Construction of Inclusive Finance Risk Assessment and Management Model Based on Big Data

Hao Li

Liaoning Communication University, Shenyang, 110136, Liaoning, China

245671265@qq.com

Keywords: Big Data; finance Risk; Management Model; inclusive finance

Abstract: The purpose of this study is to build a risk assessment and management model in inclusive finance based on big data technology, so as to improve the ability of financial institutions to identify and manage loan default risks. First of all, the paper collected a lot of data including personal loan application information, personal credit history, income, work information, etc., and preprocessed and extracted its features. Then, the paper chooses Logistic regression model as the evaluation model, and evaluates and optimizes the model by using cross-validation and other methods. The empirical analysis results show that our model performs well on the test set, with high accuracy, precision, recall rate and F1 value, and can effectively identify the risk of loan default. In addition, through the analysis of the characteristics and importance of the model, it is found that personal credit score and debt ratio have great influence on the prediction of loan default risk. To sum up, this study provides a risk assessment and management method based on big data for inclusive finance, which has important theoretical and practical significance.

1. Introduction

With the continuous development and popularization of science and technology, inclusive finance has become an important development direction in the global financial field. Inclusive finance's philosophy is to enable more people to obtain financial services including credit, savings and insurance through innovative financial products and services, especially those groups that are hard to reach by the traditional financial system [1-2]. However, there are many challenges facing inclusive finance, one of which is risk management.

The characteristics of inclusive finance determine that it faces diversified and complicated risks, including credit risk, market risk and operational risk [3]. Effective evaluation and management of these risks is not only related to the stable operation of financial institutions, but also directly affects the rights and interests of financial service objects. Traditional risk assessment and management methods are often limited by data sources, data quality and model accuracy, and it is difficult to meet the special needs of inclusive finance. The rise of big data technology provides new ideas and tools for risk management in inclusive finance [4]. Big data is massive, high-dimensional, diversified and real-time, which can help financial institutions to understand customers' credit status, market trends and operations more comprehensively and accurately, thus improving the efficiency and accuracy of risk assessment and management [5].

This paper aims to explore the construction of risk assessment and management model in inclusive finance based on big data. Firstly, the concept and development status of inclusive finance will be analyzed, then the application of big data in the financial field will be introduced, and then the model construction framework of risk assessment and management in inclusive finance based on big data will be put forward, and the effectiveness of the model will be verified through empirical analysis. Through the research of this paper, it is expected to provide more scientific and accurate risk assessment and management methods for institutions in inclusive finance, so as to promote the sustained and healthy development of inclusive finance and realize the inclusiveness and inclusiveness of financial services.

  DOI: 10.25236/etmhs.2024.029

2. Overview of inclusive finance

Inclusive finance, that is, universal financial services, means that through innovative financial products and services, more people, especially low-income people and rural residents, can obtain financial services including credit, savings, payment, insurance and so on, so as to achieve the goals of economic growth, social stability and sustainable development of individuals and families. The essence of inclusive finance is to extend financial services to groups that are hard to reach by the traditional financial system, so as to eliminate poverty and promote social and economic development [6-7].

The emergence of inclusive finance stems from the shortcomings and defects of the traditional financial service system. The traditional financial service system tends to serve high-income people and urban residents, ignoring the financial needs of rural areas and low-income groups. In addition, the traditional financial service system usually has high thresholds and risks when providing financial services to individuals with bad or no credit records, which makes it difficult for these people to obtain financing support, thus limiting their economic activities and social development.

The development of inclusive finance aims to solve these problems and provide affordable, convenient and controllable financial services for more people. The characteristics of inclusive finance include: inclusiveness. Inclusive finance is committed to serving everyone, especially those groups that are hard to reach by the traditional financial system, such as rural residents, small and micro business owners and low-income families. Accessibility, inclusive finance emphasized the universality and convenience of financial services, and made financial services more convenient and accessible through technical means and financial innovation [8]. Sustainability, inclusive finance aims to promote economic growth and social development, and realize long-term prosperity of individuals and society through the sustainability of financial services. Risk management, inclusive finance should not only meet the needs of financial clients, but also ensure that the risks of financial institutions can be controlled, so risk management is an important part of inclusive finance's development.

