Financial Engineering and Risk Management (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/ferm.2024.070522
ISSN 2523-2576 Vol. 7 Num. 5

Machine Learning and Big Data Application for Risk Assessment of IoT-enabled Financial Management of Credit

Yuchun Li[a,*], Shaojie Yang[b]

Accounting Institute, Haojing College of Shaanxi University of Science &Technology, Xianyang, Shaanxi, 712046, China
[a]ktx_19990218@163.com, [b]yangsj@snfeu.edu.cn
[*]Corresponding author

Keywords: Machine Learning, Big Data, Risk Assessment, Internet of Things

Abstract: Traditional risk models do not have a unified standard and there are loopholes in the connection. Based on machine learning and big data technology, this paper analyzes the model of applying machine learning to risk assessment and the resulting improvement in the accuracy of risk assessment. How to analyze the IoT-enabled financial management of credit in depth and accurately, develop an effective credit risk assessment model, make up for the weaknesses of the existing credit research industry is important. The evaluation report given by many machine learning algorithms has a high probability, and this high probability evaluation result is correct for making a risk assessment decision. The elimination of outliers and the selection of feature variables, the prediction accuracy rate reached 85%. At the level of left and right, the credit risk of online lending is very complicated in reality, and this accuracy has good application prospects.

1. Introduction

At present, the financial management of credit IoT lacks the corresponding risk monitoring, and the traditional risk assessment methods are difficult to use the complex and changeable credit environment. From the perspective of domestic growth, although banks currently have more detailed depositor data, they still cannot effectively distinguish between high-risk and low-risk groups. Therefore, they are unable to provide credit services for this group, which seriously affects the development of consumer credit.

The concept of risk assessment of IoT-enabled financial management of credit was first proposed by some German researchers shortly after World War II, which also indicates that risk assessment is in its infancy. Some scholars in the United States expanded on it in the 1930s and began to explore in this field. In the following two decades, there have been more and more problems in risk management, and people have almost all have designs, so attention and research in this area are becoming more and more. Silva EC believes that the main purpose of this management is to improve the success rate of the project in the process of progress and reduce the possibility of accidents [1]. Erbis S believes that the overall risk assessment of the project, the identification and

research of the risks that may be encountered, and the analysis of the countermeasures in advance [2]. Lu et al. introduced a new method in the field of risk assessment, namely Bayesian networks. Although this method was not originally designed to study risk and the research topics are different, it has similarities and equal effects, and it is modeled by its probabilistic reasoning [3].

There are also many scholars using different methods to conducted a lot of research on risk assessment of IoT-enabled financial management of credit based on machine learning and big data methods. Zhang Y created a multi-level risk coefficient prediction model, first of which is the type of danger, followed by the frequency of occurrence of various types of danger, and finally the components of each type of danger[4]. Hirsch C uses related systems and the types of external influences of the project to study whether the external influence factors of the project are applicable to the established system model, which invisibly expands the scope of use of the model and simplifies the evaluation and evaluation of the degree of influence of external factors on the project [5]. Samuel O W linked the risk of the project with factual evidence. This led to the establishment of another risk prediction system [6].

According to the current research status of financial management risk assessment methods of credit IoT at home and abroad, credit risk assessment methods mainly include statistical methods, operational research methods, non parametric analysis methods and artificial intelligence. Combined with domestic and foreign research results, it is found that the algorithm of each method is based on different research data. The longer the term, the higher the loan interest rate and the higher the rate of return, but the premise of higher returns is that the borrower can repay the principal and interest on time. As the maturity becomes longer, especially when the loan maturity is not in the same economic cycle, the borrower's repayment ability will change with the economic cycle, causing the default rate to increase.

2. Application of machine learning and big data in financial management risk assessment of credit Internet of Things

2.1 Machine learning and big data

Machine learning is a process of continuously absorbing, integrating and improving new technologies from existing knowledge. There are many factors that affect machine learning, such as sample quality. There are many typical problems in machine learning, such as regression, anomaly detection, grouping, sorting, and dimensionality reduction. Each problem is extended to an algorithm [7-8]. The process of machine learning is to first select data, create model data, verify the data, retry the data, reuse the data, and finally adjust the data. There are many scenarios in the practical application of machine learning, and improved risk reduction analysis is the most typical example[9-10]. As shown in Figure 1.

