Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 3786707, 9 pages
https://doi.org/10.1155/2022/3786707

Research Article

Research on the Method of Predicting Consumer Financial Loan Default Based on the Big Data Model

Qiuying Wang

College of Economics and Management, Shanghai Ocean University, Shanghai 201306, China

Correspondence should be addressed to Qiuying Wang; 1945101@st.shou.edu.cn

Received 19 January 2022; Accepted 22 February 2022; Published 24 March 2022

Academic Editor: Kalidoss Rajakani

With the continuous advancement of Internet technology and the continuous development of big data model applications, the data model is a method of data organization and storage, which emphasizes the reasonable storage of data from the perspective of business, data access, and use. The rapid development of information technology on the Internet and the increasing consumption level of people have created conditions for the booming development of consumer finance. Compared with traditional lending methods, more and more consumers are more inclined to choose consumption channels such as the Internet and e-commerce. Consumer finance lending is more convenient and fast. The prediction of consumer financial loan default is very important, because the inaccuracy of the prediction not only leads to the loss of profits but also infringes the rights and interests of consumers. Therefore, it is very important to propose a default prediction method in the consumer finance field with good performance based on the big data model. Simulation experiment conclusions are shown as follows: (1) the pseudo R-squared value of the model is 0.3660, indicating that the control variable can better explain the change of y. (2) The chi-square test statistic is equal to 51632.31, the degree of freedom is 68, and the corresponding $P < 0.0001$, which also shows that the entire model can significantly predict the change of y. (3) The regression coefficient of the number of loan performances is -0.207553, indicating that the number of consumer loans is negatively correlated with loan defaults. (4) The regression coefficient of the monthly loan frequency is 0.0500152, indicating that the customer applies the frequency of personal credit consumer loans which is positively correlated with loan defaults. (5) The accurate prediction ratio of the model is 86.11%, which further shows that the prediction model has a better effect.

1. Introduction

As the name implies, big data is a data set with a large amount of data. Generally, it has these characteristics: (1) capacity—the amount of data is large, and the size of the data determines the value and potential information of the data under consideration; (2) type—there are many types of data, including but not limited to text, audio, video, and pictures; (3) speed—this refers to the fast speed of data generation and acquisition; (4) low value density—the magnitude of valuable data in the data is small; and (5) authenticity—data quality factor differences in data sources and recording methods and other influencing factors will cause big differences, and this difference will greatly affect the accuracy of data analysis. Financial institutions or companies deepen the application of financial technology to provide consumers with microfinance products and more inclusive financial service technologies to meet consumer demand for financial services (such as Internet consumer lending). Consumer finance lending is a personal loan submitted by consumers to financial institutions or nonfinancial institutions when they need to increase their spending power, including but not limited to consumer goods loans, service consumer goods loans, and credit card loans, as well as a small number of vehicle mortgage loans and housing mortgage loans. Therefore, consumer finance lending is playing an increasingly important role. The development of consumer finance lending has also brought a lot of convenience to people. It is no longer necessary to complete the lending behavior in advance when purchasing loans. Various consumer behaviors can be implemented simply by choosing the consumer finance lending method when paying. Therefore, consumer finance lending will soon become an important part of people's daily consumption behavior [1–3].

In recent years, the rapid development of China's economy and social consumer goods retail has provided consumer finance with a better macroeconomic environment and market support. With the level of people's material consumption and the popularization of Internet e-commerce, the scale of online consumption transactions has increased year by year to build more diverse consumption scenarios for people. Consumer financial lending products such as Huabei, Bibai, and Baitiao are not the emergence of interruptions having deepened the vigorous development of consumer finance. 2020 is a special year for China and the whole world. In the face of the unprecedented new crown epidemic in the world, the pressure of macroeconomic growth is increasing, and the development and prevention and control of the epidemic are even more important for consumer finance. The entire financial industry in China has had a huge impact, and the world's economic development has been affected by the most serious challenge. The formation of consumer finance network lending habits is also expected to continue to accelerate the development of consumer finance [4, 5]. Consumer finance lending methods naturally become people's choice when they consume with their advantages such as high efficiency, convenience, and inclusiveness.

It is worth noting that the lending business of consumer finance companies in various countries is usually for the whole world, and the scale of consumer finance lending around the world is also increasing year by year, and the nonperforming loan rate is also a problem faced by various companies. In other words, it is imperative to build a good-performance consumer finance loan default prediction model. In recent years, with the vigorous development of the social economy, the domestic consumer finance market has developed rapidly, and the scale of various consumer finance lending businesses has continued to expand. However, consumer finance also has its flaws. The feature of no mortgage guarantee makes the risk of consumer finance companies relatively high when issuing loans, and because the lender's income stability, repayment ability, and their own ethics are uncertain, the lending business is risky. It becomes more difficult to control and predict. With the increase in the strictness of the supervision of collection methods, the cost of bad debts caused by nonperforming loans has become one of the most troublesome problems for the entire industry [6–8]. When a consumer finance company reviews a loan applicant, if it fails to make a correct assessment of its credit risk and repayment ability, it will cause serious losses.

