Hindawi Scientific Programming Volume 2021, Article ID 6276155, 9 pages https://doi.org/10.1155/2021/6276155



Research Article

Financial Credit Risk Control Strategy Based on Weighted Random Forest Algorithm

Guo Yangyudongnanxin

Business School, Hunan International Economics University, Hunan, Changsha 410205, China

Correspondence should be addressed to Guo Yangyudongnanxin; 201914040036@zknu.edu.cn

Received 1 September 2021; Revised 1 October 2021; Accepted 4 October 2021; Published 19 October 2021

Academic Editor: Rahman Ali

Copyright © 2021 Guo Yangyudongnanxin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the effectiveness of financial credit risk control, a financial credit risk control strategy based on weighted random forest algorithm is proposed. The weighted random forest algorithm is used to classify the financial credit risk data, construct the evaluation index system, and use the analytic hierarchy process to evaluate the financial credit risk level. The targeted risk control strategies are taken according to different risk assessment results. We compared the proposed method with two other methods, and the experimental results show that the proposed method has higher classification accuracy of financial credit data and the risk assessment threshold is basically consistent with the actual results.

1. Introduction

In recent years, with the rapid development of Internet finance, online credit has developed very rapidly and its participants are becoming more and more diversified. Online credit helps more and more users with emergency and short-term capital turnover. However, in the process of opening user loan business, the total amount of loans is increasing, the default probability of users is rising, and the risk of user loans is gradually revealed. Therefore, in-depth research on user loan risk and evaluation of the scientificity and rationality of the risk are extremely important for microonline credit enterprises to prevent Internet financial risk [1-3]. In the study of user loan risk prediction and analysis, it is found that the user's data have the following three important characteristics: unbalanced distribution, a large amount of noise, and high-dimensional characteristics. The user loan risk is caused by the interaction of the characteristics of multiple different dimensions of users. For the traditional statistical methods, only a single feature or a small number of features can study the relationship between user loan risk. Therefore, it is a challenge to the traditional statistical methods [4].

With the rapid development of the financial industry and the Internet industry, Internet finance, as an emerging

business model, has gradually come into everyone's sight. The online business volume of Internet finance such as medium and small loans is also increasing rapidly with the growth of Internet scale. The loan applications of individuals, businesses, and even enterprises are gradually moving towards rapidity and paperless [5]. In the early days, loan default forecasts relied entirely on manual review. This forecasting method is only suitable for small-scale credit audit, and there is nothing to do with large-scale loan audit. In order to solve the uncertainty of human factors in the process of manual audit, a method to judge whether to make a loan according to detailed rules is proposed by financial institutions. This method gives a conclusion according to the rules, so that the credit auditor can judge whether to lend money according to clear indicators. At the same time, the proposal of this method greatly reduces the requirements for the credit auditor, and ordinary staff can start credit work after relevant training. In addition, many researchers have proposed data mining models to study credit default [6].

The random forest (RF) algorithm has been very successful in general purpose classification and as a regression method. It is an machine learning algorithm that could be applied to large-scale problems that involve big data and could be easily adapted to various ad-hoc learning tasks [7–9]. The voting mechanism of random forest algorithm

assumes the same weight for all of the base classifiers, which is not always the case. Some of them may have lower weights than the others, and all of the trees may not have the same ability to make decisions. Therefore, the weighted RF algorithm performs better than RF algorithm in most cases [10, 11].

In order to control the financial credit risk more effectively, a financial credit risk control strategy based on weighted random forest algorithm is proposed.

