

Research Article

The Prediction Analysis of Peer-to-Peer Lending Platforms Default Risk Based on Comparative Models

Haifeng Guo,¹ Ke Peng,¹ Xiaolei Xu ,² Shuai Tao,³ and Zhen Wu²

¹School of Management, Harbin Institute of Technology, Harbin, China

²National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing, China

³Business Intelligence Department, Ice Kredit, Nanjing, China

Correspondence should be addressed to Xiaolei Xu; xuxiaolei@cert.org.cn

Received 12 August 2020; Accepted 24 October 2020; Published 29 November 2020

Academic Editor: Fabrizio Riguzzi

Copyright © 2020 Haifeng Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper examines the determinants of platform default risk using machine learning methods, including comprehensive models, and thus compares these models' predictive abilities. To test platform's default risk, this paper constructs three types of variables, which reflect a platform's operating characteristics, customer feedback, and compliance capability. We find that the abnormal return tends to trigger default risk significantly. However, the default risk can be minimized if a platform has positive recommendations from customers and more transparent information disclosure or is affiliated as the member of the National Internet Finance Association of China. Empirical results indicate that the CART model outperforms the Random Forests model and Logit regression in predicting platform default risk. Our study sheds light on default risk prediction and thus can improve the government regulation ability.

1. Introduction

Default risk has long been a significant risk factor to test borrowers' behaviour in Peer-to-Peer (P2P) lending. For borrower's credit risk evaluation, the study in [1] points out that the social tie has a positive influence on lending success and a negative influence on credit risk. The study in [2] also tests the impact of social ties between users from relevant platforms to measure the default risk. Borrowers with more social ties are easier to get their loans, while their default probabilities are also higher. The study in [3] investigates the role of personal guarantee in P2P marketplaces. The results reveal that the loans with guarantees and shorter time intervals between posting and closing are much easier to get. The study in [4] proposes that borrowers' default risk of the Chinese P2P lending platform Renrendai is significantly influenced by borrower's credit score and credit rate distribution.

Moreover, it is of great importance to test P2P platform's default risk. Using data from LendingClub with machine learning algorithms from 2013 to 2015, [5] outlines that P2P

lending platforms with high expected return and short payback period are more likely to have low default risk by using decision tree. The study in [6] further examines the relationship between soft information and P2P lending default risk in two European P2P lending platforms. Their experiments indicate that soft information such as the length of text, spelling mistakes, and the sentiment analysis of keywords generated from description text has a limited impact on the probability of default. Previous studies examine the platform default risk by using Probit regression and tree-based classifiers, respectively. Extending this stream of research, our study develops a comprehensive model including Logit, CART, and Random Forests algorithms to deal with credit scoring problems and test platform's default risk. Therefore, the model is optimized to obtain unbiased estimation and higher precision.

Information asymmetry is an enormous challenge in studying P2P default risk. Lenders receive information through platforms with low transparency. It appears to increase cognitive bias, which harms investors' information processing ability, departing their investment decisions from

the rational benchmarks [7]. Traditional financial institutions cope with this problem by disclosing detailed information and introducing high-quality collateral. However, it is difficult to implement such tools in P2P lending market due to the high transaction fee [8]. Most studies obtain P2P lending characteristics information in the US from the FICO score. The FICO score is widely used for investors to distinguish the creditworthiness of borrowers, along with additional information such as debt-to-income ratio and employment length to evaluate credit risk [9]. The study in [10] tests the impact of appearance-based judgments of trustworthiness based on credit grade and FICO score. Using borrowers' images to identify appearance-based impression in P2P lending market, a trustworthy appearance predicts not only the expected return but also the probability of getting a loan. According to [11], credit rating, debt-to-income ratio, FICO score, and revolving line utilization are all significant factors contributing to the probability of loan defaults. High credit rating and short repayment period effectively reduce mortality risk and default risk. The authors in [12] improve the FICO score model and design a profit scoring system. They choose the internal rate of return as profitability measurement and find that borrower's expected rate of return, indebtedness, and loan purpose are three major determinants.

However, there is no unified credit scoring system in Chinese P2P lending market. In China, platform default risk in P2P lending market is even more serious because of the lack of credit information system. People's Bank of China's credit information system is the only official way to disclose credit information, which is not accessible to all platforms. Therefore, most researchers try to find a suitable way to test credit information. The study in [13] finds that the impact of communication is strong for low credit rating borrowers using data from lending market. The study in [14] finds similar results by classifying the borrower information into loan characteristics, borrower credit, and personal details. The authors in [15] test the strong signals affecting the probability of borrowing success for PaiPaiDai. They find that acquiring verifications and borrower's history transactions are significant in both the first borrowing model and the repeated borrowing model. The characteristics of the P2P lending platform are similar to those in the credit card scoring model.

In China, the P2P industry has encountered many problems and thus accumulated serious default risk. Figure 1 shows the development of P2P lending market from 2012 to 2019 with data obtained from <http://www.wdzj.com>. The P2P lending has proliferated since 2012, and the platform growth rate peaked at 250% in 2014. However, this indiscriminate development generated considerable platform default risk. In 2015, almost half of the platforms had withdrawal problems and collapsed. In order to manage default risk effectively, ten supervisory authorities have jointly published a guideline to mandate P2P lending market operating standards and compliance rules. Along with implementing regulations, the number of P2P lending platforms shows a sustained drop, and only 344 P2P platforms remain in December 2019.