Consumer finance is a multiparty personal-oriented financial innovation business with the nature of inclusive finance. Through the research of personal default risk prediction based on big data models, it is aimed at mining the rich information hidden behind the consumer finance field, establishing strong distinguishing ability, The personal default risk early warning system with high prediction accuracy and stable operation effect, and the rapid development of related knowledge based on big data models, has laid a good technical foundation for the construction of prediction models. At the same time, through the analysis of the problems in the credit data of the specific business of credit card, consumer credit, and lending in consumer finance, in-depth research and exploration are carried out for each problem, and corresponding solutions are proposed, which can solve the default risk in the future consumer finance field. Forecast research provides strong support, and it has shifted from facing high-end business customers to a truly inclusive financial development strategy facing more ordinary people. It has very important theoretical significance and practical value [9, 10].

2. The Relevant Basic Theories of Consumer Finance Loan Defaults

2.1. Connotation and Characteristics of Consumer Finance. Consumer finance is the capital and capital financing around the consumer value chain, including credit cards, consumer credit, P2P lending, and other modes. Consumer finance is characterized by small amounts, decentralization, precision, efficiency, and emergency. It is a promotional tool for consumer products and services and a means of financial value-added. The core is to change the traditional commodity transaction model of "life-oriented" operation that realizes the allocation of consumption and financial resources across time and space and gains convenience, efficiency, and additional benefits, allowing consumers to quickly and conveniently obtain goods or services and allowing merchants to destock and increase revenue, so that merchants and financial Institutions and consumers have become a community of shared and mutually beneficial interests, gradually realizing "customization on demand, precision marketing, and mass customer acquisition."

In the "Internet +" era, business models have gradually shifted from marketing-driven to data-driven and Internet technology-driven upgrades. Internet tools are widely used, and their value advantages such as high efficiency, precision, real-time, and low cost are prominent. Therefore, the "three laws of the network" promote digitization to the greatest extent, expand market scale, improve economic governance, and reduce manufacturing and replication costs. Finally, the consumer finance business model has undergone a qualitative change [11–13]. The detailed description of the three major laws of the network is shown in Figure 1.

The following mainly introduces three modes in consumer finance [14–16], as shown in Figure 2.

2.2. Connotation and Characteristics of Default Risk. The exchange of debt in credit is always associated with credit, leverage, and risk. Low credit plus high leverage will inevitably lead to risks. The personal default risk in the context of consumer finance is mainly due to the default behavior of credit customers' inability to repay or subjective unwillingness to repay loans, which causes financial institutions to suffer the risk of capital loss. It still belongs to the research category of credit risk in essence. In a broad sense, credit risk refers to the risk of losses to the counterparty due to the counterparty defaulting in the course of credit transactions. For the narrow concept of credit risk, personal credit risk

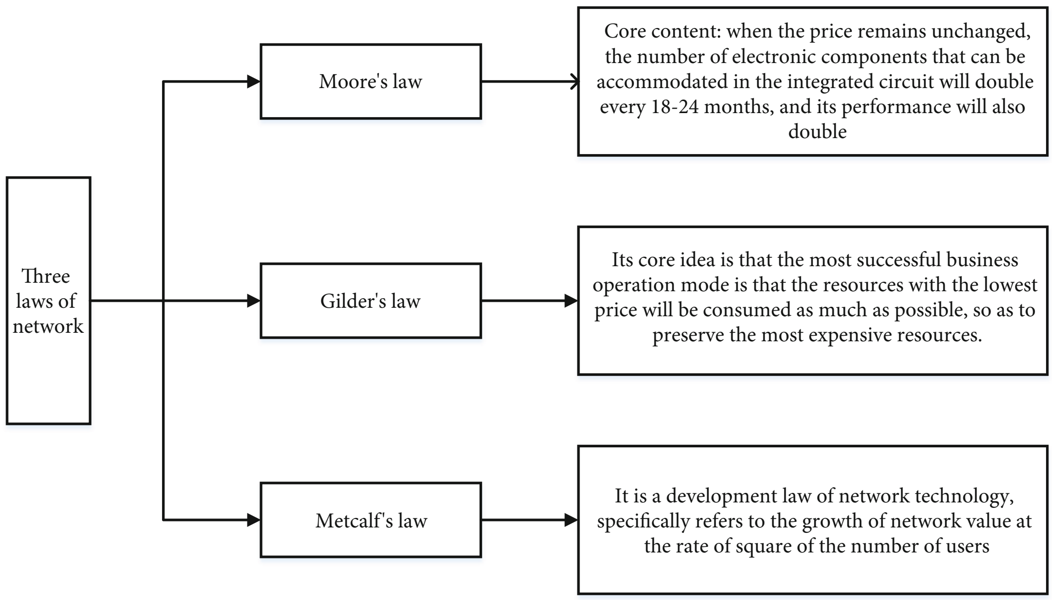

Figure 1: Three laws of network.

can also be called personal credit risk or personal default risk, which refers to the possibility that the borrower will default or suffer losses due to various reasons like unwillingness or inability to perform the contract and cause losses to the commercial bank. The nature of commercial bank loss refers to the risk of commercial bank loan default. Loan risk occurs not only in the credit check stage but also in the entire credit process: in the actual credit approval process, most of the credit check process is not very strict and comprehensive, so the possibility of nonperforming loans is increasing every day. With this in mind, a scientific and effective explanatory model must be established to evaluate and assess the creditworthiness of credit customers in order to minimize the risk of default and maximize profits.

