2–4 are repeated until one feature is left.

Step 5: h feature combinations including 1, 2, ···, h features are obtained.

3.2.2. Second feature selection based on profit

To measure the discriminative ability of the feature combination, we should not only focus on the accuracy of the feature combination but also consider whether the discrimination results yield actual profit for banks [2].

If default customers are correctly classified, the bank will avoid loss by refusing to grant them credit, which can be regarded as an expected benefit of the bank. If default customers are misclassified, the bank will incur bad debt due to the loans. If non-default customers are misclassified, the bank will refuse to grant credit, lose possible high-quality customers and lose the interest generated by the loans. The bank will earn interest income from the loans if non-default customers are correctly classified.

(1) Calculation of direct income and indirect loss of non-default customers.

For non-default customers, the direct income comes from interest income, and the indirect loss comes from the interest loss caused by the loss of customers. It is generally believed that a company has many sources of loans, and its disclosure is limited, so it is difficult to measure the amount of loans from the company's financial data. We use the ID (interest-bearing debt) to determine the benefit and loss of granting credit to non-default customers. Interest-bearing debt requires interest payments, including the company's liabilities, long-term liabilities due within one year, long-term loans, bonds payable, etc. The interest is obtained by multiplying the interest-bearing debt by the loan interest rate.

For the ith sample, the value of "interest-bearing liabilities" is set to ID_i, and the loan interest rate is set to r. If non-default customers are correctly classified, the more interest-bearing liabilities, the more interest income will be earned. If non-default customers are misclassified, the more interest-bearing liabilities there are, the more interest income will be lost.

(2) Calculation of the direct loss and indirect income of default customers.

The direct loss of default customers comes from the bad debt, and the indirect income comes from the evaded bad debt. When default customers have cash flow difficulties and refuse to repay all their loans, they will choose to pay part of the loans, and the loss given default will not be 100%.

After confirming that the quick ratio has a significant impact on default [51], we take the difference between "current liability" and "quick assets" as the direct loss estimation when granting credit to default customers. Quick assets include rapidly realizable enterprise assets such as monetary funds and accounts receivable. Current liabilities include short-term loans, accounts payable, notes payable, and other corporate liabilities that need to be repaid within a short period. The difference between current liabilities and liquid assets is the loss that companies cannot repay in a short period, which may cause a loss for commercial banks.

For the *i*th sample, the value of "interest-bearing liabilities" is set to ID_i, the value of "quick ratio" is set to QA_i, and the "current liability" is set to CL_i. The profit function constructed in this paper is as follows:

$$profit(h)=\sum_{i=1}^{a_1}(CL_i-QA_i)+\sum_{i=1}^{a_4}(ID_i\times r)-\sum_{i=1}^{a_2}(CL_i-QA_i)-\sum_{i=1}^{a_3}(ID_i\times r) \tag{10}$$

The *profit*(*h*) in Eq (10) represents the difference between the right income and the wrong cost. The first item indicates the loss avoided after identifying default customers; the second item indicates the interest income after identifying non-default customers; the third item indicates the loan loss after lending to default customers; and the fourth item means the interest loss after refusing to lend to non-default customers.

Compared with the existing research [2] that only considers the profit generated by two results of TP and FP in the confusion matrix, the profit function in this paper considers four results in Table 1, which can more comprehensively characterize the composition of bank profits.

Table 1. Profit-related confusion matrix.

Profit	Income and Loss	Predicted Status(Number)	Profit Value
profit of non-default customers	interest loss (indirect loss)	False Positive (a_3)	$-\sum_{i=1}^{a_3}(ID_i\cdot r)$
	interest income (direct income)	True Negative (a_4)	$\sum_{i=1}^{a_4}(ID_i\cdot r)$
profit of default customers	evaded bad debt (indirect income)	True Positive (a_1)	$\sum_{i=1}^{a_1}(CL_i-QA_i)$
	bad debt (direct loss)	False Negative (a_2)	$-\sum_{i=1}^{a_2}(CL_i-QA_i)$

Based on obtaining *h* feature combinations in 3.2.1, selecting the best feature combination from *h* feature combinations is necessary. The idea of the second feature selection targeting profit maximization in this paper is as follows:

Step 1: Every feature combination is substituted into six classification models, such as the support vector machine and decision tree. Four parameters of each model in the confusion matrix are obtained.

Step 2: According to Eq (10), the profit value of each classification model is obtained, and the average

profit value of six classification models is taken as the classification ability of each feature combination.
Step 3: The average profit values of h feature combinations are compared. In this paper, the feature combination corresponding to the maximum average profit value is the best.

3.3. Determination of feature weights

The contribution of each feature to discriminative ability is different. Some features with larger contributions must be given higher weights, while some features with lower contributions must be given lower weights to ensure the feature system's overall accuracy.

If h features exist in the optimal feature combination, one feature is deleted with replacement, and subsequently, the h profits containing h-1 features are calculated. The absolute differences between the h profits containing h-1 features and the profit containing an optimal feature combination are Δx_1, Δx_2, Δx_3... Δx_m. The ratio of the absolute difference to the sum of all absolute differences $\Delta x_i / \sum_{j=1}^{m} \Delta x_j$ denotes the weight of the ith feature.

3.4. Construction of credit score model

Banks consider not only whether customers default but also the default risk. Whether a customer defaults or not only indicates their ability to repay their loan on time, but it is not possible to sort the lending of customers and determine the priority lending customers. A credit score model is essential for evaluating customers' credit risk. A good credit score model should give customers with low default probability a higher credit score [52].

Let a be the offset, b be the factor, and P_i be the comprehensive probability. The credit score is:

$$Score_i = a - b \times ln\frac{P_i}{1-P_i} \tag{11}$$

The credit score in Eq (11) quantifies the customer's performance. The higher the credit score is, the lower the default probability. The comprehensive probability P_i is obtained by default probability P_{i1} and the sum of weighted features P_{i2}, P_i=0.5P_{i1}+0.5P_{i2}. The default probability P_{i1} is obtained by the model with the best classification performance among the six classification models, and P_{i2} is obtained by the optimal features and weights:

$$P_{i2} = 1 - \sum_{j=1}^{h} w_j x_{i,j} \tag{12}$$

P_{i2} in Eq (12) reflects the influence of features and feature weights on credit scores. As the credit score increases, the default probability decreases. We established a credit scorecard model [53]. When the credit score decreases by 20 points, the default rate is twice as high as that of $\frac{P_i}{1-P_i}$. When the default-to-default ratio is 1:20, the credit score is 600. The credit score of Eq (13)is redefined as follows:

$$Score_i = 600 - ln20 \times \frac{20}{ln2} - \frac{20}{ln2} \times ln\frac{P_i}{1-P_i} = 513.56 - 28.85 ln\frac{P_i}{1-P_i} \tag{13}$$

The credit score in Eq (13) differs from the existing research [52] in at least two aspects:

First, the expression for the credit score is different. Existing studies use default probability or the sum of weighted features as the credit score without recognizing that the credit score should have discriminative ability rather than a simple high or low default probability. We draw on the idea of the credit scorecard model and builds a credit score model based on comprehensive probability. On the one hand, the model has good interpretability and can explain why the credit score is not high. On the other hand, by setting the score factor, the credit score has discriminative ability.