The growing platform default risk and the sharp decline of platforms in China cast a shadow of uncertainty over the P2P lending market. Although a few platforms have made steady progress and met the criteria, the majority of platforms struggled to meet the regulatory requirements [16]. For this reason, the Chinese Internet Finance Administrative Section released several requirements in November 2019. These requirements guide some qualified P2P lending institutions to transform into small loan companies to decrease the systemic risks. Platforms without the ability to meet supervisory requirements, on the other hand, will be banned. Therefore, platform default risk measurement plays a considerable role in both minimizing the loss of lenders and maintaining the stability of the capital market during this conversion process.

Although many researchers have concentrated on default risk identification, it needs more work to identify boundary conditions along with the implementation of regulatory policy. Our paper has two contributions. The first contribution is to construct assessment determinants of default risk and figure out what factors effectively influence the operating status of P2P lending platforms. We construct three groups of variables: operating characteristics, customer feedback, and compliance capability. A second contribution is to test which model has the best prediction accuracy. Applying disequilibrium fuzzy proximal support vector machine to default risk evaluation model, borrower loan status, platforms, and policy environment are found to be the three key factors of default risk of P2P lending [17]. The study in [1, 13] indicates that Logit regression is good in risk measurement. However, machine learning algorithms such as the CART model and Random Forests model are good in feature selection. These decision-tree-based classifiers exclude the influence of outliers and reduce the ambiguity in decision-making procedure [5, 18]. Therefore, our paper will compare the Logit regression, the CART model, and the Random Forests model and figure out which model is more predictive.

Our empirical results suggest that platforms with positive customer reviews and high information disclosure quality effectively minimize information asymmetry and default risk. P2P lending platforms with abnormal return are more likely to underestimate the credit risk and default. We find that the CART model has the best predicting ability over Logit regression and Random Forests model. Our evidence is conducive to investors and regulators for optimal investments and regulatory strategies. The rest of our paper is constructed as follows. Section 2 describes the variables of the model and summarizes descriptive statistics of P2P lending platform up to July 2019. In Section 3, we present the methodologies used to predict the importance of platform default risk and report the empirical results. In Section 4, we outline the economic significance of the empirical results and make a general conclusion.

2. Data

2.1. Variable Setting. The paper examines whether some features of P2P lending platform could measure default risk

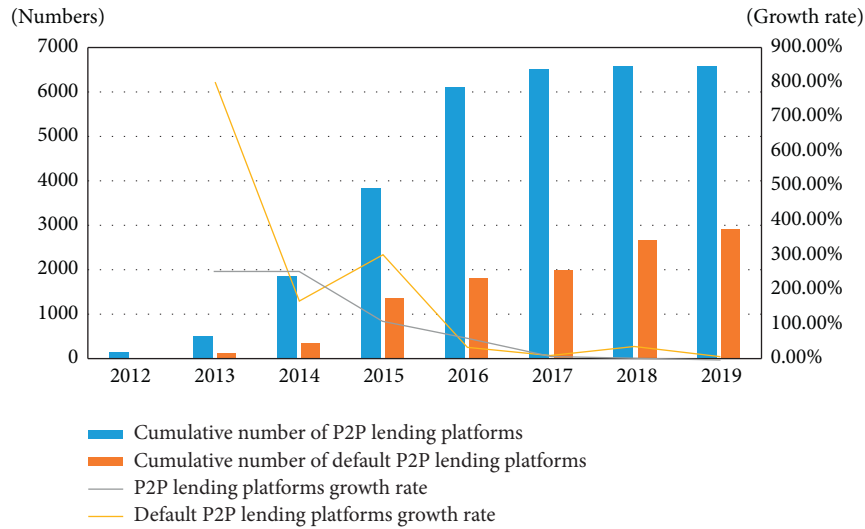


FIGURE 1: Number of online P2P lending platforms in China.

and predict the likelihood of platform default. The dependent variable, platform operating status, is equal to zero if the platform keeps operating normally and one otherwise [19, 20]. When a platform suffers from operational irregularities, it is mainly due to default events, such as running away with money or terminating the business. Under such circumstances, lenders cannot retrieve repayments from platforms.

The information about online P2P lending platform's risk is divided into three categories: operating characteristics, customer feedback, and regulatory compliance capability. Table 1 provides the definitions of all variables in different categories. The operating characteristics reveal fundamental information about platforms. Five features in the customer feedback category reflect different aspects of consumers' review of platforms. The regulatory compliance capability category is selected to inspect whether platforms follow the regulation rules and safety precautions. This category also confirms whether the platform is affiliated with the NIFA (National Internet Finance Association of China) and information disclosure statements.

2.2. Summary Statistics. In this section, we describe and summarize the descriptive statistics of the dataset, including the loan status and features of P2P lending platforms in China. We collected platform information issued by Wangdaizhijia (<http://www.wdzj.com>) up to July 2019. Some data after the initial crawling cannot be used directly. For instance, the platform duration variable, as measured by time period from year of establishment to year of data collection, needs to be transformed. Besides the dataset from Wangdaizhijia, there are several media sources containing information that reflects customer reviews. We preprocessed and screened the information and obtained a valid dataset [21]. In addition, we removed the information of bankrupt platforms and newly established platforms which have not been collected in time and meaningless data. Eventually, we get the dataset of 1283 lending platforms, 860 default platforms, and 423 platforms operating with compliance.

