



## **Internet Financial Risk Early Warning Based on Big Data Analysis**

Noura Metawa<sup>1,2</sup>, Saad Metawa<sup>2</sup>

<sup>1</sup> American University in the Emirates, Dubai, UAE

<sup>2</sup> Faculty of Commerce, Mansoura University, Egypt

Emails: [Noura.metawa@aue.ae](mailto:Noura.metawa@aue.ae) ; [s.metawa@mans.edu.eg](mailto:s.metawa@mans.edu.eg)

### **Abstract**

Internet financial risk prevention is an important area for financial risk prevention. In recent years, a series of vicious high-risk events, such as cash lending and P2P platform running, have caused a great negative impact on the reputation of the Internet financial industry, which has aroused great concern from all walks of life. Based on big data analysis technology, this paper constructs an improved algorithm model, and carries out high-precision risk warning for China's Internet financial risk. The forecast data is basically consistent with the actual situation, and the prediction accuracy reaches 90%. It can be seen that the improved model based on the decision tree algorithm has higher prediction accuracy for Internet financial risk warning. This paper systematically sorts out the risks of China's Internet finance from two dimensions: risk type and main risk. And pointed out that the current Internet finance industry in China has a large overall compliance risk, and insufficient infrastructure construction leads to fraud risks. Separate industry supervision has a regulatory vacuum, arbitrage risks are more obvious, and China's financial consumer quality is not high, Internet financial institutions Improper exemption is risky. On this basis, it is proposed to speed up the construction of a multi-integrated Internet financial risk prevention system including the internal risk control system, the industry association self-discipline system, the government administrative supervision system and the effective social supervision system.

**Keywords:** internet finance, risk early warning model, decision tree algorithm, credit evaluation system

### **1.Introduction**

Internet financial risk prevention is an important area for financial risk prevention. On the one hand, such fields are developing rapidly, but at the same time, many risks and problems have emerged in such fields. It is undeniable that some formats have tried the wrong direction of innovation. Some emerging Internet organizations lack awareness of compliance, risk awareness and consumer rights protection in their business operations, and they have not been able to establish anti-money laundering and anti-terrorism financing regulations in a timely manner as an institutional guarantee.

Many research teams at home and abroad have conducted in-depth research on Internet financial risks. In [1], the author analyzes the impact mechanism of Internet finance on the risk exposure of commercial banks

from four dimensions: risk management, operational efficiency, profitability and risk spread. The results show that the rapid development of Internet finance has improved the risk management and operational efficiency of commercial banks, offsetting their negative impact on profits and risks, reducing the bankruptcy risk of commercial banks and promoting the development of commercial banks. The risk-taking of joint-stock commercial banks has decreased, while the risk of bankruptcy of large commercial banks, city commercial banks and rural commercial banks has increased significantly. Therefore, commercial banks need to use Internet finance reasonably according to the nature of commercial banks to promote the stability of the entire financial system. Strategic needs and key scientific issues related to Internet finance, including "Internet finance risk and regulation", "Credit and risk assessment theory based on online social media data", "Behavioral financial theory for online big data", "Dynamic relationship", "The relationship between online data and asset pricing and event arbitrage theory", "Influence mechanism and risk regulation of program transactions", "Key issues in Internet financial operations management", aiming to provide guidance for researchers [2]. In [3], the author provides evidence on the economic impact of sharing by studying Internet finance, aiming to explore how Internet finance affects the relationship between commercial banks' risk appetite and monetary policy, and discusses whether this effect will occur in heterogeneous banks. There has been a change. The results show that Internet finance has changed the sensitivity of bank risk behavior to monetary policy. The influence of internet finance depends on the ownership (national or private) and scale of the bank. Unlike the sub-samples of large banks, in private banks, Internet finance has only a moderate impact on the bank's risk-bearing delivery channels for monetary policy. In [4], the author conducted a SYS-GMM test using the Internet Finance Index based on "text mining" and data from 36 commercial banks from 2003 to 2013. The results show that, first of all, the impact of Internet finance on the risk-taking behavior of Chinese commercial banks is a "U" trend. The initial development of Internet finance can help commercial banks reduce management costs and risk-taking, but then Internet finance will increase capital costs and exacerbate the bank's adventurous spirit. Second, the response of Chinese commercial banks to risky behavior is different. Large commercial banks are slow to respond, while small and medium banks are relatively sensitive.