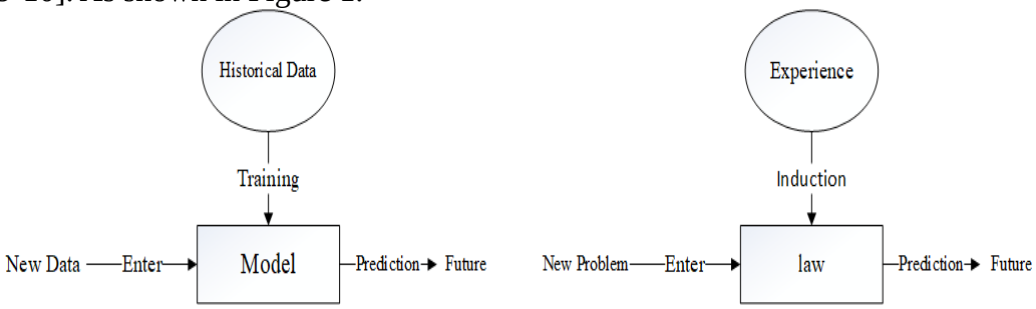

Figure 1: Comparison of machine learning (left) and human thinking (right)

In recent years, with the rapid development of the Internet society, a large amount of information has been digitized, and big data has emerged as the times require. All modern people are surrounded

by this huge amount of data. It is a large-scale, high growth and diversified information asset based on the new processing method and with stronger decision-making power, insight and program optimization power. Big data covers all aspects of us, not only production and living data, but also human beings and their environment can be quantified by data.

2.2 Internet of Things

Internet of Things system is divided into: perception layer, network layer and application layer [11-12]. As shown in Figure2.

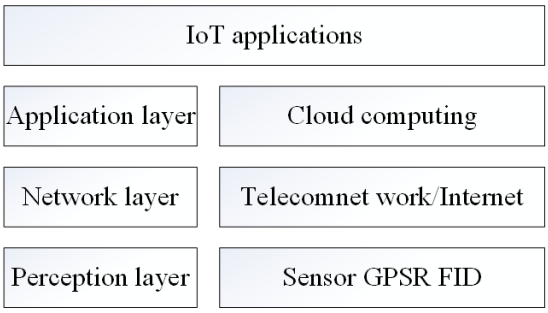

Figure 2: Internet of Things system composition diagram

(1) Virtual mode based on the Internet of Things

And the bottom-up construction the method is very simple [13-14]. As shown in Figure 3.

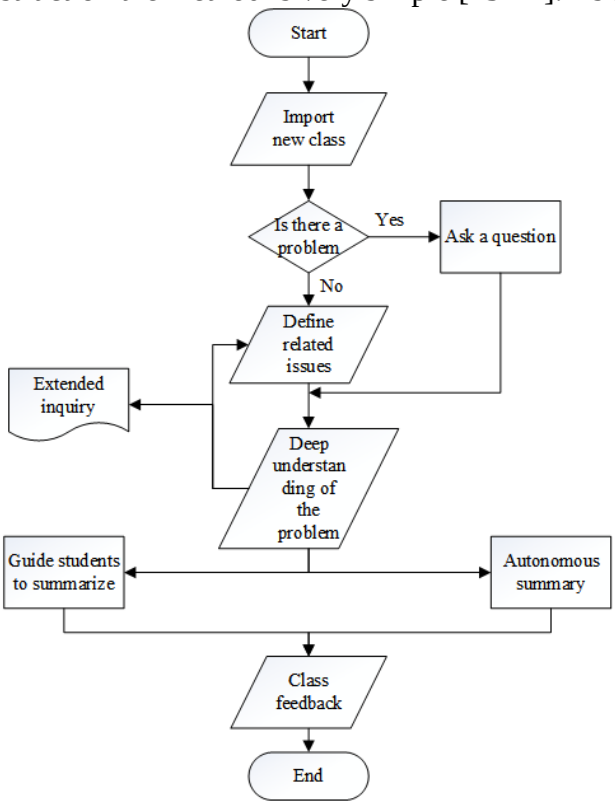

Figure 3: Virtual model diagram based on the Internet of Things

(2) IoT-based design process

Move the mobile smart device close to the NFC tag to the terminal node to receive terminal node and device information [15-16].The design process of the Internet of Things is shown in Figure 4