Table 2 describes summary descriptive statistics for all variables in the model. The platform expected return is positively correlated with platform default risk. The difference of average expected return between default platform and platform operating normally is nearly 5%. High investment period seems to decrease platform default probability. The average investment period of platforms operating normally is 6.28 months, twice larger than that of default platforms. In addition, the customer feedback scores that range from 0 to 5 are all higher than 3 for the platform operating normally. The withdrawal score and stand guard score indicate the status of platform cash flow and rigid redemption. In contrast, the means of withdrawal score, service score, and experience score of bankrupted platforms are lower than 3, and the standard deviations of these variables are significantly higher than those of platforms operating normally. It is clear that platforms operating normally get better customer feedback. However, there is no significant difference in bank depository, bid security, and safeguard mode between the platforms with these two opposite operating status. Most of the platforms operating normally have ICP registration and all of them are members of the National Internet Finance Association of China.

3. Empirical Analysis and Results

3.1. Logit Regression. Logit regression is widely used in varieties of economic domains. For example, in business creditworthiness evaluation, the Logit regression is well established to solve default problems with the highest accuracy [22]. Compared to the Probit model, the Logit does not require normally distributed independent variables, and it has stronger data processing capabilities to measure the probability of default in big data scenarios [6, 23]. In addition, the wide application of Logit in big data scenarios with different characteristics is suitable for the circumstance in P2P lending market.

We have proposed a Logit regression with credit scoring to investigate all independent variables in three aspects.

TABLE 1: Definition of variables.

Category	Variable	Description and value	
Dependent	y	Platform operating status, 0 = operating normally, 1 = operating problematically	
Independent	Operating	Rate Platform average expected rate of return (%)	
	Characteristics	exis_time	Platform duration from year of platform establishment to year of data collection
		re_capital	Platform registered capital (10,000 RMB)
		auto_invest	Whether the platform had automatic bidding, 0 = no, 1 = yes
		inv_peroid	Platform average investment period (month)
		wtd_score	Platform withdrawal score, 1 = not satisfied at all, 2 = not satisfied, 3 = normal, 4 = satisfied, 5 = very satisfied
	Customer feedback	stand_score	Platform stand guard score, 1 = not satisfied at all, 2 = not satisfied, 3 = normal, 4 = satisfied, 5 = very satisfied
		serve_score	Platform service score, 1 = not satisfied at all, 2 = not satisfied, 3 = normal, 4 = satisfied, 5 = very satisfied
		exp_score	Platform experience score, 1 = not satisfied at all, 2 = not satisfied, 3 = normal, 4 = satisfied, 5 = very satisfied
		rec_ornot	Whether the customer would recommend the platform to other people, 0 = no, 1 = yes
	Regulatory compliance capability	bank_save	Whether the platform had bank depository, 0 = no, 1 = yes
		Protection	Whether the platform had bid security, 0 = no, 1 = yes
		safe_gaurd	Whether the platform had safe guard mode, 0 = no, 1 = yes
		ICP_reg	Whether the platform had ICP registration, 0 = no, 1 = yes
asso_par		Whether the platform joined the NIFA (National Internet Finance Association of China), 0 = no, 1 = yes	
ope_data		Whether the platform disclosed operating statistics, 0 = no, 1 = yes	
safty_report		Whether the platform disclosed safety assessment report, 0 = no, 1 = yes	
fin_report	Whether the platform disclosed financial audit report, 0 = no, 1 = yes		

TABLE 2: Summary statistics.

Category	Variable	Platform operating normally				Platform operating problematically			
		Mean	Sd	Min	Max	Mean	Sd	Min	Max
Operating	Rate	10.46	2.28	5.6	21	14.49	5.98	6	74
characteristics	exis_time	4.44	0.97	2	8	4.48	1.01	2	9
	re_capital	6537.9	13873.4	500	250000	6565.5	16859.8	100	315804
	auto_inv	0.50	0.50	0	1	0.59	0.49	0	1
	inv_peroid	6.28	5.56	0.53	35.88	3.14	2.51	1	18.62
Customer feedback	wtd_score	3.32	0.57	1.3	5	2.45	0.84	1	5
	stand_score	4.00	0.39	4	5	3.12	0.90	1	5
	serve_score	3.99	0.43	2	5	2.86	0.93	1	5
	exp_score	3.96	0.42	2	5	2.82	0.91	1	5
	rec_ornot	0.87	0.34	0	1	0.13	0.34	0	1
Regulatory compliance capability	bank_save	0.33	0.47	0	1	0.35	0.68	0	1
	Protection	0.64	0.48	0	1	0.72	0.45	0	1
	safe_gaurd	0.83	0.38	0	1	0.83	0.38	0	1
	ICP_reg	0.83	0.38	0	1	0.19	0.39	0	1
	asso_par	1	0	1	1	0.11	0.32	0	1
	ope_data	0.35	0.48	0	1	0	0	0	0
	safty_report	0.06	0.24	0	1	0.02	0.14	0	1
fin_report	0.06	0.24	0	1	0.03	0.16	0	1	

Logit_model = $\ln(p/(1-p))$, where p is the number of bankrupted platforms. When platforms are operating normally, the value of p equals 0; otherwise p equals 1. In order to verify the fitting effect of the model, 20% of the data is

randomly chosen as the test dataset from our sample. After eliminating multicollinearity problem and insignificant independent variables at 1% significance level, Table 3 shows the results of Logit regression. From the empirical results,

TABLE 3: Logit regression results.