2. Related Work

The study in [12] proposed a financial credit risk assessment method based on particle swarm optimization algorithm. On the basis of fully demonstrating the supply chain financial risk characteristic index system, a binary particle swarm optimization algorithm is used to optimize the feature subset, and the support vector machine (SVM) parameters are co-optimized to obtain the financial credit risk classification assessment results. The work in [13] proposed the financial credit risk assessment method of stack noise reduction self-coding network, fully considered the correlation between data features, improved the stack noise reduction self-coding neural network model, introduced the truncated Karhunen lo è ve expansion as the noise input term, and eliminated the noise data in the financial credit risk data to obtain more effective evaluation results. The study in [14] proposed a financial credit risk assessment method based on xgbfs, which uses a series of data preprocessing methods and embedded feature selection method xgbfs (xgboost feature selection) to reduce the user's credit data dimension, train the xgboost assessment model, and finally realize the user's credit risk assessment. The work in [15] proposed that integration of supervised and unsupervised machine learning strategies will produce better results than using only one of them. They proposed a system for credit risk assessment by integrating the supervised and unsupervised learning strategies at the consensus stage and the dataset clustering stage. The study in [16] concluded that traditional approaches used for forecasting credit risks are not well suited to help the financial institutions and they need ML-based techniques for forecasting credit risk. They proposed a hybrid ensemble machine learning approach by incorporating two classic machine learning approaches, the RS (random subspace), and multiboosting. The work in [17] used the standard probit algorithm along with various machine learning algorithms like neural networks and KNN. They achieved a lower error rate with the machine learning techniques as compared to the classic methods used for credit risk assessment. The work in [18] used a loan dataset of a commercial bank to test five different machine learning algorithms including KNN, DT, RF, NB, and logistic regression for credit risk assessment. Their results show that the random forest algorithm performs better than the other algorithms tested.

3. Financial Credit Risk Control Based on Weighted Random Forest Algorithm

In order to effectively control the financial credit risk, it is necessary to accurately classify the financial credit risk level, classify the financial credit risk data by using the weighted random forest algorithm, and construct the credit risk evaluation model. Targeted control measures shall be taken for different financial credit risk levels.

3.1. Data Classification. Decision tree is a machine learning algorithm with tree structure, which is composed of root node, nonleaf node, and leaf node. Its structure is shown in Figure 1.

The decision tree is a recursive structure. The root node carries out data training and prediction from top to bottom. According to different similarity calculation standards, subsets with certain similarity in the data are divided to generate multiple branches. When reaching the leaf node of the decision tree, the division is stopped. Leaf nodes are set according to the maximum tree depth or the minimum number of leaves. One leaf node represents one classification result of the data. The decision tree can be divided into decision trees based on information entropy, information gain, information gain rate, and Gini impurity according to the different basis of feature division. Assuming that the training sample set D with N financial credit data can be divided into K categories, it establishes the calculation formulas of information entropy, information gain, information gain rate, and Gini impurity of different decision trees as follows:

$$H(D) = -\sum_{k=1}^{K} \frac{n_k}{N} \log_2 p\left(\frac{n_k}{N}\right),\tag{1}$$

where H(D) denotes information entropy, n_k denotes the number of k_th category, and p represents the probability. Similarly, for information gain, we have

$$g(D, a) = H(D) - \sum_{\nu=1}^{V} \frac{|D^{\nu}|}{D} H(D^{\nu}).$$
 (2)

For information gain rate, we have

$$g_R(D, a) = \frac{g(D, a)}{-\sum_{\nu=1}^{V} |D^{\nu}| / D \log_2(|D^{\nu}| / D)}.$$
 (3)

For Gini impurity,

Gini(D) =
$$1 - \sum_{k=1}^{K} \left(\frac{n_k}{N}\right)^2$$
. (4)

Here, n_k represents the number of kth category, ν represents the number of subsets of the sample set divided according to feature a, and the value range is [1, V] [19].

Random forest is an integrated learning algorithm based on decision tree. It randomly generates multiple irrelevant decision trees. Each decision tree learns and predicts independently. These predictions are synthesized into single prediction by voting. The category with the most votes is the model prediction result, and its result is better than the decision tree.

Suppose the input financial credit training dataset is D, the number of iterations of the decision tree is M, and the generation steps of the random forest are as follows:

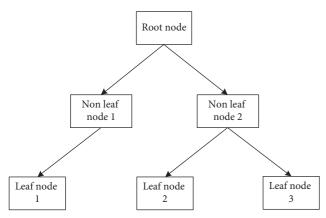


FIGURE 1: Structure diagram of the decision tree.