In [5], the author applied big data to the construction of intelligent transportation systems. The authors propose a smart model to identify unlicensed taxis. The proposed model consists of two sub-model components, the candidate selection model and the candidate refinement model. The former is used to screen out a list of coarse-grained suspected unlicensed taxi candidates. The model was evaluated using real-world data and the results were encouraging, demonstrating its efficiency and accuracy in identifying unlicensed taxis, helping the government better regulate traffic operations and reduce associated costs. In [6], the author applies big data to smart commerce. The authors assess the relevance of these trends in the current business environment through emerging evidence-based applications and their assessment of the broader business impact. The article uses BigML to examine how these two social information channels affect consumer buying decisions on social networking sites. The author conducted an empirical study that integrated the framework and theoretical models of big data analysis to demonstrate that big data analysis can be successfully combined with theoretical models to produce stronger and more effective consumer purchasing decisions. In [7], the authors attempted to use the European option evaluation tool to capture newly developed functions and perform statistical analysis on data collected from the currency option market. The author proposes a feature function method to derive closed pricing formula, and proposes a numerical solution method based on fast Fourier transform (FFT) algorithm. Finally, extensive numerical experiments were performed to validate modeling methods and numerical algorithms. The results show that the model performs well in capturing the observed properties of the market, and the FFT numerical algorithm is both accurate and efficient in processing large amounts of data. In [8], the author proposes a novel big data management method based on SDN for optimized network resource consumption (such as network bandwidth and data storage unit), using the insertion of elements in the flow table in the OpenFlow

controller based on Bloom-filter. And delete to analyze the performance of the proposed solution, which will use rules and operations to make most of the decision-making mechanisms based on network traffic classification. Using the proposed solution, developers can deploy and analyze real-time traffic behavior for future big data applications in MCE. In [9], the author applies big data to monitoring and intelligent control. The authors present a layered distributed fog computing architecture to support the integration of a large number of infrastructure components and services in smart cities of the future. The article uses an intelligent pipeline monitoring system based on fiber optic sensors and sequential learning algorithms to analyze case studies to detect events that threaten pipeline safety. The authors built a working prototype to experimentally evaluate the event detection performance of identifying 12 different events. These experimental results demonstrate the feasibility of implementing the system in the citywide in the future. In [10], the author applies big data to spectrum prediction. The authors propose a method for spectrum occupancy prediction that can be used to reduce the latency of making dynamic spectrum allocation decisions and improve the cognitive and management functions of cloud-based architectures. In addition, the author applies the ML algorithm to predict spectrum usage and compares the predicted results with the actual measured data. Considering that the accuracy of the prediction depends on the amount of data collected and the prediction time based on the BD and ML methods, the paper proposes to develop a cloud-based general processing architecture, and the proposed architecture is suitable for deployment in cognitive C-RAN.

Based on big data analysis technology, this paper constructs an improved algorithm model and provides a high degree of risk warning for Chinese Internet financial risks. This paper also uses the improved model based on decision tree algorithm to adjust and optimize the forecasting accuracy of Internet financial risk warning. The risk of China's Internet finance is systematically combed from two dimensions: risk type and main risk.

## **2.Method**

### **2.1 Financial Risk Early Warning Model Based on Decision Tree Algorithm**

The construction of the decision tree basically depends only on the training data and does not depend on other parameter settings. In the construction process, the feature selection criteria of the decision tree depends on information entropy or information gain. The so-called decision tree construction is based on the fastest descent criterion of information entropy or information gain to determine the topological structure between each feature attribute. As the core step of constructing the decision tree, the purpose of feature selection is to construct different decision paths by dividing different feature attributes. In this process, each decision branch node contains sample requirements that belong to the same category as much as possible. In general, the division selection feature is divided into three different cases:

1. For each attribute represented by a discrete numerical value, each corresponding discrete value can be regarded as a decision branch.
2. For a binary decision tree, if the feature is expressed in discrete form, a subset of the feature partition can be tested, that is, divided into two decisions according to "belonging to this subset" and "not belonging to this subset".
3. The attribute is a continuous value. At this point, determine a value as the segmentation point  $s$  and generate two branches according to  $>s$  and  $\leq s$ .