Second, the rationality of credit scores is different. Existing scholars consider default probability P_{i1} or the sum of weighted features P_{i2}. However, it is important for the credit score to consider not only the influence of the asset-liability ratio, corporate profit, and other features but also the default probability. Only considering the default probability P_{i1} easily causes the customer's credit score to fluctuate substantially, and there is a large difference between the previous year and the following year or even the opposite unreasonable situation. Only considering the sum of weighted features P_{i2} cannot reflect the impact of events such as default on credit score.

4. Empirical analysis

4.1. Data sources

(1) Data description.

In this paper, we take ST(*ST) and non-ST companies in the Shanghai and Shenzhen stock markets as the research objects. The ST(*ST) companies are treated as defaulters and other listed companies are defined as non-defaulters [54]. The ST label indicates that the listed companies' stocks have investment risks, while *ST signifies that the company's stocks are at risk of being delisted. These two situations indicate that the listed companies involve substantial risk. Ultimately, this work selects 47,171 observations of listed companies from 2000 to 2018, with a non-default sample (45,317) that is 24.44 times the size of the default sample (1,854).

Based on the classic high-frequency features of three foreign credit institutions such as Moody's, Standard & Poor's and Fitch [55], domestic financial institutions such as the Agricultural Bank of China and Bank of China, and authoritative references [56–59], a credit risk feature system for listed companies is established. The feature system includes three first-level criterion layers (432 financial factors, 119 non-financial factors, and 153 macro factors) and nine second-level criterion layers, including solvency, profitability, and operating capability.

(2) Sample processing.

The original data are normalized, and each feature is converted into a value in [0,1] [60,61]. To allow the established model to have multi-year predictive ability and to avoid the time point problem of financial statements identified by Ohlson [62], the observations at time t–m (m = 1, 2, ···, 5) before the defaults at time t are adopted.

The data at times t-1 to t-5 are defined as follows:

For each ST(*ST) company, the year when the company was marked as ST(*ST) is taken as the base time t. The observations of m years before time t, that is, the observations at time t-m (m=1, 2, 3, 4, 5) and the status at time t, are used to form five default data sets.

For each non-ST company, the number of non-ST companies with m continuous years is N_0. The number of non-ST companies in each year is determined by N_0/L (year span). The feature data x_{ij}^{t-m} of non-ST companies at time t-m (m=1, 2, 3, 4, 5) and the default status y_j^t at time t are used to construct a data set of non-ST companies.

The default and non-default companies also form five new data sets. Each time window in this paper contains 3,425 companies, consisting of 2,949 non-default companies and 476 default companies. By using 10-fold cross-validation, the samples are randomly divided into ten parts: eight for the training set, one for the validation set and one for the test set. The average of the ten classification results is taken as the final result.

4.2. Construction of the optimal feature system

In this paper, a 10-fold cross-validation method is used to test the classification performance of the feature system. According to the ratio of 8:1:1, the data are randomly divided into ten parts. Eight parts for the training set, one part for the validation set, and one part for the test set. The training set is used to adjust the model's parameters and optimize its performance. The validation set is used to obtain the optimal feature combination and weights. The test set is used for out-of-sample performance evaluation [63]. After obtaining the optimal features, they are applied to the classification model, which provides the out-of-sample default probabilities. By comparing the default probabilities with a threshold (0.5), the default status of customers can be determined, thus obtaining the accuracy of default prediction.

The model's classification performance is based on five classification criteria: Accuracy, Type-I error, Type-II error, G-mean and the Area Under Curve (AUC). In this paper, the six classification models are logistic regression [42], linear discriminant analysis [43], decision tree [44], support vector machine model [45], backpropagation neural network [46], and KNN [47], and the interest rate is 5%. The steps for obtaining the optimal feature system are shown in 3.2 and 3.3. The number of optimal features from t-1 to t-5 is shown in Table 2.

Table 2. Number of optimal features from t-1 to t-5.

Year	Number of features	Mean value of the validity indexes	Average profit
t-1	61	1.021	13.795
t-2	54	1.016	13.557
t-3	49	1.009	13.463
t-4	43	1.006	13.657
t-5	48	1.002	13.287

Through the 5-year feature frequency analysis, 43 features, such as monetary funds, notes receivable, intangible assets, and social contribution per share, have a 5-year predictive ability; 20 features, such as capital reserves, the executive shareholding ratio, whether shareholder protection is disclosed, and social donations, have a 4-year predictive ability; 15 features, such as the payout ratio before tax and whether environment and sustainable development is disclosed, have a 3-year predictive ability; and 14 features, such as the cash ratio, income tax and external guarantee balance, have a 2-year predictive ability.

4.3. Comparison of classification performance

4.3.1. Comparison of classification performance between models

The classification performance of the optimal feature combination is compared with features without profit consideration and without feature selection. The comparison results of the models' classification performance at time *t*-2 are shown in Table 3. The others have the same conclusions as time *t*-2, so only the results of time *t*-2 are listed. The five criterion values of each method are the average values of six classification models.

For the logit model, the objective function minimization is the stochastic gradient descent, the initial linear coefficient is 0, the learning rate is 0.05 and the observation weight is 1. For the LDA model, for the cost of misclassification, the default is 1, the discriminant type is linear, and the observation weight is $1/n$. For the DT model, the maximum category level is 10, the maximum number of decision splits is ($\mathbf{X}$, 1)–1, the minimum number of leaf node observations is 1, and the minimum number of branch node observations is 10. For the SVM model, the kernel function is the RBF, and the kernel scale is 2. The kernel scale is obtained through the genetic algorithm [12]. For the BPNN model, the number of network layers is 10, the learning rate is 0.1, the maximum training time is 100, the initial weight change is 0.05 and the loss function is a quadratic mean square function. For the KNN model, the number of nearest neighbours is 5, the distance is Euclidean distance, and the distance weighting is 1/distance.