3. Construction of risk assessment and management model in inclusive finance

Risk assessment and management in inclusive finance is very important for the stable operation of financial institutions [9]. Traditional risk assessment and management methods are often limited by data sources, data quality and model accuracy, and it is difficult to meet the special needs of inclusive finance. The rise of big data technology provides new ideas and tools for risk management in inclusive finance. This paper aims to propose a risk assessment and management model of inclusive finance based on big data, and elaborate it in detail. The framework of inclusive finance risk assessment and management model based on big data mainly includes several steps as shown in Figure 1.

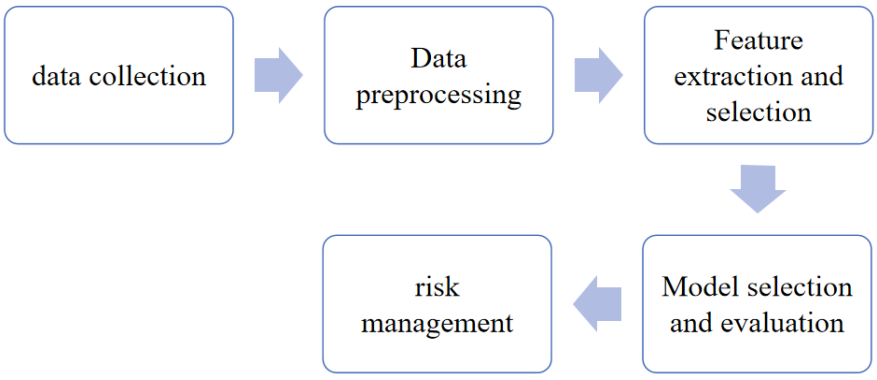

Figure 1 Construction steps of the model

Data collection, collecting all kinds of data related to inclusive finance, including customer information, transaction records, market data, etc. These data can come from internal systems of financial institutions, third-party data providers and open data sources. Data preprocessing:

preprocessing the collected data, including data cleaning, data conversion, data integration and data specification, to ensure the quality and consistency of data. Feature extraction and selection, extracting useful features from preprocessed data, and using feature selection method to screen out the most predictive features. Model selection and evaluation, choose the appropriate risk assessment model, and use cross-validation and other methods to evaluate and optimize the model. Risk management, according to the prediction results of the model, take corresponding risk management measures, including risk pricing, risk dispersion and risk monitoring.

Collected all kinds of personal information including personal loan application information, personal credit history, income, work information, and historical data of personal loan default. Clean, transform, integrate and standardize the collected data to ensure the quality and consistency of the data [10]. A series of features, such as personal credit score, debt ratio and income level, are extracted by statistical methods, and the most predictive features are screened out by feature selection method.

If there are n samples, m features, and X_{ij} represents the j-th characteristic value of the i-th sample, a series of features can be extracted by various statistical methods, and recorded as F_i. Let y_i represent the label of the ith sample and x_i represent the feature vector of the ith sample, then the importance of the feature can be calculated by the following formula:

$$importance(F_i) = \frac{variance(y_i|x_i)}{variance(y_i)} \tag{1}$$

Logistic regression model is selected as the evaluation model, and cross-validation and other methods are used to evaluate and optimize the model. Logistic regression model can be expressed as:

$$P(y=1|x) = \frac{1}{1+e^{-\beta^T x}} \tag{2}$$

Among them, y represents the label of loan default (1 represents default, 0 represents non-default), x represents the personal characteristic vector, and β represents the parameters of the model.

According to the prediction results of the model, measures such as risk pricing, risk diversification and risk monitoring can be taken to reduce the risk of loan default.