In the consumer finance market, an important reason for the default behavior of credit customers is caused by information asymmetry. For a financial market under a complete information situation, the wealthy of funds can understand and master all the information of the demanders of funds, so they can fully understand the risk factor information of the other party, and then they can fully formulate when making loan decisions. However, in the real financial environment, it is impossible for us to grasp all of its risk information. When there is a risk, it will cause the occurrence of loan default risk, causing financial institutions to incur loan losses. The occurrence of default risks is caused by loan structure. For example, financial institutions can distinguish between loan customers by adjusting loan interest rates to achieve a trade-off between risks and returns. However, interest rates are a double-edged sword. In an environment of high interest rates, financial institutions' income levels can be improved. However, as interest rates continue to increase, customers with high credit levels will not be able to withstand higher interest rates, which leads to this part. The loss of customers eventually leads to more and more customers with high default levels, which makes the probability of default risks higher. The influencing factors of consumer lending are shown in Figure 3.

2.3. Big Data Technology. With the rapid development and popularization of computer and information technology, industry application data has exploded. The society has entered the era of big data and digital economy [17–19]. As a strategic asset of operating companies and countries, big data plays a key role in the development of the data economy. Big data technology can analyze massive amounts of data and make predictions based on statistics and analysis of the data. Since consumer loans are more decisions based on customer credit status than collateral and guarantees, accurate assessment of consumer credit levels is important for reducing information asymmetry in financial transactions, reducing credit risks and transaction costs. In the Internet age, massive amounts of data have become an important thrust to promote the development of Internet consumer financial services. In the Internet era, big data technology and the credit investigation industry have begun to integrate deeply [20, 21]. Data acquisition, mining, and analysis capabilities have gradually become important indicators for evaluating the reliability of the credit reporting system. The development of big data technology has opened up a new credit investigation channel for the Internet consumer finance platform. Big data credit investigation has gradually become an important means to promote the accelerated development of the Internet consumer finance industry. With the deepening of big data technology in the Internet finance industry, Internet consumer finance platforms have begun to apply big data technology in the field of credit investigation. The classification of big data technologies is shown in Figure 4.