The core content of the construction decision tree is the criterion of feature selection. The feature selection criterion is a heuristic method that divides the data partition  $D$  of the training set of a given class mark into the individual class, which determines the topology and the splitting point. The choice of  $s$ . In information theory, entropy measures the uncertainty of random variables. Assuming  $D$  is a data set, the entropy of  $D$  is expressed as:

$$info(D) = -\sum_{i=1}^m p_i \log_2(p_i) \quad (1)$$

Where  $p_i$  quantization describes the probability values of the  $i$ -th category in data set  $D$ . Here, the entropy value represents the average amount of information contained in the data and the category labels in  $D$ . The conditional entropy  $info_A(D)$  characterizes the uncertainty of the random variable  $D$  under the condition that the known random variable  $A$  is known. Assuming that the training data  $D$  is divided by the attribute  $A$ , under this division method, the conditional entropy of  $D$  is:

$$info_A(D) = \sum_{j=1}^n \frac{|D_j|}{|D|} info(D_j) \quad (2)$$

Information gain quantification describes the degree of uncertainty in category  $D$  under different values of known feature  $A$ . In general, in the learning process of the decision tree, the information gain is equivalent to the mutual information of the class and feature in the training data set (ie, the difference between entropy and conditional entropy).

$$gain(A) = info(D) - info_A(D) \quad (3)$$

Based on this, the ID3 algorithm calculates the corresponding information gain for all possible features (internal nodes) and selects the feature with the largest information gain as the node.

For the sake of simplicity, the feature attributes are discretized. In fact, the log density and the friend density are continuous attributes. The ID3 algorithm is also applicable when the feature attribute is a continuous value. In this case, heuristics can be used to select features as well as segmentation points. Taking the  $j$ -th feature and all its possible values  $s$  as the segmentation variable and the segmentation point, splitting  $D$  and calculating the information gain, the point with the smallest information gain is called the best split point of this attribute, and its information gain is used as this information gain of the feature is then selected as the feature of the node with the feature with the smallest information gain.

## 2.2 The basis of risk control theory based on big data analysis

### (1) Big data risk control and characteristics

So far, although many financial institutions in China have actively participated in the team that supports small and medium-sized enterprises, there are still a large number of institutions, including financial institutions of various Internetes, which have not really understood the status quo of SMEs. There is a more obvious mentality of traditional borrowing, especially for individual industrial and commercial households. This has led to the rapid loss of a large number of potential lending customers, which has an impact on the sustainable development of the industry. In this context, through the rational use of big data risk control technology can effectively solve this problem, in the big data-based environment, financial institutions can comprehensively collect the credit situation of borrowers in various platforms or institutions, and enhance

their credit ability. The Internet finance industry has taken the lead in trying to use big data analysis to establish a system for comprehensive analysis and evaluation of participants' credit ratings, investment preferences, and risk tolerance.

Diversity is a very important feature of big data risk control. It can also be said to be its primary feature. In addition to customer information from traditional banks, it also uses personal techniques such as reptiles to collect personal social and corporate grievances from the Internet. Supplementary data, some of the risk control platform even integrated the data sources from the government departments such as industry and commerce, taxation, power supply, customs, etc., and made a good complement to the accurate portrait of the capital demand side. The second feature of big data risk control is the application of advanced technology architecture. The machine can automatically analyze multi-source heterogeneous data in a few seconds and give the borrower's credit comprehensive reference value, which is undoubtedly more than the traditional credit. It is more convenient and objective for the approving officer to issue loans based on personal experience.

## (2) The role of big data risk control

According to the current situation in China, in order to achieve continuous innovation in the social governance system and continuous improvement of the market economy, it is necessary to strengthen the construction of the social credit system so as to better promote the overall development of the local economy and society. Credit risk control based on big data technology is the basic project of social credit system construction, and its important role is as follows:

### I. Carry out financial innovation and open up financial markets

For an industry or a company, to maintain a long-term development momentum, we must pay attention to innovation. In the context of big data, financial institutions, with their advantages and characteristics, can better realize the exploration of the potential of the industrial market, thus promoting the innovative development of the entire industry. In the future, under the influence of the continuous development of big data finance, enterprises participating in Internet finance can continuously explore the role of big data and apply it to product service and industrial chain innovation development. Specifically, on the one hand, based on big data finance, the chain of financial services is appropriately extended to enable it to expand from a single supply chain to the entire industry chain; on the other hand, it can also expand services. The scope of the object is set as the focus point, so that individuals or families can be gradually included in the ranks of the service objects.

### II. Eliminate information asymmetry between the two sides of the transaction, reduce transaction costs

The ultimate goal of financial products is to be able to convert into the payment of funds for financial participants. Regardless of financial services or products, credit is inevitably a matter of great concern to both parties. According to the previous industry model, traditional financial service institutions, including various Internet financial institutions, investors, or their partners, often need to spend a lot of manpower, material resources and financial resources on the credit review and risk tolerance of the financing party. Wait for resources, but the results are not necessarily very reliable. At the same time, for the financing party, it is also necessary to judge the financial institution's business scale, credibility, financial strength and other related information, but the degree of understanding of the information channel is very vague, and the result is a clear information asymmetry in the information of both parties to the transaction, which leads to a decline in the trust value of financial institutions.