Variables	Coef.	St dev.	Z value	P value	P _{difference/average}
Const	3.6970	0.756	4.892	≤0.001	
Rate	0.1934	0.041	4.741	≤0.001	0.0053
inv_peroid	-0.1674	0.039	-4.316	≤0.001	-0.0028
wtd_score	-0.7899	0.181	-4.367	≤0.001	-0.0044
rec_ornot	-0.9424	0.259	-11.376	≤0.001	0.0509
ICP_reg	-1.8549	0.259	-7.149	≤0.001	-0.0069

platform expected rate of return, investment period, withdrawal score, consumer recommendation, and platform affiliated as the member of the NIFA are the five determinants with significant impact on platform default risk. The equation of Logit regression is calculated as

$$\begin{aligned} \text{Logit}_{\text{model}} = & 3.6970 + 0.1934 \text{ rate} - 0.1674 \text{ inv_peroid} \\ & - 0.7899 \text{ wtd_score} - 0.9424 \text{ rec_ornot} \\ & - 1.8549 \text{ ICP_reg.} \end{aligned} \quad (1)$$

As expected, the coefficient of platform expected return is positive. It is worth pointing out that the platforms with abnormal expected return seem to generate more default risk than what they can control. The strategy to attract more investors with high expected rate of return also leads to higher default risk exposure. Notably, Table 3 illustrates that the estimated coefficients for investment period, platform withdrawal score, the recommendation from customer, and ICP registration are all negative. The table also shows that the probability of default decreases by 5% from platform with no recommendation to the platform with positive customer review when other independent variables remain at their average level. Based on the result of regression, we conclude that high withdrawal score and positive consumer recommendation effectively alleviate information asymmetry. In addition, longer platform average investment period increases platform default risk, potentially leading to the failure to retrieve investors' money back. ICP registration is a permission released by local financial regulatory authorities. In accordance with the financial regulations, ICP registration motivates P2P lending platforms to fulfil the responsibilities of self-discipline and help in regulating their market behaviour. Platforms with ICP registration entail lower platform default risks by preventing adverse selection problems with the third-party guarantee.

3.2. CART and Random Forests Models. Decision tree learning is a popular supervised learning algorithm for building a binary tree structure with each corresponding split at the node of a tree branch. As a data mining method, decision tree learning produces a set of rules to solve both classification and regression tasks [5]. CART and Random Forests are the most widely used methods to test the nonlinear relationship between predictive factors and default risk. The process of building a decision tree is a divide-and-conquer approach. Based on the test condition for the associated feature, the root node of the decision tree

corresponds to the entire training data and each node split corresponds to a partitioning of the available data at that node.

There are two critical issues in decision tree learning: how to choose the appropriate split at each node and how many levels are there at each tree branch. Within the context of Random Forests model, which are collections of decision trees, splitting is done according to Gini Index, which is described below. The number of levels in each decision tree branch is controlled by an algorithm parameter [24]. The Gini Index for internal tree node is computed as below, where the probability of sample in k^{th} category is p_k :

$$\begin{aligned} \text{Gini}(p) &= \sum_{k=1}^K p_k(1 - p_k), \\ &= 1 - \sum_{k=1}^K p_k^2. \end{aligned} \quad (2)$$

Classification and Regression Tree (CART) model solves both classification and regression issues and avoids the advantages brought by overfitting. It is a binary recursive segmentation technique, which consists of a series of binary trees. The root nodes in the CART model represent the input variable and the leaf nodes represent the predicted output variable. The bootstrapping process built could be summarized as tree building with recursive segmentation and tree pruning process with verification sets. The primary principle of this algorithm is to find the largest segmentation point of Gini Index in each binary tree and distinguish the purity of training data when the nodes split. Although the greater the importance coefficient implies the greater the impact on the dependent variable, it is noticeable that the importance value of the coefficient represents the influence of independent variables on Gini Index directly rather than on the dependent variable.

Random Forests (RFs) method is another kind of popular decision tree algorithm with bootstrap aggregating technique. This kind of method builds up several decision trees and decomposes the branches and nodes through randomized split attributes. It further improves the model accuracy and decreases the variance through averaging effectively [24]. When building these trees, the candidate split is chosen by a random selection from the full set of attributes. The split is allowed to use only one of these attributes, and a fresh selection of attributes is made at each split. In each tree, splitting is to be continued until the tree reaches a certain depth. Due to the randomness of variable selection, this algorithm is not sensitive to multicollinearity [25]. Random Forests model also avoids overfitting problem, correcting the shortcomings of the training dataset. In addition, Random Forests model effectively predicts the relative importance of each factor. However, the construction of a large number of decision trees slows down the progress of algorithm, leading to a slower model fitting speed.

The results concerning importance of coefficients with CART model and Random Forests model are presented in Table 4. Table 4 reports that the importance of coefficients in different categories diverges between CART and Random

TABLE 4: The importance of coefficients.