1). The proposed validity index+Profit
2). The proposed validity index+AUC
3). Validity index proposed by Liu et al. [1]+Profit
4). Validity index proposed by Liu et al. [1]+AUC
5). Validity index proposed by Peña et al. [64]+Profit
6). Validity index proposed by Peña et al. [64]+AUC
7). Sequential Forward Selection [11]+Profit
8). Sequential Forward Selection [11]+AUC
9). Sequential Backward Selection [50]+Profit
10). Sequential Backward Selection [50]+AUC
11). LASSO [65]+Profit
12). LASSO [65]+AUC
13). Profit function proposed by Maldonado et al. [9]
14). Profit function proposed by Garrido et al. [66]
15). Profit function proposed by Kozodoi et al. [2]

Table 3. Classification comparison at time *t*-2.

Method	Accuracy	Type-I error	Type-II error	G-mean	AUC	Profit
1	91.17%	3.63%	41.49%	73.58%	89.55%	**13.80**
2	**91.26%**	3.08%	25.14%	72.80%	89.54%	13.38
3	91.15%	3.81%	39.58%	75.61%	**89.84%**	13.60
4	89.54%	3.03%	**25.10%**	62.88%	86.73%	13.46
5	91.03%	3.12%	44.95%	72.30%	87.68%	13.56

Continued on next page

Method	Accuracy	Type-I error	Type-II error	G-mean	AUC	Profit
6	91.12%	4.48%	36.06%	**77.14%**	91.21%	13.27
7	90.24%	3.23%	50.30%	66.22%	82.84%	13.29
8	88.93%	5.30%	47.20%	68.04%	88.32%	12.79
9	89.88%	**2.75%**	55.67%	63.42%	85.71%	13.66
10	88.76%	4.41%	53.04%	64.17%	87.20%	13.15
11	88.13%	6.50%	44.46%	69.03%	88.58%	12.65
12	86.58%	6.76%	54.56%	61.14%	83.25%	12.39
13	89.32%	4.66%	47.84%	68.67%	85.45%	12.93
14	87.85%	5.60%	52.48%	63.64%	85.96%	12.60
15	87.67%	5.70%	53.11%	63.05%	85.41%	12.70

Fifteen feature selection methods are compared in Table 3. Feature selection considering profit can significantly improve predictive ability and obtain the maximum profit. The predicted profit values at *t*-1 to *t*-5 are larger than those of the models that do not consider profit. Compared with the validity indices proposed by Liu et al. [1] and Peña et al. [64], the validity index proposed in this paper has higher classification accuracy and profit, which shows that the features selected by the proposed validity index have higher default identification ability to separate the two types of samples. Sequential backward selection [50] is used in the feature selection process. By comparison, the backward selection algorithm based on the validity index has higher classification accuracy and profit than the backward selection algorithm. By comparison with other feature selection methods, such as sequential forward selection [11] and LASSO [65], the effectiveness of the proposed feature selection method is further verified.

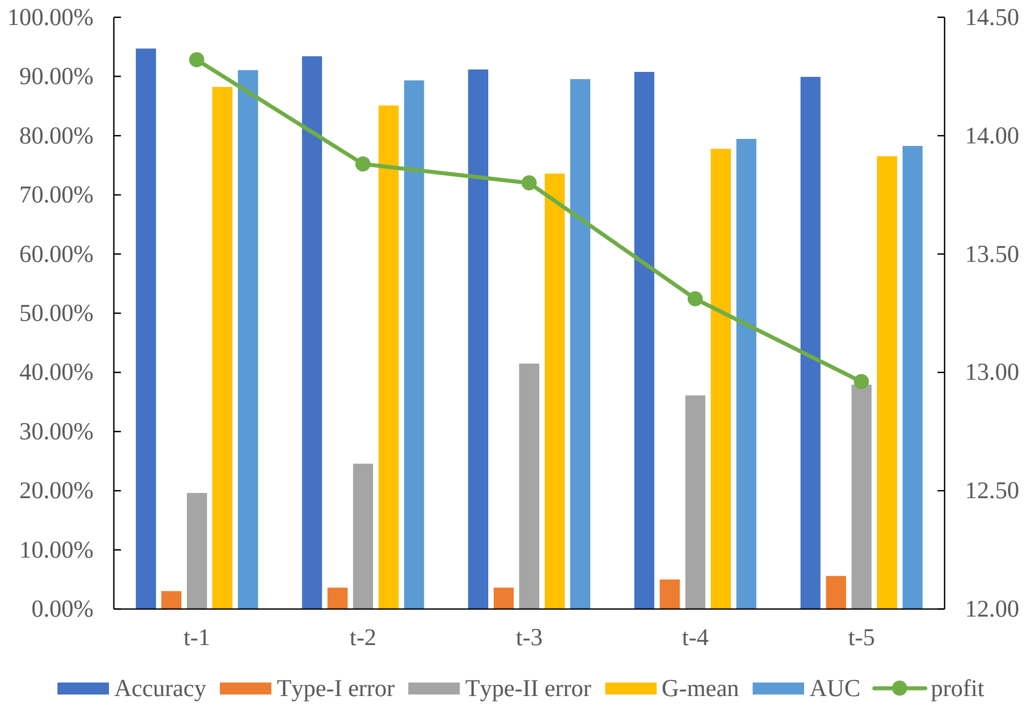

Figure 1. Comparison of classification performance and profit at times *t*-1 to *t*-5.

The following conclusions can be drawn from the comparative analysis of the model classification performance in Figure 1.

First, the order according to the average accuracy is *t*-1> *t*-2>*t*-3> *t*-4>*t*-5. The model proposed in this paper has the strongest predictive ability in the next year, followed by the next two years, and the worst in the next five years. The results fully conform to the law of "the longer the prediction time, the lower the prediction accuracy" [67]. The possible reason is that the longer the time span of feature data X and label Y, the weaker the correlation between them is, leading to worse model classification performance.

Second, after averaging the profits of the five-time windows, the order is *t*-1> *t*-2>*t*-3> *t*-4>*t*-5. The feature combination selected by the model proposed in this paper has the maximum profit-predicting ability in the next year, followed by the next two years, and the worst in the next five years. The results conform to the law that the longer the prediction time, the lower the prediction profit. There are two possible reasons: one is that the accuracy meets this standard due to features and samples, and the other is that lower Type-I and Type-II errors determine a higher profit.