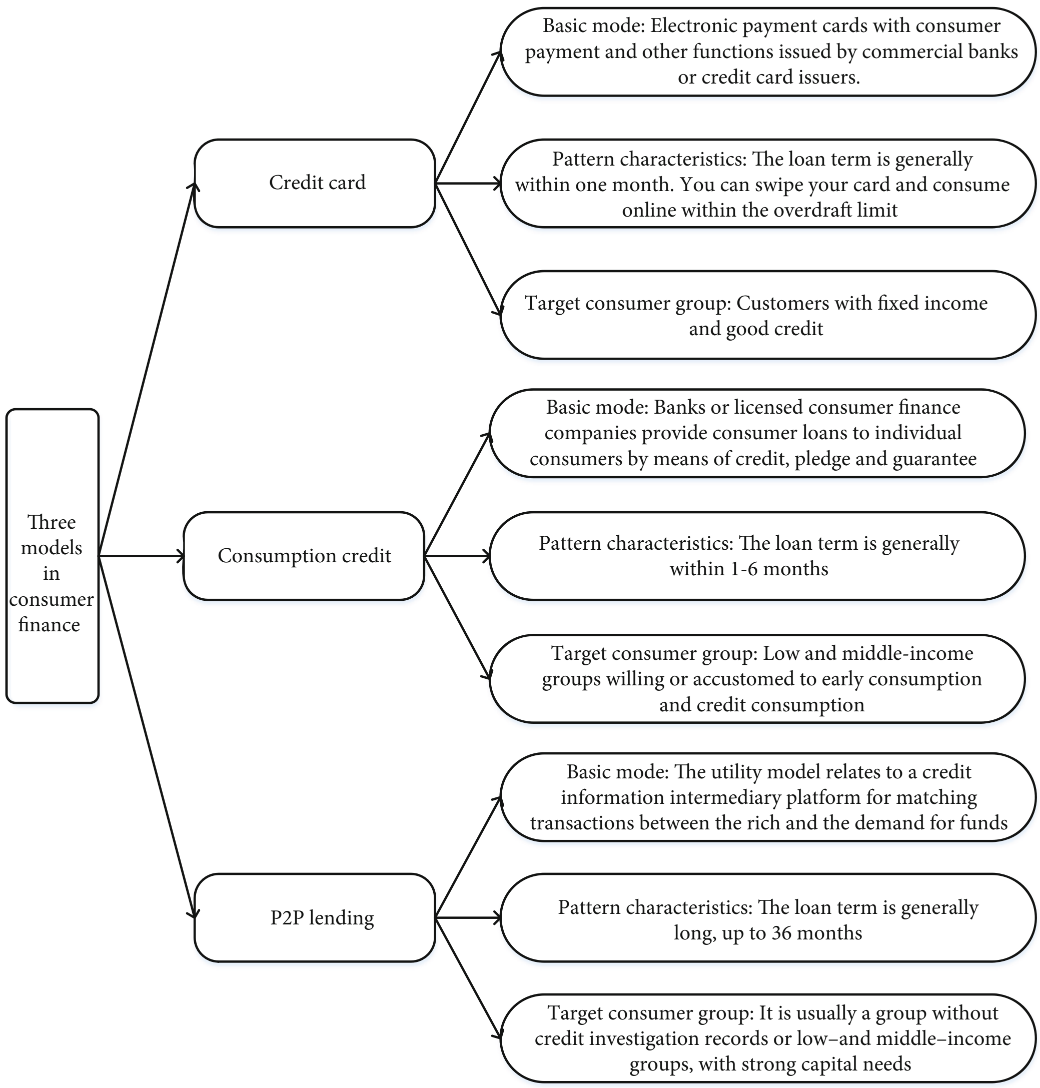

Figure 2: Three modes of consumer finance.

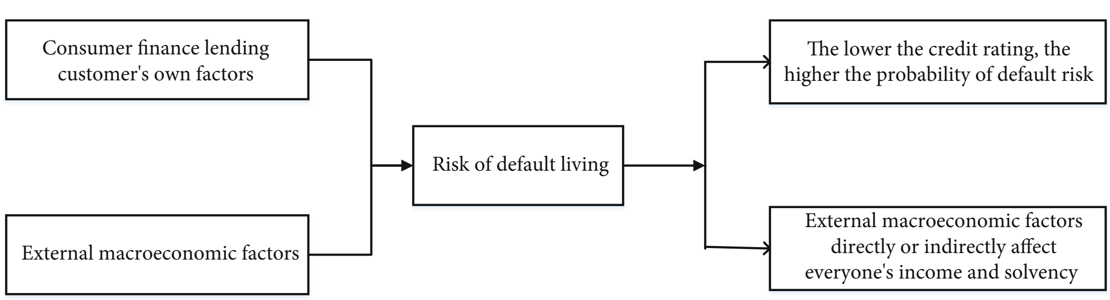

Figure 3: Influencing factors of default risk.

3. The Application of the Big Data Model to the Prediction of Loan Default in Consumer Finance

3.1. Logistic Regression Model. Logistic regression is essentially linear regression [22]. However, the value range of ordinary linear regression is the real number domain, which cannot well measure the probability of an event. Logistic regression is based on ordinary linear regression and normalizes the predicted value by using a function to make the predicted value The value is in the interval (0,1); this function is called the logistic function (logistic function), also called the sigmoid function (sigmoid function). The function expression is as follows:

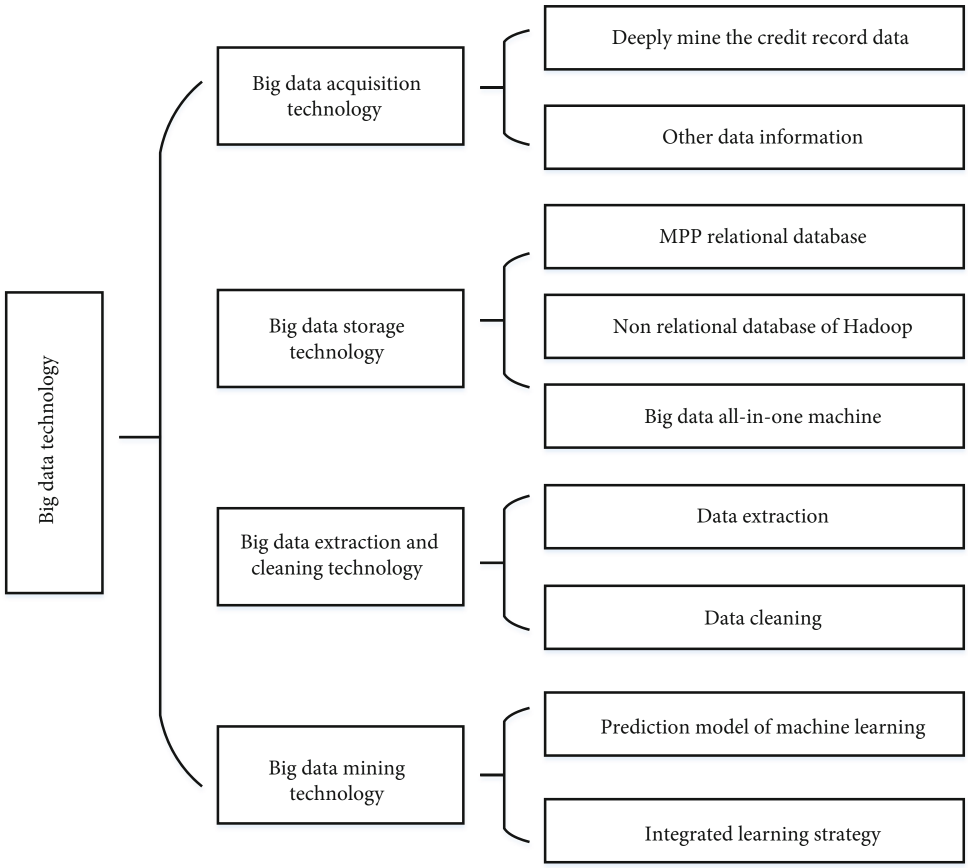

Figure 4: Big data technology classification.