Category	Variable	The importance of coefficient	
		CART	Random Forests
Operating	Rate	0.00575833	0.03720244
characteristics	exis_time	0.00054727	0.00653306
	re_capital	0.00000000	0.00935029
	auto_invest	0.00000000	0.00161401
	inv_peroid	0.01377586	0.03078771
Customer feedback	wtd_score	0.00771625	0.04652210
	stand_score	0.00192677	0.08167963
	serve_score	0.07383511	0.07247281
	exp_score	0.00964909	0.12250993
	rec_ornot	0.11605336	0.11150608
Regulatory compliance capability	bank_save	0.00000000	0.00542350
	Protection	0.00215937	0.00133869
	safe_gaurd	0.00000000	0.00082923
	ICP_reg	0.00025593	0.05372013
	asso_par	0.72992238	0.35057620
	ope_data	0.03840029	0.06492066
	safty_report	0.00000000	0.00224815
fin_report	0.00000000	0.00076537	

Forests models. For CART model, the importance of coefficients in regulatory compliance capability is more than 0.7. The coefficient of customer feedback information is 0.2, nearly twice that of operating characteristics. More precisely, the top five characteristics are the platform affiliated as the member of the NIFA, recommendation from investors, platform service score, operating data disclosure, and platform average investment period. The goodness of fit using the CART model is 0.9154; that is, the model achieves a prediction accuracy of 91.54%.

The result of coefficient importance with Random Forests model is similar to that of CART model. The importance of coefficients in regulatory compliance capability information, customer feedback information, and operating characteristics information are 0.48, 0.43, and 0.09, respectively. The descending order of top five important coefficients with Random Forests model is platform affiliated as the member of the NIFA, platform experience score, recommendation from investors, platform stand guard score, and platform service score. The cross-validation result with *Python* software shows that the model's goodness of fit is 0.9203. In other words, the independent variables in the model contribute 92.03% explanatory power for platform default risk prediction.

4. Model Assessment

The prediction performance measurement is an essential step to evaluate the accuracy of machine learning. In the binary models, the error rate is widely used to measure predictive power and performance. The confusion matrix is a performance analysis table with four different combinations of predicted and actual values, which helps to better understand the errors in the classification. The records in dataset are collected in a matrix according to the real category and classification model prediction category. The row of this matrix represents the true value, and the column represents the predicted value. The form of confusion matrix

is shown in Figure 2. However, the confusion matrix only reflects the amount of actual and predictive data. To evaluate the trade-off relationship between overfitting and underfitting problems in decision tree classifiers, our paper selects several performance indicators to test the accuracy, precision, sensitivity, and specificity of the model as suggested by [26]. These performance indicators are shown in Table 5.

AUC-ROC curve is one of the most significant evaluation indicators to test the performance of machine learning models. The ROC curve describes a graphical trade-off relationship between true positives and false positives, balanced at equal error rate (ERR). The ERR is positively correlated with model performance. This 2D curve plots the performance of binary classifiers under threshold options such as false acceptance rate (FAR) and false rejection rate (FRR). The AUC (Area Under Curve) is the area enclosed by the ROC curve. As the ROC curve plots above function $y = x$, AUC is usually more than 0.5. If AUC = 1, the classifier has perfect predicting power and the true value of each sample could be predicted correctly. If $0.5 < \text{AUC} < 1$, the classifier has certain predicting power under threshold settings. If AUC = 0.5, the classifier is complete randomly. We compared the performances of Logit regression, CART, and Random Forests models according to AUC-ROC curve. The results for all the coefficients are shown in Figures 3(a)–3(d).

The AUC of operating characteristics information with Logit regression is 0.68, compared to 0.75 for CART model and 0.77 for Random Forests model. The AUC for customer feedback information with Logit regression, CART model, and Random Forests model are 0.85, 0.90, and 0.91. In regulatory compliance capability information, the AUC for two decision tree methods are both 0.97 and Logit regression shows the worst performance. CART model has the highest overall AUC at 0.99, whereas Random Forests model has the second highest AUC at 0.98. The overall AUC for Logit regression is the lowest at 0.96. Therefore, CART model has the best prediction performance based on ROC-AUC.

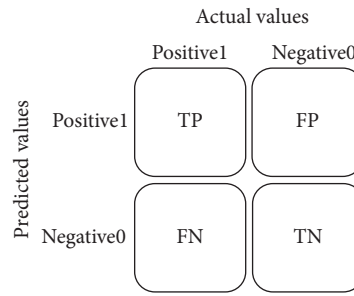


FIGURE 2: The confusion matrix.

TABLE 5: Performance indicator.

Name	Explanation
Accuracy (ACC)	Calculated as $(TP + TN)/(TP + TN + FP + FN)$
Precision (PPV)	Calculated as $TP/(TP + FP)$
Sensitivity (TPR)	Calculated as $TP/(TP + FN)$
Specificity (TNR)	Calculated as $TN/(TN + FP)$
FPR	Calculated as $100\% - TN/(TN + FP)$
PR curve	Precision-recall curve (<i>x</i> -axis: PPV versus <i>y</i> -axis: TPR)
ROC curve	Receiver operating characteristics curve (<i>x</i> -axis: FPR versus <i>y</i> -axis: TPR)