- (1) Conduct the mth sampling on the sample training set, where m is an integer and the value range is [1, M]. Randomly collect n times to obtain the training set D_m containing n samples.
- (2) When the decision tree is divided into nodes, all n input variables do not fully participate in node splitting, but $k(k \le n)$ random feature variables are randomly selected, the value of k is generally $2 \log_2 n + 1$, the best feature among the k features is split as a node, and the mth decision tree G_m generated by classification and regression tree CART algorithm is trained.
- (3) The CART decision tree is divided into features based on Gini impurity. When the Gini impurity is smaller, the representative impurity is smaller, and its features are better. Finally, M CART decision trees are generated to form a random forest.
- (4) Determine the type of data by calculating the number of votes. The flow of random forest algorithm is shown in Figure 2.

The decision tree algorithm is fast and easy to understand, but it is easy to over fit, and when dealing with unbalanced data, the feature division tends to choose features with more values. Compared with the decision tree algorithm, random forest has high accuracy and is not easy to over fit, but in unbalanced datasets, the classification accuracy of a few classes is still not high. Aiming at the classification problem of unbalanced datasets, an accurate classification method based on weighted random forest lifting is proposed.

In order to minimize the overall error rate, the machine learning algorithm will ignore the classification of classes with less data, and it is easy to form a model conducive to the classification of most classes during training [2]. In order to improve this situation, the weighted random forest follows the idea of cost sensitive learning, increases the influence of fewer classes by giving more weight to fewer classes, and balances the relationship between samples, which can make the generated model more suitable for unbalanced data and improve the accuracy of classification results of a few data.

Class weight is mainly reflected in the following:

(1) The growth process of decision tree, the reduction Δgi of weighted Gini impurity GI is used to find the optimal division feature. The larger the value, the smaller the representative impure, and the better the separation result. The calculation formula is

$$GI(J) = \frac{\sum_{i=1}^{K} (n_i W_i)^2}{\sum_{i=1}^{K} n_i W_i},$$

$$\Delta gi = GI(J) - GI(J_L) - GI(J_R).$$
(5)

Here, K represents the total number of categories, J represents the sample set at the nonseparated node, J_L represents the separated left node sample set, J_R represents the separated node sample set, n_i represents the number of various samples in the node, and W_i represents the class weight value assigned to each class.

(2) When determining the category at the leaf node, the final classification result c is determined by combining the weighted voting of each decision tree. The calculation formula is

$$c = \arg\max_{i} (n_i W_i). \tag{6}$$

Through the above calculation, the classification of financial credit data for subsequent financial credit risk assessment is completed.

3.2. Financial Credit Risk Assessment

3.2.1. Construction of Financial Credit Risk Evaluation Index (1) Principles for Construction of the Evaluation Index System

(1) Principle of comprehensiveness

The evaluation index system of personal credit risk of commercial banks must cover all factors affecting credit risk control. On the one hand, the impact of personal factors on credit risk should be considered; on the other hand, the impact of social environmental factors of lenders should be considered. Only by ensuring the comprehensiveness of the evaluation index system we can ensure the accuracy of credit evaluation of borrowers [20, 21].

(2) System optimization principle

Ensuring the comprehensiveness of the evaluation index system does not mean that the more indicators the better. The selection of the index system must take into account the system optimization principle, that is, select important indicators, and eliminate those redundant indicators with the same impact level, so that the whole index system can fully reflect the actual situation of the lender and form an interconnected unity.

(3) Operability

The selection of personal credit risk evaluation indicators must be combined with the actual situation, select those indicators that can obtain the actual data, and ensure the reliability and accuracy of these data.

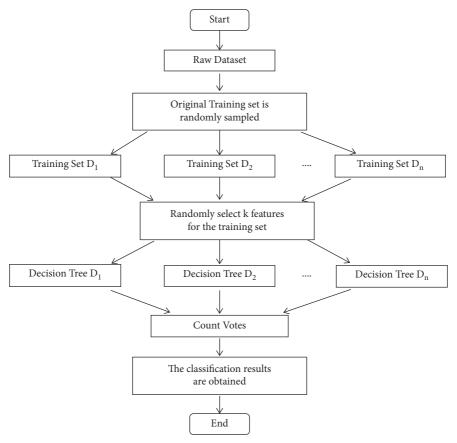


FIGURE 2: Flow chart of the random forest algorithm.