$$f(x) = \frac{1}{1+e^{-z}}. \quad (1)$$

Then, the default probability of consumer finance loan P is

$$P = f(Z) = \frac{1}{1+e^{-Z}}, \quad (2)$$

where Z is shown in the following expression:

$$Z = w_0 + w_1x_1 + w_2x_2 + \cdots + w_nx_n = W^TX. \quad (3)$$

The probability that the observed sample of financial lending is a default sample is

$$p_i = P(y_i = 1|X). \quad (4)$$

For each observation sample, its observation probability is

$$p(y_i) = p_i^{y_i}(1-p_i)^{(1-y_i)}. \quad (5)$$

Obtain the likelihood function of the default prediction model:

$$L(W) = \prod_i^m p(y^i) = \prod_i^m p_i^{y_i}(1-p_i)^{(1-y_i)}. \quad (6)$$

Take the logarithm of both sides to get the log-likelihood function:

$$\ln(L(W)) = \sum_i^m y_i \ln(p_i) + (1-y_i)\ln(1-p_i). \quad (7)$$

The logistic regression equation expression is

$$p(y=1|X) = \frac{\exp\left(\beta_0 + \beta^TX\right)}{1+\exp\left(\beta_0+\beta^TX\right)}. \quad (8)$$

3.2. Lasso-Logistic Regression Model. The loss function of the regression model is as follows:

$$\text{loss} = -\ln(L(W)) + \lambda\|W\|_1, \quad (9)$$

which can be obtained:

$$\|W\|_1 = \sum_{j=0}^n |w_j|. \quad (10)$$

By minimizing the loss function to find the default parameters of consumer finance loans,

$$\widehat{W} = \arg\min_{W} \sum_{i=0}^{m} [-y_i \ln (p_i) - (1 - y_i) \ln (1 - p_i)] + \lambda \sum_{j=0}^{n} |w_j|. \tag{11}$$

3.3. Decision Tree Model. The definition of the purity reduction index is as follows:

$$\triangle i = i(0) - \left[\frac{n_1}{n_0} i(1) + \frac{n_2}{n_0} i(2) + \frac{n_3}{n_0} i(3) + \frac{n_4}{n_0} i(4)\right]. \tag{12}$$

The Gini coefficient is defined as

$$1 - \sum_{j=1}^{r} p_j^2 = 2\sum_{j<k} p_j p_k. \tag{13}$$

The information entropy value is defined as

$$H(p_1, p_2, \cdots, p_r) = -\sum_{i=1}^{r} p_i \log_2(p_i). \tag{14}$$

The Pearson chi-square test is

$$\chi_v^2 = \sum \frac{(0 - E)^2}{E}, \tag{15}$$

which can be drawn:

$$v = (r - 1)(B - 1). \tag{16}$$

The logworth value can be expressed as follows:

$$\text{logworth} = -\log_{10}(P). \tag{17}$$

The reduction index of purity is as follows:

$$i(t) = \frac{1}{n_t} \sum_{j=1}^{n_t} (y_{it} - \bar{y}_t)^2. \tag{18}$$

The F test is as follows:

$$F = \left(\frac{\text{SS}_{\text{Between groups}}}{\text{SS}_{\text{Within group}}}\right)\left(\frac{n - B}{B - 1}\right) \sim F_{B-1,n-B}. \tag{19}$$

The ID3 algorithm calculates the information entropy of the node:

$$\text{info}(D) = -\sum_{i=1}^{m} p_i \log_2 p_i. \tag{20}$$

The expected information required to classify node samples is calculated as follows:

$$\text{info}_A(D) = \sum_{j=1}^{v} \left[\left(\frac{|D_i|}{|D|}\right) * \text{info}(D_i)\right]. \tag{21}$$

The calculated information gain is as follows:

$$\text{Gain}(A) = \text{info}(D) - \text{info}_A(D). \tag{22}$$

3.4. Random Forest Model. The marginal function is calculated as follows:

$$mg(X, Y) = av_k(I(h_k(X) = Y)) - \max_{i \neq Y} av_k(I(h_k(x) = j)). \tag{23}$$

The generalization error is calculated as follows:

$$\begin{aligned} PE^* &= P_{X,Y}(mg(X, Y) < 0), \\ mr(X, Y) &= P(h_k(X) = y) - \max_{\substack{j \neq Y \\ j=1..c}} P(h_k(X) = j). \end{aligned} \tag{24}$$

Then, the random forest model [23] correctly classified the probability estimation of the consumer finance loan default prediction as follows:

$$Q(x, y_i) = \frac{\sum_k I\left(h_k(x) = y_j, (x, y) \in O_k(x)\right)}{\sum_k I(h_k(x), (x, y) \in O_k(x))}. \tag{25}$$