Finally, by comparing the average accuracy and the average profit, it is found that the year with the high profit is not high in accuracy. For example, the profit at time *t*-3 ranks third, but the G-mean ranks last. This shows that the impact of classification performance on profit is not significantly positively correlated but is also affected by other factors, such as interest-bearing liabilities, current liabilities, quick assets, interest rate, etc. Type-II error has a more significant impact on profit. Once the customer defaults, the bank will suffer a more significant loss.

4.3.2. Robustness tests

(1) Data sources.

In this paper, we use three credit datasets, namely, Chinese Farmer, Germany, and Taiwan [68–70], listed in Table 4, to verify the robustness of the proposed model. If the profit and accuracy proposed in this paper are also higher than those of the other models in the above three datasets, the feature selection model is robust.

Chinese Farmer dataset. There are 2,044 observations and 43 features, including 1,816 non-default customers (label 0) and 228 default customers (label 1).

Germany dataset. There are 1,000 observations and 20 features, including 700 non-default customers (label 0) and 300 default customers (label 1).

Taiwan dataset. There are 30,000 observations and 23 features, including 23,364 non-default customers (label 0) and 6,636 default customers (label 1).

The Chinese Farmer dataset contains the loan amount (A), interest rate (r), and outstanding principal and interest (O). The interest loss in Table 1 is $-A\times r$, the interest income is $-A\times r$, the evaded bad debt is O, and the bad debt is O. Because the Germany and Taiwan datasets only contain the loan amount (A) and interest rate (r) set as 5%, the interest loss is $-A\times r$ and the interest income is $-A\times r$. For the evaded bad debt and bad debt, the evaded bad debt is O, and the bad debt is O [71,72].

Table 4. Descriptive statistical analysis of three datasets.

Dataset	Number of features	Sample size	Default samples	Non-default samples
Chinese Farmer	43	2,044	228	1,816
Germany	20	1,000	300	700
Taiwan	23	30,000	6,636	23,364

(2) Comparison of accuracy and profit in different datasets.

The experimental results in Table 5 show that the accuracy of Model 1 (the proposed model) is not the highest, as the accuracy, Type-II error, and G-mean are lower than those of Model 3 in the Chinese Farmer dataset. The accuracy and Type-II error are lower than those of Model 2 in the Germany dataset. The accuracy is lower than that of Model 13, and the G-mean is lower than that of Model 2 in the Taiwan dataset. However, the model in this paper has the highest profit among all 15 models.

There is an inconsistency between accuracy and profit in Table 5, i.e., the model with the highest accuracy does not necessarily obtain the most loan profit. This result is related to the construction of the objective function. The classic classification models take the minimum error as the objective function, guaranteeing high accuracy. In contrast, our model takes the maximum profit as the objective function, ensuring a high-profit value.

Table 5. Comparison of accuracy and profit in different datasets.

Method	Accuracy	Type-I error	Type-II error	G-mean	AUC	Profit
Chinese Farmer						
1	93.20%	1.08%	42.20%	75.44%	**95.44%**	**14.40**
2	92.44%	0.88%	48.50%	71.33%	95.21%	14.25
3	**94.72%**	3.02%	**19.62%**	**88.23%**	91.06%	14.32
4	90.16%	5.12%	38.93%	76.06%	78.12%	13.03
5	91.97%	2.56%	41.89%	75.01%	93.04%	14.18
6	84.50%	9.14%	57.78%	55.43%	84.40%	12.60
7	91.88%	2.53%	43.41%	74.10%	93.14%	14.18
8	92.47%	0.92%	48.32%	71.26%	95.17%	14.01
9	92.18%	2.41%	41.37%	75.58%	92.99%	14.17
10	89.08%	**0.38%**	75.94%	48.54%	92.44%	13.36
11	93.23%	1.22%	41.20%	76.12%	94.61%	14.32
12	91.97%	2.56%	41.89%	75.01%	93.04%	14.18
13	85.14%	10.01%	44.25%	69.13%	86.91%	11.71
14	94.63%	3.12%	19.57%	88.15%	91.12%	14.38
15	89.90%	5.08%	40.74%	74.91%	77.09%	12.95
Germany						
1	92.76%	**1.55%**	42.33%	75.19%	**96.29%**	**14.13**
2	**93.43%**	3.62%	**24.57%**	85.10%	89.35%	13.58
3	91.04%	1.79%	52.68%	67.82%	92.58%	13.96
4	91.56%	3.63%	38.14%	77.12%	91.72%	13.89
5	92.20%	1.72%	45.16%	73.13%	91.24%	13.98
6	**93.43%**	1.93%	35.38%	79.52%	96.17%	13.92

Continued on next page

Method	Accuracy	Type-I error	Type-II error	G-mean	AUC	Profit
Germany						
7	88.99%	2.41%	64.22%	58.66%	86.45%	13.46
8	90.36%	5.12%	38.16%	76.44%	78.36%	12.58
9	90.39%	1.62%	58.63%	63.63%	91.94%	13.83
10	91.15%	3.53%	41.70%	74.77%	90.51%	13.53
11	91.97%	3.57%	35.74%	78.58%	94.02%	13.37
12	93.20%	3.63%	26.45%	84.18%	89.37%	13.34
13	92.76%	3.90%	27.68%	83.32%	92.48%	13.19
14	88.76%	1.96%	68.81%	**55.17%**	93.12%	13.72
15	90.01%	2.58%	56.05%	65.03%	70.41%	13.56
Taiwan						
1	88.47%	1.70%	27.40%	84.48%	90.93%	**13.83**
2	91.66%	2.98%	30.04%	82.39%	**93.16%**	12.88
3	88.93%	1.90%	20.86%	**88.11%**	89.84%	13.78
4	89.20%	2.17%	26.68%	84.69%	87.57%	13.20
5	86.95%	2.64%	28.06%	83.69%	79.67%	13.24
6	90.13%	5.36%	37.81%	76.72%	78.42%	12.71
7	91.18%	4.86%	28.45%	82.51%	80.98%	12.99
8	91.27%	3.43%	**18.88%**	87.51%	90.52%	13.34
9	86.89%	**1.49%**	84.22%	39.43%	72.25%	13.82
10	92.59%	4.32%	26.48%	83.87%	88.22%	13.11
11	89.43%	2.78%	58.96%	63.17%	92.33%	13.65
12	90.77%	5.00%	36.09%	77.92%	79.45%	13.31
13	**93.28%**	3.74%	25.09%	84.92%	89.76%	13.13
14	90.10%	5.65%	36.08%	77.66%	79.13%	12.77
15	89.37%	2.82%	59.19%	62.98%	92.49%	13.56

4.4. Comparative analysis of credit qualifications

The key to credit risk management is to measure the default risk, and the credit score model is an important means to quantify the default risk. The credit score model is directly related to the default risk, which can affect banks' bad debts. In this paper, we construct a credit score model to measure the credit risk of Chinese listed companies and compare and analyze the credit qualifications of different company sizes and industries.