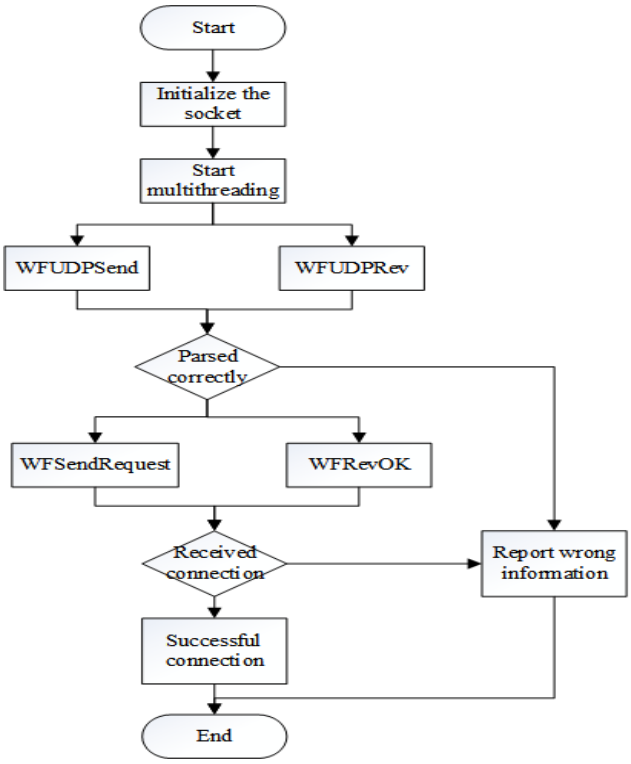

Figure 4: Design flow chart based on the Internet of Things

2.3 Logistic Regression Model

In the modern society where cash is king, cash flow is the lifeblood of enterprises. Risk has a certain embryonic period before it forms harm. This paper uses the Logistic regression model to predict financial health, the Logistic regression model can be expressed as:

$$p(y=1|x)=\pi(x)=\frac{1}{1+e^{-g(x)}} \tag{1}$$

$$g(x)=w_0+w_1x_1+\cdots+w_nx_n \tag{2}$$

$$p(y=0|x)=1-p(y=1|x)=1-\frac{1}{1+e^{-g(x)}}=\frac{1}{1+e^{g(x)}} \tag{3}$$

$$\frac{p(y=1|x)}{p(y=0|x)}=\frac{p}{1-p}=e^{g(x)} \tag{4}$$

$$\ln(\frac{p}{1-p})=g(x)=w_0+w_1x_1+\cdots+w_nx_n \tag{5}$$

$$L(w)=\prod_1^n(\pi(x_i))^{y_i}(1-\pi(x_i))^{1-y_i} \tag{6}$$

The maximum likelihood estimation is to find the parameter $w_0,w_1,\cdots,w_n$ so that $L(w)$ gets the maximum value.

2.4 Application of big data and machine learning in credit financial risk management

Traditional software tools cannot collect data of this size within a certain time frame. For acquisition, management and analysis, new processing concepts and technical tools are needed. There is also a large amount of data about policyholders in the credit IoT financial market. Various data about policyholders can be collected through the Internet, including identity information, interpersonal relationships, consumer behavior, credit evaluation, etc., and then based on big data technology, the credit of creditor carry out an assessment to predict the corresponding insurance risk [17].

The target variable is whether the credit IoT-Enanled financial enterprises underwrites or not, and various information of the applicant as an explanatory variable, including age, height, education, consumption behavior, and business records. Random forest is a kind of ensemble algorithm.

As China has gradually developed into an important market concerned by credit IoT-Enanled financial enterprises, the market competition of Chinese Internet enterprises is becoming increasingly fierce.

Accounts receivable risk (credit risk): in the increasingly fierce market competition, credit sales have become an important means for enterprises to expand sales, resulting in accounts receivable.

Inventory management risk (capital occupation risk): inventory risk refers to the possibility that the inventory value will decrease due to price changes, product obsolescence, natural loss and other reasons when an enterprise owns inventory.

2.5 Risk Assessment

(1) Information intermediary model

Budget management is an important part. The budget of an enterprise is actually the financial plan of an enterprise, which can reflect the expected indicators and resource allocation of the enterprise engaged in business activities in a specific period of time. An enterprise shall coordinate its strategy, business plan and budget well, select appropriate budget management methods according to the characteristics, business strategy and management level of its industry, correctly identify, evaluate and manage the possible risks in each link of budget management, and make good use of the concepts, ideas and methods of modern scientific budget management in budget management.