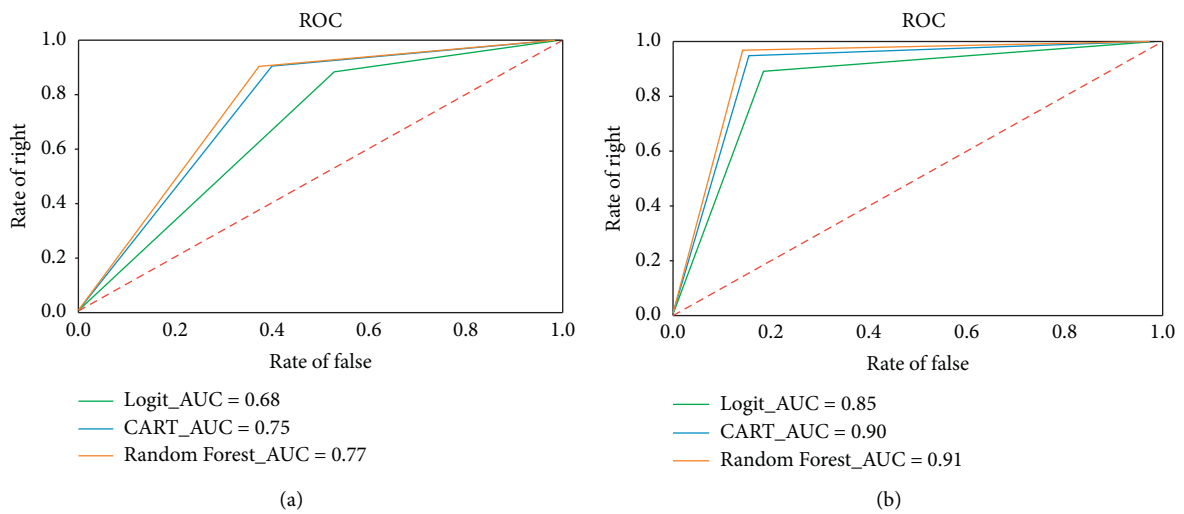


FIGURE 3: Continued.

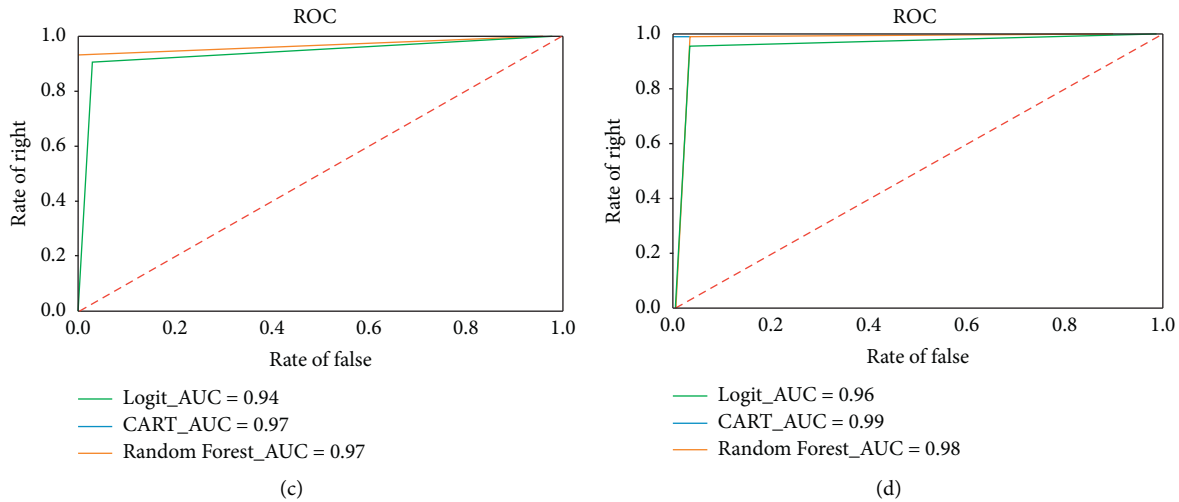


FIGURE 3: (a) ROC curve in operating characteristics category. (b) ROC curve in customer feedback category. (c). ROC curve in regulatory compliance capability category. (d) ROC curve of all variables.