With the comprehensive promotion and application of big data finance, financial institutions can better obtain the relevant credit information of the service targets, and minimize the investigation cost, so that the

information asymmetry between the borrowers and the borrowers can be solved. Under the influence of the continuous wave of the Internet, the level of informationization in various industries is continuously improving, and the collection and sharing of basic information data for each subject will become possible.

### 2.3 Internet Financial Personal Credit Evaluation Model Based on BP Neural Network

The search direction  $p_0$  of all conjugate gradient algorithms starts from the steepest descent direction  $g_0$  (the negative direction of the gradient), and then uses the line search to determine the weight and the threshold ( $X$ ) along the current search direction, where  $p$  is the search direction, the parameter  $\alpha$  is used to reduce the gradient of the search direction. Next, the next search direction is determined by the conjugate direction of the previous two search directions. The specific expression is as follows:

$$p_{\theta} = -g_0, \quad X_{k+1} = X_k + \alpha_k P_k, \quad p_k = -g_k + \beta_k p_{k-1} \quad (4)$$

As long as the objective function is continuous, the BP network of the single hidden layer is sufficient to map the relationship. The data processed in this paper is not complicated, and the number of samples is 150 groups. Therefore, the number of BP neural network layers used in this paper is sufficient for Handling problems, namely the input layer, the output layer, and the hidden layer. The design of the number of hidden layer nodes is an important part, and their determination will closely affect the accuracy of the final output. Excessive number of nodes will reduce the training efficiency, and secondly, it will cause "excessive matching". When the error cannot be guaranteed to be optimal, the network will adapt to the new data. Too few nodes will make the network not complete. The information ends when the training is lacking and the ability to explain the problem is lacking.

This paper decides to use the network structure growth method based on different empirical formulas, and gradually increases from a small number of nodes, and calculates the average convergence step length under different number of hidden layer nodes to determine the optimal setting. The transfer function of the hidden layer node adopts a simple calculation and a micro-simultaneous nonlinear Log-sigmoid function. The Log-sigmoid formula is as follows:

$$\text{logsig}(n) = \frac{1}{\exp(-n) + 1} \quad (5)$$

At the end of this paper, only one indicator of the credit risk level of each test borrower is predicted, so the number of nodes in the output layer is one. The error function is:

$$E = \frac{1}{2} \sum_{k=1}^L (T_k - o_k)^2 \quad (6)$$

Where  $T$  is the desired output and  $o$  is the actual output. The output layer node uses the Tan-sigmoid formula as the transfer function:

$$\text{tansig}(n) = \frac{2}{(1 + \exp(-2 * n)) - 1} \quad (7)$$

The stability of a BP neural network depends on the speed of algorithm learning. In order to get a faster convergence speed, the learning rate can be set higher, but this is at the expense of increased training error; the learning speed is too slow and will be extremely large. Different use cases, different research settings,

different settings make the neural network vary widely. In most cases, system stability considerations take precedence over learning time considerations, and the learning rate is chosen to be between 0.01 and 0.8. Therefore, this article focuses on making the network minimize the prediction error and setting the learning rate to 0.05.

### 3. Experiment

#### 3.1 Data source

This model is applicable to the more complex social sciences. Based on the characteristics of Internet financial risks, this model is used to quantify the indicators and measure Internet financial risks. This study used the questionnaire survey method to conduct questionnaire surveys for 80 financial industry practitioners. A total of 80 questionnaires were sent out and 78 were returned, of which 76 were valid questionnaires, the effective rate was 95%, and the questionnaire results were in confidence interval. In the questionnaire, each risk of each level is evaluated separately. The participants participate in the scoring of each risk factor at each level. In this study, the fuzzy judgment matrix is numerically determined by the 1-9 scale method. See Table 1 for details.

**Table 1 Risk Assessment Questionnaire 1-9 Scale Method**

Pij	Scaling
Pi is as important as Pj	1
Pi is slightly more important than Pj	3
Pi is more important than Pj	5
Pi is significant important than Pj	7
Pi is extremely important than Pj	9
Pi and Pj importance between the two indicators	2、4、6、8
The importance of Pi relative to Pj	Countdown to the above scale

#### 3.2 Evaluation index system

This paper elaborates on the characteristics of Internet financial risk. In this part, this paper constructs a risk assessment index system to support the Internet financial risk in the above theory, in order to better identify and pay attention to various risks. This paper carefully summarizes the reference research report at home and abroad, and designs the index system of this research based on the existing research results. A total of 10 first-level indicators and 29 second-level indicators are selected. See Table 2 for details. The first-level indicators include traditional financial risks such as credit risk, market risk, operational risk, liquidity risk, legal risk, money laundering risk, crime risk, fraud risk, etc., as well as information technology risks, data risks and information security risks. The second-level indicators were selected on the basis of the first-level indicators to select technical vulnerability risks, data defect risks, and information security legal risks to be measured separately.