4. Simulation Experiment

4.1. Experimental Data Selection and Variable Interpretation. The initial data set has a total of 138650 samples, of which 46,500 defaulted persons account for approximately 33.54% and a total of 92,150 nondefaulting persons account for approximately 66.46%. The initial data set has too many variables. Variable screening is performed. In order to ensure the quality of the data, the processing of abnormal samples, missing values and outliers, processing of irrelevant variables, processing of duplicate information variables, processing of low-information variables, and other variables are carried out to ensure the quality of the data. Standardization of processing and data processing, and finally 136,750 samples were determined for final experimental verification. In order to facilitate measurement and analysis and to make the model more stable and reduce the risk of model overfitting, this paper performs data quantification and data binning for some variables, as shown in Table 1.

The number of variables is determined as annual income, loan amount, loan period, loan interest rate increase ratio, average monthly wages, number of loan performances, actual occupied credit ratio, monthly loan frequency, gender, age, occupation, education level, and marriage. Model analysis and verification of 13 variables include conditions. The definition of each variable is shown in Table 2.

Table 1: Selection and construction of control variables.

Indicator name	Actual value of index	Quantitative value of index
Gender	Male and female	Male—0, female—`1
Age	Actual value	25, 35, 45, 50, 60, 70
Occupation	Civil servants, employees of public institutions; other industry staff, soldiers, workers, farmers, civil	01, 02, 03, 04, 05, 06, 07, 08, 09, 10
Education level	Unknown, doctor and above, master postgraduate, undergraduate, technical secondary school, high school, junior high school	01, 02, 03, 04, 05, 06
Marital status	Unknown, unmarried, married	Unknown—1 Unmarried—2 Married—3
Annual income	Actual value	1, 2, 3, 4, 5, 6, 7, 8, 9
Loan amount	Actual value	1, 2, 3, 4, 5, 6, 7, 8, 9
Loan term	Actual value	Actual value
Lending rate	Actual value	Actual value
Ratio monthly repayment to income	Actual value	Set to 1 below 20% and add 1 for each subsequent 10% increase
Average monthly payroll	Actual value	Divided by reference to annual income
Number of loan performance	Actual value	Actual value
Monthly loan frequency	Actual value	Actual value

Table 2: Variable definition.

Indicator name	Variable name	Is it included in the model	Variable type
Default flag	y	Yes	Explained variable
Gender	$X1$	Yes	Control variable
Age	$X2$	Yes	Control variable
Occupation	$X3$	Yes	Control variable
Education level	$X4$	Yes	Control variable
Marital status	$X5$	Yes	Control variable
Annual income	$X6$	Yes	Control variable
Loan amount	$X7$	Yes	Control variable
Loan term	$X8$	Yes	Control variable
Proportion of loan interest rate	$X9$	Yes	Control variable
Actual occupied credit ratio	$X10$	Yes	Control variable
Average monthly payroll	$X11$	Yes	Control variable
Number of loan performances	$X12$	Yes	Control variable
Monthly loan frequency	$X13$	Yes	Control variable

4.2. Model Comparison. To compare models, we use the three indicators of accuracy, precision, and recall to evaluate [24, 25]. The specific definitions are as follows: accuracy is the proportion of all samples with correct predictions to the total sample; accuracy is the correct prediction as default The proportion of the sample size in the total number of samples predicted to be in default; the recall rate is the proportion of the number of samples correctly predicted to be in default to the total number of samples that are actually in default. Divide 80% of the 136,750 samples determined in the previous section into training sets and 20% into validation sets according to the principle of random sampling (that is, 109,400 samples are divided into training samples and 27,350 samples are divided into verification samples). Model training and verification are performed on each model, and the specific verification results of the verification set are shown in Figure 5.

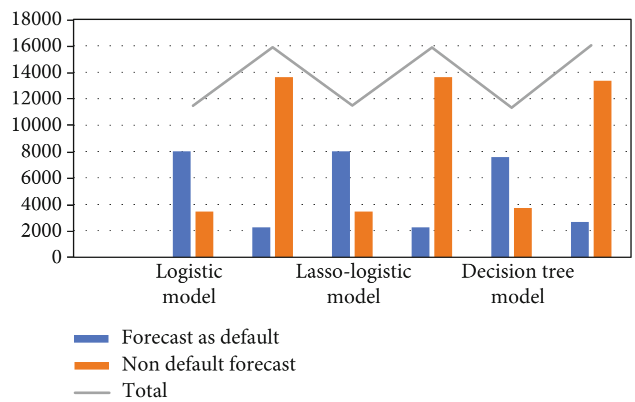

Figure 5: Prediction results of each model validation set.