4.4.1. Analysis of companies' credit qualifications in different sizes

The proposed validity index+Profit is the model with the highest profit from t-1 to t-5. The samples at t-m (m = 1, 2, ···, 5) are trained on. The samples of Chinese listed companies in 2018 are substituted into the five classification models trained by the decision tree to obtain the default probability from 2019 to 2023. The mean value of each customer's five-year default probability is taken as the final default probability P_{i1}. P_{i2} is obtained by the mean of the sum of weighted features. The credit score of each customer is calculated by Eq (13), and the credit qualifications in different sizes are further shown in Table 6.

Table 6. Companies' credit qualifications in different sizes.

Company size	Sample size	Credit score	*Mann-Whitney U*		
			large	medium	small and micro
large	2,481	523.51	—	-7.52***	-7.72***
medium	351	533.61		—	1.33
small and micro	593	532.35			—

According to size, the companies are divided into large, medium and small and micro companies [73]. The average credit scores from high to low are medium-sized, small and micro-sized and large-sized. The *Mann-Whitney U* test is used to verify whether there is a significant difference between two different sizes of companies. It is found that there are significant differences between large companies and the other two types of companies, while there is no significant difference between medium-sized companies and small and micro companies.

4.4.2. Analysis of credit qualifications in different industries

We divided the observations into 18 industries according to the China Securities Regulatory Commission's 2012 industry classification guidelines [74]. This article only shows the differences in credit qualifications among the ten industries with large samples (greater than 50).

1). Manufacturing Industry
2). Information Transmission, Software and Information Technology Service Industry
3). Wholesale and Retail Industry
4). Real Estate Industry
5). Electricity, Heat, Gas, Water Production and Supply Industry
6). Construction Industry
7). Transportation, Storage, and Postal Industry
8). Financial Industry
9). Mining Industry
10). Culture, Sports, and Entertainment Industry

The credit scores in column 3 in Table 7 show that the credit qualifications of the information transmission, software, information technology services industry & culture, sports, and entertainment industries are good. The manufacturing and electricity, heat, gas, water production and supply industries are mediocre. The financial industry and real estate industry are poor. A *Mann-Whitney U* test between different industries shows significant differences in credit qualifications between most industries, such as the financial industry and all other industries, the manufacturing industry and the wholesale and retail industry. However, there is no difference in credit qualifications among some industries, such as the manufacturing industry and the electricity, heat, gas, water production and supply industry.

Table 7. Companies' credit qualifications in different industries.

Industry	Sample size	Credit score	*Mann-Whitney U* 1	2	3	4	5	6	7	8	9	10
1	2,163	527.18	—	2.19**	3.38***	6.23***	-0.04	4.39***	3.59***	7.03***	-2.38**	-0.22
2	244	529.75		—	4.28***	6.68***	-1.211	5.03***	-4.45***	7.26***	3.10***	-0.73
3	161	522.75			—	2.70***	-2.20**	-1.49	-0.81	4.28***	-0.36	-1.82*
4	123	517.30				—	4.30***	-0.75	-1.58	-2.37**	-1.38	3.41***
5	104	525.85					—	3.17***	2.69***	5.19***	-1.75*	-0.30
6	99	520.17						—	-0.66	2.61***	-0.87	2.73***
7	90	520.07							—	3.41***	-0.27	-2.18**
8	77	511.08								—	2.93***	4.77***
9	74	522.30									—	-1.79*
10	56	528.23										—

4.4.3. Analysis of credit qualifications based on different production factors

According to different production factors, we divide industries into labor-intensive, capital-intensive and technology-intensive industries [75–78]. This article presents the credit scores of these three industries and examines whether there are significant differences.

Table 8. Companies' credit qualifications in different production factors.

Production factors	Sample size	Credit score	*Mann-Whitney U* labor-intensive	capital-intensive	technology-intensive
labor-intensive	791	526.43	—	2.54**	-5.44***
capital-intensive	904	522.33		—	-6.55***
technology-intensive	565	530.67			—

The credit scores in Table 8 from high to low are technology-intensive industries, labor-intensive industries, and capital-intensive industries. The *Mann-Whitney U* test shows that there are significant differences between any two of the three industries.

5. Conclusions and future work

This work offers the following primary conclusions:

(1) By taking the minimum validity index and the maximum profit as the objective function, the features selected from 3,425 observations of listed companies at times *t*-1 to *t*-5 are 91, 84, 79, 73 and 58, respectively. Forty-three features, such as monetary funds and social contribution per share, have a 5-year predictive ability; 20 features, such as capital reserve and social donation, have a 4-year predictive ability; 15 features, such as whether to disclose environment and sustainable development, have a 3-year predictive ability; and 14 features, such as cash ratio and external guarantee balance, have a 2-year predictive ability.

(2) Compared with 14 methods, the features selected by our proposed method can obtain the maximum profit to ensure classification performance. The processed *t*-1 to *t*-5 data can indicate whether there is a default in the next five years and the profits generated by bank decisions; that is, it can provide at least five years of predictions.

(3) The listed companies' difference test ranks companies' credit qualifications: medium companies > larger companies > small and micro companies. The credit qualifications of the information transmission, software, and information technology services industry and the culture, sports, and entertainment industry are good. The manufacturing industry and the electricity, heat, gas, water production and supply industry are mediocre. The financial industry and real estate industry are poor. The credit qualifications from high to low are technology-intensive industries, labor-intensive industries, and capital-intensive industries. The *Mann-Whitney U* test shows that there are significant differences between any two of the three industries.

The proposed two-stage feature selection method provides a new perspective to obtain the status of Chinese listed companies in a timely manner. This will help lending institutions better identify high-risk customers, reduce default risks and increase loan profitability, thereby enhancing the competitiveness of the entire banking industry.

However, the proposed framework still has some shortcomings, which also represent a direction for future research. (i) Further studies may include using a method for imbalanced data, such as undersampling and oversampling. (ii) Further studies may use fuzzy-based models because crisp data are insufficient to tackle uncertain situations.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

We gratefully acknowledge financial support from the General Program of National Social Science Foundation of China (21BJY065).