**Table 2 Evaluation Index System of Internet Financial Risk**

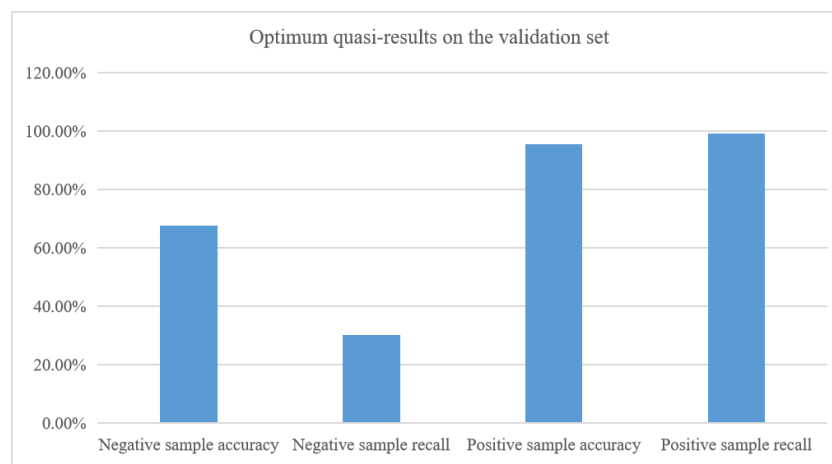
Target layer	Primary indicator	Secondary indicators
Internet financial risk (S)	Credit risk (P1)	Own credit risk (p11), Credit risk of others (p12)
	Market risk (P2)	Interest rate risk (p21), exchange rate risk (p22), stock price (p23), commodity price (p24)
	Liquidity risk (P3)	Asset mismatch risk (p31), maturity mismatch risk (p32), fund balance risk (p33)
	Operational risk (P4)	Internal staff (p41), institutional processes (p42), technological innovation (p43), external events (p44)
	Information security risk (P5)	Information security legal risk (p51), privacy information disclosure risk (p52), reputation risk (p53)
	Money laundering risk (P6)	Management framework vulnerability risk (p61), assessment mechanism risk (p62), monitoring capability risk (p63)
	Information technology risk (P7)	Technical loopholes (p71), management flaws (p72), human factors (p73)
	Fraud risk (P8)	Moral Hazard (p81), False evidence risk (p82)
	Data risk (P9)	Data defect (p91), network security (p92)
	Compliance risk (P10)	Narrow laws and regulations (p101), local government regulations (p102), normative documents of other administrative agencies (p103)

**4.Results and Discussions**

**4.1 Internet financial tax risk analysis**

Through the empirical test of more than 12 million tax data of the Internet finance industry in A city in 2016-2018, the results show that the risk assessment of the big data platform can effectively identify the negative in the verification set after the feature extraction modeling analysis in the training set. The sample, efficiency and ability have been greatly improved compared with the traditional financial institutions relying on manual judgment, and the practical effect of the model is better. The test results of the two major taxes are as follows:

(1) Value added tax

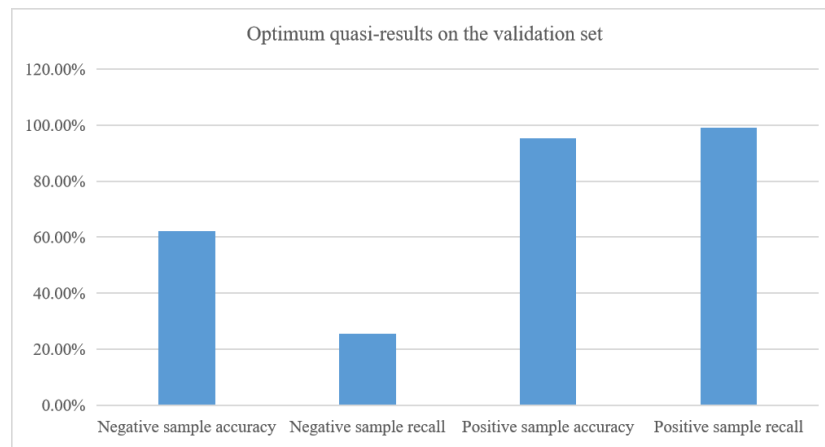


**Figure 1 Optimal results on the validation set**

**Table 1 Number of samples in different probability intervals**

Probability interval	Number of test samples	% of total test samples
(0.5,0.6]	109	3.13%
(0.6,0.7]	256	7.36%
(0.7,0.8]	185	5.32%
(0.8,0.9]	109	3.13%
(0.9,1.0]	504	14.49%