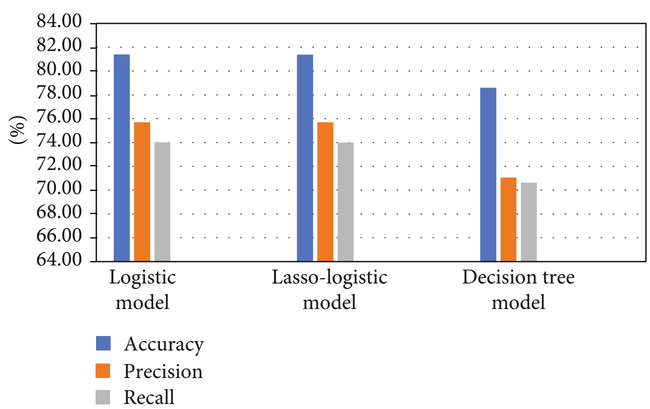

Figure 6: Evaluation index results of each model.

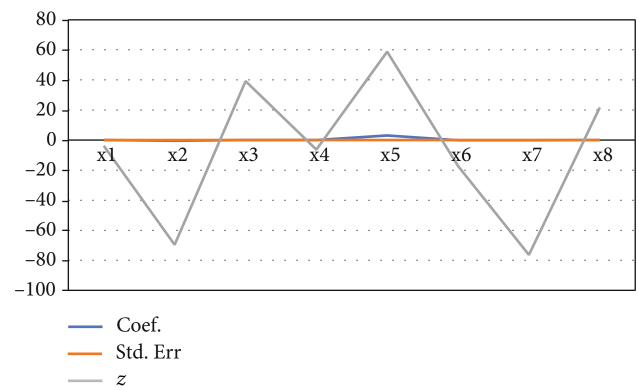

Figure 7: Logistic regression result figure.

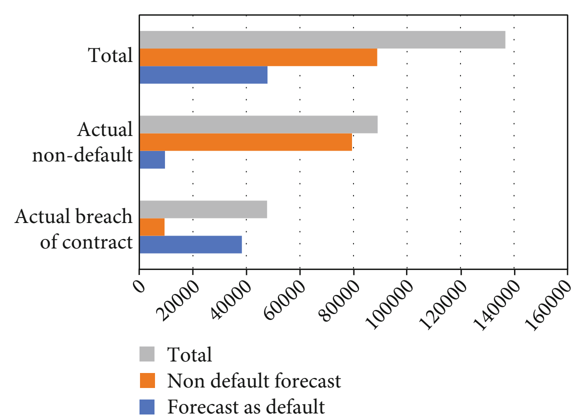

Figure 8: Specific prediction results of the model.

4.3. Model Prediction. According to the above results, the accuracy, precision, and recall indicators of each model were calculated, and the results are shown in Figure 6.

From the comprehensive point of view of the evaluation indicators of each model, the logistic regression model has the best predictive effect. This article will use the logistic regression model for follow-up empirical research.

4.4. Model Application. The detailed description of the prediction result of the logistic regression model is shown in Figure 7.

From the regression results, the pseudo R-squared value (Pseudo R2) of the model is 0.3660, indicating that the control variables included in the model can better explain the changes in the explained variables. The chi-square test statistic LR chi2 is equal to 51632.31, the degree of freedom is 68, and the corresponding $P < 0.0001$, which also shows that the entire model can significantly predict the change of the explained variable. As shown in Figure 8, the ratio of model's accurate prediction is $(38304 + 79447)/136750 =$ 86.11%, which further shows that the prediction model has a better effect and has strong reference significance and guiding value for the prediction of consumer financial loan default. The regression coefficient of the number of loan performances is -0.207553, indicating that the number of consumer loans is negatively correlated with loan defaults; that is, every time a customer applies for personal credit consumer loans, the logarithm of the customer's loan default ratio will decrease by 0.2007553. The regression coefficient of monthly loan frequency is 0.0500152, indicating that the frequency of customers applying for personal credit consumer loans is positively correlated with loan defaults; that is, the frequency of customers applying for personal credit consumer loans increases once a month, and the logarithm of the ratio of loan defaults will increase by 0.0500152.

5. Conclusion

This article researches on the method of predicting the default of financial consumer lending to make the big data model better fit the prediction of consumer lending default. And use a logistic regression model to predict the data. Finally, (1) the pseudo R-squared value (Pseudo $R2$) of the model is 0.3660, indicating that the control variables included in the model can better explain the changes in the explained variables. (2) The chi-square test statistic LR chi2 is equal to 51632.31, the degree of freedom is 68, and the corresponding $P < 0.0001$, which also shows that the entire model can significantly predict the change of the explained variable. (3) The regression coefficient of the number of loan performances is -0.207553, indicating that the number of consumer loans is negatively correlated with loan defaults;

that is, for every increase in the number of times a customer applies for a personal credit consumer loan, the logarithm of the ratio of the customer's loan default will decrease by 0.2007553. (4) The regression coefficient of monthly loan frequency is 0.0500152, indicating that the frequency of customers applying for personal credit consumer loans is positively correlated with loan defaults; that is, the frequency of customers applying for personal credit consumer loans increases once a month, and the logarithm of the ratio of loan defaults will increase by 0.0500152.5. The accurate prediction ratio of the model is $(38304 + 79447)/136750 = 86.11\%$, which further shows that the prediction model has a better effect.