If some indicators cannot obtain the original data, or the obtained data are inaccurate, it has no practical significance in the evaluation process and may even cause deviation in the evaluation results.

(4) Dynamic principle

After the completion of the personal credit risk evaluation index system, it is not invariable. It can change dynamically according to the actual situation. Specifically, the indicators can be adjusted according to the lender's personal basic situation or changes in the economic environment.

(5) Principle of legality and fairness

Legitimacy means that the establishment of the index system must be consistent with the national macroeconomic policies and comply with the provisions of relevant laws and regulations; fairness means that the establishment of personal credit risk evaluation index system is objective and fair. It is based on objective facts, which can scientifically and accurately reflect the basic attributes of commercial banks and loan applicants and does not favor either party.

The personal credit risk evaluation index system is selected by analyzing the evaluation index system established by relevant research institutes at home and abroad, drawing on the relevant experience of domestic and foreign experts on personal credit risk evaluation research and combining

the possibility of obtaining correct data for each specific index, a set of evaluation index system including personal basic information. The second level index system of personal credit risk evaluation was based on three first level indexes of economic status and reputation status. The specific index system is shown in Table 1.

(2) Basic Personal Information Indicators

(1) Age

Generally, the probability of default of the lender decreases with the increase in age. Borrowers under the age of 20 may have no fixed economic income, but there are very few borrowers in this age group, so the risk of default is not high; at the age of 35–55, the lender in this age group has generally made achievements in his career, and his income is relatively stable, and his ability to repay the loan is generally strong; after the age of 55, with the growth of age, the expenditure on medical treatment and life may gradually increase, and the default risk of this part of people may be relatively high.

(2) Marital status

Statistics show that there is not much difference in loan default rate between unmarried borrowers and married borrowers, and married borrowers are slightly lower.

(3) Gender

Serial number	Primary index	Secondary index	Risk factor marker
		Age	u_{11}
1	Basic personal information	Marital status	u_{12}
		Gender	u_{13}
		Education level	u_{14}
		Occupation type	u_{15}
		Duties	u_{16}
2	Economic situation	Monthly income	u_{21}
		Ratio of average monthly repayment to monthly disposable income	u_{22}
		Loan term	u_{23}^{-1}
		Loan amount	u_{24}
		Bank account records	u_{25}
3	Personal honor status	History of loans, overdrafts, and guarantees	u_{31}
		Personal judicial records	u_{32}^{31}
		Other bad reputation records	u_{22}

TABLE 1: Credit risk evaluation index system.

Combining the two groups of statistical data on male and female loan default rates, it can be found that this group of statistical data shows that the male lender default rate is higher than the female lender default rate, and the difference in default rate of this group of data is about 6%.

(4) Education level

Usually, with the continuous improvement of the lender's education, the probability of default will decrease. The personal quality of the lender is generally positively correlated with the level of education. The lender with a high level of education may have stronger self-restraint ability, so his personal reputation is relatively high.

(5) Occupation type

The occupation type of the applicant is closely related to the bank's decision whether to lend to him. The unemployed will not become the loan object of the commercial bank. Borrowers have different occupations, and their income also varies greatly. Borrowers with good career prospects and stable income will generally become potential customers of commercial banks.

(6) Position

The loan default rate is also related to the position of the lender. The lender with relatively high position may pay more attention to personal reputation, so the probability of default may be low.

(3) Economic Indicators.

(1) Monthly income

It is of great significance to judge the default risk of loan applicants. Generally speaking, it is considered that the higher the monthly income, the lower the default risk, otherwise the higher. When the loan applicant has the same willingness to repay, the higher the income level, the stronger the repayment ability. There is a positive correlation between the income and the borrower's repayment ability.

(2) Proportion of monthly repayment in monthly disposable income

The proportion of monthly repayment in monthly disposable income is also an important indicator to measure the lender's repayment ability. The monthly repayment amount refers to the amount of loan interest and principal repaid by the lender to the bank every month, and the monthly disposable income is the remaining income after the lender's monthly income minus the amount of loan interest and principal repaid to the bank every month. Theoretically, the ratio value fluctuates between 0.1. The smaller the ratio value, the better the financial situation of the borrower.