(2) Corporate income tax



**Figure 2 Optimal quasi-results on the validation set**

**Table 2 Number of samples in different probability intervals**

Probability interval	Number of test samples	% of total test samples
(0.5,0.6]	95	2.51%
(0.6,0.7]	10	0.26%
(0.7,0.8]	36	0.95%
(0.8,0.9]	24	0.63%
(0.9,1.0]	83	2.19%

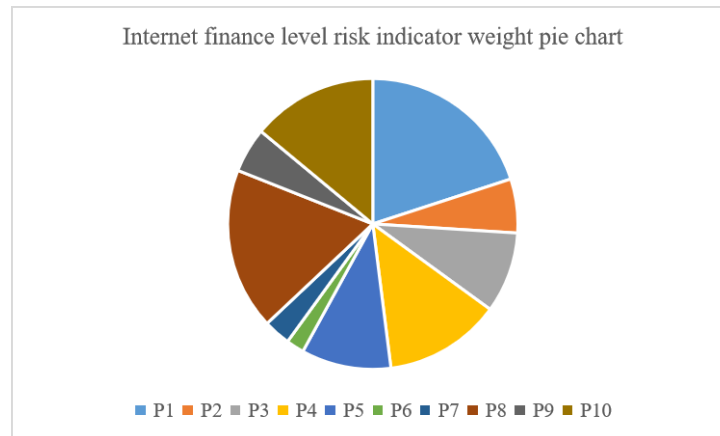
Through the collection, integration, modeling and analysis of various types of data, we monitor the credit risk of pre-credit and loan in the lending process of the Internet financial SMEs through the integrated operation of the big data platform, and the machine continuously learns to accurately identify the risks. The reason is to develop targeted solutions through the analysis of development trends to achieve a project management process that controls and resolves the credit risk of Internet financial institutions. Big data technology is used to monitor credit risk. Its risk model is more objective, sensitive and meticulous than traditional expert scoring system such as Delphi method and AHP. The machine can automatically optimize the selection of indicator systems and the construction of models from the constant changes of data, and constantly adapt to the new means of fraudulent loans; and can find the possibility of generating risks in real time from the slight change of data, quickly identify the risks after the loan and advance early warning, help financial institutions do a good job after the loan.

Judging from the current situation, the traditional Internet financial institutions' credit approval for SMEs with financing needs basically depends on the subjective experience judgment of the credit department staff. It not only lacks uniform standards but also has low efficiency, mainly reflected in different credit approvals. The results of the application for approval by the same loan company may not be consistent.

After the approval of the SMEs, the project may have missed the best window period. The continuous development of big data technology provides multi-source data support for financial institution credit approval. The data collected by the big data platform from government agencies and the Internet can effectively supplement the risk information of SME customers. The big data risk assessment model established in this study can effectively supplement the credit approval of financial institutions, embed the big data system into the internal information construction of financial institutions, and realize the direct decision-making reference for credit approval personnel through the system automation operation process. Big data system scoring model scores higher than certain conditions can be set to pass the credit application directly, below some scores can be set to automatically reject directly, only the model scores between the two customers, by the professional approval loan officer perform a manual discriminant review. Through the automated operation of the big data platform, the time for approval of financing credit for SMEs can be greatly reduced, and the probability of operational errors for manual approval can be reduced to a certain extent.

#### 4.2 Internet financial evaluation indicators risk prediction

Figure 3 more intuitively interprets the weight ratio of various Internet financial risk indicators, as follows.