Data Availability

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

Conflicts of Interest

The author declared that he has no conflicts of interest regarding this work.

References

[1] H. Y. Wang, R. Zhu, and P. Ma, "Optimal subsampling for large sample logistic regression," *Journal of the American Statistical Association*, vol. 113, no. 522, pp. 829–844, 2018.

[2] H. Y. Wang, "Divide-and-conquer information-based optimal subdata selection algorithm," *Journal of Statistical Theory and Practice*, vol. 13, no. 3, pp. 1–19, 2019.

[3] M. Ai, Y. Jun, H. Zhang, and H. Y. Wang, "Optimal subsampling algorithms for big data regressions," *Statistica Sinica*, vol. 31, pp. 749–772, 2021.

[4] W. Zhuo, Z. He, M. Zheng, B. Hu, and R. Wang, "Research on personalized image retrieval technology of video stream big data management model," *Multimedia Tools and Applications*, vol. 2, pp. 1–18, 2021.

[5] D. Laura and T. Chiara, "Big data and model-based survey sampling," 2020, https://arxiv.org/abs/2002.04255.

[6] M. E. Hamzaoui and F. Bensalah, "Toward a knowledge-based model to fight against cybercrime within big data environments: a set of key questions to introduce the topic," in *Machine Intelligence and Big Data Analytics for Cybersecurity Applications. Studies in Computational Intelligence*, Y. Maleh, M. Shojafar, M. Alazab, and Y. Baddi, Eds., vol. 919, pp. 75–101, Springer, Cham, 2021.

[7] K. Xie, K. Ozbay, D. Yang, C. Xu, and H. Yang, "Modeling bicycle crash costs using big data: a grid-cell-based Tobit model with random parameters," *Journal of Transport Geography*, vol. 91, p. 102953, 2021.

[8] Y. Liu, L. Wang, Y. Lin et al., "The Influence of built environment on the spatial distribution of housing price: based on multiple big data and hedonic model," *Journal of Landscape Research*, vol. 13, no. 4, p. 5, 2021.

[9] I. M. El-Hasnony, S. I. Barakat, M. Elhoseny, and R. R. Mostafa, "Improved feature selection model for big data analytics," *IEEE Access*, vol. 8, pp. 66989–67004, 2020.

[10] Y. Gao, X. Chen, and X. Du, "A big data provenance model for data security supervision based on PROV-DM model," *IEEE Access*, vol. 8, pp. 38742–38752, 2020.

[11] P. Ma, M. W. Mahoney, and Y. Bin, "A statistical perspective on algorithmic leveraging," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 861–911, 2015.

[12] M. W. Mahoney and P. Drineas, "CUR matrix decompositions for improved data analysis," *Proceedings of the National Academy of Sciences*, vol. 106, no. 3, pp. 697–702, 2009.

[13] P. Drineas, M. Mahoney, S. Muthukrishnan, and T. Sarlós, "Faster least squares approximation," *Numerische Mathematik*, vol. 117, no. 2, pp. 219–249, 2011.

[14] X. Ma and S. Lv, "Financial credit risk prediction in Internet finance driven by machine learning," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8359–8367, 2019.

[15] G. Chen and S. Li, "Research on location fusion of spatial geological disaster based on fuzzy SVM," *Computer Communications*, vol. 153, pp. 538–544, 2020.

[16] Y. Chen, "Research on the credit risk assessment of Chinese online peer-to-peer lending borrower on logistic regression model," DEStech Transactions on Environment, Energy and Earth Sciences, 2017.

[17] A. Byanjankar, M. Heikkila, and J. Mezei, "Predicting credit risk in peer-to-peer lending: a neural network approach," in *IEEE Symposium Series on Computational Intelligence*, pp. 719–725, Cape Town, South Africa, 2015.

[18] D. Cui, "Financial credit risk warning based on big data analysis," *Metallurgical and Mining Industry*, vol. 7, no. 6, pp. 133–141, 2015.

[19] T. Chen and G. Carlos, "XGboost: a scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

[20] G. B. Chen and S. Li, "Network on chip for enterprise information management and integration in intelligent physical systems," *Enterprise Information Systems*, vol. 15, no. 7, pp. 935–950, 2021.

[21] E. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, 1968.

[22] R. C. West, "A factor-analytic approach to bank condition," *Journal of Banking & Finance*, vol. 9, no. 2, pp. 253–266, 1985.

[23] W. N. Pugh, E. P. Daniel, S. John, and J. Jahera, "Antitakeover charter amendments: effects on corporate decisions," *Journal of Financial Research*, vol. 15, no. 1, pp. 57–67, 1992.

[24] J. Ericsson, J. Kris, and O. Rodolfo, "The determinants of credit default swap premia," *Journal of Financial and Quantitative Analysis*, vol. 44, no. 1, pp. 109–132, 2009.

[25] G. Arminger, D. Enache, and T. Bonne, "Analyzing credit risk data: a comparison of logistic discrimination, classification tree analysis, and feedforward networks," *Social Science Electronic Publishing*, vol. 12, no. 2, pp. 293–310, 1997.