(3) Loan term

The loan term refers to the time period from the time when the commercial bank issues the loan to the lender to the time when the lender pays off all the loan and interest. The length of the loan term is closely related to the loan interest rate. Generally, the longer the loan term, the higher the loan interest rate. The probability of default is relatively high for customers with long loan term than short loan term. The longer the loan term, the worse the liquidity and the higher the risk.

(4) Loan amount

The loan amount is closely related to the loan applicant's application amount and the applicant's credit rating. Generally, lenders with relatively small loan amount have relatively low probability of default, while customers with high probability of default are mainly concentrated in the range with relatively high loan amount. Therefore, including the loan amount in the personal credit risk evaluation

index system is of great significance for commercial banks to adopt appropriate risk aversion measures according to the loan amount.

(5) Bank account records

Bank account records mainly refer to the records of customers opening accounts and handling bank cards in this commercial bank. Generally, the default risk of users who do not open an account in the loan application bank is higher than that of the lender who has opened an account.

(4) Personal Reputation Indicators

(1) Historical credit history

Historical credit records refer to the past loan repayment, credit card overdraft information, guarantee, and other information of the loan applicant. Through the review of the loan applicant's historical credit records, we can preliminarily judge the credit situation of the loan applicant and the probability of default risk in the future.

(2) Personal judicial records

The review of personal judicial records is mainly to investigate whether the loan applicant has a criminal record in the past, which is only a supplementary indicator.

(3) Other records of reputation damage

This includes personal reputation indicators other than historical credit records and personal judicial records, including whether the lender has tax evasion, arrears of water, electricity costs, etc.

3.2.2. Financial Credit Risk Assessment Based on Principal Component Analysis. The principal component analysis method is used for financial credit risk assessment. The principal component analysis method is the basic means of multivariable data analysis. Its main idea is map a group of multivariable information to several principal components, analyze and process the principal components as the overall information, and there is no linear correlation between the principal components; therefore, the dimension of the original data is reduced, the redundant information is effectively eliminated, and the noise in the original information is effectively suppressed. It is especially suitable for dealing with the situation that there are many and complex original data samples. It can effectively reduce the dimension and denoise and improve the convergence accuracy and risk assessment accuracy of financial and credit data.

If *X* is an $m \times n$ -dimensional data sample set, the input is $X(n \times m)$ and the output is $Y(n \times m)$. The kth principal component is $P_k = (P_{1k}, P_{2k}, ..., P_{nk})^T$, k = 1, 2, ..., m; the principal component matrix (i.e., principal component score matrix) is the crossproduct of X and principal component load:

$$T = XP. (7)$$

Here, the single variable in the score matrix is $T_{ij} = \sum_{k=0}^{m} P_{kj} X_{ik}$ and the $T_j = \sum_{k=0}^{m} X_k P_{kj}$ variable is j.

(1) Data Standardization Processing. The original financial loan risk data shall be processed as follows:

$$X'_{ij} = \frac{X_{ij} - M_{j}}{S_{j}},$$

$$M_{j} = \frac{1}{n} \sum_{1}^{n} X_{ij},$$

$$S_{j} = \sqrt{\left[\frac{1}{n-1} \sum_{1}^{n} (X_{ij} - M_{j})^{2}\right]}.$$
(8)

Here, X_{ij}' represents the standardized new financial credit risk data, and M_j and S_j represent the arithmetic mean and standard deviation of a column of the original financial credit risk data, respectively.

(2) Calculation of Characteristic Matrix D of the Original Matrix

$$D = X^T X. (9)$$

The covariance is as follows:

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1p} \\ d_{21} & d_{22} & \cdots & d_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ d_{p1} & d_{p2} & \cdots & d_{pp} \end{bmatrix},$$
(10)

$$d_{ij} = \frac{\sum_{k=1}^{n} (x_{ki} - \overline{x}_i) (x_{kj} - \overline{x}_j)}{\sqrt{\sum_{k=1}^{n} (x_{ki} - \overline{x}_i)^2 \sum_{k=1}^{n} (x_{kj} - \overline{x}_j)^2}}.$$

(3) Calculation of Eigenvector P and Eigenroot

$$DP = P\lambda.$$
 (11)

When only the *j*th eigenvalue is considered, there is $DP_j = P_j \lambda_j$, that is, $|D - \lambda_j I| = 0$. Solving λ and arranging the periods in order of size, that is, $\lambda_1 \ge \lambda_2 \ge, \ge \lambda_p \ge 0$. Then, the eigenvector P corresponding to each eigenvalue can be obtained, and the eigenequation can be solved.