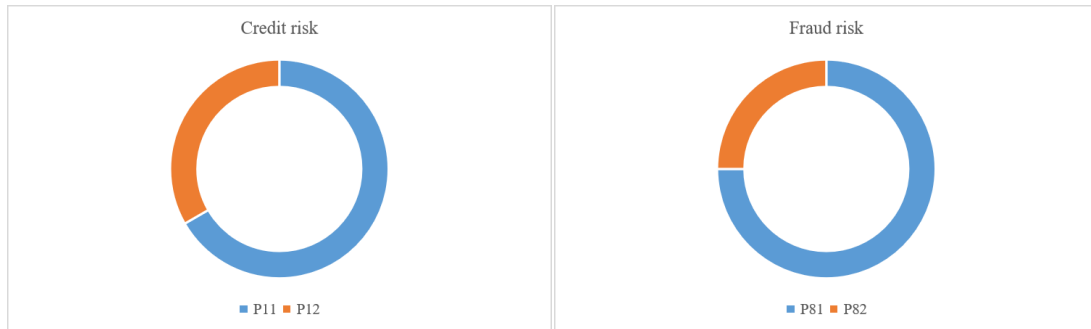


**Figure 3 Internet finance level risk indicator weight pie chart**

It can be seen from the pie chart of the first level risk index weight of Internet Finance in Figure 3 that the first level risk of Internet finance is ranked from high to low as follows: compliance risk, credit risk, fraud risk, operation risk, information security risk, liquidity risk, market risk, data risk, information technology risk and money laundering risk. The secondary risks of Internet Finance rank from high to low as follows: normative document risk, moral risk, self owned credit risk, other related credit risk, internal personnel risk, narrow sense laws and regulations risk, false evidence risk, privacy information disclosure risk, capital balance risk, incomplete local government regulations risk, system process risk, network security risk, commodity price risk, information security legal risk, term mismatch risk, technology vulnerability risk, external event risk, stock price risk, evaluation mechanism risk, data defect risk, technology innovation risk, asset capital mismatch risk, interest rate risk, reputation risk, monitoring ability risk, management defect, exchange rate risk, human factor risk, management framework vulnerability risk.

From the above analysis, it can be seen that the compliance risks of various Internet finance businesses in China are more prominent. In the process of the special rectification activities of Internet financial risks, Internet financial institutions have generally experienced the pain of compliance transformation. Some small-scale and relatively weak-employed institutions took the initiative to suspend operations and end their

operations, achieving a benign and orderly exit. What needs special attention is those who have repeatedly violated regulations in the course of historical management, and have a large risk factor, but there are institutions that continue to operate. The difficulties in achieving transformation are large, and the compliance risk is a big challenge.



(a) Credit risk

(b) Fraud risk

Figure 4 Credit risk and fraud risk information map

At present, China's Internet finance industry infrastructure construction is seriously inadequate, leading to serious information asymmetry in the industry, and fraud risks and credit risks are prominent. First, most of the institutions have not yet docked with the credit system, and the current regulations are inadequate. The fraud and default costs of some illegal operators are low and cannot be paid enough attention. Second, up to now, the national statistical department has not included the market operation of such institutions, and the existing third-party data has a lack of definition, coverage and accuracy, especially the basic gap in the flow of funds.

**4.3 Analysis of the accuracy of Internet financial risk warning**

The following figure shows the comparison between the prediction error of the algorithm model for Internet risk warning and the prediction of the traditional model.

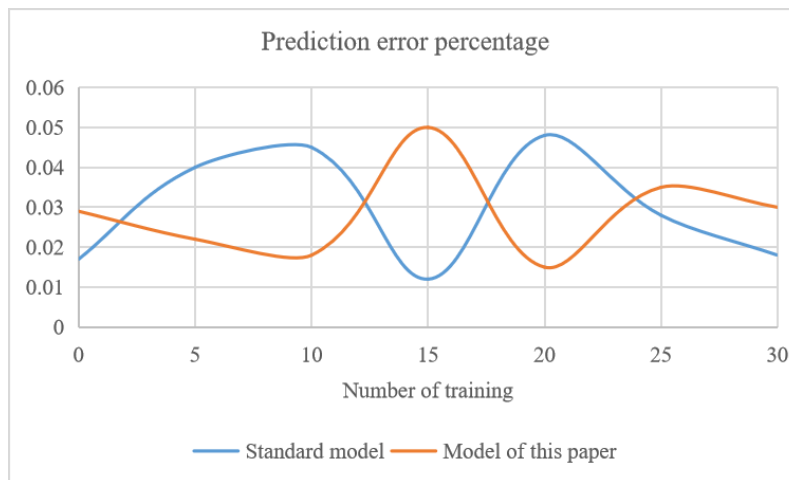


Figure 5 Comparison of the accuracy of early warning errors in Internet financial risks

Comparing the accuracy results and graphs of the two curves, it can be seen that the standard algorithm model is trained by the training group and then predicted by the experimental group. The relative error is 3% as the standard limit. There are 10 groups exceeding the error. The accuracy rate is only 67%, and the error of 8 groups is more than 4%, and the prediction result is not very satisfactory. The improved algorithm model in this paper is also bounded by 3%, and the prediction data is basically consistent with the actual situation, and the prediction accuracy reaches 90%. It can be seen that the improved model is more accurate for the prediction of Internet financial risk warning.