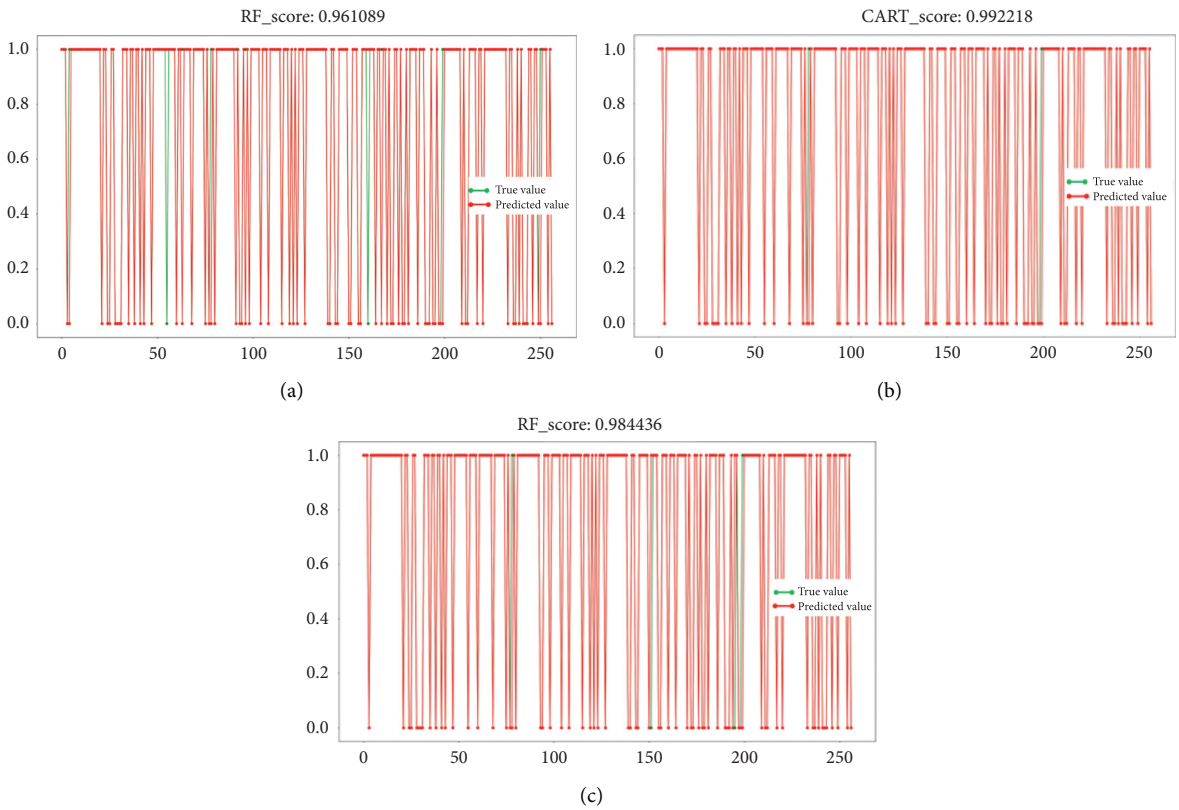


FIGURE 4: (a) Predicted score of Logit regression. (b) Predicted score of CART model. (c) Predicted score of Random Forests model.

It is common practice to fit a model using training data and then to evaluate the performance on a test dataset. A test set in machine learning is a secondary or tertiary dataset used to evaluate the machine learning program after it has been trained. After testing the data in each model, we get the predicted value based on the test set and compare the predicted value with the true value. The predicted results are shown in Figures 4(a)–4(c). When

the red line covers the green line, the predictive performance of the model is positively correlated with the distance between the red line and green line. As shown below, the predicted score of the Logit regression is 0.77. The predicted score of CART model is 0.99, slightly outperforming Random Forests model. Thus, we declare that the CART model is the best overall classifier compared to Logit regression and Random Forests model.

5. Conclusions

In this paper, we explore the factors predicting P2P lending platform default risk in China. We present three machine learning methods, Logit regression, CART model, and Random Forests model, to derive insights into platform operating status prediction. Our empirical analysis selected 18 features in three groups: operating characteristics, customer feedback, and regulatory compliance capability. The results suggest that platform abnormal expected rate of return is a crucial factor that contributes to default risk. When the expected return is greater than 14%, the probability of platform default is higher than 50%. This paper provides evidence that high interest rate could be treated as a signal of the platform's poor solvency. The results also imply that most of the platforms operating normally join the NIFA and disclose operating reports regularly. With the help of third-party supervision, platforms reduce their moral hazard problems and information asymmetry. The positive recommendation from investors has more explanatory power for the decrease of platform default risk. Potential investors would build confidence in the P2P lending platform based on the customer's review.

Furthermore, the proposed CART model shows better predictive ability with the highest AUC and prediction score, indicating the effectiveness of platform default risk prediction. However, there are a few limitations in our research. Due to the lack of critical information, we exclude some platform operating characteristics in model construction. This may lead to the omission of important variables and decrease the prediction accuracy. Another limitation is that the sample size of the training set is still insufficient for machine learning classifiers. We may increase training data to improve the accuracy of the model in further research.

To sum up, we can conclude that the default risk of P2P platforms in China could be predicted with machine learning algorithms. The extent of platform operating status appears to be most reflected by platform customer feedback information. In the operating characteristics aspect, the results in this paper enable investors to take precautions of platforms whose expected return is above the normal level. Platforms that join the NIFA and disclose information regularly effectively mitigate the information asymmetry. There are also some inconsequential features such as the platform safeguard mode and financial audit report disclosure. We encourage the use of sound credit scoring models, rooted in machine learning techniques, to increase the predictive ability. By following the features of platform default risk, investors tend to act more rationally in judging P2P lending platforms. Our research also provides solid empirical support for supervisors to identify platforms that have relative probabilities to default in the future.

Our results have some specific implications for P2P lending market regulation in the compliance transformation period. Firstly, according to the empirical results from CART model, membership of the NIFA is a good criterion for a platform to show a lower probability of default. Therefore, the regulatory authorities should make detailed regulation guidelines and encourage platforms to meet the requirements of NIFA, which would improve the market

entry threshold and reduce information asymmetry. Secondly, the consumer's evaluation is very important, which indicates that the feedback of consumers could be adopted as a rule for future regulatory instrument. The regulation authorities may make some rules that highlight consumer's evaluation as one of the platform's business qualities. Finally, evaluation of P2P lending platform default risk in a regular period would help the regulatory authorities to develop a healthy ecological environment for P2P industry's management in China.

Data Availability

The data used to support this study are available at <https://shuju.wdzj.com/>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (71773025, 71532004, and 71850031) and the National Key R&D Program of China (2019YFC0850105 and 2020YFB1006104).