(4) Calculation of the Contribution Rate and Total Contribution Rate of Single Principal Component. The principal component contribution rate and total contribution rate are

$$\frac{\lambda_j}{\sum_{i=1}^p \lambda_i}, \\
\frac{\sum_{i=1}^m \lambda_i}{\sum_{i=1}^p \lambda_i}.$$
(12)

(5) Selection of the Principal Components That Can Replace the Original Sample Data. The calculation formula of the principal component is

Serial	Age	Education	Occupation	Monthly	Loan	Loan	Historical credit	Output
number		level	type	income	term	amount	history	value
1	24	4	4	300	10	100000	0	0
2	40	3	5	5500	15	200000	0	0
3	35	5	5	6000	20	250000	1	0
4	23	3	3	2500	15	100000	3	1
5	33	3	4	3800	15	150000	0	0
6	50	1	3	2500	10	120000	1	1
7	32	3	1	10000	30	300000	1	1
8	28	4	6	3300	5	60000	2	1
9	49	1	3	3200	10	130000	2	1
10	26	5	4	4500	20	300000	0	0
30	38	2	4	3800	15	200000	1	0
31	42	2	3	4100	20	350000	0	1
32	27	4	3	2000	30	300000	2	1
33	22	4	6	1700	30	200000	0	0
34	40	4	1	20000	5	800000	0	1
35	25	3	3	1900	30	3500	2	1
36	25	4	3	2100	20	300000	0	0
37	40	3	5	4300	20	330000	0	0
38	38	5	4	5500	30	450000	0	0
39	21	3	3	1600	30	250000	0	1
40	26	4	3	2200	30	400000	3	1

TABLE 2: Training samples.

$$l_{ij} = p(z_i, x_j) = \sqrt{\lambda_i} p_{ij}, \quad (i, j = 1, 2, ..., p).$$
 (13)

After the load of each principal component is obtained, the score matrix of financial credit risk evaluation is obtained:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ z_{11} & z_{12} & \cdots & z_{2m} \end{bmatrix}.$$
(14)

According to the calculation results of financial credit risk, the evaluation of financial credit risk can be completed, so as to formulate targeted financial credit risk control strategies.

4. Experimental Verification

4.1. Experimental Setup. We used Python's machine learning library Scikit-learn [22] to do our experiments. Scikit-learn is an open source machine learning library written in Python that provides efficient implementations of various machine learning algorithms. It takes advantage of Python interactivity and modularity to supply fast and easy prototyping [22, 23]. The ML library is used in Windows environment, with a standalone system having specification of Intel® Core™ i3-4010U CPU @ 1.70 GHz, to simulate the credit risk assessment problem.

4.2. The Dataset. The sample data selected in this paper are the real data from the personal credit database of a branch of a commercial bank. In this paper, 75 data samples are randomly selected in the personal credit database. Among them, there are individual data missing or the data are obviously in line with

the actual situation. After removing these invalid samples, 61 valid samples are determined. There are 40 learning samples and 21 test samples. The training samples are shown in Table 2.

The quantitative table of qualitative indicators is shown in Table 3.

In order to fully verify the performance of the proposed method, comparative verification experiments are carried out. The experimental scheme is set as follows: taking the accuracy of financial credit risk data classification and risk assessment as the experimental comparison index, the proposed method is compared with the financial credit risk assessment method based on particle swarm optimization algorithm proposed in [12] and the financial credit risk assessment method based on stack noise reduction self-coding network proposed in [13].