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. C. Liu, W. Wang, M. Konan, S. Wang, L. Huang, Y. Tang, et al., A new validity index of feature subset for evaluating the dimensionality reduction algorithms, *Knowl.-Based Syst.*, **121** (2017), 83–98. https://doi.org/10.1016/j.knosys.2017.01.017
2. N. Kozodoi, S. Lessmann, K. Papakonstantinou, Y. Gatsoulis, B. Baesens, A multi-objective approach for profit-driven feature selection in credit scoring, *Decis. Support Syst.*, **120** (2019), 106–117. https://doi.org/10.1016/j.dss.2019.03.011
3. F. Chen, F. Li, Combination of feature selection approaches with SVM in credit scoring, *Expert Syst. Appl.*, **37** (2010), 4902–4909. https://doi.org/10.1016/j.eswa.2009.12.025

4. M. Doumpos, J. R. Figueira, A multicriteria outranking approach for modeling corporate credit ratings: An application of the Electre Tri-nC method, *Omega*, **82** (2019), 166–180. https://doi.org/10.1016/j.omega.2018.01.003
5. D. Mateos-García, J. García-Gutiérrez, J. C. Riquelme-Santos, On the evolutionary weighting of neighbours and features in the k-nearest neighbour rule, *Neurocomputing*, **326** (2019), 54–60. https://doi.org/10.1016/j.neucom.2016.08.159
6. F. N. Koutanaei, H. Sajedi, M. Khanbabaei, A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring, *J. Retail. Consum. Serv.*, **27** (2015), 11–23. https://doi.org/10.1016/j.jretconser.2015.07.003
7. S. Lessmann, B. Baesens, H. V. Seow, L. C. Thomas, Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research, *Eur. J. Oper. Res.*, **247** (2015), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030
8. S. A. Sridharan, Volatility forecasting using financial statement information, *Account. Rev.* **90** (2015), 2079–2106. https://doi.org/10.2308/accr-51025
9. S. Maldonado, J. Pérez, C. Bravo, Cost-based feature selection for support vector machines: An application in credit scoring, *Eur. J. Oper. Res.*, **261** (2017), 656–665. https://doi.org/10.1016/j.ejor.2017.02.037
10. P. Bertolazzi, G. Felici, P. Festa, G. Fiscon, E. Weitschek, Integer programming models for feature selection: New extensions and a randomized solution algorithm, *Eur. J. Oper. Res.*, **250** (2016), 389–399. https://doi.org/10.1016/j.ejor.2015.09.051
11. Y. Xia, C. Liu, Y. Li, N. Liu, A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring, *Expert Syst. Appl.*, **78** (2017), 225–241. https://doi.org/10.1016/j.eswa.2017.02.017
12. S. Jadhav, H. He, K. Jenkins, Information gain directed genetic algorithm wrapper feature selection for credit rating, *Appl. Soft Comput.*, **69** (2018), 541–553. https://doi.org/10.1016/j.asoc.2018.04.033
13. N. Arora, P. D. Kaur, A Bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment, *Appl. Soft Comput.*, **86** (2020), 105936. https://doi.org/10.1016/j.asoc.2019.105936
14. W. Gu, M. Basu, Z. Chao, L. Wei, A unified framework for credit evaluation for internet finance companies: Multi-criteria analysis through AHP and DEA, *Int. J. Inf. Tech. Decis.*, **16** (2017), 597–624. https://doi.org/10.1142/S0219622017500134
15. Z. Li, N. Hou, J. Su, Y. Liu, Model of credit rating of micro enterprise based on fuzzy integration, *Filomat*, **32** (2018), 1831–1842. https://doi.org/10.2298/FIL1805831L
16. A. Karaaslan, K. Ö. Özden, Forecasting Turkey's credit ratings with multivariate grey model and grey relational analysis, *J. Quant. Econ.*, **15** (2017), 583–610. https://doi.org/10.1007/s40953-016-0064-1
17. X. Zhu, J. Li, D. Wu, H. Wang, C. Liang, Balancing accuracy, complexity and interpretability in consumer credit decision making: A C-TOPSIS classification approach, *Knowl.-Based Syst.*, **52** (2013), 258–267. https://doi.org/10.1016/j.knosys.2013.08.004
18. H. Chen, T. Li, X. Fan, C. Luo, Feature selection for imbalanced data based on neighborhood rough sets, *Inform. Sciences*, **483** (2019), 1–20. https://doi.org/10.1016/j.ins.2019.01.041
19. D. Panday, R. C. de Amorim, P. Lane, Feature weighting as a tool for unsupervised feature selection, *Inform. Process. Lett.*, **129** (2018), 44–52. https://doi.org/10.1016/j.ipl.2017.09.005