4. Empirical analysis

After constructing the inclusive finance risk assessment and management model based on big data, the study proceeds with an empirical analysis to validate the model's effectiveness. The analysis utilizes datasets comprising personal loan application details, individual credit histories, income, and employment information, encompassing a specific proportion of loan default cases and instances of regular repayments. The dataset is partitioned into a training set and a test set, with the former utilized for model training and parameter optimization, while the latter is employed to assess the model's performance. The evaluation model chosen is logistic regression, and the model undergoes evaluation and refinement via cross-validation and other methodologies.

Table 1 Performance of the model on the test set

Performance index	value
accuracy	0.95
precision	0.92
Recall	0.94
F1 value	0.93

Table 1 provides four performance indexes of the model on the test set, such as accuracy,

precision, recall and F1 value. According to the numerical values in the table, it can be concluded that the model performs well in identifying the risk of loan default, and all performance indicators are high, which shows that the model has good forecasting ability.

The model achieves an accuracy of 0.95 on the test set, indicating that it correctly predicts 95% of the samples. This high accuracy demonstrates the model's effectiveness in predicting overall loan default risk, with prediction results aligning well with real-world scenarios. Additionally, the model exhibits a precision of 0.92 on the test set, meaning that 92% of samples predicted as loan defaults by the model are indeed instances of loan default. This high precision indicates the model's ability to accurately identify loan default risks and minimize misjudgments. Furthermore, the model achieves a recall rate of 0.94 on the test set, indicating that it successfully identifies 94% of loan default samples. This high recall rate demonstrates the model's capability to detect loan default cases effectively and mitigate instances of missed judgments. The F1 value of the model on the test set is 0.93, which is the harmonic average of precision and recall rate, and the accuracy and coverage of the model are considered comprehensively. The high F1 value indicates that the model has achieved a balanced performance in the identification of loan default risk, that is, it can capture positive samples while ensuring the accuracy.

The performance of the model on the test set is good, with high accuracy, precision, recall and F1 value, which shows that the model has good forecasting ability and can effectively identify the risk of loan default. These results are of great significance to financial institutions, which can help them to formulate more accurate risk management strategies and reduce the risks caused by loan default.

Figure 2 shows the results of feature importance analysis of inclusive finance risk assessment and management model based on big data. Personal credit score is of great importance, which means that personal credit score is one of the important factors to predict the risk of loan default. Credit score usually reflects the borrower's credit history and credit risk, and a lower credit score may indicate that the borrower has poor credit and is more prone to default. The debt ratio is also of high importance. Debt ratio refers to the ratio of the borrower's total debt to its total income. A higher debt ratio may indicate that the borrower's financial situation is poor and it is difficult to bear additional debts, thus increasing the risk of loan default. Income level is also important for the prediction of loan default risk. The income level usually reflects the borrower's repayment ability. A higher income level may mean that the borrower has enough income to repay the loan, which reduces the possibility of default.