4.3. Evaluation Criteria. To evaluate the weighted random forest algorithms on the selected dataset, accuracy is considered as the evaluation criterion, where accuracy is the rate of correct predictions made by the model:

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN}.$$
 (15)

4.4. Classification Accuracy. We tested the algorithm on 40 test samples of the dataset. The samples contained 22 negative samples and 18 positive samples. Out of 40, 39 samples were classified correctly, achieving an accuracy rate of 97.5 for the particular dataset. Confusion matrix of the weighted RF algorithm for credit risk assessment is shown in Table 4.

The comparison results of financial credit risk data classification accuracy of the three methods are shown in Figure 3.

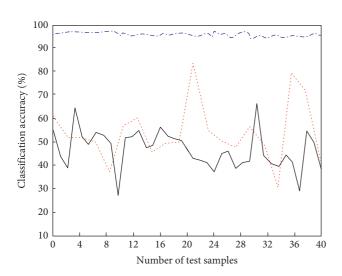
TABLE 3:	Quantitative	table of	of a	ualitative	indicators.

	Index	Characteristics	Quantized value
		Graduate degree or above	5
	Educational background	Bachelor degree	4
		College degree	3
		High school/technical secondary school degree	2
Quantitative table of individual		Compulsory education and below	1
credit risk evaluation indicators		Management personnel	5
		Technician	4
	Occupation type	Employee/worker	3
		Farmer	2
		Other	1

Table 4: Confusion matrix of the weighted RF algorithm for credit risk assessment.

8

TN = 21	FP = 1
FN = 0	TP = 18



- ·-·- Proposed method
- Particle swarm optimization based collaborative optimization algorithm
- Stack based noise reduction self coding network

FIGURE 3: Comparison of the proposed method with two other methods.

From the comparison results of classification accuracy of financial and credit risk data shown in Figure 3, it can be seen that the classification accuracy of the proposed method is always higher than the two traditional comparison methods, and the classification accuracy of the method in this paper is always maintained at about 95%.

4.5. Accuracy of Financial Credit Risk Assessment. The comparison results of financial credit risk assessment accuracy of the three methods are shown in Figure 4.

By observing the accuracy results of financial credit risk assessment shown in Figure 4, it is not difficult to see that the risk threshold of the proposed method is basically consistent

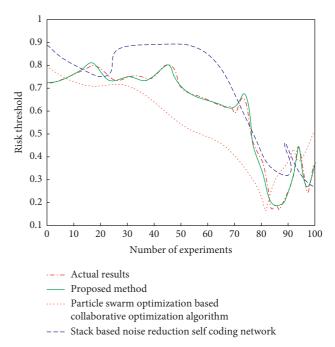


FIGURE 4: Comparison of the accuracy of the three methods.

with the actual results, while the risk threshold of the two literature comparison methods is quite different from the actual value.

5. Conclusion

The awareness of credit risk and virtuous asset cycle forces China's commercial banks to establish an effective credit risk monitoring and control platform. From the perspective of the virtuous cycle of commercial bank loans, establishing a scientific and reasonable credit risk evaluation index system and realizing the credit risk evaluation model of commercial banks has a certain practical significance for the benign development of the financial industry, the healthy development of China's national economy, and the prevention of the global credit crisis. Therefore, a financial credit risk control strategy based on the weighted random forest algorithm is proposed to verify the performance of the method from both theoretical and experimental aspects. This method has high accuracy in financial credit risk data classification and financial credit risk assessment. Specifically, compared

with the method based on particle swarm optimization, the classification accuracy of financial and credit risk data is significantly improved and always maintained at more than 95%, and compared with the method based on stack noise reduction self-coding network, the risk threshold is consistent with the actual results.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported by the Provincial Soft Science Key Project (2013ZK2024 and 2014GK3147) of the Science and Technology Department of Hunan Province.