(2) Guarantee model

The platform will further cooperate with third-party guarantee agencies or other companies to provide guarantees for the safety of investors' funds. As shown in Figure 5.

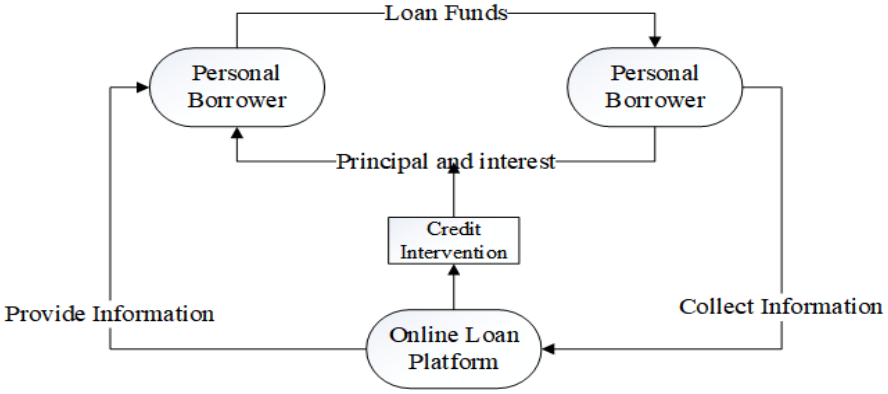

Figure 5: Guarantee mode flow chart

(3) Asset-like securitization model

Different from the above two models, the asset-like securitization model specifically includes third-party professional investors, and third-party professional investors divide the loan amount and maturity into different combinations in order to better attract investors. As a result, assets with more flexible maturities or more flexible amounts have been formed. There are mainly the following two steps: The third-party professional investor first lends funds to individuals or companies that need financing as the first investor, thereby forming debt assets. Then the third party splits and combines the amount and maturity of the debt assets with larger amount or longer maturity to form a combined asset, and sell the combined asset to investors.

3. Risk Assessment of IoT-enabled Financial Management of Credit

3.1 Test Subject

Our country's online lending data has not yet been made public. Some documents have obtained some loan data with the help of web crawler technology, but the information that can be used is extremely limited. Many foreign platforms provide researchers with data download interfaces, with rich data content, which can better characterize the credit status of borrowers. We selected 16,000 samples and divided them into four groups. Three of them used different machine learning methods to assess their risks. The samples of the control group used traditional risk assessment models to assess their risks, and then obtained experimental results for analysis and analysis. Provide risk control for the online lending business of this business, and predict the possibility of default through the information generated by users accessing the mobile network before lending. In the process of modeling, a series of characteristics related to user credit generated from the user's mobile data are formed, so as to be used for modeling similar products.

3.2 Experimental Method

Based on factors such as data integrity and sufficiency, 16,000 loan records on the X platform were collected. The risk assessment level is divided into written off, grace period (15 days), overdue (16-30 days), overdue (31-90 days) and normal. We choose four evaluation indicators: basic information, repayment ability, bad credit history and external credit status. It is generally believed that the earlier the overdue occurs, the smaller the impact than the later.

3.3 Gather Data

This article uses Cora data set and IMB data set. Cora data set is used as a common entity analysis and evaluation data set. For example, the entity analysis process based on Markov logic network adopts Cora. This article uses Cora data set to compare the experimental results with the entity analysis based on Markov logic network.

4. Risk Assessment Test of Credit IoT-Enabled Financial Management Based on Machine Learning and Big Data

4.1 Evaluation Index System Based on Index Reliability Testing

Reliability refers to the stability and reliability of the questionnaire results are shown in Table 1.

It can be seen from Table 1 that the effects of the four indicators set in this experiment on the risk assessment are all acceptable (α>0.7). There is no absolute pros and cons between the models.

Table 1: Summary Table of Reliability Test Results

Group	Index combination	alpha coefficient(α)
Binomial Logistic Regression	Basic Information	0.8447
	ability to pay debts	
	Bad credit history	
	External credit status	
Naive Yebes	Basic Information	0.7644
	ability to pay debts	
	Bad credit history	
	External credit status	
IBk	Basic Information	0.7224
	ability to pay debts	
	Bad credit history	
	External credit status	
Traditional risk assessment model	Basic Information	0.7382
	ability to pay debts	
	Bad credit history	
	External credit status	

4.2 Four Indicators by Different Risk Assessment Systems

Here we test and analyze the results of sample processing on different algorithmic model assessment risk systems. The result is shown in Figure 6.