References

- [1] M. Lin, N. R. Prabhala, and S. Viswanathan, "Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending," *SSRN Electronic Journal*, vol. 59, no. 1, pp. 17–35, 2013.
- [2] S. Freedman and G. Z. Jin, "The information value of online social networks: lessons from peer-to-peer lending," *International Journal of Industrial Organization*, vol. 51, pp. 185–222, 2017.
- [3] S. Agarwal, Y. Li, C. Liu, and J. Zhang, "Personal guarantee and Peer-to-Peer lending: evidence from China," *Social Science Electronic Publishing*, vol. 11, no. 2015, p. 2015, 2015.
- [4] K. Zhang and X. Chen, "Herding in a P2P lending market: rational inference OR irrational trust?" *Electronic Commerce Research and Applications*, vol. 23, pp. 45–53, 2017.
- [5] V. Kumar, S. Natarajan, S. Keerthana, K. M. Chinmayi, and N. Lakshmi, "Credit risk analysis in peer-to-peer lending system," *Institute of Electrical and Electronics Engineers*, in *Proceedings of the 2016 IEEE International Conference on Knowledge Engineering and Applications (ICKEA)*, pp. 193–196, Nanyang Ave, Singapore, September 2016.
- [6] G. Dorflleitner, C. Priberny, S. Schuster et al., "Description-text related soft information in peer-to-peer lending - evidence from two leading European platforms," *Journal of Banking & Finance*, vol. 64, pp. 169–187, 2016.
- [7] S. Ayal, D. Bar-Haim, and M. Ofir, "Behavioral biases in peer-to-peer (P2P) lending," *Behavioral Finance*, vol. 367, p. 367, 2019.
- [8] X. Lin, X. Li, and Z. Zheng, "Evaluating borrower's default risk in peer-to-peer lending: evidence from a lending platform in China," *Applied Economics*, vol. 49, no. 35, pp. 3538–3545, 2017.
- [9] H. Tang, "Peer-to-peer lenders versus banks: substitutes or complements?" *The Review of Financial Studies*, vol. 32, no. 5, pp. 1900–1938, 2019.

- [10] J. E. Teich, H. Wallenius et al., "Borrower decision aid for people-to-people lending," *Decision Support Systems*, vol. 49, no. 1, pp. 52–60, 2010.
- [11] R. Emekter, Y. Tu, B. Jirasakuldech, and M. Lu, "Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending," *Applied Economics*, vol. 47, no. 1, pp. 54–70, 2015.
- [12] C. Serrano-Cinca and B. Gutiérrez-Nieto, "The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending," *Decision Support Systems*, vol. 89, pp. 113–122, 2016.
- [13] J. J. Xu and M. Chau, "Cheap talk? The impact of lender-borrower communication on peer-to-peer lending outcomes," *Journal of Management Information Systems*, vol. 35, no. 1, pp. 53–85, 2018.
- [14] Y. Liu, Q. Zhou, X. Zhao, and Y. Wang, "Can listing information indicate borrower credit risk in online peer-to-peer lending?" *Emerging Markets Finance and Trade*, vol. 54, no. 13, pp. 2982–2994, 2018.
- [15] S. Cai, X. Lin, D. Xu, and X. Fu, "Judging online peer-to-peer lending behavior: a comparison of first-time and repeated borrowing requests," *Information & Management*, vol. 53, no. 7, pp. 857–867, 2016.
- [16] M. Wang, X. Zheng, M. Zhu, and Z. Hu, "P2P lending platforms bankruptcy prediction using fuzzy SVM with region information," in *Proceedings of the 2016 IEEE 13th International Conference on E-Business Engineering (ICEBE)*, pp. 115–122, Macau, China, November 2016.
- [17] H.-Z. Ma and X.-R. Wang, "Influencing factor analysis of credit risk in P2P lending based on interpretative structural modeling," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 19, no. 3, pp. 777–786, 2016.
- [18] B. Luo and Z. Lin, "A decision tree model for herd behavior and empirical evidence from the online P2P lending market," *Information Systems and E-Business Management*, vol. 11, no. 1, pp. 141–160, 2013.
- [19] Y. Ge, H. Guo, H. Fung, and K. Guang, "CEO effects on the IPO market under different policy regimes: evidence from the Chinese SME board," *International Review of Economics & Finance*, vol. 63, pp. 163–175, 2019.
- [20] H. Guo, Y. Ge, C. Hsu, and H. Fung, "How did the elimination of the window guidance policy affect IPO performance in China? A stochastic dominance analysis," *Emerging Markets Finance and Trade*, vol. 1, pp. 1–15, 2019.
- [21] H. Guo, Y. Wang, X. Qiao, and Y. Ge, "Do CEO attributes create value in IPO price revision? Evidence from China," *Emerging Markets Finance and Trade*, 2020.
- [22] T.-C. Tang and L.-C. Chi, "Predicting multilateral trade credit risks: comparisons of Logit and Fuzzy Logic models using ROC curve analysis," *Expert Systems with Applications*, vol. 28, no. 3, pp. 547–556, 2005.
- [23] H. P. Tserng, P.-C. Chen, W.-H. Huang, M. C. Lei, and Q. H. Tran, "Prediction of default probability for construction firms using the logit model," *Journal of Civil Engineering and Management*, vol. 20, no. 2, pp. 247–255, 2014.
- [24] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [25] M. Malekipirbazari and V. Aksakalli, "Risk assessment in social lending via random forests," *Expert Systems with Applications*, vol. 42, no. 10, pp. 4621–4631, 2015.
- [26] C.-L. Huang, M.-C. Chen, and C.-J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert Systems with Applications*, vol. 33, no. 4, pp. 847–856, 2007.