References

- F. Assef, M. T. Steiner, P. J. Steiner Neto, and D. G. D. B. Franco, "Classification algorithms in financial application: credit risk analysis on legal entities," *IEEE Latin America Transactions*, vol. 17, no. 10, pp. 1733–1740, 2019.
- [2] A. Kalimashi, Y. Ahmeti, and A. Ahmeti, "The impact of audited financial statements on credit risk assessment in the banking system of kosovo," *Period 2014–2018*, vol. 8, no. 1, pp. 119–128, 2020.
- [3] Z. Niu, "Visual analytics for networked-guarantee loans risk management," in *Proceedings of the IEEE Pacific Visualization* Symposium (Pacific Vis), IEEE, Kobe, Japan, April 2018.
- [4] S. Khemakhem and Y. Boujelbene, "Predicting credit risk on the basis of financial and non-financial variables and data mining," *Review of Accounting and Finance*, vol. 17, no. 3, pp. 316–340, 2018.
- [5] T. Grebenar, "Behavioural model of assessment of probability of default and the rating of non-financial corporations," p. 56, The Croatian National Bank, Croatia, 2018.
- [6] S.-F. Chen, G. Charkaborty, L. H. Li, and T. Lin, "Credit risk assessment using regression model on P2P lending," *International Journal of Applied Science & Engineering*, vol. 16, no. 2, pp. 149–157, 2019.
- [7] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.
- [8] A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintha, and S. Kundu, "Improved random forest for classification," *IEEE Transactions on Image Processing*, vol. 27, no. 8, pp. 4012–4024, 2018.
- [9] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine*, vol. 47, no. 1, pp. 31–39, 2017.
- [10] M. Shahhosseini and G. Hu, "Improved weighted random forest for classification problems," in *Proceedings of the In*ternational Online Conference on Intelligent Decision Science, Springer, Istanbul, Turkey, August 2020.
- [11] S. Xuan, G. Liu, and Z. Li, "Refined weighted random forest and its application to credit card fraud detection," in Proceedings of the International Conference on Computational Social Networks, 2018.

[12] K. N. Ohei and R. Brink, "Employability distress: factors affecting ICT graduates' employability and work integration in the workspace," *Research in World Economy*, vol. 12, no. 1, pp. 138–149, 2021.

- [13] D. Yang, "Individual credit risk assessment based on stacked," Denoising Autoencoder Networks, vol. 46, no. 10, pp. 7–13, 2019.
- [14] B. Dufour, F. Petrella, and N. Richez-Battesti, "Understanding social impact assessment through public value theory: a comparative analysis on work integration social enterprises (WISEs) in France and Denmark," Working Papers, vol. 41, no. 27, pp. 112–138, 2020.
- [15] W. Bao, N. Lianju, and K. Yue, "Integration of unsupervised and supervised machine learning algorithms for credit risk assessment," Expert Systems with Applications, vol. 128, pp. 301–315, 2019.
- [16] Y. Zhu, L. Zhou, C. Xie, G.-J. Wang, and T. V. Nguyen, "Forecasting SMEs' credit risk in supply chain finance with an enhanced hybrid ensemble machine learning approach," *International Journal of Production Economics*, vol. 211, pp. 22–33, 2019.
- [17] A. Signoretti and S. Sacchetti, "Lean HRM practices in work integration social enterprises: moving towards social lean production. Evidence from Italian case studies," *Annals of Public and Cooperative Economics*, vol. 76, no. 32, pp. 87–99, 2020.
- [18] Y. Wang, Y. Zhang, Y. Lu, and X. Yu, "A comparative assessment of credit risk model based on machine learning --a case study of bank loan data," *Procedia Computer Science*, vol. 174, pp. 141–149, 2020.
- [19] U. P. Shukla and S. J. Nanda, "Designing of a risk assessment model for issuing credit card using parallel social spider algorithm," *Applied Artificial Intelligence*, vol. 33, no. 3, pp. 191–207, 2019.
- [20] X. Chen and M. Pan, "Research on the credit risk assessment of peer-to-peer lending borrower based on logistic regression model," *DEStech Transactions on Environment Energy and Earth*, vol. 6, p. 03, 2019.
- [21] M. Soui, I. Gasmi, S. Smiti, and K. Ghédira, "Rule-based credit risk assessment model using multi-objective evolutionary algorithms," *Expert Systems with Applications*, vol. 126, pp. 144–157, 2019.
- [22] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [23] A. Abraham, F. Pedregosa, M. Eickenberg et al., "Machine learning for neuroimaging with scikit-learn," Frontiers in Neuroinformatics, vol. 8, p. 14, 2014.