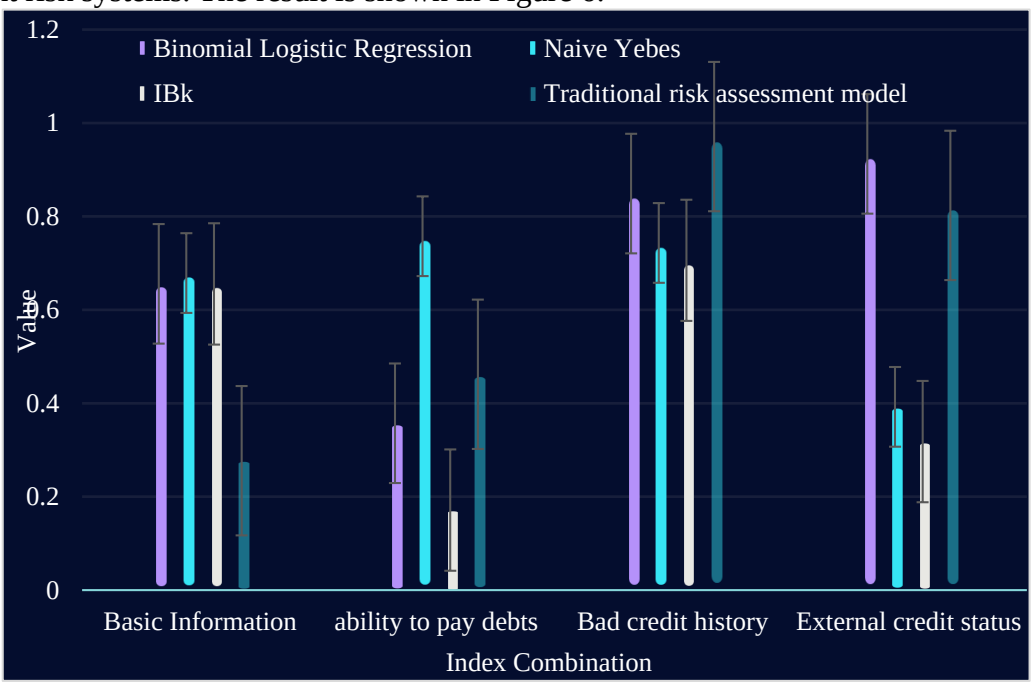

Figure 6: Analysis diagram of sample processing results of different risk assessment systems

It can be seen from Figure 6 that the loan period is mainly 36 periods. It is generally believed that the longer the period. However, the premise of higher returns is that the borrower can repay the principal and interest on time. As the maturity becomes longer, especially when the loan maturity is

not in the same economic cycle, the borrower's repayment ability will change with the economic cycle, causing the default rate to increase.

4.3 Results of Sample Processing by Different Risk Assessment Systems

(1) Risk assessment system based on binomial logistic regression algorithm

The result is shown in Figure 7. At the same time, and the mean, standard deviation, standard error of the mean and other data are obtained, as shown in Table 2.

Table 2: Logistic regression algorithm risk assessment system on the result analysis table

	Mean	Standard deviation	Standard error of the mean
Basic Information	-1.34	8.347	1.682
ability to pay debts	-2.73	7.324	1.685
Bad credit history	0.32	7.685	1.679
External credit status	0.44	6.924	1.680