20. Y. O. Serrano-Silva, Y. Villuendas-Rey, C. Yáñez-Márquez, Automatic feature weighting for improving financial Decision Support Systems, *Decis. Support Syst.*, **107** (2018), 78–87. https://doi.org/10.1016/j.dss.2018.01.005
21. M. Mercadier, J. P. Lardy, Credit spread approximation and improvement using random forest regression, *Eur. J. Oper. Res.*, **277** (2019), 351–365. https://doi.org/10.1016/j.ejor.2019.02.005
22. M. M. Chijoriga, Application of multiple discriminant analysis (MDA) as a credit scoring and risk assessment model, *Int. J. Emerg. Mark.*, **6** (2011), 132–147. https://doi.org/10.1108/17468801111119498
23. L. Kao, C. Chiu, F. Chiu, A Bayesian latent variable model with classification and regression tree approach for behavior and credit scoring, *Knowl.-Based Syst.*, **36** (2012), 245–252. https://doi.org/10.1016/j.knosys.2012.07.004
24. N. Mahmoudi, E. Duman, Detecting credit card fraud by modified Fisher discriminant analysis, *Expert Syst. Appl.*, **42** (2015), 2510–2516. https://doi.org/10.1016/j.eswa.2014.10.037
25. S. Y. Sohn, D. H. Kim, J. H. Yoon, Technology credit scoring model with fuzzy logistic regression, *Appl. Soft Comput.*, **43** (2016), 150–158. https://doi.org/10.1016/j.asoc.2016.02.025
26. M. S. Colak, A new multivariate approach for assessing corporate financial risk using balance sheets, *Borsa Istanb. Rev.*, **21** (2021), 239–255. https://doi.org/10.1016/j.bir.2020.10.007
27. N. Dwarika, The risk-return relationship and volatility feedback in South Africa: a comparative analysis of the parametric and nonparametric Bayesian approach, *Quant. Financ. Econ.*, **7** (2023), 119–146. https://doi.org/10.3934/QFE.2023007
28. Y. Guo, Y. Bai, C. Li, Y. Shao, Y. Ye, C. Jiang, Reverse nearest neighbors Bhattacharyya bound linear discriminant analysis for multimodal classification, *Eng. Appl. Artif. Intel.*, **97** (2021), 104033. https://doi.org/10.1016/j.engappai.2020.104033
29. N. Chukhrova, A. Johannssen, Fuzzy regression analysis: systematic review and bibliography, *Appl. Soft Comput.*, **84** (2019), 105708. https://doi.org/10.1016/j.asoc.2019.105708
30. A. Khashman, Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes, *Expert Syst. Appl.*, **37** (2010), 6233–6239. https://doi.org/10.1016/j.eswa.2010.02.101
31. S. Maldonado, C. Bravo, J. López, J. Pérez, Integrated framework for profit-based feature selection and SVM classification in credit scoring, *Decis. Support Syst.*, **104** (2017), 113–121. https://doi.org/10.1016/j.dss.2017.10.007
32. A. Bequé, S. Lessmann, Extreme learning machines for credit scoring: An empirical evaluation, *Expert Syst. Appl.*, **86** (2017), 42–53. https://doi.org/10.1016/j.eswa.2017.05.050
33. X. Zhang, Y. Han, W. Xu, Q. Wang, HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture, *Inform. Sciences*, **557** (2021), 302–316. https://doi.org/10.1016/j.ins.2019.05.023
34. M. Ala'raj, M. F. Abbod, M. Majdalawieh, Modelling customers credit card behaviour using bidirectional LSTM neural networks, *J. Big Data*, **8** (2021), 69. https://doi.org/10.1186/s40537-021-00461-7
35. F. Zhao, Y. Lu, X. Li, L. Wang, Y. Song, D. Fan, et al., Multiple imputation method of missing credit risk assessment data based on generative adversarial networks, *Appl. Soft Comput.*, **126** (2022), 109273. https://doi.org/10.1016/j.asoc.2022.109273

36. S. Asadi, S. E. Roshan, A bi-objective optimization method to produce a near-optimal number of classifiers and increase diversity in Bagging, *Knowl.-Based Syst.*, **213** (2021), 106656. https://doi.org/10.1016/j.knosys.2020.106656
37. Y. C. Chang, K. H. Chang, G. J. Wu, Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions, *Appl. Soft Comput.*, **73** (2018), 914–920. https://doi.org/10.1016/j.asoc.2018.09.029
38. Y. Xia, J. Zhao, L. He, Y. Li, X. Yang, Forecasting loss given default for peer-to-peer loans via heterogeneous stacking ensemble approach, *Int. J. Forecasting*, **37** (2021), 1590–1613. https://doi.org/10.1016/j.ijforecast.2021.03.002
39. F. Shen, X. Zhao, G. Kou, F. E. Alsaadi, A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique, *Appl. Soft Comput.*, **98** (2021), 106852. https://doi.org/10.1016/j.asoc.2020.106852
40. J. Forough, S. Momtazi, Ensemble of deep sequential models for credit card fraud detection, *Appl. Soft Comput.*, **99** (2021), 106883. https://doi.org/10.1016/j.asoc.2020.106883
41. A. Belhadi, S. S. Kamble, V. Mani, I. Benkhati, F. E. Touriki, An ensemble machine learning approach for forecasting credit risk of agricultural SMEs' investments in agriculture 4.0 through supply chain finance, *Ann. Oper. Res.*, **2021** (2021), 1–29. https://doi.org/10.1007/s10479-021-04366-9
42. C. Jiang, W. Xiong, Q. Xu, Y. Liu, Predicting default of listed companies in mainland China via U-MIDAS Logit model with group lasso penalty, *Financ. Res. Lett.*, **38** (2021) 101487. https://doi.org/10.1016/j.frl.2020.101487
43. J. Donovan, J. Jennings, K. Koharki, J. Lee, Measuring credit risk using qualitative disclosure, *Rev. Account. Stud.*, **26** (2021), 815–863. https://doi.org/10.1007/s11142-020-09575-4
44. N. Camanho, P. Deb, Z. Liu, Credit rating and competition, *Int. J. Financ. Econ.*, **27** (2022) 2873–2897. https://doi.org/10.1002/ijfe.2303
45. H. Zhang, Y. Shi, X. Yang, R. Zhou, A firefly algorithm modified support vector machine for the credit risk assessment of supply chain finance, *Res. Int. Bus. Financ.*, **58** (2021), 101482. https://doi.org/10.1016/j.ribaf.2021.101482
46. Z. Ma, W. Hou, D. Zhang, A credit risk assessment model of borrowers in P2P lending based on BP neural network, *PLOS one*, **16** (2021), e0255216. https://doi.org/10.1371/journal.pone.0255216
47. W. Hou, X. Wang, H. Zhang, J. Wang, L. Li, A novel dynamic ensemble selection classifier for an imbalanced data set: an application for credit risk assessment, *Knowl.-Based Syst.*, **208** (2020), 106462. https://doi.org/10.1016/j.knosys.2020.106462
48. F. O. Sameer, M. R. A. Bakar, A. A. Zaidan, B. B. Zaidan, A new algorithm of modified binary particle swarm optimization based on the Gustafson-Kessel for credit risk assessment, *Neural Comput. & Applic.*, **31** (2019), 337–346. https://doi.org/10.1007/s00521-017-3018-4
49. J. Traczynski, Firm default prediction: A Bayesian model-averaging approach, *J. Financ. Quant. Anal.*, **52** (2017), 1211–1245. https://doi.org/10.1017/S002210901700031X
50. Y. Zhou, W. Zhang, J. Kang, X. Zhang, X. Wang, A problem-specific non-dominated sorting genetic algorithm for supervised feature selection, *Inform. Sciences*, **547** (2021), 841–859. https://doi.org/10.1016/j.ins.2020.08.083