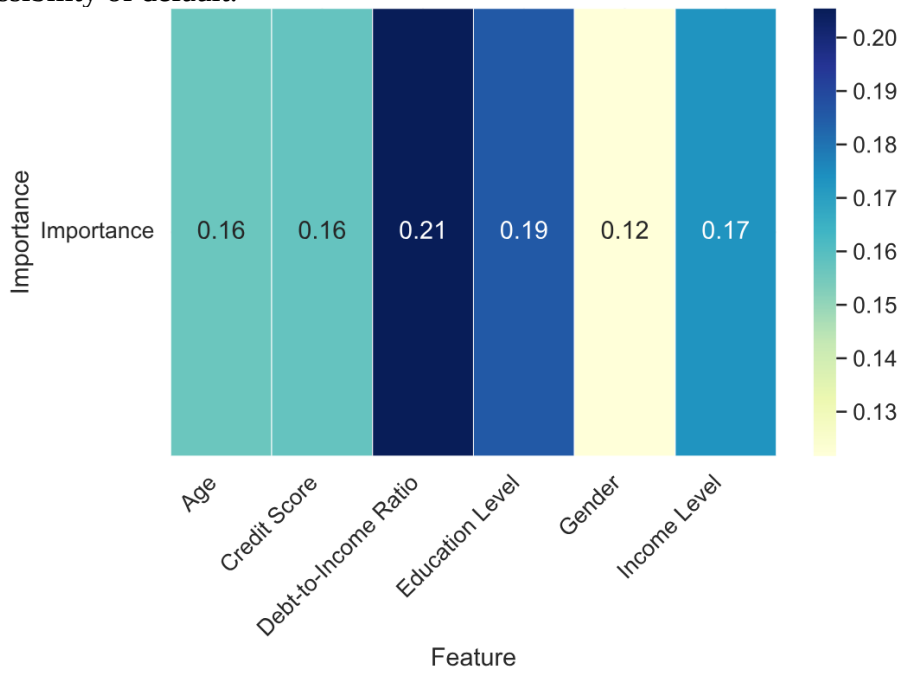

Figure 2 Feature importance analysis

To sum up, personal credit score, debt ratio and income level are the characteristics that have great influence in predicting loan default risk. Therefore, in the field of inclusive finance, it is very important to fully evaluate and manage these characteristics, so as to reduce the risk of loan default and ensure the stability of the financial system and the interests of borrowers.

5. Conclusion

Based on big data technology, this study constructs a risk assessment and management model of inclusive finance, aiming at improving the prediction accuracy and risk management efficiency of financial institutions for loan default risk. The model performs well on the test set: through empirical analysis, we find that the accuracy, precision, recall and F1 value of the model are all high on the test set, which shows that the model has good forecasting ability and generalization ability, and can effectively identify the risk of loan default. By analyzing the characteristics and importance of the model, it is found that personal credit score and debt ratio have great influence on the prediction of loan default risk. These results provide an important reference for financial institutions, and help them to evaluate the credit risk of loan applicants more accurately and formulate more effective risk management strategies. Inclusive finance risk assessment and management model based on big data has the potential of practical application, which can provide more reliable risk management tools for financial institutions. However, it is also realized that there is still room for improvement in the model, such as further optimizing model parameters, adding more characteristic variables and improving data processing methods, so as to improve the prediction performance and stability of the model. The future research direction can focus on the further optimization of the model and the expansion of application scenarios, so as to better meet the actual needs of inclusive finance.

References

[1] Hao, Y. (2021). Digital inclusive finance risk prevention based on machine learning and neural network algorithms. Journal of Intelligent and Fuzzy Systems, 2021(2), 1-11.

[2] Peng, H. , Lin, Y. , & Wu, M. (2022). Bank financial risk prediction model based on big data. Scientific programming, 2022(5), 2022.

[3] Zhang, M. , Luo, D. , & Su, Y. (2022). Drought monitoring and agricultural drought loss risk assessment based on multisource information fusion. Natural Hazards, 2022(1), 111.

[4] Liu, Y. , Lu, J. , Mao, F. , & Tong, K. (2019). The product quality risk assessment of e-commerce by machine learning algorithm on spark in big data environment. Journal of Intelligent & Fuzzy Systems, 37(4), 4705-4715.

[5] Henriksen-Bulmer, J. , Faily, S. , & Jeary, S. (2019). Privacy risk assessment in context: a meta-model based on contextual integrity. Computers & Security, 2019(82), 82.

[6] Lin, S. S. , Shen, S. L. , Zhou, A. , & Xu, Y. S. (2021). Risk assessment and management of excavation system based on fuzzy set theory and machine learning methods. Automation in Construction, 122(4), 103490.

[7] Tan, L. , Yao, W. , Chen, F. , & Li, L. (2020). Economic loss assessment of tropical cyclones based on bibliometric data analysis. Tropical Conservation Science, 13(1), 895.

[8] Cao, Q. (2023). An enterprise financial data leakage risk prediction based on arima-svm combination model. International Journal of Applied Systemic Studies, 2023(3), 10.

[9] Li, S. , & Quan, Y. (2019). Financial risk prediction for listed companies using ipso-bp neural network. International Journal of Performability Engineering, 15(4), 1209-1219.

[10] Song, Y. , & Peng, Y. (2019). A mcdm-based evaluation approach for imbalanced classification methods in financial risk prediction. IEEE Access, 2019(99), 1.