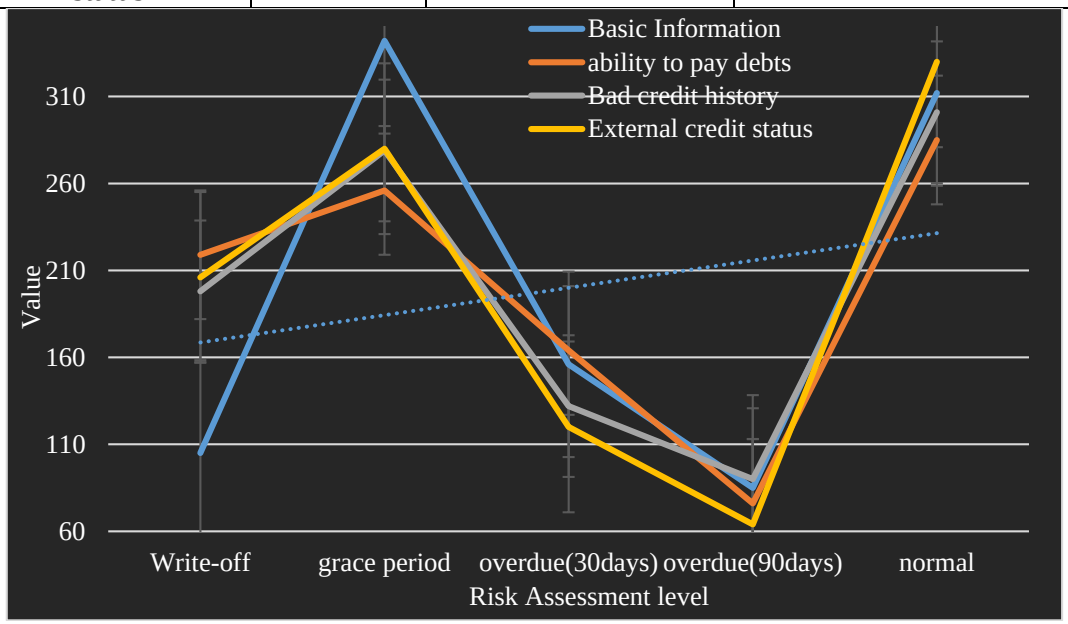

Figure 7: Analysis of the results of the risk assessment system based on the binomial logistic regression algorithm

It can be seen from Table 2 that at the same time each sample is tested in pairs, the standard error level of the mean is 0.262, indicating that for a borrower, there are currently a certain number of credit card accounts that need to be repaid. His level of default risk is high. It can be seen from Figure 7 that the number of credit accounts opened in a short period of time reflects the current capital circulation of borrowers, especially those with a short credit history. The length of the account duration reflects the historical credit management time of the borrower. Multiple borrowing applications will reveal that the borrower's recent funds are relatively tight and the possibility of late repayment is greater.

(2) Risk assessment system based on Yebes algorithm

Each sample is tested on the selected samples, and the mean, standard deviation, standard error of the mean and other data are obtained, as shown in Table 3.

Table 3: Analysis table of the result analysis of the risk assessment system of Yebes algorithm

	Mean	Standard deviation	Standard error of the mean
Basic Information	2.47	9.642	1.426
ability to pay debts	1.36	8.473	1.424
Bad credit history	-0.32	6.756	1.429
External credit status	-0.92	8.279	1.431

It can be seen from Table 3 that at the same time each sample is tested in pairs, the standard error level of the mean is 0.004, indicating that in the credit business, the binary data is often unbalanced, because defaults are always less, but once it happens, it will cause serious negative effects. As can be seen from Figure 7, the credit rating of the risk assessment is very strong.

(3) Risk assessment system based on IBk algorithm

Data such as the mean, standard deviation, and standard error of the mean are obtained, as shown in Table 4.

Table 4: IBk algorithm risk assessment system analysis table of results

	Mean	Standard deviation	Standard error of the mean
Basic Information	1.49	11.624	1.965
ability to pay debts	1.42	11.523	1.923
Bad credit history	-1.33	9.643	1.275
External credit status	-1.28	9.523	1.243

It can be seen from Table 4 that at the same time each sample is tested in pairs, the prediction accuracy rate has reached at a level of around 85%, the credit risk of online lending is very complicated in reality, and this accuracy has good application prospects.

5. Conclusions

The credit risk assessment model is an important support tool for modern banks to provide credit loans. It can greatly reduce the bank's non-performing loan interest rate, increase bank profits and hedge risks. At the same time, it can promote the development of services after paying consumer credit in the telecommunications, retail and tourism industries, which is of great significance for maintaining the stability of the IoT financial market and avoiding financial crises and financial risks. The credit risk assessment model mentioned in this article can be applied to the credit risk assessment of financial institutions, and its effectiveness has been proven in some banking applications. In short, with the development of economy and society, the expansion of financial institutions' activities, the increasing demand for credit consumption and the increasingly important demand for credit risk assessment, the use of machine learning and big data technology to create credit model capacity has unique advantages and comparisons. The importance and value of high research value use has broad application prospects.