51. Y. Zhu, L. Zhou, C. Xie, G. Wang, T. V. Nguyen, Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach, *Int. J. Prod. Econ.*, **211** (2019), 22–33. https://doi.org/10.1016/j.ijpe.2019.01.032
52. G. Chi, B. Meng, Debt rating model based on default identification: Empirical evidence from Chinese small industrial enterprises, *Manage. Decis.*, **57** (2019), 2239–2260. https://doi.org/10.1108/MD-11-2017-1109
53. A. Bequé, K. Coussement, R. Gayler, S. Lessmann, Approaches for credit scorecard calibration: An empirical analysis, *Knowl.-Based Syst.*, **134** (2017), 213–227. https://doi.org/10.1016/j.knosys.2017.07.034
54. R. Geng, I. Bose, X. Chen, Prediction of financial distress: An empirical study of listed Chinese companies using data mining, *Eur. J. Oper. Res.*, **241** (2015), 236–247. https://doi.org/10.1016/j.ejor.2014.08.016
55. R.P. Baghai, B. Becker, Reputations and credit ratings: Evidence from commercial mortgage-backed securities, *J. Financ. Econ.*, **135** (2020), 425–444. https://doi.org/10.1016/j.jfineco.2019.06.001
56. N. Chai, B. Wu, W. Yang, B. Shi, A multicriteria approach for modeling small enterprise credit rating: evidence from China, *Emerg. Mark. Financ. Tr.*, **55** (2019), 2523–2543. https://doi.org/10.1080/1540496X.2019.1577237
57. L. Li, J. Yang, X. Zou, A study of credit risk of Chinese listed companies: ZPP versus KMV, *Appl. Econ.*, **48** (2016), 2697–2710. https://doi.org/10.1080/00036846.2015.1128077
58. M. Livingston, W. P. Poon, L. Zhou, Are Chinese credit ratings relevant? A study of the Chinese bond market and credit rating industry, *J. Bank. Financ.*, **87** (2018), 216–232. https://doi.org/10.1016/j.jbankfin.2017.09.020
59. M. S. Uddin, G. Chi, M. A. A. Janabi, T. Habib, Leveraging random forest in micro-enterprises credit risk modelling for accuracy and interpretability, *Int. J. Financ. Econ.*, **27** (2022), 3713–3729. https://doi.org/10.1002/ijfe.2346
60. B. Meng, G. Chi, Evaluation index system of green industry based on maximum information content, *Singap. Econ. Rev.*, **63** (2018), 229–248. https://doi.org/10.1142/S0217590817400094
61. Z. Li, S. Liang, X. Pan, M. Pang, Credit risk prediction based on loan profit: Evidence from Chinese SMEs, *Res. Int. Bus. Financ.*, **67** (2024), 102155. https://doi.org/10.1016/j.ribaf.2023.102155
62. J. A. Ohlson, Financial ratios and the probabilistic prediction of bankruptcy, *J. Account. Res.*, **18** (1980), 109–131. https://doi.org/10.2307/2490395
63. D. G. Kirikos, An evaluation of quantitative easing effectiveness based on out-of-sample forecasts, *National Accounting Review*, **4** (2022), 378–389. https://doi.org/10.3934/NAR.2022021
64. M. Peña, M. Cerrada, D. Cabrera, R.-V. Sánchez, Fast feature selection based on cluster validity index applied on data-driven bearing fault detection, *2020 IEEE ANDESCON*, Quito, Ecuador, 2020, 1–6. http://doi.org/10.1109/ANDESCON50619.2020.9272146
65. Y. Zhou, M. S. Uddin, T. Habib, G. Chi, K. Yuan, Feature selection in credit risk modeling: an international evidence, *Econ. Res.-Ekon. Istraž.*, **34** (2021), 3064–3091. http://hdl.handle.net/10.1080/1331677X.2020.1867213
66. F. Garrido, W. Verbeke, C. Bravo, A Robust profit measure for binary classification model evaluation, *Expert Syst. Appl.*, **92** (2018), 154–160. https://doi.org/10.1016/j.eswa.2017.09.045

67. T. M. Luong, H. Scheule, Benchmarking forecast approaches for mortgage credit risk for forward periods, *Eur. J. Oper. Res.*, **299** (2022), 750–767. https://doi.org/10.1016/j.ejor.2021.09.026
68. C. Bai, B. Shi, F. Liu, J. Sarkis, Banking credit worthiness: Evaluating the complex relationships, *Omega*, **83** (2019), 26–38. https://doi.org/10.1016/j.omega.2018.02.001
69. M. Z. Abedin, C. Guotai, F. E. Moula, A. S. Azad, M. S. U. Khan, Topological applications of multilayer perceptrons and support vector machines in financial decision support systems, *Int. J. Financ. Econ.*, **24** (2019), 474–507. https://doi.org/10.1002/ijfe.1675
70. Q. Lan, X. Xu, H. Ma, G. Li, Multivariable data imputation for the analysis of incomplete credit data, *Expert Syst. Appl.*, **141** (2020), 112926. https://doi.org/10.1016/j.eswa.2019.112926
71. S. Wu, X. Gao, W. Zhou, COSLE: Cost sensitive loan evaluation for P2P lending, *Inform. Sciences*, **586** (2022), 74–98. https://doi.org/10.1016/j.ins.2021.11.055
72. N. Kozodoi, J. Jacob, S. Lessmann, Fairness in credit scoring: Assessment, implementation and profit implications, *Eur. J. Oper. Res.*, **297** (2022) 1083–1094. https://doi.org/10.1016/j.ejor.2021.06.023
73. X. Su, S. Zhou, R. Xue, J. Tian, Does economic policy uncertainty raise corporate precautionary cash holdings? Evidence from China, *Account. Financ.*, **60** (2020), 4567–4592. https://doi.org/10.1111/acfi.12674
74. L. He, L. Zhang, Z. Zhong, D. Wang, F. Wang, Green credit, renewable energy investment and green economy development: Empirical analysis based on 150 listed companies of China, *J. Clean. Prod.*, **208** (2019), 363–372. https://doi.org/10.1016/j.jclepro.2018.10.119
75. V. Hlasny, Market and home production earnings gaps in Russia, *National Accounting Review*, **5** (2023), 108–124. https://doi.org/10.3934/NAR.2023007
76. Y. Huang, Y. Ma, Z. Yang, Y. Zhang, A fire sale without fire: An explanation of labor-intensive FDI in China, *J. Comp. Econ.*, **44** (2016), 884–901. https://doi.org/10.1016/j.jce.2016.04.007
77. Z. Zhao, K. H. Zhang, FDI and industrial productivity in China: Evidence from panel data in 2001-06, *Rev. Dev. Econ.*, **14** (2010), 656–665. https://doi.org/10.1111/j.1467-9361.2010.00580.x
78. Y. Zhang, L. Ma, Board faultlines, innovation strategy decisions, and faultline activation: Research on technology-intensive enterprises in Chinese A-share companies, *Front. Psychol.*, **13** (2022), 855610. https://doi.org/10.3389/fpsyg.2022.855610