**Training speed:** Before the algorithm is improved, 105 iterations are needed to converge to the set error. The improved model only needs 19 iterations to converge to the set error. Through the improvement of the standard model algorithm, the network convergence speed A considerable increase was mentioned. **Training accuracy:** Before the algorithm is improved, the relative error is 3%, and there are 10 groups exceeding the error. The overall prediction accuracy is only 67%, and more than 4% of the groups account for 80% of the error group. The result is not very good. After the algorithm is improved, the prediction data is basically consistent with the actual situation, and the prediction accuracy reaches 90%. It can be seen that the use of the improved risk warning prediction is more accurate. **Simulation training:** The model simulation training results show that the improved model predicts the credit risk level and is consistent with the actual situation. The model does not have the problem of confusing users with high credit ratings with users with low credit ratings, and the matching of the credit rating of the simulation results with the actual situation is also quite high, reaching 96%.

## 5. Conclusions

Based on big data analysis technology, this paper constructs an improved algorithm model, and carries out high-precision risk warning for China's Internet financial risk. The forecast data is basically consistent with the actual situation, and the prediction accuracy reaches 90%. It can be seen that the improved model used in this paper is more accurate for the prediction of Internet financial risk warning. This paper draws the following conclusions:

This paper systematically sorts out the risks of China's Internet finance from two dimensions: risk type and main risk. And pointed out that the current Internet finance industry in China has a large overall compliance risk, and insufficient infrastructure construction leads to fraud risks. Separate industry supervision has a regulatory vacuum, arbitrage risks are more obvious, and China's financial consumer quality is not high, Internet financial institutions Improper exemption is risky. On this basis, it is proposed to speed up the construction of a multi-integrated Internet financial risk prevention system including the internal risk control system, the industry association self-discipline system, the government administrative supervision system and the effective social supervision system.

In the short-term, China should establish the Internet Finance Professional Committee of the Financial Stability Development Committee of the State Council to make up for the inadequacy of the existing separate supervision system and accelerate the construction of a multi-level supervision sandbox system. In the long run, China should explore the macro-prudential policy framework for Internet finance, fully develop regulatory technology, and strengthen international coordination of Internet financial regulation.

## References

- [1] Liu, Z., & University, X. (2016). Research on the influence of internet finance on commercial banks' risk-taking. *Finance & Trade Economics*, 37(4), 71-85.
- [2] Stepp, M. A., Zhu, L., & Cranfill, R. (2016). Changes in beta 4 integrin expression and localization in vivo in response to corneal epithelial injury. , 37(8), 1593.

- [3] Haishu, Q. , Meilin, C. , & Yue, X. . (2018). The effects of the sharing economy: how does internet finance influence commercial bank risk preferences?. *Emerging Markets Finance and Trade*, 54(13), 3013-3029.
- [4] Guo, P. , & Shen, Y. . (2016). The impact of internet finance on commercial banks' risk taking: evidence from china. *China Finance and Economic Review*, 4(1), 16.
- [5] Yuan, W, Deng, P, Taleb, T, Wan, JF, & Bi, CF. (2016). An unlicensed taxi identification model based on big data analysis. *IEEE Transactions on Intelligent Transportation Systems*, 17(6), 1703-1713.
- [6] Xuemei Tian, & Libo Liu. (2017). Does big data mean big knowledge? integration of big data analysis and conceptual model for social commerce research. *Electronic Commerce Research*, 17(1), 169-183.
- [7] Shuang Xiao, Shi-Hua Ma, Guo Li, & Samar K. Mukhopadhyay. (2017). European option pricing with a fast fourier transform algorithm for big data analysis. *IEEE Transactions on Industrial Informatics*, 12(3), 1219-1231.
- [8] Rajat Chaudhary, Gagangeet Singh Aujla, Neeraj Kumar, & Joel J.P.C. Rodrigues. (2018). Optimized big data management across multi-cloud data centers: software-defined-network-based analysis. *IEEE Communications Magazine*, 56(2), 118-126.
- [9] Bo Tang, Zhen Chen, Gerald Hefferman, Shuyi Pei, Tao Wei, & Haibo He. (2017). Incorporating intelligence in fog computing for big data analysis in smart cities. *IEEE Transactions on Industrial Informatics*, 13(5), 2140-2150.
- [10] Baltiiski, P. , Iliev, I. , Kehaiov, B. , Poulkov, V. , & Cooklev, T. . (2016). Long-term spectrum monitoring with big data analysis and machine learning for cloud-based radio access networks. *Wireless Personal Communications*, 87(3), 815-835.