Acknowledgment

Research on the Reform and Innovation of the Joint Cultivation Model of Government, Industry, Academia and Research for New Liberal Arts Talents (2021JG004).

References

[1] Jan M A, Khan F, Alam M, et al. A payload-based mutual authentication scheme for Internet of Things. Future Generation Computer Systems, 2017, 92(MAR.):1028-1039.

[2] Liu X, Zhao S, Liu A, et al. Knowledge-aware Proactive Nodes Selection approach for energy management in Internet of Things. Future Generation Computer Systems, 2017, 92(MAR.):1142-1156.

[3] Krotov V. The Internet of Things and new business opportunities. Business Horizons, 2017, 60(6):831-841.

[4] Raja N B, Ihsan Necla, et al. Correction to: Landslide susceptibility mapping of the Sera River Basin using logistic regression model. Natural Hazards, 2018, 91(3):1423-1425.

[5] Suriya K, Niwattisaiwong S. Adaptive Logit Model: Enhancement of Predictability in Logistic Regression. International Journal of Intelligent Technologies and Applied Stats, 2019, 12(2):169-181.

[6] Zhang D, Ren N, Hou X. An improved logistic regression model based on a spatially weighted technique (ILRBSWT v1.0) and its application to mineral prospectivity mapping. Geoscientific Model Development, 2018, 11(6):2525-2539.

[7] Lixin Z, Long W. Construction of the Logistic Regression Estimation Model in Early Warning on Pure Financial Indicators. Open Cybernetics & Systemics Journal, 2016, 10(1):292-293.

[8] Chen J, Tao Y, Wang H. Big data based fraud risk management at Alibaba. Journal of Finance & Data ence, 2015, 1(1):1-10.

[9] Gupta S, Provenzale D, Regenbogen S E, et al. NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 3.2017. Journal of the National Comprehensive Cancer Network: JNCCN, 2017, 15(12): 1465-1475.

[10] Cvitić, I., Peraković, D., Periša, M., & Stojanović, M. D. (2021). Novel Classification of IoT Devices Based on Traffic Flow Features. Journal of Organizational and End User Computing (JOEUC), 33(6), 1-20. http://doi.org/ 10.4018/JOEUC.20211101.oa12

[11] Zhang, B., Wen, K., Lu, J., & Zhong, M. (2021). A Top-K QoS-Optimal Service Composition Approach Based on Service Dependency Graph. Journal of Organizational and End User Computing (JOEUC), 33(3), 50-68. http://doi.org/10.4018/JOEUC.20210501.oa4

[12] Khalaf, O.I., Abdulsahib, G.M. Optimized dynamic storage of data (ODSD) in IoT based on blockchain for wireless sensor networks. Peer-to-Peer Netw. Appl. (2021). https://doi.org/10.1007/s12083-021-01115-4

[13] Zhao, H., Chen, P.-L., Khan, S. and Khalafe, O.I. (2021), "Research on the optimization of the management process on internet of things (Iot) for electronic market", The Electronic Library, Vol. 39 No. 4, pp. 526-538. https://doi.org/10.1108/EL-07-2020-0206

[14] S Hariharan, Monika Gupta, Improving Cloud-based ECG Monitoring, Detection and Classification using GAN, Fusion: Practice and Applications, Vol. 2, No. 2, (2020) : 42-49 (Doi : https://doi.org/10.54216/FPA.020201)

[15] Samia Mandour, Ibrahim el-henawy, Kareem Ahmed, An Improved Equilibrium Optimizer Algorithm for Tackling Global Optimization Problems, Fusion: Practice and Applications, Vol. 3, No. 1, (2021) : 01-28 (Doi : https://doi.org/10.54216/FPA.030101)

[16] Selçuk Topal, Ferhat Tas, Said Broumi, Oguz Ayhan Kirecci, Applications of Neutrosophic Logic of Smart Agriculture via Internet of Things, International Journal of Neutrosophic Science, Vol. 12, No.2, (2020) : 105-115 (Doi : https://doi.org/10.54216/IJNS.120205)

[17] Lu H. (2023). Empirical Test of Credit Risk Assessment of Microfinance Companies Based on BP Neural Network. International Journal of Information Technologies and Systems Approach, 16(3), 1-14.